



HAL
open science

Fingerprint approach using macrocyclic chemical nose sensors for disease diagnostics

Monica Swetha Bosco

► **To cite this version:**

Monica Swetha Bosco. Fingerprint approach using macrocyclic chemical nose sensors for disease diagnostics. Chemical Sciences. Université Paris Cité, 2024. English. <NNT : 2024UNIP5113>. <tel-04873486v2>

HAL Id: tel-04873486

<https://hal.science/tel-04873486v2>

Submitted on 21 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Université Paris Cité

Ecole doctorale : ED 563 - Médicament, Toxicologie, Chimie, Imageries

Laboratoire : Unité des Technologies Chimiques et Biologiques pour la Santé
(UTCBS)-UMR 8258

Fingerprint approach using macrocyclic chemical nose sensors for disease diagnostics

Par **Monica Swetha BOSCO**

Thèse de doctorat de : Interface Chimie-Biologie

Dirigée par Nathalie GAGEY-EILSTEIN

Présentée et soutenue publiquement le 16 octobre 2024

Devant un jury composé de:

Dr. Clémence ALLAIN, <i>DR</i>	<i>Université Paris-Saclay</i>	Rapportrice
Pr. Arnaud GAUTIER, <i>PU</i>	<i>Sorbonne Université</i>	Rapporteur
Dr. William PEVELER, <i>MCF-HDR</i>	<i>University of Glasgow</i>	Examineur
Pr. Benoit COLASSON, <i>PU</i>	<i>Université Paris Cité</i>	Examineur
Dr. Florence MAHUTEAU-BETZER, <i>DR</i>	<i>Institut Curie</i>	Invitée
Dr. Nathalie GAGEY-EILSTEIN, <i>MCF-HDR</i>	<i>Université Paris Cité</i>	Directrice de thèse

To the one who gave me roots to grow and wings to fly,

The foundation of every success I've achieved—this is for you, அம்மா."

ACKNOWLEDGEMENTS

"Alone we can do so little; together we can do so much."

— Helen Keller

The success and final outcome of this thesis were made possible through the guidance and assistance of many individuals. I would like to take this opportunity to express my deepest gratitude to all those who have supported me throughout my Ph.D journey

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Nathalie Gagey-Eilstein, for giving me this opportunity. Her unwavering support, invaluable guidance, and continuous encouragement have been instrumental in shaping both the direction and success of this work. I am deeply thankful for her insights, expertise, and patience, which were vital in navigating through the challenges I encountered throughout this journey and in fostering my growth as a researcher.

I would also like to extend my sincere appreciation to the members of my thesis committee —Dr. Clémence Allain, Dr. Arnaud Gautier, Dr. William Peveler, Dr. Benoit Colasson and Dr. Florence Mahuteau-Betzer—for their thoughtful feedback, constructive criticism, and valuable suggestions throughout this process. Their input has significantly enriched the quality of this work.

I am deeply indebted to my collaborators for their significant contributions to this research work. Your insights and expertise have been invaluable and this project has greatly benefitted from our collaboration. I want to extend my special thanks to Dr. Florence Mahuteau Betzer, Dr. Delphine Martin-Naud, and Dr. Claire Beauvineau for generously providing the necessary compounds for the project and for assisting with the experiments using the pipetting robot. I am also thankful to Dr. Stephanie Descroix, Dr. Monica Araya-Farias, and Dr. Giacomo Groppero for their help in setting up the droplet microfluidic platform and for providing me with the training needed to utilize this system. Additionally, I acknowledge Dr. Jean-Francois Gaucher for his expertise and time spent on the ITC experiments, as well as for the valuable discussions that influenced the progress of this project. The statistical analysis conducted in this research would not have been possible without the support of Dr. Yves Rozenholc and Zeki Topçu. I also extend my gratitude to Dr. Vassilis Tsatsaris for granting us access to the APHERESE cohort and for his valuable scientific input on this project.

Additionally, I would like to extend my heartfelt thanks to the members of my Comité de suivi individual (CSI), Dr. Claire Mangeney and Dr. Stephanie Descroix, whose consistent feedback and guidance have been key in refining and enhancing this project over the past three years.

I am also grateful to the funding agencies and organizations that supported this research, including the Agence Nationale de la Recherche (ANR), whose financial assistance made this work possible.

Additionally, I extend my thanks to the ED-563 MTCI Doctoral School and Université Paris Cité, Santé, for providing the resources necessary for me to pursue this Ph.D.

I am deeply grateful to the Unité de Technologies Chimiques et Biologiques pour la Santé (UTCBS) laboratory and all its members for providing me an enjoyable and intellectually stimulating environment. Special thanks go to Dr. Nathalie Mignet and Dr. Cyrille Richard for their invaluable support during my transition to this lab, which greatly facilitated my adjustment and enabled me to continue my work smoothly. I would also like to express my heartfelt appreciation to my fellow Ph.D. students and Master's students, for their camaraderie and the time we spent together; their support was crucial during challenging times. Additionally, I am thankful to Carlos Gonzalez Galindo for his contributions to this project through his experiments and scientific discussions.

On a personal note, I would like to extend my heartfelt thanks to my family—Leo, Daddy, Sanjana, and Prashanth—for their unwavering belief, constant encouragement, prayers, and love, which have been the bedrock of my strength throughout this journey. To the two little angels of our family, Seana and Valfrid, your immense love and constant cheerleading have brightened every single day. Last but certainly not least, to my special angel, Mummy—your presence has been the guiding light of my journey, now and always. I hope I have made you proud.

I would also like to express my deepest gratitude to my family away from home at the Maison de l'Inde. You have filled these three years of my life with unforgettable memories that I will cherish forever. Thank you all for your care, concern, love, and unwavering support. I will forever be indebted to each one of you.

Divya, Sundar, Rohini and Mahalakshmi, மிக்க நன்றி

Ajjath, വളരെ നന്ദി

Mukul and Simran, खूप धन्यवाद

Akansha, ଧନ୍ୟବାଦ

Sakshi, Vani, Vanshika and Sahil, धन्यवाद

As I reflect on this journey, I am deeply moved by the collective support, encouragement, and wisdom I have received. Each person mentioned here has played a crucial role in shaping this work and my experience. I am truly fortunate and grateful for the extraordinary network of support that has accompanied me throughout this Ph.D journey. Thank you all for being an integral part of this achievement.

Thank you,
Monica Swetha Bosco

**Fingerprint approach using macrocyclic
chemical nose sensors for disease
diagnostics**

Présentée par

Monica Swetha BOSCO

Thèse de Doctorat

PUBLICATIONS

1. “Bimodal recognition ‘Chemical Nose Sensor’ and microfluidic device for pattern recognition of proteins and diseases”

[Under revision- *ACS Applied Biomaterials*]

Bosco, M. S.; Naud-Martin, D.; Gonzalez-Galindo, C.; Auvray, M.; Araya-Farias, M.; Gropplero, G.; Rozenholc, Y.; Topçu, Z.; Gaucher, J.-F.; Tsatsaris, V.; Descroix, S.; Mahuteau-Betzer, F.; Gagey-Eilstein, N.*

2. “A methodological study for the diagnostic of the SARS-CoV-2 infection in human serum with a macrocyclic sensor array”

Sensors and Diagnostics, **2024**, 3, 863-871.

Bosco, M. S.; Topcu, Z.; Pradhan, S.; Sossah, A.; Tsatsaris, V.; Vauloup-Fellous, C.; Agasti, S.S.; Rozenholc, Y.*; Gagey-Eilstein, N.*

ORAL COMMUNICATIONS AND POSTERS

Oral Communications:

1. **REcontres en Chimie Organique Biologique (RECOB19)** - March 2024, Aussois, France
“Fingerprint approach using macrocyclic chemical nose sensors”
2. **Jeunes Chercheuses et Jeunes Chercheurs de la FR FédCUP**- November 2023, Paris, France
-Awarded prize for best oral communication
“Fingerprint approach using macrocyclic chemical nose sensors – Applications in disease diagnosis”
3. **FHU Prema Journée Jeunes Chercheurs** - July 2022, Paris, France
“Fingerprint approach using macrocyclic chemical nose sensors to predict outcomes in Preeclampsia”

Posters:

“Fingerprint approach using macrocyclic chemical nose sensors – Applications in disease diagnosis”

1. **Supr@Paris, French Supramolecular Chemistry Congress** – May 2024, France
2. **FHU Prema Journée Jeunes Chercheurs** - July 2023, Paris, France
3. **2023, MTCI Scientific Days** - March 2023, Le Meé-sur-Seine, France
4. **30th SCT – Young Research Fellows Meeting** – February 2023, Paris, France

5. **3rd PSL Symposium** – Institut Curie – January, 2023, Paris, France
6. **Innovative and Chemical tools and systems for sensing and imaging of life** - December 2022, Paris, France

WORKSHOP

1. **“Chemical Nose/Tongue for clinical diagnosis and monitoring of health”** - Université Paris Cité, March, 2023

LIST OF ABBREVIATIONS

VOCs: Volatile organic compounds	CNPs: carbon nanoparticles
ORCs: Olfactory receptor cells	TNBC: triple negative breast cancer
ORs: Olfactory receptors	PKU: Phenylketonuria
RGB: Red, green and blue	IgG: immunoglobulin G
SERS: Surface Enhanced Raman Scattering	PKU: Phenylketonuria
IDA: Indicator displacement assay	PE: Preeclampsia
ABA: Associative binding assay	FITC: fluorescein isothiocyanate
CDs: Cyclodextrins	WHO: World Health Organization
C[n]As: Calixarenes	RF: Random forest model
CB[n]s: Cucurbiturils	PPV: Positive predictive value
AIE: Aggregation induced emission	NPV: Negative predictive value
AuNPs: Gold nanoparticles	OOB: Out-of-the-bag
AuNCs: Gold nanoclusters	LOO: Leave-One-Out
AgNTs: silver nanotriangles	LOD: limit of detection
AgNRS: silver nanorods	LDA: Linear discriminant analysis
GFP: Green fluorescent proteins	PCA: Principal Component Analysis
CPs: Conjugated polymers	HCA: Hierarchical Clustering Analysis
BSA: bovine serum albumin	ITC: Isothermal titration calorimetry
HSA: human serum albumin	AUC: Area under curve
DNA: Deoxyribonucleic acid	POC: Point-of-care
ssDNA: single stranded DNA	sFlt-1: soluble fms-like tyrosine kinase 1
FRET: Förster resonance energy transfer	PIGF: placental growth factor
HOMO: Highest occupied molecular orbital	ROI: Region of interest
LUMO: Lowest unoccupied molecular orbital	CB[7]-FL: CB[7]- fluorophore conjugate
ELISA: enzyme-linked immunosorbent assay	
AD: Alzheimer's disease	
MOFs: metal organic frameworks	
RT-PCR: Reverse transcription polymerase chain reaction	
SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2	
ROC: Receiver Operating Characteristic curves	

Titre : Un nez Chimique macrocyclique pour une approche diagnostique non spécifique et sans a-priori

Résumé: Les compositions biomoléculaires des fluides corporels reflètent la gravité et la progression des maladies, offrant ainsi de précieuses indications diagnostiques. Cette thèse étudie une approche de « nez chimique » non spécifique et sans à-priori, inspirée du système olfactif humain, afin de créer des empreintes uniques qui peuvent être liées à la composition du fluide corporel étudié. En développant des réseaux de capteurs optiques basés sur l'interaction hôte-invité entre les cucurbit[n]uriles (CB[n]) et les dérivés triphénylamine multibranchés (TPA), cette étude présente une approche permettant d'établir le profil des changements protéomiques globaux dans les états pathologiques. Dans ce système, les CB[n] servent d'éléments de reconnaissance, tandis que les dérivés du TPA servent d'éléments de transduction. Ces TPA sont affectés par le confinement structural causé par les interactions hôte-invité avec les CB[n], ce qui entraîne des changements optiques par le biais d'interactions compétitives avec les analytes. L'efficacité de cette approche est démontrée dans les réseaux de capteurs fluorescents et colorimétriques, avec l'identification de protéines et la prédiction d'états pathologiques tels que la phénylcétonurie et la prééclampsie. La grande dimension des données générées par ces réseaux de capteurs ont été analysées à l'aide de méthodes statistiques avancées, l'automatisation étant réalisée grâce à un pipeline rationalisé basé sur le programme R. L'intégration de plateformes automatisées telles que la microfluidique en gouttelettes et les robots de pipetage améliore encore les capacités de diagnostic à haut débit du système, soulignant son potentiel pour le diagnostic des maladies infectieuses.

Mots clefs: Détection basée sur des réseaux, chimie supramoléculaire, spectroscopie de fluorescence, analyse statistique multivariée, microfluidique

Title: Fingerprint approach using macrocyclic chemical nose sensors for disease diagnostics

Abstract: The biomolecular compositions of body fluids reflect the severity and progression of diseases, offering valuable diagnostic insights. This thesis investigates a non-specific, hypothesis-free, serum-based "chemical nose" approach, inspired by the human olfactory system, to create unique analyte fingerprints that can be tied back to the body fluid composition. By developing optical sensor arrays based on host-guest interactions between cucurbit[n]urils (CB[n]) and multibranch triphenylamine (TPA) derivatives, this study presents an approach for profiling global proteomic changes in diseased states. In this system, CB[n]'s function as recognition elements, while TPA derivatives serve as transduction elements. These TPA components are affected by the structural confinement caused by host-guest interactions with CB[n], leading to changes of their optical properties through competitive interactions with analytes. The effectiveness of this approach is demonstrated in both fluorescence and colorimetric sensor arrays, with successful identification of proteins and prediction of diseased states such as phenylketonuria and preeclampsia. The high-dimensional data generated by these sensor arrays were analyzed using advanced statistical methods, with automation achieved through a streamlined R program-based pipeline. The integration of automated platforms like droplet microfluidics and pipetting robots further enhances high-throughput diagnostic capabilities of the system, highlighting its potential for diagnosis of multifactorial diseases and large-scale data analysis in chemical sensing.

Keywords: Array-based sensing, Supramolecular Chemistry, Fluorescence Spectroscopy, Multivariate statistical analysis, Microfluidics

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
PUBLICATIONS	xi
ORAL COMMUNICATIONS, POSTERS AND WORKSHOP	xi
LIST OF ABBREVIATIONS	xv
TITLE, ABSTRACT, KEYWORDS	xvii
TABLE OF CONTENTS	xix

Chapter 1

Introduction to supramolecular host-guest systems and their utility in array-based sensing for biomedical applications

1.1. Sensing systems in nature	27
1.2. Array-based sensing	29
1.2.1. Recognition elements for array-based sensing	31
1.2.2. Transduction mechanisms for array-based sensing	33
1.2.3. Design criteria for array-based sensing	34
1.2.4. Pattern-recognition by multivariate statistical analysis	36
1.3. Supramolecular macrocyclic systems for sensing	40
1.3.1. Structural properties of common macrocyclic host systems	40
1.3.2. Differential sensing using macrocyclic host systems	43
1.4. Biomedical application of array-based sensors	49
1.4.1. Discrimination of basic biomolecules: Proteins	51
1.4.2. Discrimination and evaluation of cellular characteristics	51
1.4.3. Discrimination and evaluation of body fluids for disease diagnosis	53
1.5. State of the Art: Biomedical Applications of array-based sensing	54
1.6. Conclusions	62

Chapter 2

Bimodal recognition by host-guest chemical nose platform: Design strategy and characterization of system based on cucurbit[n]uril and triphenylamine derivatives

2.1. Introduction	67
2.2. Macrocyclic family of Cucurbit[n]urils (CB[n])	67
2.2.1. Structural and thermodynamic properties of CB[n]	68

2.2.2. Properties of CB[n] for sensing applications	72
2.2.3. Analyte binding properties of CB[n] homologues	73
2.3. Triphenylamine derivatives: structure and properties	75
2.3.1 Structural properties	75
2.3.2 Photophysical properties	76
2.3.3 Interactions with (bio)molecules	77
2.4. Results and Discussion: TPA-CB[n] fluorescent array-based sensor	79
2.4.1. Design strategy for construction of sensor array	79
2.4.2. Diversity of library of triphenylamine derivatives	80
2.4.3. Characterization of TPA-CB[n] complexation	82
2.4.4. Construction and characterization of a fluorescent based sensor array	89
2.4.5. Interaction of TPA and CB[7] with (bio)molecules	101
2.5. Conclusions	108
2.6. Materials and Methods	109
2.6.1 Synthesis and characterization spectra of unpublished TPA derivatives	109
2.6.2 Protein visualization and determination of solvent exposed hydrophobic residues on protein surfaces	112
2.6.3 UV- Visible and fluorescence spectroscopy	112
2.6.4. Isothermal Titration Calorimetry	113

Chapter 3

Data management and statistical analysis for pattern recognition in array-based sensing

3.1. Introduction	117
3.2. From data management to statistical treatment of fingerprints	118
3.2.1. Design of sensing experiments	119
3.2.2. Management of generated fluorescence response data	122
3.2.3. Multivariate statistical analysis of fingerprint response data	124
3.3. Conclusions	137

Chapter 4

Optical fingerprinting for diagnostic applications: Proteins and disease models

4.1. Introduction	141
4.2. Results and Discussion: Protein fingerprinting	142

4.2.1. Sensitivity of TPA-CB[7] sensor array toward proteins	142
4.2.2. Discrimination of proteins in physiological buffer	145
4.2.3. Discrimination of proteins in complex matrix: serum and treated serum	151
4.3. Fingerprinting of disease models: Phenylketonuria	155
4.3.1. Phenylketonuria: State of the art diagnosis	155
4.3.2. Setup of PKU sample cohort using artificial urine diluent	157
4.3.3. Optical fingerprinting of L-phenylalanine in artificial urine	158
4.4. Fingerprinting of disease model: Preeclampsia	158
4.4.1. Preeclampsia: State of the art diagnosis	160
4.4.2. Fingerprint of PE samples by four TPA with CB[7]	161
4.5. Conclusion	165
4.6. Materials and Methods	166
4.6.1. Physiochemical properties of protein used in the discrimination study	166
4.6.2. Protocol for discrimination of spiked proteins (PBS, serum, treated serum)	167
4.6.3. Depletion of human serum	167
4.6.4. Preeclampsia: Clinical data of patient samples from APHERESE cohort	168
4.6.5. Ethical Declaration and biological samples	169

Chapter 5

Optimization of analytical platform for throughput analysis: use of droplet microfluidic platform and pipetting robot

5.1. Introduction	173
5.2. Results and Discussion: Droplet-based microfluidic platform	173
5.2.1. Experimental set-up of droplet microfluidic platform	174
5.2.2. Optimization of the microfluidic platform for the TPA-CB[n] working model	176
5.2.3. Optimization of protein-sensing element interactions	179
5.2.4. Discrimination of proteins on droplet microfluidic platform	183
5.2.5. Development of microfluidic chip for improvement of droplet merging strategy	185
5.3. Pipetting robot platform	188
5.3.1. Protein Discrimination using the TPA-CB[7] sensor array with pipetting robot platform	189
5.4. Conclusion	191

5.5. Materials and Methods	192
5.5.1. Analysis of fluorescence fingerprint generated by the droplet microfluidic platform	192

Chapter 6

Host-guest based colorimetric sensor array for discrimination of pharmaceutical compounds

6.1. Introduction	199
6.2. Results and Discussion: TPA-CB[n] colorimetric sensor array	200
6.2.1. Characterization of TPA-CB[n] complexes	200
6.2.2. Optical fingerprinting by TPA-CB[n] sensor array	207
6.3. Conclusion	216
6.4. Materials and Methods	217
6.4.1. Protocol for discrimination of analytes by colorimetric sensor array	217
6.4.2. R codes for LDA-PCA routine	217

Chapter 7

Macrocyclic fluorescence-based sensor array:

A methodological study for the diagnosis of the SARS-CoV-2 infection in human serum

7.1. Presentation of the study	221
7.1.1 Abstract	221
7.1.2. Covalently linked vs. Host-guest based sensing strategy	222
7.1.3. Background of developed CB[7]-fluorophore sensor array	223
7.1.4. Diagnostic application of CB[7]-FL sensor array	224
7.1.5. Publication (<i>Sensors and Diagnostics</i> , 2024)	226

Chapter 8

General Discussion: Future outlook and perspectives

8.1 Introduction	251
8.2. Comprehensive overview of the project	251
8.3. Future perspectives of the project	254
8.4. Conclusions and future directions	255

Appendix

Résumé	261
<i>Chapter 2</i>	267

Appendix 1: Characterization of TPA derivatives	267
<i>Chapter 4</i>	282
Appendix 2: R-code for analysis of 14 proteins in PBS, serum and treated serum	282
Appendix 3: R-code for analysis of different concentrations for PKU model	284
Appendix 4: R-code for analysis of different concentrations of PE samples	286
<i>Chapter 5</i>	288
Appendix 5: R-code for analysis of proteins on droplet microfluidics platform	288
Appendix 6: Protocol for sensing of protein on 384-well plate using the pipetting protocol	289
<i>Chapter 6</i>	296
Appendix 7: R-code for PCA analysis of 8 analytes with 11-channel sensor array	296
Appendix 8: R-code for LDA analysis of 8 analytes with 11-channel sensor array	296
Appendix 9: Optimization routine using PCA and LDA for biomolecule discrimination with sensors	297
References	301

Chapter 1

***Introduction to supramolecular host-guest systems and their utility
in array-based sensing for biomedical applications***

1.1. Sensing systems in nature

Nature serves as a remarkable model of inspiration when it comes to sensory systems and their associated mechanisms. These systems have empowered living organisms to perceive and process variations in their environment, thereby providing an estimate of vital information regarding their surroundings by responsive reactions to stimuli¹. The varieties of sensory responses allow such active systems to detect, discriminate and localize environmental objects in addition to assisting in communication amongst species members. For instance, the mechanism of echolocation in nocturnal animals allows them to navigate and hunt in conditions of darkness by emitting and listening for echoes of ultrasonic waves². In humans, proprioception, often termed as the sixth sense, allows for individuals to perceive position and movement of body parts, crucial for coordinated movement and overall balance³. Such responses are essential for the performance and maintenance of life sustaining activities. Biological sensor systems are therefore highly sensitive and no organism can be imagined to function effectively without them. Each of these systems has evolved and continues to adapt to fulfill different purposes based on exposure to specific environments. For example, some animals have developed over time, defense mechanisms to sense and protect themselves from threats, such as the chameleon's ability to camouflage or the emission of fluorescence by soft-bodied organisms like octopuses⁴. For years, scientists have investigated these systems to gain essential insights into the underlying mechanism of sensory uptake, transformation and transduction⁴. These sensors although small, are energy efficient and highly sensitive in nature. They exhibit a nature of redundancy, where numerous sensory receptor organs operate in parallel, each comprising dozens of receptor cells, such as the photoreceptor rod and cone cells of the human retina, that facilitate vision. This parallel sampling and processing of sensory information serves to enhance the signal-to-noise ratio by averaging and decreases the risk of errors resulting from the malfunction or loss of individual sensory elements⁵. The study of these systems by scientists has also helped to uncover various impressive physical principles that have provided rich inspiration for the design of artificial sensors for revolutionizing various scientific fields, from medical diagnosis to environmental monitoring and beyond.

Humans are recognized for possessing multiple sensory tools that govern their five primary senses — sight (vision), hearing (audition), taste (gustation), smell (olfaction) and touch (somatosensation). Among these, humans being visual beings tend to prioritize their sense of sight over all other senses; a recent quantitative study of people aged 16-30 revealed that slightly over half of the participants would choose to give up their sense of smell to keep a piece of technology⁶. However, their sense of smell is perhaps the most powerful ability that they rely on for their safety and health and those who lose this sense, often realize its crucial role in daily functioning. The human olfactory system demonstrates remarkable efficiency in recognizing and distinguishing thousands of distinct odors, including both individual and intricate combinations, mere seconds after exposure. This system exhibits a dynamic range spanning from high concentrations down to the parts-per-billion range. The specificity of the

olfactory system, as compared to other sensory systems in nature, does not arise from specific receptors for specific analytes, resembling the traditional lock-and-key model of substrate-enzyme interactions. Instead, it relies on the pattern recognition of combined responses gathered from several hundred olfactory receptors that can bind numerous odorants with specific affinities. While mammalian genetics encodes for approximately 1000 different types of olfactory receptors, with humans having ~400 active receptors⁷, the system is still capable of discrimination of thousands of scents by using each of these receptors in a cross-reactive manner. The integration of electric responses sent to the brain generates a highly complex, multidimensional pattern that encodes the identity of an odorant.

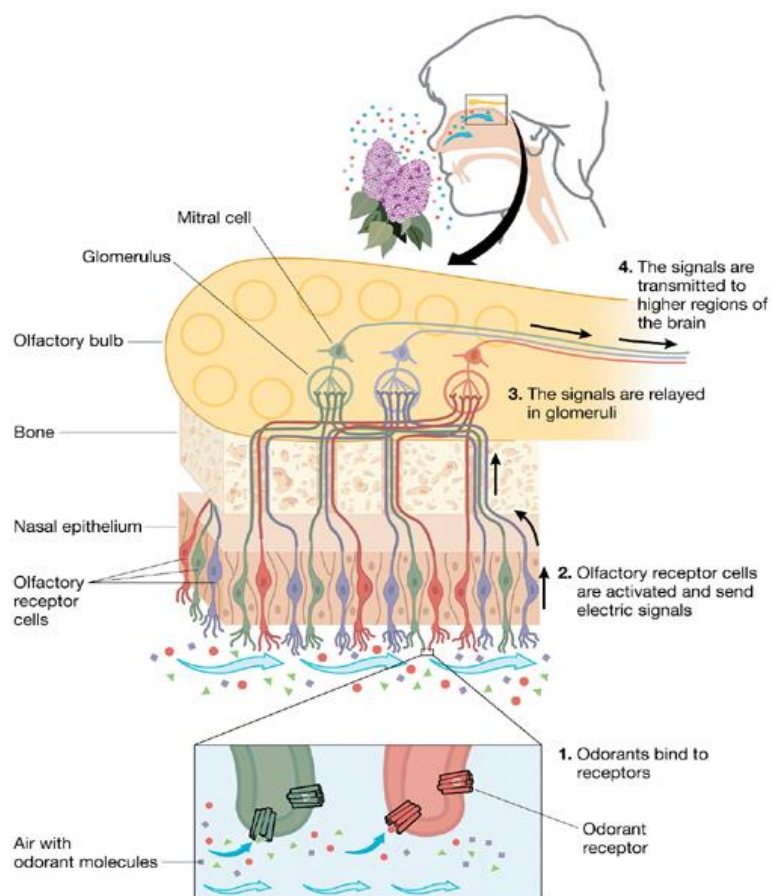


Figure 1.1: Structure and mechanism of transmission of olfactory information to the brain in the human olfactory system (adapted from Rinaldi, A.⁸).

The olfaction process firstly begins with a regulated “sniffing step” to prime itself for detecting odors. During the sniffing step, the low molecular weight volatile organic compounds (VOCs) that constitute the odorous molecules are inhaled into the olfactory epithelium and partitioned into the mucus layer, enabled by the fast and turbulent air flow generated by the normal average breathing rate. The mucous layer covering the epithelium protects the sensory olfactory receptor cells (ORCs). Notably, as most odorants are hydrophobic, the transport of molecules through aqueous mucus may necessitate small, low-molecular weight soluble proteins called odorant binding proteins (OBPs). The ORCs

possess transducer bipolar neurons consisting of dendrites at one end (in the epithelium) and axons at the other (in the olfactory bulb). The dendritic end of the ORCs hosts non-motile cilia that contain the olfactory receptors (ORs), which are seven-domain transmembrane receptors, also referred to as G-protein coupled receptors. The binding of odorant molecules to the ORs triggers inwardly depolarizing currents that activates the action potentials of the receptor cells. All the cells containing each receptor type converge on the olfactory bulb of the brain, boosting the signal-to-noise ratio through signal summation. These signals are then transmitted to the higher-level processing centers in the brain (olfactory cortex), where they are analyzed to enable identification of the encountered odorant (Figure 1.1). The brain does not interpret the singular response of an olfactory receptor, but rather the pattern arising from their collective interaction with the VOCs in order to discern an odor. This necessitates a training phase to associate a detection pattern to an object, so that it is perceived and memorized to enable the persistent discrimination of the multiple odors over time⁹.

Therefore, the mammalian olfactory system, in its entirety, is a highly complex yet remarkably sensitive and efficient system; and scientists have devoted extensive efforts over many years to mimic its capabilities in artificial systems. This has led to the development of artificial olfaction-based sensing systems, known as ‘electronic noses’. These systems replicate the selective combinatorial abilities of the human olfactory system using an array of distinct chemical materials that physically adsorb gaseous analytes and undergo modulations in their electrical properties upon exposure to a vapor containing a mixture of entities. The first artificial olfaction based system was developed by Persaud & Dodd in 1982, who constructed an electronic nose with three metal oxide semiconductor transducers to sense and discriminate among similar odor mixtures based on steady-state sensor signals¹⁰. While the performance of such sensors are primarily judged by quantitative properties such as sensitivity, resolution, reproducibility and in some cases reversibility, the most crucial aspect remains to be sensor selectivity, which ensures detection of the target analyte amidst other interfering chemical species. Since then, significant advancements have been made in this field to harness the efficiency of nature’s system in artificial sensor designs, not only in gaseous conditions but also in solution phase. This progress has led to the introduction of the terms ‘chemical nose’ and ‘chemical tongue’ systems, with the former primarily facilitating detection in the gas phase, while also being adapted to the solution phase and the latter solely designed for detection in the liquid phase only. The common foundational principle underlying the design of these differential sensing systems resulted in this methodology being referred to as ‘array-based sensing’.

1.2. Array-based sensing

The ‘array-based sensing’ approach mirrors the fundamental principles of the human olfactory system, employing a set of cross-reactive, non-specific sensors that interact individually with multiple analytes, and through an output signal proportional to the interaction, generate a unique pattern of

response for each analyte. Consequently, this multidimensional pattern enables molecular recognition by comparison to a predetermined library of responses facilitated by pattern recognition. In contrast to the traditional lock-and-key design criteria that is adopted in antibody and aptamer-based systems, this approach adopts a more flexible, hypothesis free, differential and non-specific design strategy, allowing for maximum chemical diversity with minimal number of sensor elements within the array to detect a wide cross-section of analytes (Figure 1.2). These pattern recognition-based sensors are broadly referred to as chemical nose, chemical tongue or E-tongue systems¹¹.

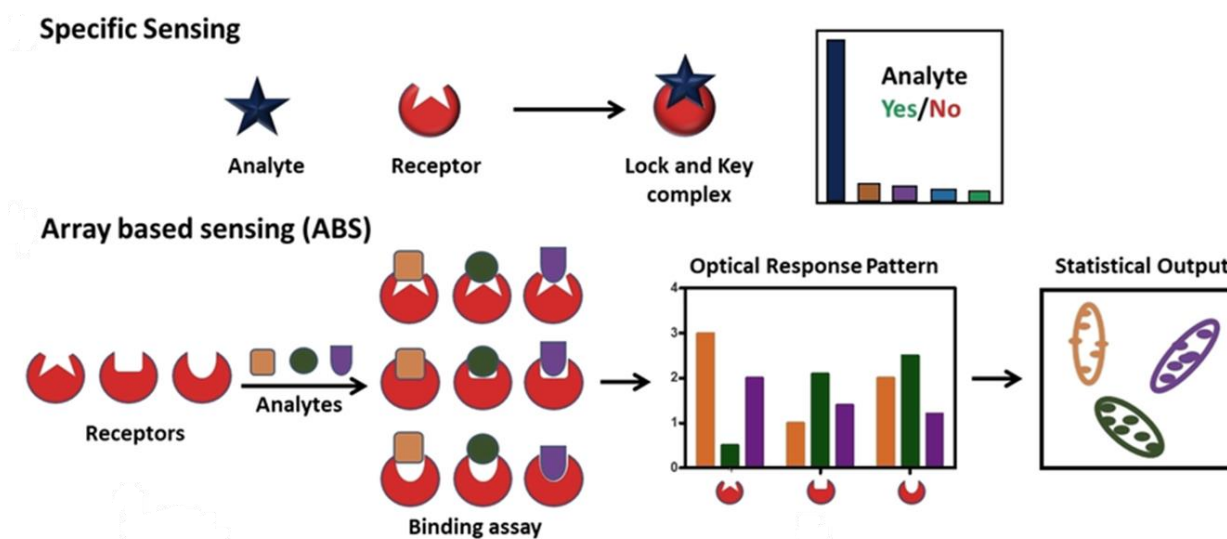


Figure 1.2: Schematic representation of specific, lock-and-key sensing and differential array-based sensing (adapted from Behera, P. et al.¹²).

The chemical nose sensors, operating on the principle of a lock fitting a number of imperfect keys, are often susceptible to interference from chemicals that are either structurally or chemically similar to the target analytes. It is precisely this interference or cross-reactivity that these types of differential sensors exploit to generate a unique fingerprint that allows the identification and classification of different analytes. Consequently, the elements of the array are not required to exhibit individual selectivity (i.e. ability to detect one analyte in the presence of other analytes in the sample) towards a given analyte and this flexibility enables the sensing of complex analytes or analyte mixtures that have not been exhaustively characterized.

The chemical nose sensor array is assembled by combining multiple sensor elements. Each sensor element comprises of two components that are responsible for two essential processes involved in the sensing: the *recognition element* and the *transduction element* (Figure 1.3). The recognition element binds with the target analytes through nonspecific and cross-reactive interactions. Various weak forces, such as electrostatic, hydrophobic, π - π , hydrogen bonding and van der Waals interactions, facilitate the recognition element to obtain combinatorial information on all the target analytes¹³. It is also imperative that this element does not rely on specific binding sites, but instead offers a high density of multiple

cross-reactive binding sites for interaction with the analyte, particularly in the case of large and conformationally dynamic analytes. This allows for selectively extracting maximal information from such analytes while ignoring the effect of small molecules that are not stable in concentration, such as salts, thus generating unique fingerprints of these complex analytes¹⁴. Moreover, the chemical stability of the recognition element is required to achieve reproducible response patterns with the sensor array. The transduction element transforms these binding events between the recognition element and the analytes into measurable output signals and are typically classified as optical or non-optical in nature¹³. While the optical sensors encompass fluorometric, colorimetric, chemiluminescent, scattering-based or refraction-based transducers, the non-optical sensors include thermal, mass, potentiometric, amperometric or conductometric sensors, among others^{15,16}.

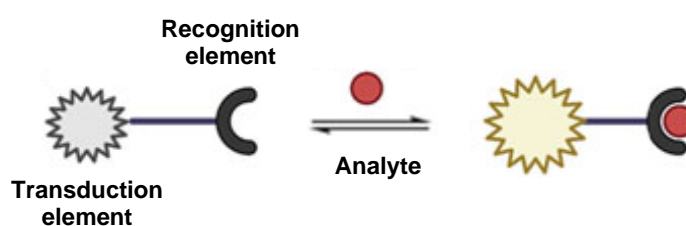


Figure 1.3: Schematic representation of individual sensing elements forming the sensor array, with recognition and transduction elements (adapted from Lim, S. et al.¹⁷).

1.2.1. Recognition elements for array-based sensing

Chemical nose sensors interact with the surrounding media by creating an interface between target analytes and the recognition elements of the array. Careful consideration is essential when selecting suitable materials for the recognition elements. While purely biological elements may suffer from instability under ambient sensing conditions, synthetic organic and inorganic materials offer greater stability for the constructed array. Below are the different groups of recognition elements described in the literature, such as nanomaterials, macrocyclic molecules, supramolecular assemblies, polymeric networks etc.

1) **Nanomaterials**: They offer a dual advantage when used as recognition elements; they possess diverse surface properties to interact with a wide range of samples, and they can also be surface-modified through fabrication with external recognition units to facilitate binding with target analytes.¹² Nanomaterials, available in various sizes, shapes and functionalities, have a high surface-to-volume ratio, providing more sites for interaction with analytes with requiring less sensor material and enhancing the sensitivity of the sensing process¹⁸. The most advantageous property of these materials is their tunable core properties, achievable through surface engineering techniques applied during nanomaterial synthesis, such as adhering, capping, stabilizing, or doping. Additionally, surface

functionalization can be introduced post-synthesis via covalent or non-covalent interactions to obtain divergent surface properties. Furthermore, the intrinsic optoelectronic properties of the nanomaterials can be exploited to function as the transducer of the sensor element, providing electromagnetic outputs that are induced or modified upon analyte binding. For instance, the variation in plasmonic properties of AuNPs can be directly used as a sensor readout¹⁹. Changes in the size and aggregation properties of AuNPs upon analyte binding result in shifts in plasmon band positions, while the tunable fluorescence emission of metal nanoclusters²⁰ and quantum dots^{21,22} can also serve as effective readouts. AuNPs have also been combined with DNA, fluorophores, and enzymes and the promotion of aggregation of AuNP or their displacement from the AuNP, triggers signals that provide unique patterns for protein identification^{23,24,25,26}.

2) **Conjugated polymers:** Polymeric recognition elements have been extensively used in the development of array-based sensors due to their robustness and versatility. Controlled synthesis using diverse monomeric units allows for variations in composition, dimension, and molecular architecture²⁷. This synthetic control facilitates easy functionalization with transducer elements like fluorophores and enables fine-tuning of hydrophobic and electrostatic properties for analyte interactions and control over water solubility. Their highly delocalized structures and unique optical and electronic properties enable efficient energy transfer processes, amplifying signals more effectively than small molecule fluorophores. Their sensitivity towards conformational and environmental changes makes them ideal for detecting subtle variations upon binding to different target analytes. Additionally, tunable polymer size allows for the creation of optimally-sized recognition surfaces with multiple interaction sites, ensuring the necessary cross-reactivity for analyte discrimination²⁸. Conjugated polymers are particularly appealing in biological sensing as they enable multiple recognition events over large areas, such as on proteins, bacteria, and cell surfaces. These multivalent interactions in polymer design open opportunities for inexpensive diagnostic tools for biological macromolecules. The most important polymers for sensor array applications are poly(para-phenylene ether) PPEs and poly(para-aryleneethynylene) (PAEs)^{29,30}, polyfluorenes^{31,32}, polythiophenes^{33,34} and mixed co-polymers of these polymer types³⁵.

3) **Macrocyclic molecules:** A wide array of synthetic macrocyclic molecules including crown ether, cyclodextrins, calixarenes, cavitands, pillarenes and cucurbiturils, have been synthesized and investigated extensively. These macrocyclic receptors feature hydrophobic cavities of varying molecular dimensions, serving as recognition sites for a large and diverse group of target analytes. They interact with the analytes via intermolecular forces such as hydrogen bonds, salt bridges, π - π stacking, van der Waals forces, and hydrophobic interactions³⁶. These molecular receptors are produced using highly controlled synthesis protocols, resulting in much more reproducible production compared to other recognition elements such as polymers and nanoparticles which are not highly stable and are

highly heterogeneous in nature. Furthermore, their synthetic versatility and strong binding affinity for a broad spectrum of analytes, ranging from basic metal ions to complex organic molecules, have been extensively adopted for various sensing applications. Their rigid architectures with distinct conformational preferences like the π -electron-rich cavities, and separate ion-binding sites, provide possibilities for derivatization of these systems with suitable transducers for the development of the sensor arrays.

1.2.2. Transduction mechanisms for array-based sensing

The transduction element in array-based sensing converts interaction strengths between recognition elements and target analytes into measurable signals, enabling the readout of the sensing process. These elements operate on various principles, including optical, electrochemical, and thermometric methods. The choice of transduction mechanism significantly impacts the sensitivity, selectivity, and overall performance of the sensor array. Keeping in line with the scope of this thesis, optical transducers will be discussed in details. There are a wide range of optical transduction mechanisms, including absorbance, fluorescence, diffusion, scattering, refraction, chemiluminescence or photoluminescence. Different regions of the electromagnetic spectrum are employed with different transduction techniques, and numerous parameters are assessed for each method, encompassing intensity, lifetime, polarization, quantum yield and quenching efficiency.

Colorimetric: It involves quantitatively measuring absorbance or reflectance, allowing for the easy visual observation of output signals. Traditional sensors typically utilize the three- channel visible range of red, green and blue (RGB). However, recent sensors have also adopted the use of larger number of channels within a narrower spectral range, often incorporating infrared and ultraviolet regions of the spectrum. Nonetheless, this method is primarily used by analyzing the discreet RGB regions or selecting the highest absorbance peaks in UV-visible spectra.

Fluorescence: This mechanism has emerged as the most widely adopted method primarily due to its higher sensitivity as compared to other methods. As a result, even small quantities of the sensor are sufficient for analysis, which is highly advantageous when working with valuable analyte samples for clinical applications. It encompasses a range of measurable parameters such as intensity, anisotropy, lifetime, emission and excitation spectra, fluorescence decay and quantum yield, providing substantial flexibility in adopting this optical method as a transduction mechanism¹⁶. Most commonly the quenching of intrinsically fluorescent molecules in the presence of an analyte is measured as the sensing output signal. In fluorescence, the measurement of output light is made, thereby providing a better signal-to-noise ratio as compared to absorbance measurements where the output is based on the ratio of light absorption. However, when an analyte quenches the fluorescence of the sensor, subtracting the large background signal of the fluorescent molecule in the absence of the analyte may result in low

sensitivity, particularly if the quenching effect is minimal. In array methods, this substantial background is usually less of an issue because the practical comparison is often made between the fluorescence responses of different analytes to the sensor, rather than solely comparing the fluorescence response of the analyte to the sensor's emission alone. Nevertheless, turn-on fluorescence approaches are a more preferable approach, where the emission intensity of a fluorescent molecule increases in the presence of an analyte. The measurement of the signal relative to a dark background provides a much better sensitivity³⁷. Due to the limited availability of commercial fluorescent dyes, designing fluorescent sensor arrays involves leveraging various parameters from fluorescence spectra, such as peak shifts and/or spectral shape changes. Additionally, selecting multiple wavelengths from a single sensing element is employed to increase the number of dimensions provided by the sensing elements.

Fluorescent solution-based sensor arrays adopt the use of the microtiter plates (96 wells or more per plate) along with spectrophotometric plate readers for high-throughput signal quantification. These plate readers are equipped with optical components that deliver excitation light of a specific wavelengths to each sample well and detect the specific wavelengths of the light emitted from each sample well. These convenient setup conditions further make such fluorescent assays a preferred choice for the design of chemical sensor arrays¹³.

Surface Enhanced Raman Scattering (SERS): Vibrational spectroscopy is a valuable transduction mechanism for organic analytes, as they typically exhibit infra-red and/or Raman signals, which often serve as characteristic spectroscopic fingerprints for different analytes³⁸. However, the low efficiency of inelastic photon scattering often results in weak signals, limiting its direct application in analyte detection³⁹. The presence of nanoparticles or rough surfaces enhances the Raman scattering by the creation of a local electromagnetic field, whose enhancement is induced by the plasmon resonance of the surfaces referred to as SERS^{40,41}. The enhancement of the Raman signal is dependent on the size, shape, orientation and aggregation state of the nanoparticle. The significant enhancement in detectable signal, coupled with the unique molecular fingerprints generated, has made SERS a powerful tool for the multiplex detection of analytes in array-based sensing.

1.2.3. Design criteria for array-based sensing

Selection of the appropriate combination of recognition and transduction elements is essential for addressing various applications, particularly for creating sensitive systems aimed at rapid analyte identification and efficient recognition in challenging and complex media. These elements can be combined by covalent bonding or by weak non-covalent interactions, each with its own set of advantages and disadvantages. Sensor designs that employ covalent bonding offer enhanced stability, increased sensitivity by minimization of signal loss during transduction and an improved signal-to-noise ratio owing to reduced chances of non-specific interactions that cause background noise. They also demonstrate efficient detection across different analyte concentration ranges with minimal effects on

performance upon dilution. On the other hand, utilizing weak non-covalent interactions provides flexibility in independently optimizing the recognition and transduction elements, allowing for sensor development tailored to different analytical needs. In many such configurations, the transducer can be reused with different recognition elements, making this approach cost-effective and adaptable for different target analytes. Additionally, the dynamic range of the sensor can be expanded by varying the affinity of the transducer with respect to different target analytes. This ultimately also allows improvements to be made iteratively to the design of the sensor with great ease and contributes to the modular nature of such sensor arrays. Since the overall array is highly cross-reactive in nature, individual sensor elements may often generate very low output information. Therefore, simultaneous analysis by multiple sensor elements becomes crucial to provide a differential response, but this could

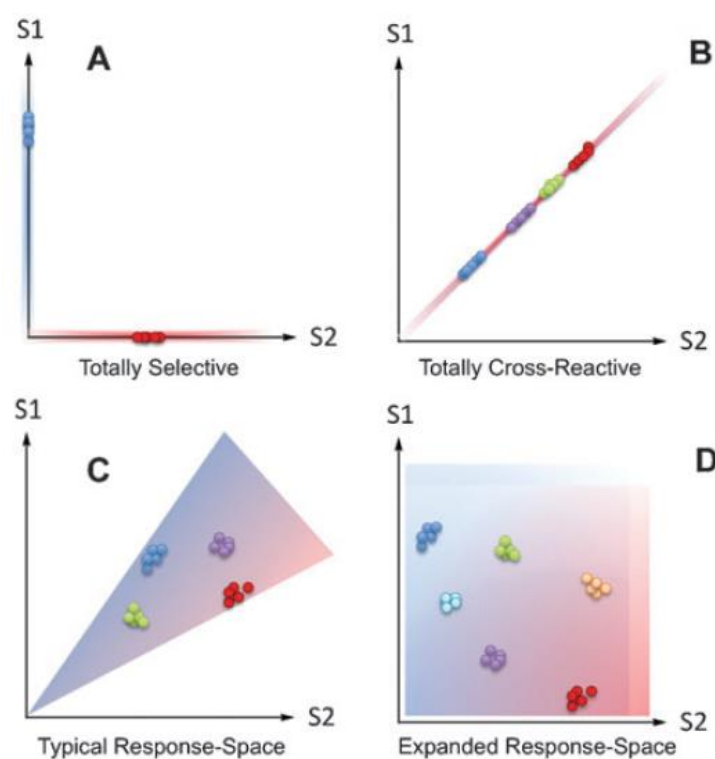


Figure 1.4: Schematic representation of the response-space of two sensors $S1$ and $S2$. (A) $S1$ and $S2$ are 100% selective. (B) $S1$ and $S2$ are 100% cross-reactive to a group of analytes. (C) $S1$ and $S2$ are cross-reactive, but have certain selectivity towards some analytes. (D) $S1$ and $S2$ are cross-reactive and have enhanced selectivity towards key analytes, hence expanding the potential response-space (adapted from Anzenbacher, P. et al. ⁴²).

invariably lead to noise amplification in practical sensing systems. This underscores the importance of finding a middle ground between selective and cross-reactive sensor arrays. Anzenbacher *et al.*, has very clearly addressed this issue with reference to the response space generated by a hypothetical sensor array with two sensors under different conditions. While complete selectivity achieves space resolution,

it restricts the ability of the two sensor elements to respond only to the analytes they are selective for. In the case of complete cross-reactivity, multiple analytes can be detected, but the associated lack of resolution makes data extraction difficult. However, the intermediate condition where the two sensors are cross-reactive in their response towards certain analytes with a certain extent of bias towards some key components, leads to the increase in the space resolution of the array (Figure 1.4). Hence, this semi-selectivity of the sensing elements could potentially be the key component that increases the discriminatory power of the sensor array and allows a significant reduction in the number of sensing elements comprising the array⁴². Thus, the optimal number of sensing elements in an array is typically determined by striking a balance between acquiring sufficient information and maintaining redundancy to minimize noise, and avoiding overfitting during statistical analysis.

1.2.4. Pattern-recognition by multivariate statistical analysis

The interaction between sensing elements and analytes generates fingerprint patterns unique to each analyte and the core of an array-based sensor lies in converting these patterns into interpretable data. To achieve this, multivariate statistical analysis tools that systemically identify trends and predictability in data have been used. In general, there are two classes of statistical methods that are adopted for the analysis of cross-reactive sensor arrays: descriptive and classification methods²⁸.

Descriptive analysis employs unsupervised techniques, where the class of the analyte being detected (such as a protein, a cell line, or a disease state) is not specified and the analysis methodology is used to identify trends and extract elements that cluster the data as effectively as possible. This technique is unbiased towards the analyte class and therefore employs simple algorithms to describe general trends in the data. It provides a qualitative evaluation of the extent of separation between the classes and the relationship among the different classes identify redundant sensing elements. This type includes the Principal Component Analysis (PCA) and Hierarchical Clustering Analysis (HCA) methodologies. While this approach is useful for evaluating overall trends in the dataset, it is less effective for making predictions, thus limiting the sensor's ability to predict unknown analyte patterns.

In contrast, classification analysis uses supervised techniques that employ provided analyte class information to distinguish the fingerprint data according to these classes. An algorithm is developed from the primary “training” dataset, which consists of samples with known class, which is then tested on an independent secondary “test” dataset, where the class information is withheld from the model. The success in classifying this second dataset indicates the model's accuracy in identifying unknown samples and is referred to as the “classification accuracy”¹¹ The most common type is the Linear Discriminant Analysis (LDA) methodology.

Thus, in general classification techniques are supplemented by initial descriptive analysis so as to optimize the sensor array before performing the classification of the tested analytes. The most

commonly utilized statistical techniques for the analysis of array-based sensing data are discussed below³⁷:

Principal Component Analysis (PCA):

This method reduces the dimensionality of a dataset by consolidating the data into eigenvalues and eigenvectors. The eigenvalues represent the variance, with the first principal component (PC) showing the maximum variance, followed by others with decreasing variance. Consequently, PCA consolidates the dataset into a lower-dimensional space and ranks the new dimensions by importance. Plots of the first few PCs facilitate easy visualization and interpretation of the high dimensional datasets¹⁵. Thus, this method is used to evaluate the importance of different receptors in the assay. A visual inspection of the PCA plot enables the detection of close clustering between data points that represent the same analyte, but also presents good separation between data points that represent a different analyte class (Figure 1.5).

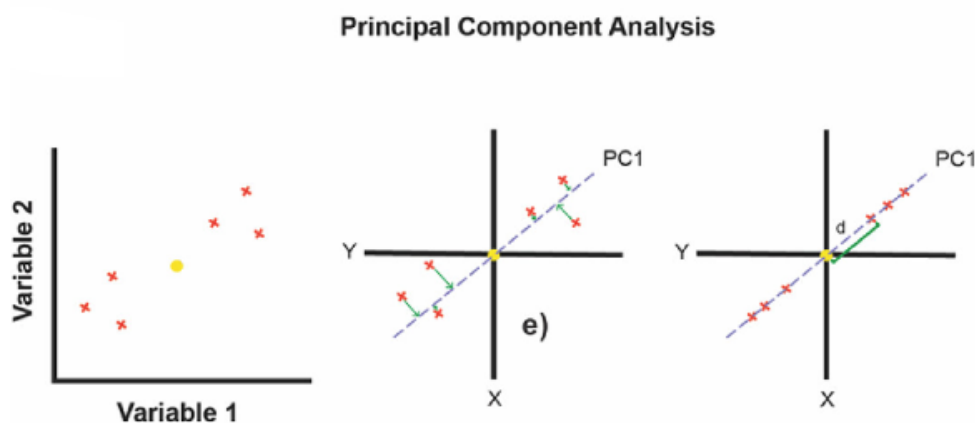


Figure 1.5: Schematic representation of PCA method of determining the center of the sensor data, projecting points onto a new vector, and calculating the maximum variance and thus best-fitting line, (adapted from Mitchell, L. et al.¹²).

Hierarchical Clustering Analysis (HCA):

This method clusters data points based on their relative distances, often using the Euclidean distance metric, which measures the distance between two data points in N-dimensions (N being the number of different sensor responses). In chemical array analysis, HCA groups analytes hierarchically based on differences in response patterns. This can be achieved through an agglomerative ("bottom up") approach, starting with each data point as its own cluster and merging similar clusters (Figure 1.6), or a divisive ("top down") approach, starting with all data points in one cluster and splitting them until each data point is isolated. This results in a graphical representation in the form a dendrogram. When using this method, the researcher must interpret the clustering significance based on the sensing

application, as the adopted algorithm is only capable of either merging or separating the data into clusters based on whether the approach adopted is agglomerative or divisive respectively.

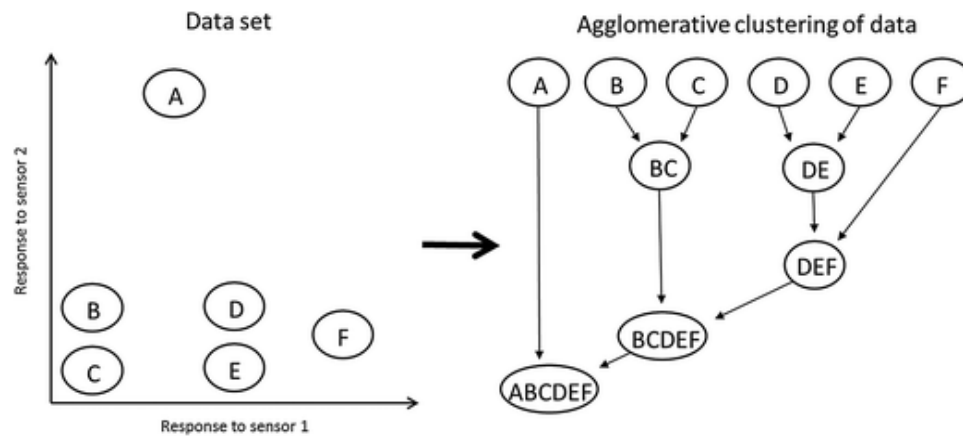


Figure 1.6: Agglomerative clustering uses the Euclidean or simple distance between data points in hierarchical clustering analysis as the basis for grouping the data points (adapted from Diehl, K. L. et al.³⁷).

Linear Discriminant Analysis (LDA):

This method is used to classify data and assign unknown analytes to their appropriate classes. The method uses array data and analyte classes as inputs to calculate discriminant functions that maximize separation between classes while minimizing within-classes separation (Figure 1.7). A successful LDA plot shows good separation between different classes. For prediction, a training set with known analyte responses is provided to the algorithm. The algorithm learns from this set, enabling it to assign unknown analytes to classes based on response similarity to the training set, thereby determining the array's predictive power.

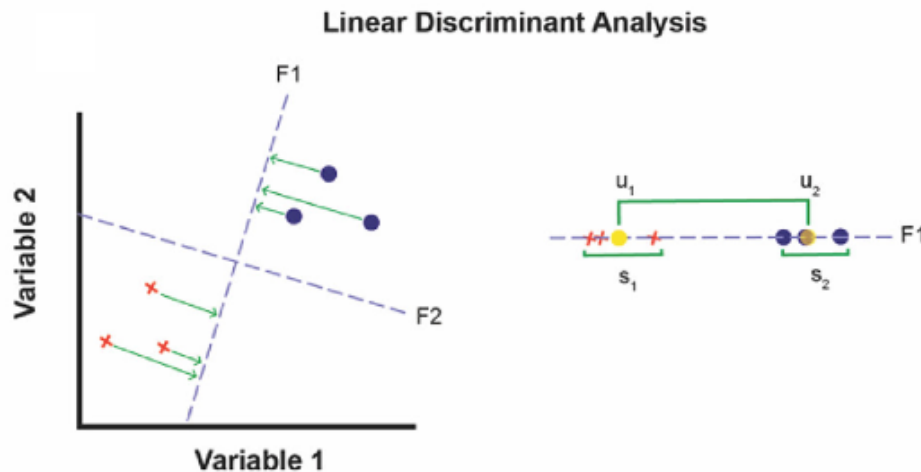


Figure 1.7: Schematic representation of LDA method of projecting points onto a new vector $F1$ that fulfills the criteria of maximizing the ratio of between-class to within-class variance (adapted from Mitchell, L. et al.¹²)

Artificial Neural Network: This technique consists of layers that can be trained to produce a desired response pattern. It includes an input layer that is data from the array (e.g., fluorescence counts or RGB color values), one or more hidden layers which are user determined and are adjusted to suit the requirement of the sensor system, and an output layer which provides the desired analysis of the system (e.g., analyte identification). The training process involves the hidden layers which are adjusted to maximize the accuracy of the desired output from the given inputs. Once trained, the ANN can predict the identity of an unknown analyte based on its data set. An example of ANN includes support vector machines (SVM).

Decision Trees and Random Forest Model: Decision trees are a supervised classification technique that mimics a tree structure, where the root and nodes serve as branching points. In decision trees, nodes correspond to specific attributes in a dataset, while the branches represent different value ranges for those attributes. These value ranges function as partitioning criteria, dividing the dataset based on the characteristic represented by each node. Decision trees are effective for handling non-linear relationships and interactions between variables, making them particularly useful in complex classification tasks. However, they are prone to overfitting, especially with noisy or small datasets⁴³.

To address the limitations of single decision trees, random forest model was developed, which combine multiple decision trees, each built on different subsets of the data and features. This ensemble method reduces overfitting and improves classification accuracy and robustness. In a random forest, each tree contributes to the classification, with the final class determined by majority vote. This method increases the model's stability and generalizability, making it a dependable choice in discriminant analysis. While

decision trees offer a simple solution for discriminant analysis, random forests provide a more robust and precise alternative by combining the power of multiple decision trees.

1.3. Supramolecular macrocyclic systems for sensing

The field of Supramolecular Chemistry gained prominence following the Nobel Prize in Chemistry awarded to Lehn, Cram and Pederson in 1987, for their contributions in the development of structure-specific interactions of high selectivity⁴⁴, advancing chemistry beyond the molecule. While the use of non-covalent interactions for sensing predates the formal establishment of supramolecular chemistry, the discovery and exploration of this field has further accelerated the exploration of supramolecular-based sensing systems. This led to the emergence of *recognition-based sensing* strategies that utilize non-covalent interactions and molecular recognition for sensor and analyte interaction. The *recognition-based sensors* are further categorized based on the binding mode between the receptor (host) and the indicator (guest), which could be covalently or non-covalently linked. While the former involves a covalent connection between the indicators and the receptor, wherein analyte binding to the receptor results in output signal change, the latter relies on host-guest interactions among the receptor, the indicator, and the analyte, and embodies the popular chemical sensing method of indicator displacement assays (IDA)⁴⁵. A variety of macrocyclic synthetic receptors including crown ether, cyclodextrins, calixarenes, cavitands, pillarenes and cucurbiturils, have been adopted for this purpose. The weak non-covalent interactions associated with these systems impart the feature of reversibility, allowing convenient dissociation and reassembly of the supramolecular systems with minimal energy expenditure³⁶. These systems in most cases are chemically stable and their cavity size, shape, number and directionality of interactions can be tailored in accordance with the supramolecular principles of complementarity, enabling structural recognition and preorganization for enhanced binding capabilities⁴⁶. These molecular receptors have been extensively combined with fluorescent indicators due to the high sensitivity, convenience and rapid response time for real-time detection associated with fluorescence spectroscopy⁴⁷.

1.3.1. Structural properties of common macrocyclic host systems

While numerous macrocyclic compounds have emerged since the first synthesis of crown ethers by Charles Petersen in 1967, four of these compounds have been extensively investigated: cyclodextrins (CDs) as natural macrocycles, and the other three artificial macrocycles, namely calixarenes (C[n]As), cucurbiturils (CB[n]s) and pillarenes. These adaptable compounds show considerable potential in various fields such as molecular separation, catalysis, sensing, disease diagnosis and drug delivery⁴⁸⁻⁵⁰. Particularly, they have gained considerable popularity in the biomedical field on account of their excellent biocompatibility and capacity for designing stimuli-responsive supramolecular systems.^{49,51-53}

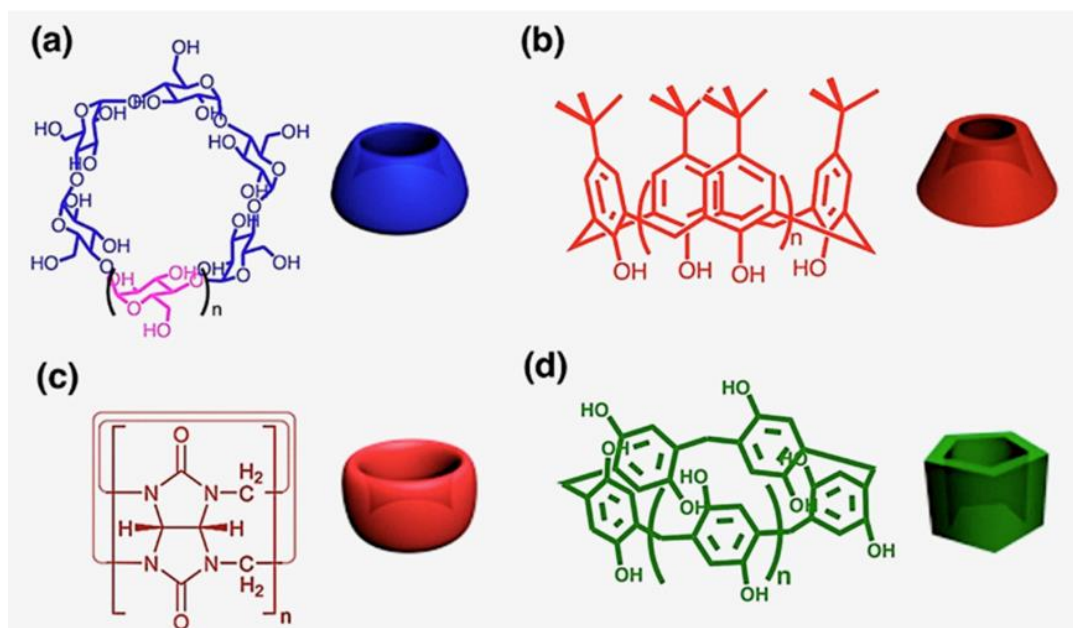


Figure 1.8: Chemical structures and structural illustration of common macrocyclic hosts: a) cyclodextrins, $n = 1-3$, b) calixarenes, $n = 1-3$, c) cucurbiturils, $n = 5-8, 10, 13-15$, d) pillararenes, $n = 1-11$ (adapted from Lou, X.-Y et al.⁵⁴).

The properties of the four most commonly investigated macrocyclic systems are given below^{45,48}:

a) Cyclodextrins (CDs)

This series of macrocyclic molecules, first isolated by Villiers in 1891 from bacterial starch digestion products⁵⁵, comprises α -1,4 glycosidic bond linked oligosaccharides. They are produced in significant quantities through enzymatic degradation of α -glucosidase and α -amylase in starch precursors like potato, rice and corn, making them cost-effective for various applications. The most commonly used cyclodextrins (CDs) are α -, β -, and γ -CD, containing 6, 7, and 8 α -d-glucopyranose units, respectively. They have a truncated cone structure with two sides (primary & secondary) and a hollow cavity. The cavity size depends on the number of glucopyranose units, while all forms have a consistent depth of 0.78 nm with the hydroxyl groups of the glucopyranose units oriented outwards, making them highly hydrophilic. However, their hydrophobic interior cavity is covered by glycosidic oxygen and methinic C-H units, facilitating the complexation of non-polar molecules through hydrophobic and van der Waals interactions. This enables encapsulation of guest molecules in aqueous conditions. Their natural availability confers them with good water solubility, bioavailability, biocompatibility and non-toxicity towards biological systems,^{56,57} leading to extensive research on the preparation of CD-based supramolecular systems for biological applications.

b) Calix[n]arenes (C[n]As)

Calixarenes represent the third generation of macrocyclic molecules following crown ethers and CDs. Synthesized from phenols and formaldehyde, CAs feature phenolic units linked by methylene bridges. Even-numbered C[n]As (n=4, 6, 8) are synthesized in a one-step process and are easily purified, making them widely studied. Odd-numbered C[n]As (n=5, 7, 9) are harder to synthesize and less common. CAs have a corn-shaped structure with a hollow cavity and two rims: a hydrophilic lower rim with phenolic oxygen and a hydrophobic upper rim with methyl groups. Their cavity sizes vary based on the number of phenolic units. The rims can be functionalized to tailor recognition properties towards specific analytes. Complex formation is driven by an array of forces including hydrophobic effects, π - π stacking, cation- π , CH- π and by hydrogen bonding and electrostatic interactions, which are enabled by modifying the substituent groups at the upper rim. Unlike CDs, the bare macrocycles are less usable for biomedical studies due to their lower solubility and higher hydrophobicity. Therefore, CAs are often modified by sulfonation, carboxylic acid conjugation and functionalization by polar groups to prepare water soluble forms⁵⁸⁻⁶⁰. These water soluble derivatives exhibit low toxicity and are biocompatible for biomedical applications such as drug solubilization, biosensing and nanomedicine⁶¹⁻⁶⁵, thus serving as a viable alternative to CDs and having wide-range applications.

c) Cucurbit[n]urils (CB[n]s)

Cucurbit[n]urils are macrocyclic containers composed of glycouril units linked by methylene bridges, resembling the shape of a pumpkin. They feature two hydrophilic carbonylated rims and a hydrophobic cavity, with a consistent depth of 0.91nm across all members, while the diameter varies depending on the number of monomeric glycouril units. They distinguish themselves from other macrocycles by their remarkable host-guest binding constants, which surpass those of CDs and CAs⁶⁶. The structural features of the CBs provide them with high affinity toward cationic guests⁶⁷, while nonpolar fragments of the guest molecules are stabilized by complex formation within the hydrophobic cavity⁶⁸. Additionally, positively charged species can be encapsulated by ion-dipole interactions facilitated by the carbonyl entry and hydrogen bonding between protonated molecules and the carbonyl portals. Variation in the portal size leads to different recognition properties among the different members. The remarkable host-guest properties of CBs allow them to interact with diverse analytes in aqueous media, including drug molecules, amino acids, peptides, carbohydrates, and proteins⁶⁹.

d) Pillar[n]arenes

Pillararenes represent a new class of [1_n] paracyclophane-like macrocycles developed by Ogoshi *et al.*, in 2008. They are synthesized by the condensation of 1,4-dialkoxybenzenes with paraformaldehyde, resulting in hydroquinone units linked by methylene bridges at para-position. Pillararenes have a pillar-shaped structure with alkoxy groups at the edges of both cavities. Their rigid, electron-rich structures and tunable cavity sizes make them excellent for molecular recognition applications^{70,71}. In comparison

to traditional hosts, pillar[n]arenes offer several advantages due to their highly symmetrical and rigid structures, which afford selective binding to guests. Furthermore, they allow for facile functionalization with different substituents on all benzene rings or selectively on specific positions, enabling the fine-tuning of their host-guest binding properties. Their solubility in organic solvents further makes them a suitable supplement to CDs and CBs with comparable cavity size⁷². The versatility of functionalization of their aromatic units and their exceptional host-guest properties have propelled pillararenes into the spotlight, garnering increasing attention in the field of host-guest chemistry^{73,74}.

1.3.2. Differential sensing using macrocyclic host systems

Recognition based sensors primarily utilize non-covalent interactions between the analyte and the recognition element of the sensor to facilitate detection. There has been a focused effort to identify new, simpler base materials for building such sensor libraries. Over the past few decades, supramolecular chemistry has emerged as a refined approach for building multifunctional materials, making it a preferred choice for the development of chemical nose sensor systems. The effectiveness of using macrocyclic host-guest systems for differential sensing has been illustrated in several examples in literature. These studies highlight the use of the broad-spectrum recognition properties of various macrocycles to enhance cross-reactivity and improve information extraction, leading to innovative applications and significant advancements in this field⁴⁵. These recognition-based sensor systems are further classified based on the nature of the binding mode between the recognition and transduction element as follows:

Recognition-based sensing – covalently linked type

The sensing systems formed by covalently linking the transducer to the macrocycle, demonstrates excellent stability and capability to detect analytes across a wide concentration range. It removes the dependence of their association on concentration, with a resistance towards dilution and competitive binding of non-target analytes⁷⁵.

The following section provides a few examples of this kind of system:

- A fluorescent sensor array composed of calixpyrrole chemosensors embedded in hydrophilic polyurethane films was constructed for the detection of anionic carboxylate drugs. The sensors utilized a common octamethylcalix[4]pyrrole (OMCP) receptor covalently conjugated to seven different fluorophores. The structural design of the sensors initiates an intramolecular partial charge transfer (IPCT) cascade, due to the presence of electron withdrawing groups in the fluorophores attached to the OMCP via a vinyl branch. The IPCT results in anion binding-induced changes in fluorescence, resulting in the detection of fourteen carboxylate drugs. In addition to these seven sensors, a tripodal turn-on fluorescent sensor prepared from (2,4,6-triethyl-1,3,5-trimethylamino) benzene with selectivity for

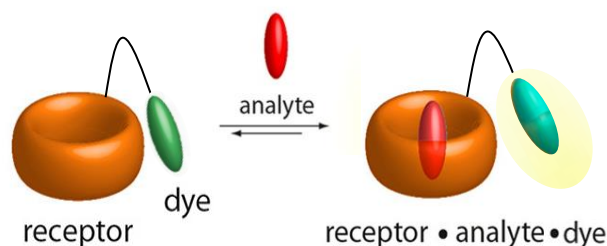


Figure 1.9: Illustration of recognition-based sensing, with covalent bonding between the fluorophore and the recognition receptor (adapted from Krämer, J. *et al.*⁷⁶).

aliphatic carboxylates and phosphates was employed to increase signal variability of the sensor array. This sensor remains flexible in the resting state, but forms a stable bowl-shape complex upon interaction with anions making it rigid and increasing the radiative dissipation of its excited state energy by fluorescence turn-on signals. Despite the seven chemosensors sharing similar receptor units, the varied covalently attached fluorophores facilitated the classification and quantification of the carboxylate drugs in human urine due to their differential responses⁷⁷.

- In the field of protein structure sensing, Agasti *et al.*, have adopted a direct sensing strategy by synthesizing a library of CB[7] molecules covalently linked to a group of fluorescent dyes. This strategy leverages the optical response of the fluorophores to their nearby microenvironment when CB[7] recognizes the chemical and topological features of diverse protein structures. This sensor detected conformational changes in proteins and differentiated the diverse self-assembled forms of specific amyloid- β aggregates. Thus, providing a versatile platform for analyzing misfolding and aggregation states in pathology-associated proteins⁷⁸.

- A well-designed pillararene-based sensor array has also been developed using a similar covalent strategy for the recognition of highly similar neurotransmitters. Tian *et al.*, created a supramolecular fluorescent probe where the upper and lower edges of pillar[5]arene were covalently modified with boric acid and coumarin-derivatized fluorescent recognition groups, respectively. This sensor was designed to discriminate three major categories of neurotransmitters namely catechol-type, monoamine-type, and charged neurotransmitters. The boric-acid-derived naphthalimide group interacts with catechol-type neurotransmitters through a condensation reaction, while the aldehyde-based coumarin derivative forms a Schiff base with the monoamine-type neurotransmitters, and the charged neurotransmitters are attracted by the electron-rich cavity of pillararene through electrostatic interaction. This careful molecular design allows this single sensing molecule to generate different fluorescence patterns at two wavelengths by means of these diverse interactions, facilitating the identification of neurotransmitters in biofluids, living neurons, and tissues⁷⁹.

- Other “turn-on” sensors have also been developed using cyclodextrin derivatives where fluorescence increases by the formation of host-analyte complexes. These sensors use a rigid spacer

between CDs and a fluorophore, preventing self-inclusion complexes and resulting in suppressed fluorescence due to the hydrophilic environment. When a hydrophobic guest molecule is included in the CD cavity, the fluorophore is positioned closer to a more hydrophobic environment, thereby enhancing its fluorescence emission. “Turn-on” type fluorescent chemical sensors such as 4-amino-7-nitrobenz-2-oxa-1,3-diazole-CD⁸⁰, naphthol-CD⁸¹ and hydroxyquinoline-CD⁸¹ conjugates have been well-studied.

Recognition-based sensing – non-covalently linked (IDA)

This approach employs the non-covalent interactions between macrocycle receptors, optical transducers and analytes. This mainly involves the indicator displacement approach (IDA) that has been the most commonly adopted strategy for differential sensing with macrocyclic systems. In a standard IDA setup, the receptor initially binds with an indicator to form the sensing ensemble. Upon introducing the analyte, the indicator is displaced from the ensemble. Typically, the free and bound indicators exhibit distinct optical properties (colorimetric or fluorescent), leading to a measurable signal change for the detection of analytes.

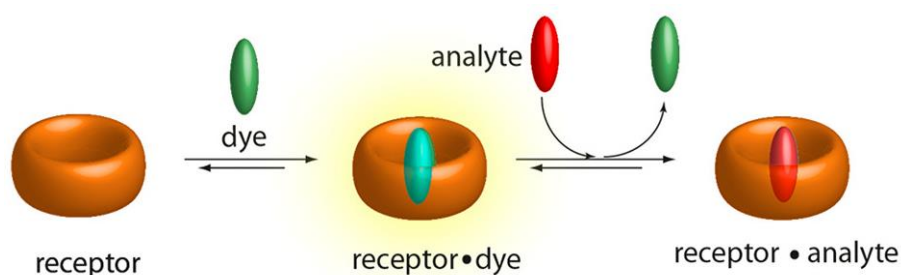


Figure 1.10: Illustration of recognition-based sensing, with non-covalent interactions between the fluorophore and the recognition receptor (adapted from Krämer, J. et al.⁷⁶).

- Hof *et al.*, constructed a simple chemical sensor array using different sulfonated CAs and fluorescent dyes based on IDA. Two types of sensor arrays were developed either employing a single calixarene host PSC6 (p-sulfonatocalix[6]arene) and a dye molecule, or using a set of three calixarene hosts (PSC6, PSC4, and PSC4(Br)) and the dye molecule, with fluorescence fingerprints generated by varying pH and/or organic co-solvent conditions. These arrays were initially validated by the recognition of structurally similar lysines, followed by those with varying degrees of methylation, using dye displacement facilitated by incorporation of the modified lysine residues into the PSC through cation- π interactions incorporated. This sensor array has shown promise in distinguishing lysine methylation statuses in synthetic peptides and histone peptides bearing trimethyllysine modifications but having different peptide sequences. Additionally, it effectively discriminates the post-translational modifications associated with arginine⁸².

- Recent advancements in the use of self-folding and deep cavitands as synthetic receptors, have led to their use in cavitand-based sensing applications. A fluorescence displacement system with a self-aggregating and water soluble cavitand was proposed for detecting trimethylated lysine peptides and determining histone demethylase activity. The strong affinity between a tetracarboxylated cavitand and a selected fluorophore was employed for the selective detection of trimethylated peptide targets. Fluorescence quenching in this system arises from encapsulation by the macrocyclic cavitand and the aggregation of complexes, thereby bringing the fluorophore moieties into close proximity. The addition of trimethylated lysine was observed to cause a significant recovery of the fluorescence signal, allowing the systems to differentiate between unmethylated and trimethylated peptides, based on the replacement of hard NH_3^+ ions with softer, less hydrogen-bonded NMe_3^+ . This characteristic further allows the system to monitor histone demethylation processes. Upon the introduction of a histone demethylase and its cofactors, the demethylation process causes a decrease in the observed fluorescence signal. The demethylated product has weak affinity for the cavity and cannot compete with the probe, thereby enabling the sensing of this process^{83,84}.
- The rapid identification and classification of nucleic acid secondary structures is both necessary and challenging. To address this, a deep cavitand-based sensor array was developed to differentiate G-quadruplex (G4) topologies. This innovative approach utilized five synthetic cavitand hosts and two cationic fluorescent dyes as reporter pairs, forming a 12-component sensor array. The dyes engage in a competitive binding equilibrium between cavitands and DNA, sometimes even forming a ternary complex with the dye, G4 and cavitand. These complex and cross-reactive interactions generate a unique fluorescence response for each G4 structure, allowing the easy differentiation of various topologies and even different G4 structures of the same fold type. This work introduces a novel concept in sensing, moving beyond the traditional IDA-based host guest complexation to include a wider range of recognition and sensing mechanisms, ultimately yielding more comprehensive and satisfactory results⁸⁵.
- A more recent example that is not based on the above described common macrocycles involves the use of an array of computationally designed de novo peptides as alternative synthetic receptors for differential sensing. These self-assembling α -helical barrels (α HBs) are oligomers of five or more helical peptides that form coiled structures with central solvent-accessible channels. These channels can be predictably altered in size, shape and chemistry and can contain environment-sensitive dyes that fluoresce upon binding. When exposed to analytes, these dye-loaded barrels cause unique fluorophore displacement patterns. These fluorimetric fingerprints have been used to train machine-learning models that correlate the patterns with specific analytes for discrimination of various biomolecules, drinks, and diagnostically relevant biological samples. While α HBs are similar to other synthetic receptors in their ability to bind various analytes, they also offer the advantages of being water-soluble, thermally stable, and reliably constructed based on design rules and established sequence-to-structure relationships.

• In addition to these examples, the development of multifunctional units for sensor arrays has been a major advancement in the field. Coassemblies of cyclodextrins with various amphiphilic calixarenes have been utilized to create distinct receptors. Upon complexation with corresponding dyes, these receptors have been formulated into versatile sensing units. This type of sensing unit effectively combines the recognition and assembly properties inherent in supramolecular chemistry. The tunability of these sensing units, including component ratios and environmental factors, allows for the construction of numerous sensor array units using a limited number of compounds, significantly reducing the synthetic workload. The coassemblies incorporating two macrocycles exhibit heteromultivalent recognition, making them suitable for analyzing biological macromolecules and complex systems⁸⁶.

Fluorophore -appended macrocycle sensing or Intramolecular Indicator Displacement Assay (IIDA)

In addition to these general sensing systems, there has been an emergence of a slightly different approach, involving covalent fluorophore-appended macrocycle sensor arrays. These arrays feature host-fluorophore conjugates which via a flexible linker form self-inclusion complexes. In the presence of a competitive guest, the fluorophore is excluded from the inside to the outside of the cavity. While the host-fluorophore conjugate exhibits strong emission properties in the self-inclusion state due to the encapsulation within the hydrophobic environment of the macrocyclic cavity, the exclusion of the fluorophore from the cavity to the bulk weakens the emission and generates a “turn-off” response⁸⁷. Cyclodextrin systems carrying covalently tethered fluorophores such as dansyl⁸⁸, naphthyl⁸⁹, pyrene⁹⁰ and anthracene⁹¹ have been extensively studied for a long time by many researchers and a number of such turn-off fluorescent chemical sensors have been reported.

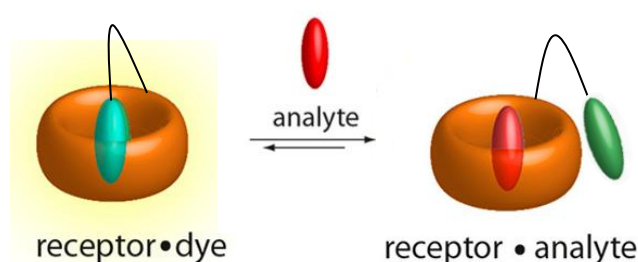


Figure 1.11: Illustration of intramolecular indicator displacement assays with covalent bonding between recognition receptor and fluorophore and non-covalent interaction based displacement of analyte (adapted from Krämer, J. et al.⁷⁶).

- More recently, Biedermann *et al.*, have extensively developed this approach and have reported the use of the CB[7] macrocycle system for the development of such unimolecular host-dye conjugates with hydrophilic and flexible linkers. They firstly designed and synthesized novel CB[7]-berberine chemosensors for the selective micromolar detection of Parkinson's drug amantadine in urine and saliva over potential interferents such as polyamines (cadaverine, spermidine, and spermine) or the steroid nortestosterone⁹². The study paved the way for the group to explore the use of this approach in designing differential sensing systems. They synthesized a novel nitrobenzofuran functionalized CB[7]-dye conjugate (CB[7]-NBD) that was reported for the differential sensing of fourteen bioorganic analytes via salt-adaptive behavior in biofluids such as urine, saliva and blood serum. The selectivity of this array was introduced and tuned by the addition of metal cations, which in combination with the non-charged and weakly binding NBD dye, can engage in secondary interactions when higher binding analytes are bound in the cavity of CB[7]. Unique response of the sensor-analyte complexes was observed upon the addition of different concentration and type of salts. This led to the development of a salt-addition assay, which could be transferable to other macrocyclic receptors with inherent inorganic anionic or cationic binding tendencies⁷⁵.
- This strategy was also adopted by Nau *et al.*, for the development of a FRET-based specific DNA sensing system using CB[7] host covalently attached to a fluorescein dye as the energy acceptor. DAPI, a non-covalently bound dye and a known DNA minor groove binder, served as the energy donor. When DAPI was encapsulated within the CB[7] host, the close proximity of the tethered fluorescein resulted in FRET emission upon excitation at 360 nm. Upon addition of DNA, DAPI bound to the DNA instead of CB[7], disrupting the FRET process and altering fluorescence emissions. This macrocyclic system thus enabled the quantitative detection of DNA via ratiometric fluorescence sensing⁹³.

Thus, the diverse macrocyclic receptors exhibit rich binding characteristics, offering numerous opportunities for the development of array-based sensing systems. The systems formed by covalently linking the optical transduction element to the macrocycle, eliminates the concentration dependency of their association, thereby enabling the detection of analytes across a broad concentration range. Conversely, non-covalent linkage-based sensing systems offer advantages such as compatibility with commercially available probes, modular construction, and enhanced adaptability for improved pattern recognition. Both types of macrocyclic sensing systems have been employed to detect and differentiate a variety of analytes, including ions, small molecules, and biological macromolecules. Thus, these growing advancements demonstrate the significant potential of macrocyclic receptors in designing and creating effective sensing platforms for a wide range of applications.

1.4. Biomedical application of array-based sensors

The versatility of cross-reactive sensor arrays has resulted in their use across various fields, including the detection of explosives, monitoring of industrial effluents like volatile organic compounds, tracking environmental pollutants, and ensuring quality control of food, beverages, and drugs. However, this thesis will focus on the application of array-based sensing systems in the field of biomedical sciences, where numerous applications related to human health have emerged.

The creation of reliable assays for recognizing multiple analytes using specific receptors is fundamental to the field of biomedical sciences. Conventional lock-and-key strategies that have been adopted include capture of the concentration of specific biomarkers or bioanalytes by enzyme-linked immunosorbent assays (ELISA), histopathology or bioimaging. These enzyme and antibody assays rely on the costly development of specific antibodies^{94,95} by the prior identification and clinical validation of pertinent biomarkers present in invasively or non-invasively collected biological samples. Emerging “omics” methods applied to body fluids or biological specimens provide information on a large number of biomarkers or bioanalytes expression levels and, combined with bioinformatic tools, offer characteristic signatures useful for identification. However, these assays are time-consuming, expensive, and labor-intensive, often requiring modifications such as surface immobilization of biomarkers, that reduce their affinity and specificity⁹⁶. Despite their theoretically high specificity, their performance often drops under unpredictable cross-reactive biological environments, including unprecedented variations in the target such as sequence changes, deletions or insertions.⁹⁷ In contrast, selectivity-based array sensor systems use molecular receptors that are cost-effective, easily scalable, and tunable for recognizing a wide range of targets. These systems offer a complementary yet powerful alternative to the use of antibodies, proteins and aptamers for molecular recognition⁹⁸. A primary advantage of the differential sensing approach over traditional biosensors that target specific biomarkers is the potential to discriminate complex mixtures of analytes (as in the case of a mixture of odorant molecules). Notably, the selectivity of these synthetic receptors allows the possibility for the distinction of multiple structurally similar analytes from one another⁹⁹⁻¹⁰¹. This is facilitated by the cross-reactivity of the sensing elements that form the sensor array, avoiding the challenges associated with designing high specificity based sensing elements. The differential sensor array considers the simultaneous interaction of multiple analytes with an entire system, enabling the monitoring of overall changes in complex mixtures, a feature which is particularly well-suited for the analysis of biological systems. Cross-reactive arrays have the advantage of also being hypothesis-free, where the individual interactions between sensors and analytes do not need to be fully understood to interpret results. This facilitates the rapid development of sensor arrays to address the analytical challenges in the field of biomedical sciences²⁸. These sensing arrays have thus been adopted to address more and more complex

challenges faced in biomedical applications, from protein discrimination to disease diagnostic, as reviewed below (Figure 1.12).

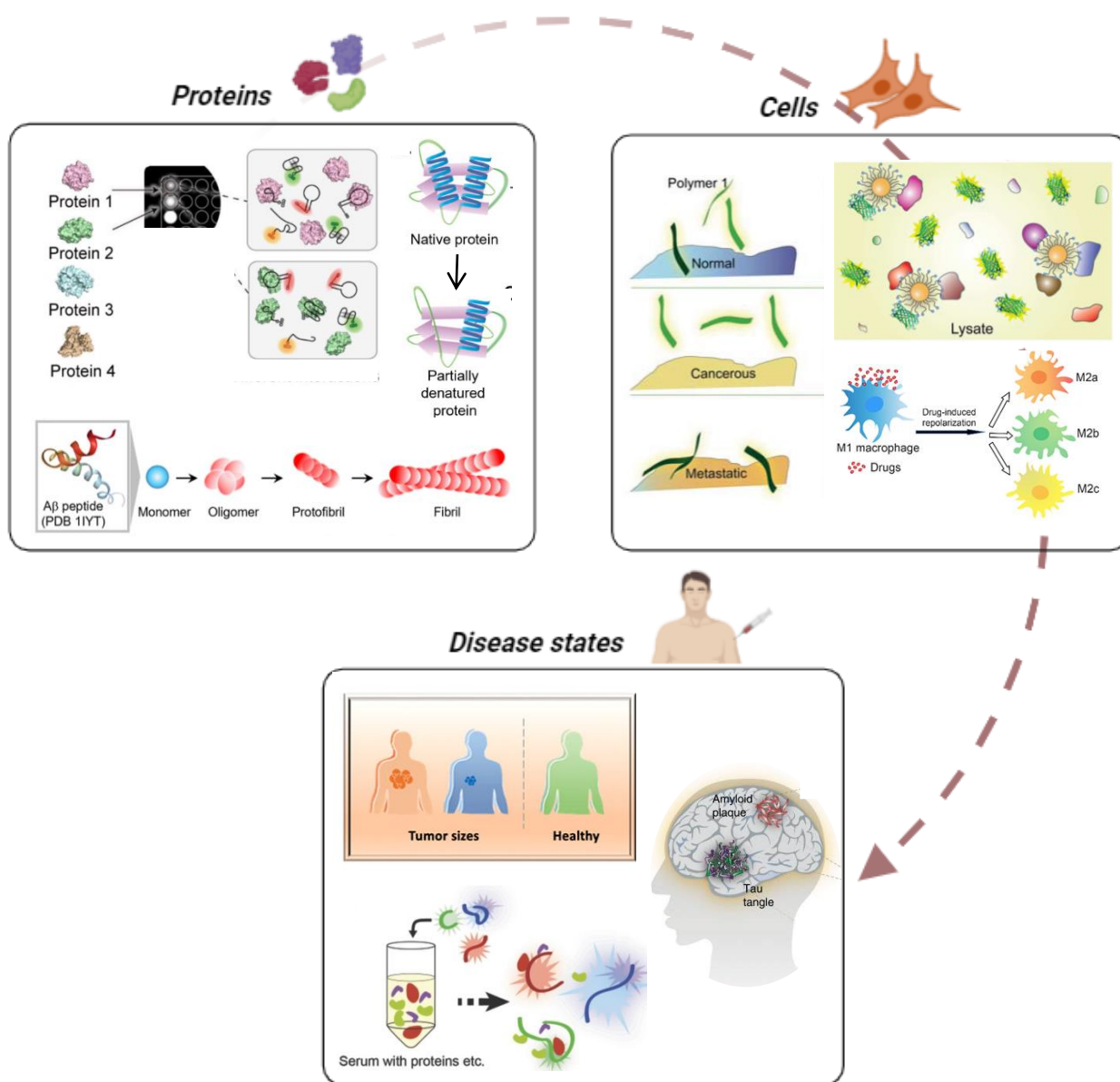


Figure 1.12: Array-based sensing for biomedical applications: From protein detection to cellular profiling and tackling challenges of disease diagnosis in human body fluids.

1.4.1. Discrimination of basic biomolecules: Proteins

Proteins are essential components of the human body, playing vital roles in maintaining various metabolic processes. Their physiological functions, such as recognition, enzymatic activity, ion transport, and signal transduction, highlight their importance. Their content, abundance and structure in human body fluids are finely tuned to preserve the physiological conditions of human beings. Abnormal protein concentration levels in biofluids such as blood, serum, urine or in cells and tissues provide essential information regarding the onset of a pathological condition. Thus, monitoring global protein changes is critical in medical diagnosis, staging, prognosis or treatment selection^{102,103}.

Therefore, the development of array-based sensors was initiated for the detection of these important biomolecules. The developed arrays have used fluorogenic sensors combined with various recognition elements to create unique response patterns upon binding of the target proteins. While some methods use multi-element cross-reactive receptors for pattern recognition, others create environment-sensitive sensor arrays that respond to changes in solvents, probe concentration, pH, or ionic concentrations. Additionally, multi-wavelength cross-reactive single-system sensors are also being employed¹⁰⁴.

In general, the detection of proteins using sensor arrays has been done in simple physiological buffers, where interference from other biomolecules is minimal. This allows the array to be trained in the detection and classification of a wide range of analytes under simple and straightforward matrix conditions. To simulate the complex conditions found in biological mixtures, where multiple interfering bioanalytes, including non-target proteins, are often present in much higher concentration, sensor arrays have been tested with proteins spiked into more complex matrices such as human blood and serum samples. Furthermore, the detection of variable concentration ranges of individual proteins associated with disease states could also be effective for the identification and monitoring of progress of a disease state. In addition to monitoring individual protein levels, it is also crucial to detect the formation of misfolded proteins, their aggregates, and other higher-order structures, which could enable the early diagnosis of diseases associated with such protein structure abnormalities, like Alzheimer's disease⁷⁸. Thus, the overall progress in array-based sensors has enabled the differential sensing of proteins under various practical conditions, addressing key challenges in biomedical research and clinical practice¹⁰⁴.

1.4.2. Discrimination and evaluation of cellular characteristics

The absence of universal biomarkers due to perturbations such as disease or exposure to environmental stimuli makes characterizing cells challenging. The diversity and complexity of cellular surfaces along with their total bio-molecular content makes them ideal targets for array-based sensing, which is a crucial tool for understanding global proteomic changes, rapid diagnostic assessment and disease treatment. The key advantage being that once a training dataset is established, it becomes feasible to develop evaluation systems to identify cellular statuses that have not yet been elucidated¹³.

Array-based cell sensing successfully classifies a wide range of cells, including both prokaryotes like bacteria and eukaryotes¹¹.

Identification of cancerous cell subtypes: Diagnosis of certain cancers such as the aggressive triple-negative breast cancer, is challenging due to the poor or absent expression of associated biomarkers^{105,106}. Early and accurate diagnosis is essential for effective treatment, leading to the development of array-based sensors to detect and differentiate non-cancerous, cancerous, and metastatic cell lines, offering an alternative to traditional biomarker strategies. These systems have also been modeled to detect circulating tumor cells (CTCs), which are associated with cancer metastasis and have poor prognosis using conventional antibody-based techniques¹⁰⁷. Sensor arrays have detected these CTC's at low concentrations by spiking cancer cell lines into matrices with peripheral blood mononuclear cells¹⁰⁸. While most array systems target cells with genotypic differences; by differentiating between different cancer cell lines or between metastatic and non-metastatic lines, systems capable of discriminating cells from diverse genetic backgrounds as well as varying degrees of metastasis have also been tested¹¹.

Monitoring cellular differentiation processes of stem cells: Stem cells, known for their ability to differentiate into different cell types, are increasingly used in clinical trials, particularly in cancer treatment and regenerative medicine. They are steadily becoming fundamental in addressing challenges associated with modern medicine. Array-based sensors have been developed to non-invasively detect these cells, including targeting cancer stem cells (CSCs). These cells form small solid tumors that possess stem cell characteristics like self-renewal, differentiation and tendency to populate new tumors. They remain resistant towards chemotherapy and cause the recurrence of tumors and metastasis, thus making them key therapeutic targets. While most cell sensing studies employ cell-damaging techniques to analyze cells, noninvasive analytical approaches which do not require analysis of cells directly are preferred. Therefore, characteristic components secreted by cells into the culture medium are being extensively used as analytes. These secreted molecules, collectively known as the "secretome," are crucial for identifying cellular characteristics^{109,110}. Thus, pattern-based sensing systems have utilized this secretome to monitor cellular differentiation without damage to cells¹¹¹.

Cell-based screening for chemotherapeutics and therapeutic discovery: Cell-based screening methodologies are essential tools in drug discovery and risk assessments, offering a cost-effective and practical alternative to animal testing¹¹². Cell viability assays predict cellular dysfunctions or death from acute drug exposure¹¹³, but long-term, low dose drug exposures cause subtler cellular modifications and phenotypic responses¹¹⁴. These phenotypic signatures serve as sensitive indicators of drug effects, surpassing conventional cell-viability and biomarker methods, which are costly and require multi-step processing, limiting their use in high-throughput detection¹¹⁵. Hypothesis-free array-based sensing enhances sensitivity to early and subtle phenotypic changes. Thus, this approach is used for high-

throughput investigations in chemical safety assessments and for evaluating drug efficacy and off-target effects for drug discovery¹¹.

1.4.3. Discrimination and evaluation of body fluids for disease diagnosis

Points-of-care (POC) diagnostics are rapidly expanding to create simple, portable, reliable and cost-effective bioassay readouts for clinical biomarkers/analytes. This approach aims to improve the efficiency and accessibility of healthcare with a gradual shift towards the realization of personalized medicine¹¹⁶. However, achieving this goal necessitates minimizing size and operational complexity while ensuring analytical efficiency meets clinically relevant detection thresholds. Developing comprehensive arrays for POC diagnostics requires detecting diverse biomarkers and analytes, each with dissimilar technical requirements. Ideally, the sensor design would rely only on one common technique for assay readouts, enabling direct analysis of body fluids such as blood, urine, saliva, sweat samples etc. with minimal user intervention and simple readouts for easy analysis. Most current clinical diagnostics fail to meet these criteria because conventional methodologies are often complex, labor-intensive, and require specialized training to operate expensive instruments¹¹⁷.

The major analytes in diagnostics typically fall into three groups: proteins, nucleic acids, and small molecules. Of these, proteins like enzymes, antibodies, hormones are commonly targeted as analytes in POC diagnostics. The low concentration of these protein biomarkers (pM- μ M range) in various body fluids requires detection by methods like the enzyme-linked immunosorbent assay (ELISA), which has emerged as the gold standard over the past few decades, utilizing antibodies for their specificity and affinity. However, antibodies present challenges such as stability issues, variability between batches, and high production costs, limiting their suitability for POC assays. Therefore, there's a growing focus on approaches that enable high-throughput and simultaneous monitoring of multiple proteins rather than individual ones to ensure accurate diagnostics^{26,118}. Thus cross-reactive sensor arrays have emerged as viable alternatives for high-throughput and hypothesis-free detection of bio-analytes that stage a diseased state. Initial investigations utilized six different Au nanoclusters (NC) with varied surface properties and fluorescence characteristics to discriminate between proteins. This array was subsequently evaluated for clinical relevance by effectively distinguishing serum samples from breast cancer, rectal cancer, and healthy individuals without overlap¹¹⁹. Subsequent research has expanded on these findings, developing sensor arrays aimed at identifying, categorizing, and distinguishing various diseases in bodily fluids.

1.5. State of the Art: Biomedical Applications of array-based sensing

• Rotello *et al.*, have extensively developed array-based sensing strategies based on gold nanoparticles and polymers to address the biomedical applications discussed previously. Firstly they developed a prototype sensor using six cationic AuNP's with different structures and an anionic poly (p-phenyleneethynylene) (PPE) polymer (Figure 1.13a)²⁵. The electrostatic complexation of the AuNP's and polymer results in the quenching of the polymer fluorescence through energy transfer ("OFF state"). The addition of protein analytes disrupts the AuNP-polymer by competitive binding, leading to a recovery of the polymer fluorescence ("ON state"). The differential interaction between t

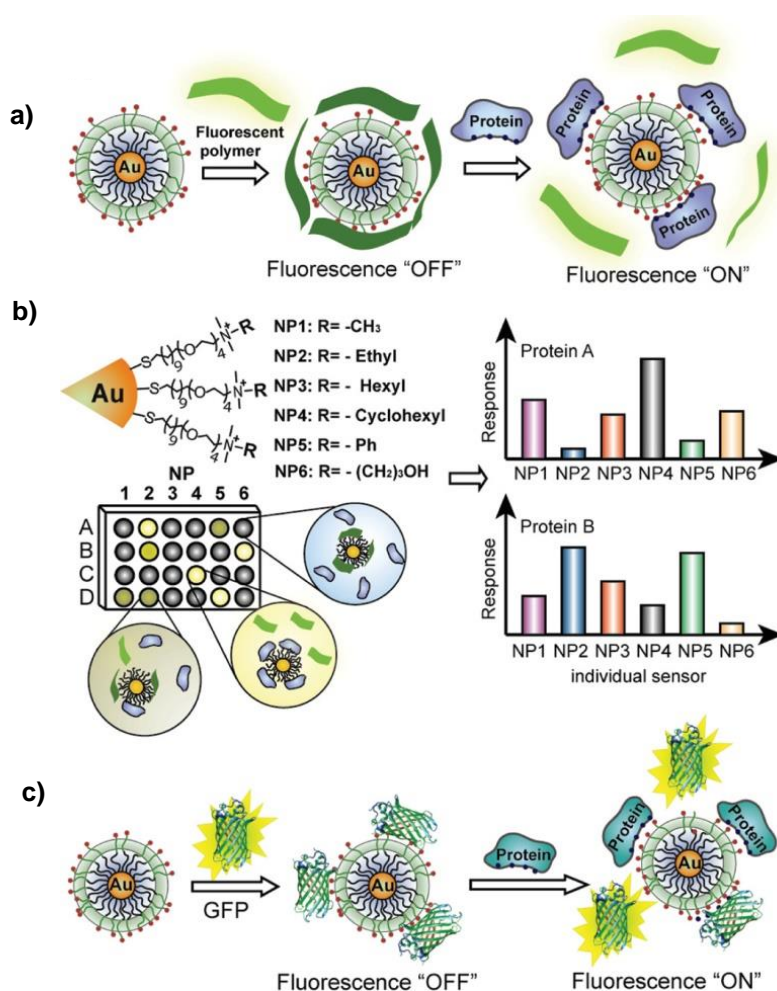


Figure 1.13: Schematic illustration of array-based sensors based on AuNP-fluorescent polymer/GFP conjugates. a) Competitive binding between protein and quenched polymer-AuNP complexes leads to the restoration of fluorescence b) Generation of fluorescence fingerprint response patterns for individual proteins. c) Competitive binding between protein and nanoparticle-GFP complexes leading to the fluorescence 'ON' state (adapted from Saha, K. *et al.*¹⁹).

the proteins and nanoparticles results in the generation of a fingerprint fluorescence response pattern for each protein (Figure 1.13b) that was characterized by LDA. Using the same principle, AuNP-green fluorescent protein complexes were designed to detect and identify proteins at very low concentration (500 nM) in undiluted human serum (Figure 1.13c)¹¹⁸. Furthermore, the AuNP-conjugated polymer systems have been employed for rapid and effective differentiation between normal, cancerous, and metastatic cells^{120–122}. The fluorescence responses analyzed by LDA were capable of distinguishing (1) different cell types; (2) normal, cancerous, and metastatic human breast cells; and (3) isogenic normal, cancerous, and metastatic murine epithelial cell lines.

- The group also reported an enzyme-AuNP conjugate based array for protein detection with improved sensitivity through enzymatic activity. In this approach the AuNP are combined with β -galactosidase (β -Gal) through electrostatic interaction, inhibiting the β -Gal enzymatic activity. The addition of proteins, displaces the β -Gal from the AuNPs thereby restoring its enzymatic activity towards a fluorogenic substrate 4-methylumbelliferyl- β -D-galactopyranoside, resulting in an amplified readout of the protein analyte binding²⁶(Figure 1.14).

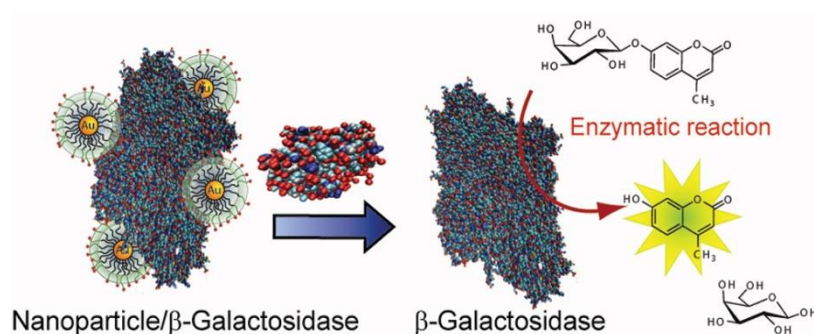


Figure 1.14: Schematic illustration of β -galactosidase (β -Gal) and cationic AuNPs based sensor array. The displacement of β -Gal from the β -Gal/AuNP complex by protein analytes, restores the catalytic activity of β -Gal toward the fluorogenic substrate 4-methylumbelliferyl- β -D-galactopyranoside, resulting in an amplified signal for detection (adapted from Miranda, O. R. et al ²⁶).

- To demonstrate the utility of array-based sensing for profiling of the mechanism of chemotherapeutic drug action, they also developed a multichannel sensor platform. The sensor consists of a three-channel anionic fluorescent protein platform [Enhanced green fluorescent protein (EGFP), Enhanced blue fluorescent protein (EBFP), tandem dimer Tomato (tdTomato)] complexed to a cationic benzyl-functionalized AuNP. These complexes upon interaction with cells generate a turn-on fluorescence signal by the release of the fluorescent proteins (Figure 1.15). This nanosensor has been utilized to screen 15 chemotherapeutics with different molecular mechanisms to generate a training set of fluorescence fingerprints. The sensor displayed an overlap for drugs with similar mechanisms and

could efficiently discriminate between apoptotic and necrotic drug groups. In addition, the sensor could also predict unknown mechanisms and determine mechanistic correlations between individual drugs and their combinations¹²³

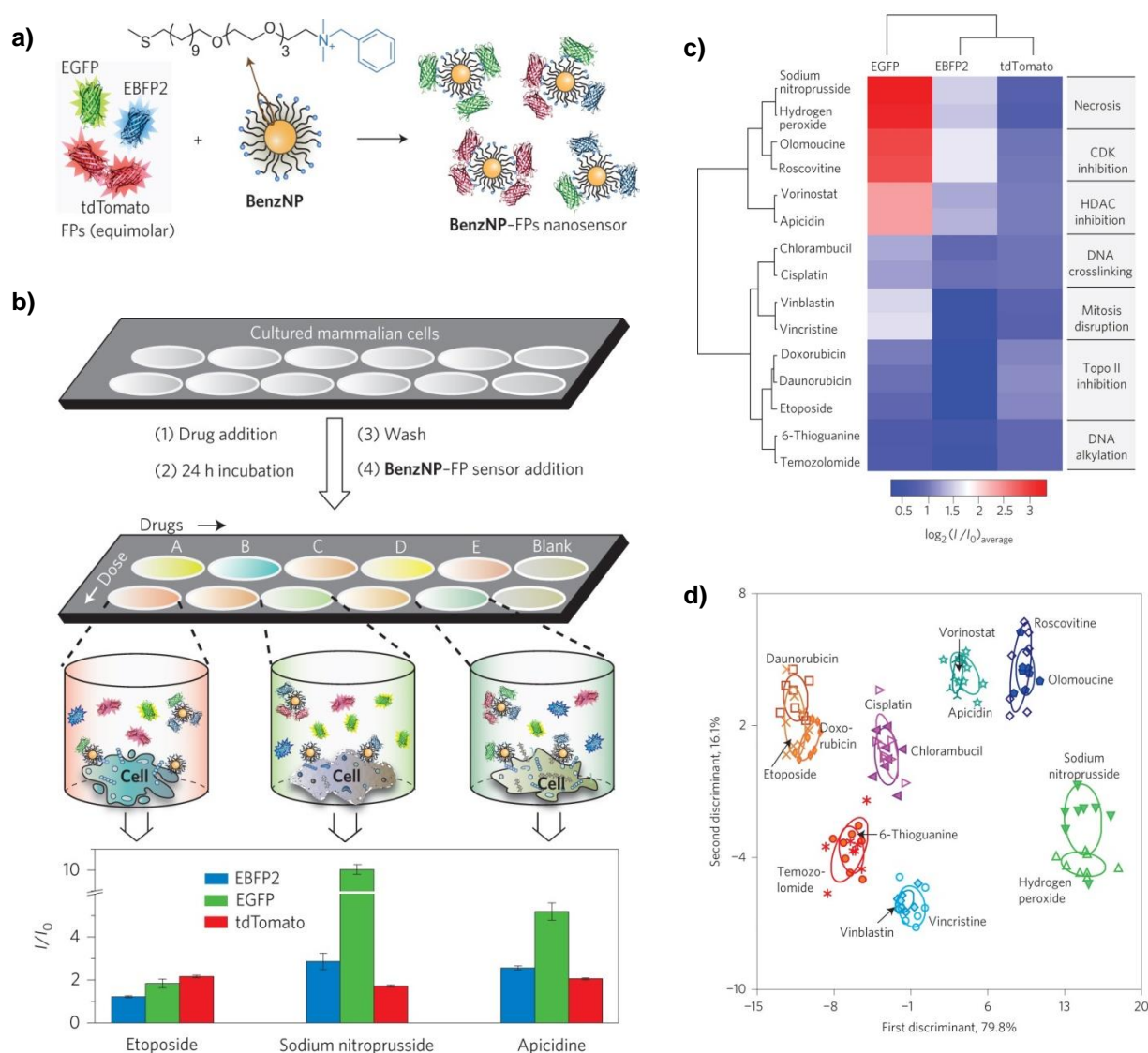


Figure 1.15: a) Schematic illustration of three channel nanosensor fabricated from benzyl-functionalized AuNP- fluorescence protein complexes b) Drug screening workflow adopted with the nanosensor array to generate fluorescence responses for pattern recognition. c) Heatmap and HCA performed on $\log_2(I/I_0)$ average of the fluorescence responses, showing the degree of association of the tested drugs. d) LDA-based clustering of fluorescence responses of reference drugs (adapted from Rana, S. et al.¹²⁴).

Furthermore, the group expanded the application of the array-based sensors towards developing a robust, multiplexed fluorescent polymer-based sensor platform for detecting liver fibrosis from serum samples with clinically relevant accuracy (Figure 1.16). This sensor array is composed of a poly(oxanorborneneimide) (PONI) random copolymer scaffold featuring benzoate monomers for dye molecule conjugation. Three PONI-polymer sensing elements, each bearing pyrene, dapoxyl and PyMPO dyes, were utilized to generate a four-channel sensor array from a single fluorescence measurement. The array displayed variations in fluorescence intensity upon the addition of protein analytes due to changes in pH, ionic strength and supramolecular interaction of the dyes. The array was initially tested for the successful identification of protein spiked in human serum samples. Subsequently, it was evaluated for its ability to fingerprint liver fibrosis from patient serum samples. A training model was developed for the diagnostic test of healthy, mild-moderate or severe fibrosis using a train set of 50 samples, achieving an accuracy of 60%, while the remaining 15 samples of the test set were identified by the model with an accuracy of 66.7%. The robustness, accuracy and sensitivity of this array-based approach, without the need for complex instrumentation, makes it a favorable method as compared to other biomarker-based detection methods. Thus providing a new avenue for establishing a point-of-care disease diagnostic approach for liver fibrosis¹²⁵.

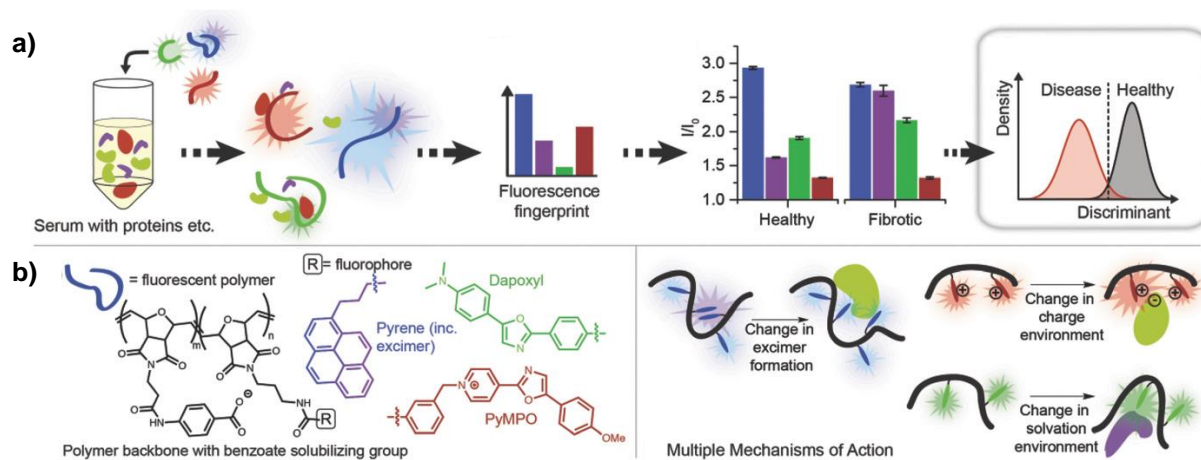


Figure 1.16: a) Schematic illustration of fluorescence fingerprint generation for discriminant analysis for liver fibrosis detection from serum-polymer interaction using an array-based sensing approach. b) Polymer structure with three conjugated fluorophores and mechanism of action of the responsive fluorophores generating a four channel sensing array (adapted from Peveler, W. et al.¹²⁵).

Agasti *et al.*, utilized different fluorophore-labeled single stranded DNA (ssDNA) complexed with cationic surface-functionalized AuNPs to create a robust multichannel array-based sensing platform. The total protein content of cells in different states, obtained from lysates, interacts

differentially with the cationic AuNP due to their unique surface amino acid residue signatures. Initially, the interaction with the AuNP quenches the fluorescence of the ssDNA, but the presence of protein lysates competes with this interaction, regenerating the fluorescence fingerprint response. This enables the discrimination of cells based on their proteomic signatures. The identification of cell states by this multichannel sensor was further investigated by employing small-molecule autophagy modulators to induce alterations in the global cellular state. The high accuracy of discrimination obtained between inducers and inhibitors further demonstrated the potential of such multichannel sensing systems for high-throughput drug screening¹²⁶.

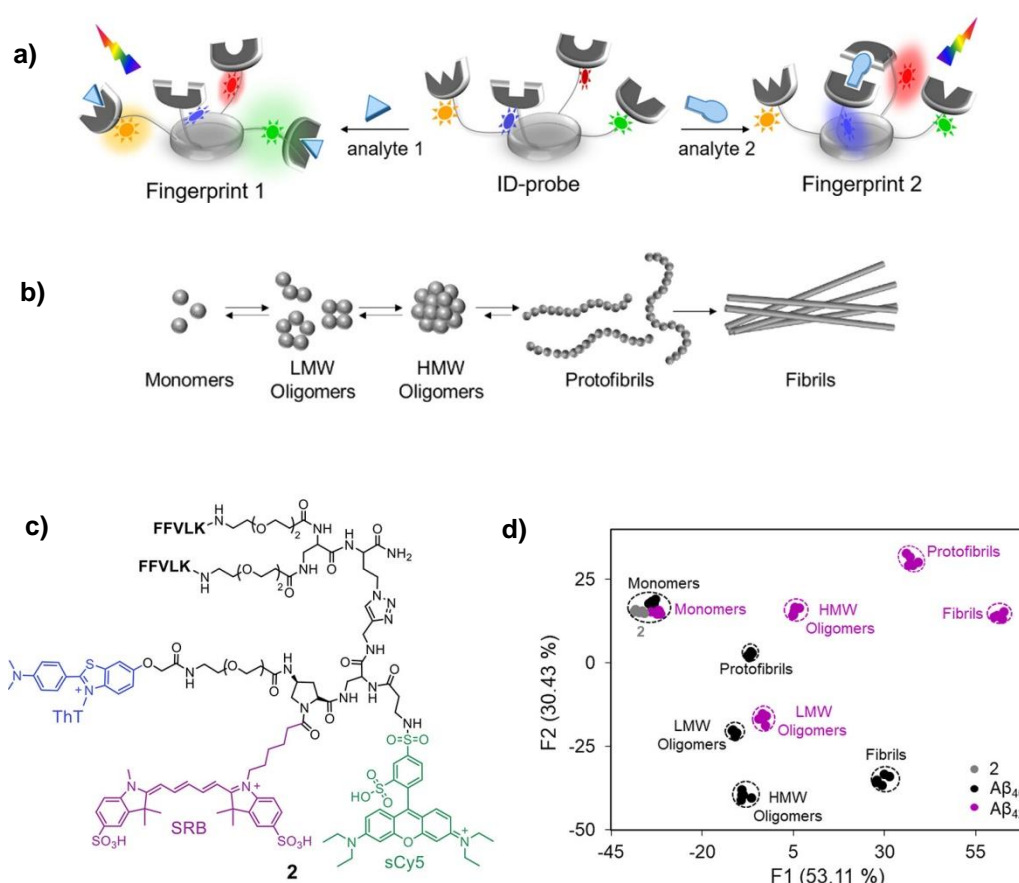


Figure 1.17: a) Schematic illustration of the operating principle of the turn-on unimolecular fluorescent ID probes. b) Schematic illustration of the A β aggregation states c) chemical structure of ID probe (**2**) d) the LDA map of the patterns generated by **2** upon interaction with the different aggregation states monitored at seven representative emission channels (adapted from Motiei, L. et al.¹²⁷).

- Margulies *et al.*, initiated the development of unimolecular combinatorial fluorescent molecular sensors, termed as ID-probes that integrate various non-specific fluorescent probes that emit at different wavelengths onto a single molecular platform. The differential binding of analytes to the different fluorophores generates unique emission fingerprints (ID) for analyte identification and classification (Figure 1.17a). A ID-probes consisting of an amino proline scaffold appended to three fluorophores,

namely thioflavin T (ThT), sulforhodamine B (SRB), and sulfo-Cy5 (sCy5) that serve as FRET donor, acceptor/donor and acceptor respectively was designed (Figure 1.17c). Two site-specific amyloid recognition groups: bis-KLVFF peptide and thioflavin T (ThT) that can interact with A β aggregates were incorporated into the ID-probe¹²⁸. The changes occurring in the various A β aggregation states were effectively differentiated in a single solution by this probe (Figure 1.17b, d). In addition aggregates that were generated from different alloforms, through distinct pathways or from distinct amyloidogenic proteins were also successfully discriminated¹²⁸. The group also designed a unimolecular fluorescent probe that incorporated four fluorescent dyes (nitrobenzoxadiazole, Nile red, cyanine 5.5, and cyanine 7) and three specific binders for various protein families. This probe was synthesized to enable the analysis of specific protein populations in their natural environment and their combination in complex mixtures¹²⁹.

- Jiang *et al.*, fabricated an array-based sensing system comprising of 12 fluorescent molecules developed using conjugated polymers (functionalized PPE units) and small molecular dyes (Nile red, anilinonaphthalene-8-sulfonic acid (ANS)). The interaction of the protein with these sensing elements via electrostatic and hydrophobic interactions perturbs their fluorescence properties, generating unique fingerprint patterns. While Nile red is sensitive to the polarity of its environment and interacts with the hydrophobic residues of the proteins, ANS majorly engages through electrostatic interactions between its negatively charged sulfonate groups and the positively charged amino acids of the target proteins. A proof-of-concept was established by testing this sensor with proteins known to be related with urinary diseases such as HSA, transferrin, lysozyme, IgG, myoglobin and acid phosphatase in physiological buffer. The sensor's fluorescence intensity fold change was analyzed by pattern recognition algorithms to generate a discrimination accuracy of 100%. Evaluating each sensing element's contribution to this performance allowed the original 12-unit sensor array to be optimized down to a 3-unit array. This optimized array was further utilized for the quantitative analysis of single proteins and protein mixtures (binary and ternary) in both PBS and urine, with good discrimination and correlation with varying protein concentrations. Subsequently, the array was applied to urine samples collected from clinical patients with different urinary disease such as tubular injury (TI), nephropathy (Nep), diabetic nephropathy (DN) and systemic lupus erythematosus (SLE) along with samples from healthy patients (Figure 1.18). The sensor array exhibited high accuracy of discrimination among the different diseases and enabled precise prediction of unknown urine samples in a blind test. This development marks a significant advance in non-invasive, point-of-care diagnostic technologies, offering a rapid, sensitive, and cost-effective method for detecting and monitoring urinary system diseases through the analysis of urine samples¹³⁰.

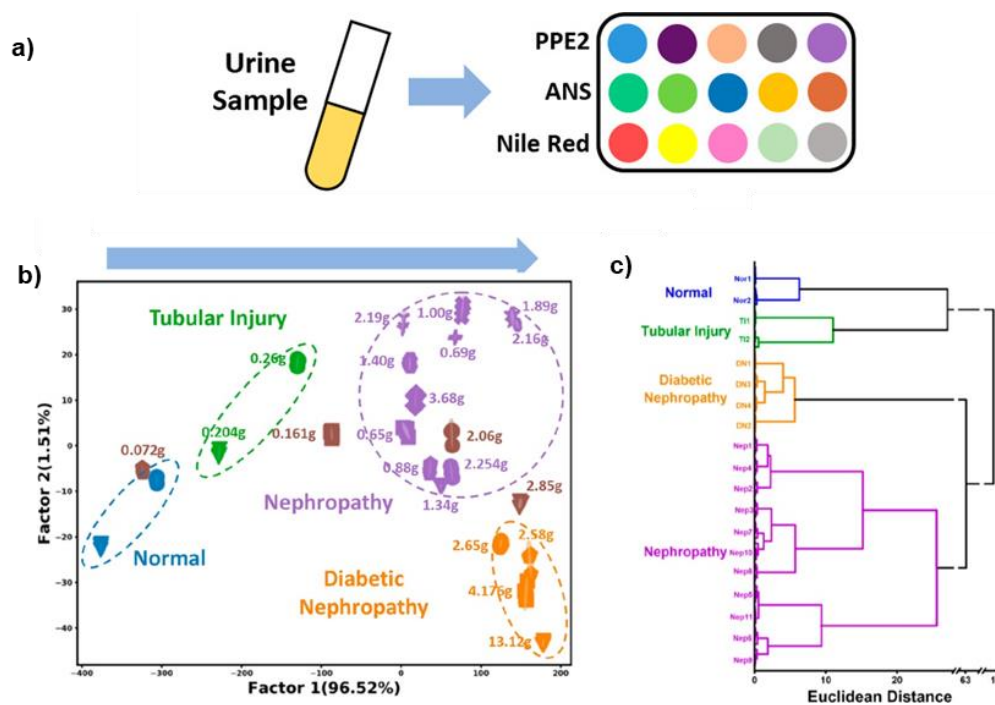


Figure 1.18: a) Schematic illustration of the three-unit fluorescence sensor array which interacts differentially with proteins in urine to realize the discrimination of protein profiles of urine samples. Discrimination of urine samples from patients suffering from urinary system diseases and healthy people as presented by b) a LDA plot and c) hierarchical clustering dendrogram based on the fluorescence variations of the optimized sensor array with urine samples (adapted from Shen, Y. et al.¹³⁰).

- Shouzhuo *et al.*, describe the creation of a multichannel fluorescence sensor array for sensing human gut microbiota using a recognition engineering strategy. The array employs antimicrobial agent (vancomycin, bacitracin, and lysozyme) functionalized gold nanoclusters (AuNCs) with different emission wavelengths and gluconamide-modified Ti_3C_2 MXenes (transition metal carbides) as signal reporters. Each antimicrobial agent targets specific components on bacterial cell walls, such as D-Ala-D-Ala moieties, pyrophosphate groups, and polysaccharides. Ti_3C_2 MXenes, a two-dimensional nanomaterial, efficiently quench the fluorescence of nearby AuNCs through energy transfer due to their strong light absorption and excellent quenching effect. When the AuNCs and Ti_3C_2 MXenes are combined, their close interaction brought about by hydrogen bond interactions leads to fluorescence quenching of the AuNCs through energy transfer. In the presence of gut-derived bacteria, the fluorescence signals of the AuNCs change due to competitive binding between the AuNCs and bacteria towards MXenes (Figure 1.19a). The varying fluorescence responses create a discriminative fingerprint map of the gut microbiota using pattern recognition algorithms. This sensor array also successfully distinguished colorectal cancer (CRC) patients from healthy individuals in a cohort of 20 samples,

demonstrating its potential as a robust and simple platform for clinical analysis and cancer diagnostics (Figure 1.19b-d).

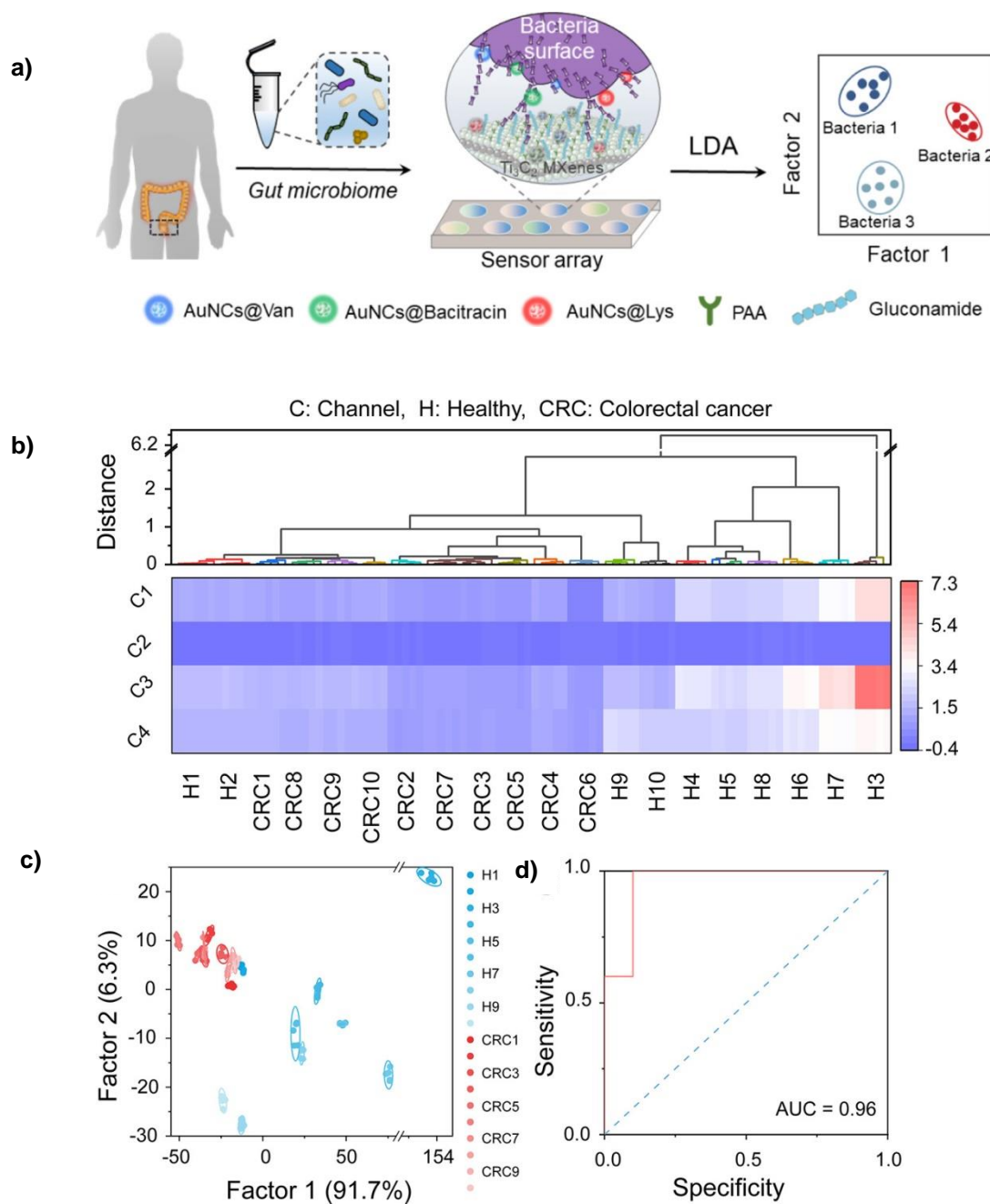


Figure 1.19: a) Schematic illustration of multichannel fluorescence sensor array for gut microbiota sensing. b) Heat map and resulting HCA dendrogram of the fluorescence fingerprints of the feces samples from colorectal cancer patients and healthy individuals (C1-C4: sensor channels) c) Canonical score plot of LDA for the discrimination of individual CRC patients and healthy individuals. d) Receiver operator curve indicating the performance of the diagnostic ability of the sensor array towards CRC (adapted from Liu, Z. et al.¹³¹).

1.6. Conclusions

Cross-reactive chemical nose sensing arrays have seen rapid advancements over the past few decades, greatly enhancing their ability for the identification and discrimination of structurally similar analytes in both simple and complex media. The hypothesis-free nature of this strategy has established its applicability as a mainstream tool for evaluation of analytes that are challenging to differentiate by conventional analytical methods. These chemical nose-based sensing arrays have been employed to create unique fingerprints of proteins, cells and for the diagnosis of diseased state from healthy ones. While traditional, specific sensing approaches continue to be fundamental to biological assays; array-based sensing is gradually becoming more significant in biomedical sciences. The performance of these cross-reactive sensors is constantly being improved by the adoption of advanced design strategies and multifaceted statistical methodologies, offering benefits such as ease of operation, low cost and high-throughput detection. Nevertheless, there still remain few challenges that need to be addressed for these arrays to achieve versatile and practical applicability, and current research in this field is focused on overcoming these issues.

The fabrication of sensing elements that produce differential responses, along with the selection of appropriate statistical models, is crucial for the development of high-quality and efficient sensor arrays. The utilization of multivariate analysis to analyze developed probe designs will enable a better understanding of sensor design principles and will aid in the creation of more efficient cross-reactive sensor receptors. The diversity of available chemical materials will benefit from this feedback, facilitating the extension of the sensor library. When combined with high-throughput and automated analytical platforms, this approach can be scaled up to the level of big data analysis, enabling the identification and classification of complex analytes. To translate this sensing technology to clinical settings, it is essential to combine the outputs of sensor arrays with specific sensors that employ antibodies, aptamers and similar specific recognition elements. Additionally, adopting “omics” methodologies is crucial for enhancing the applicability of this technology in clinical diagnostics. These integrations will lead to the understanding of what the sensor array is actually interacting with, leading to mutual improvements in both biomarker discovery and sensor array functionality. Such a strategy paves the way for large scale and rigorous screening for disease fingerprinting, which, when paired with robust statistical analysis, can yield more reliable conclusions. Finally, efforts are needed for the miniaturization and commercialization of these sensor arrays. This will enable the development of portable, accurate and integrated systems that will contribute to the transition from lab-based research to simple, robust, real-time point-of-care diagnostics in clinical settings. Thus, while numerous challenges remain in this field of sensing, there are also exciting developments and promising science and technology waiting for the near future, with significant application potential in various fields including biomedicine, food safety, environmental monitoring and so on.

In this thesis, the synthetic host-guest interaction directed by the macrocycle cucurbit[n]uril has been employed to develop selective chemical nose sensor arrays for application in disease diagnosis. Chapter 2 explores the design, conception and characterization of a bimodal recognition driven fluorescence sensor array for pattern recognition of proteins. Chapter 3 will provide an overview of the statistical approaches and streamlined methodologies that will be adopted in the discrimination studies of the TPA-CB[n] sensor arrays. Chapter 4 addresses the use of the developed sensor for the optical fingerprinting of body fluids, establishing proof-of concept studies for diagnosis of disease models: Phenylketonuria (PKU) and Preeclampsia (PE). Chapter 5 highlights the preparation and optimization of analytical platforms for high-throughput sensing analysis. Chapter 6 delves into the further development of a colorimetric sensor array and its potential applicability in the discrimination of pharmaceutical products. Finally, Chapter 7 demonstrates the utility of a library of cucurbit[7]uril-fluorophore conjugates in developing a methodology study towards the diagnosis of SARS-Cov-2 in human serum. The final Chapter 8 provides an outline of the entire project and addresses the future perspectives of the project and the field of array-based sensing.



Chapter 2

Bimodal recognition by host-guest chemical nose platform: Design strategy and characterization of system based on cucurbit[n]uril and triphenylamine derivatives

2.1. Introduction

Array-based sensing has rapidly emerged as a useful approach for profiling macromolecular analytes in complex biosystems. This strategy requires the design of spatially separated sensing elements, each equipped with distinct recognition and transduction components, to provide multiple outputs required for fingerprinting various analytes. The recognition elements are expected to provide a broad spectrum of interaction modes with the analytes, while the transduction elements are expected to translate these binding interactions into interpretable output signals, thereby maximizing sensitivity to subtle changes in the complex patterns generated by different analytes. Thus, the core of chemical nose sensing lies in the meticulous and efficient design of the sensing elements to address specific analytical challenges. This design process is crucial for achieving non-specific, reliable and reproducible hypothesis-free sensing.

Host-guest chemistry has evolved as a tool of interest to enable the development of such versatile and robust sensing elements for array-based sensing. Host receptors are engineered to interact with analytes through a range of non-covalent interactions, including hydrogen bonding, hydrophobic interactions, and electrostatic forces. The guest molecules, on the other hand, are modified to function as transducers in these systems. This modification typically involves incorporating optically active molecules, which can be either covalently or non-covalently linked to the macrocyclic receptors. This design allows for sensitive and effective real-time monitoring of fingerprints generated by the analytes. Consequently, the integration of host-guest chemistry into array-based sensing platforms offers improved flexibility in sensor design and adaptability to a wide range of analytical contexts. We propose to utilize the molecular recognition properties of the macrocyclic receptor family of cucurbit[n]urils as recognition elements, in combination with the conjugated system of triphenylamine derivatives as transduction elements to design and optimize an optical chemical nose sensing array with application in diagnosis of complex disease models.

2.2. Macrocyclic family of Cucurbit[n]urils (CB[n])

Cucurbit[n]urils (CB[n]) are a family of rigid macrocyclic host molecules formed by the condensation reaction between glycouril and paraformaldehyde under acidic conditions (HCl/H₂SO₄, 80-100°C). This reaction was first reported by Eberhard Meyer and Robert Behrend in 1905^{132,133}. However, it wasn't until 1981 that the product was crystallized from the condensation reaction, featuring glycouril with methylene groups acting as bridges between adjacent units¹³⁴. The resemblance of this structure to a pumpkin (belonging to the *cucurbitaceae* family) resulted in the naming of this family as cucurbit[n]uril, with n indicating the number of glycouril building blocks that constitute the macrocycle. While CB[6] garnered significant attention in the 1980s and 1990s for its ability to facilitate

non-covalent binding, it was mainly between 2000 to 2002 that Kim¹³⁵ and Day^{136,137} modified the existing reaction conditions and synthesised the other members of the family i.e. CB[5], CB[7], CB[8] and CB[10] (Figure 2.1). The modified synthesis under milder and kinetically controlled conditions yields CB[6] as the major product (~60%), while the other members are obtained in lower yields (CB[5] (~15%), CB[7] (~15%) and CB[8] (~5%)), along with trace quantities of other higher order members of the family¹³⁸.

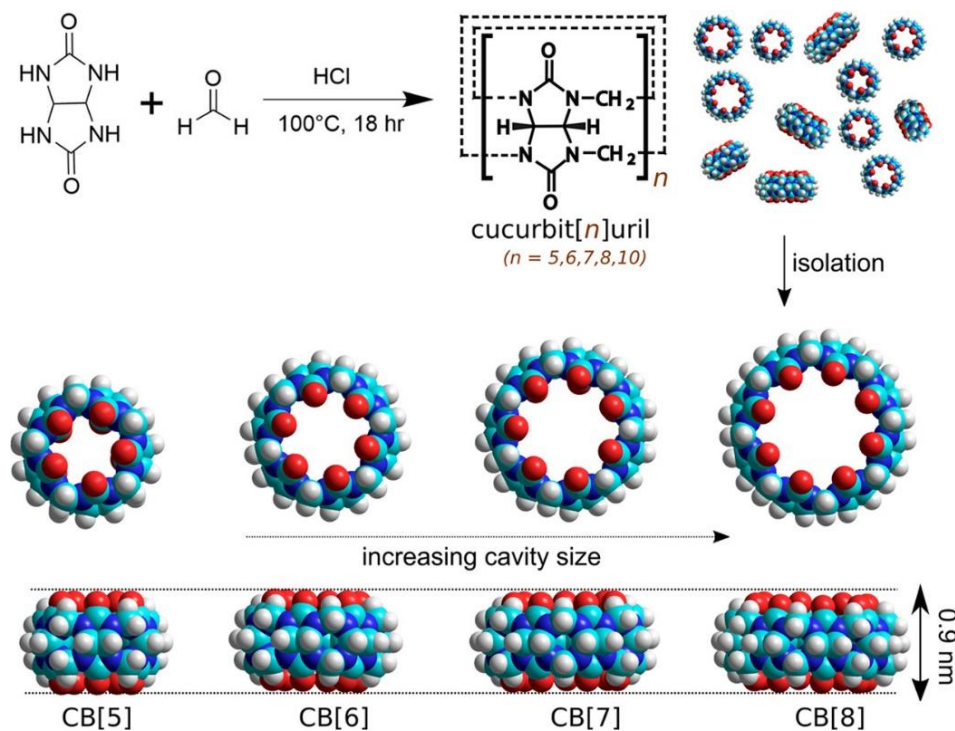


Figure 2.1: Synthesis of CB[n] homologues by condensation of glycouril and paraformaldehyde. Space-filling models of CB[5] –CB[8] shows an increase in size while maintaining a constant height across the macrocycles (adapted from Barrow, S.J. *et al.*¹³⁸).

2.2.1. Structural and thermodynamic properties of CB[n]

Structural properties of CB[n] homologues

The CB[n] family features characteristic structural properties; they are highly symmetric pumpkin shaped structures with open ends. The two portals of the barrel structure are lined with ureido carbonyl groups that provide entry to the internal cavity which is hydrophobic in nature. CB[n] molecules possess an equatorial plane of symmetry, making the two portals identical, with the portal being almost 2 Å narrower than the middle of the cavity. This structural feature creates a constrictive environment ensuring the slow dissociation of molecules enclosed within the inner cavity¹³⁹, whose size spans and exceeds that of the cyclodextrins especially in the case of higher members of the family¹⁴⁰. The electrostatic potential map of CB[n] displays high electron density along the carbonyl groups,

highlighting their cation-receptor functionality. The absence of functional groups or electrons pointing towards the inside of the cavity explains the high hydrophobicity of the cavity¹⁴¹. The rigid structure of these macrocycles facilitates the easy measurement of the cavity parameters, where all the CB[n] ($n = 5-8$ and 10) have the same height at 9.1 Å, but show increasing width with increase in the number of monomeric units. This cavity enables the encapsulation of smaller guest molecules based on complementarity in size and volume with the CB[n] to form inclusion host-guest complexes. While the lower order homologues (CB[5], CB[6], CB[7]) form 1:1 binary complexes, the higher order members of the series form 1:2 homoternary or 1:1:1 heteroternary complexes. The host-guest complexes are stabilized by variable modes of interactions which are both polar and non-polar in nature. The polar interaction includes hydrophilic interactions such as ion-dipole, hydrogen bonding, dipolar interactions between the carbonyl groups and polar moieties of guest molecules, while the nonpolar interactions involve hydrophobic reactions of the cavity with hydrophobic guest moieties (Figure 2.2).

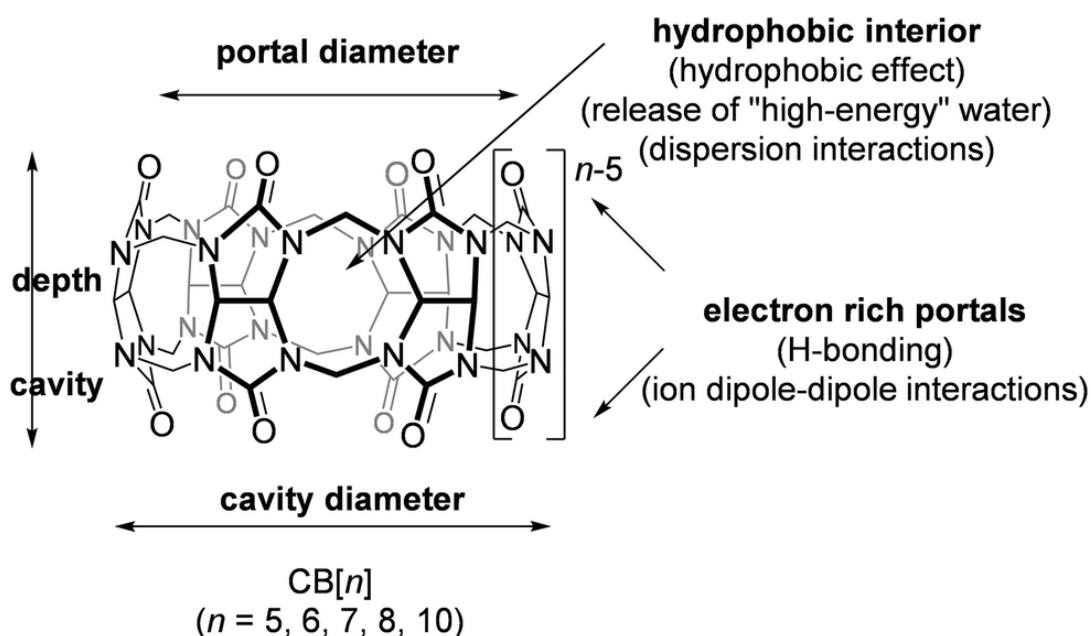


Figure 2.2: Structural properties of CB[n] homologues depicting regions important for molecular recognition (adapted from Alešković, M. and Šekutor, M.¹⁴²).

Rebek and Mecozzi established the concept of packing coefficient (PC), representing the ratio of the size of the guest to the size of the host cavity volume, which is an estimate of the steric goodness of fit for the formation of host-guest complexes¹⁴³. A PC value of 55% has been primarily identified to give the best binding affinities for the selected host-guest pair, while other values are associated with lower affinities. This concept has been successfully applied to predict the stability of CB[n] inclusion complexes with representative sets of known guest molecules. An average value of 47% for CB[5],

58% for CB[6], 52% for CB[7], and 53% for CB[8] were obtained, aligning with the ideal packing solution of 55%¹⁴⁴.

Role of water in CB[n] host-guest complexation

CB[n] has been exploited for applications in different fields of research based on their ability to form association complexes with various guests, including small molecules, drugs, amino acids, peptides, saccharides, dyes, hydrocarbons, per fluorinated hydrocarbons and even high molecular weight molecules such as proteins (e.g., human insulin)^{140,145}. However, the first guest molecule that was proven to be experimentally encapsulated within the hydrophobic cavity of CB[6] was water. Depending on the specific member of the homologous series, CB[n] can form inclusion complexes with 2 (CB[5]) to 22 (CB[10]) water molecules. Various studies have highlighted the significance of host-guest size complementarity, hydrophobic effect (entropic gain from water release upon binding) and ion-dipole interactions in rationalizing CB[n] complexation^{140,146,147}. Despite this, the exceptionally strong binding of the CB[n] complexes, even with uncharged guests, remained unclear.

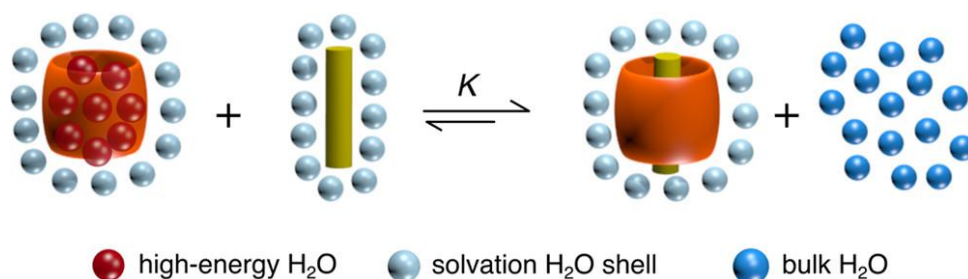


Figure 2.3: Schematic illustration of release of high energy water molecules from the cucurbit[n]uril cavity upon binding of hydrophobic guest molecules (adapted from Barrow, S.J. *et al.*¹³⁸).

Biedermann *et al.*, investigated the presence of additional factors, specifically the confinement of water molecules within the CB[n] cavity prior to guest inclusion. These confined water molecules possess properties that are very different from the bulk. These molecules exhibit high energy as a result of the limited tendency for hydrogen bond formation and due to the weak dispersion interactions with the weakly polar CB[n] cavity. In aqueous conditions the release of these high-energy water molecules from the cavity of the host molecule upon complexation with a guest, primarily contributes to the overall hydrophobic nature of the CB[n] cavity and the high-affinity binding of such host-guest complexes. Molecular dynamic (MD) simulations and isothermal titration calorimetry (ITC) was adopted to demonstrate this solvent effect, highlighting, the release of enthalpically and entropically unfavourable high-energy water, from the CB[n] cavity upon host-guest complexation. This release of the frustrated water molecules to the bulk, increases the overall enthalpy of the host-guest association (Figure 2.3), which is supported by the exothermic binding results obtained in the ITC titrations of CB[7] host with neutral solvents such as DMSO, DMF, acetone etc.

Thermodynamics of CB[n] host-guest complexation

The most remarkable feature of the CB[n]s is their high affinity for certain organic molecules, often exceeding the benchmark of the strongest non-covalent interaction ever measured of avidin-biotin at approximately 10^{15} M^{-1} ^{148,149}.

Binding enthalpy: While the subtler nature of CB[n] host-guest interactions with even neutral guest molecules was studied (ferrocene, adamantane and bicycle[2.2.2]octanes)¹⁵⁰, it was observed that the enthalpy of interaction for the positively charged derivatives of these guests remained nearly identical despite significant differences in their total charges. This phenomenon can be attributed to strong coulombic attraction between the positively charged substituents and the negatively charged CB[n] rims, counterbalanced by the loss of solvation energy upon host-guest binding. Thus, ion-dipole interactions in water are not the primary driving force of the host-guest interaction and the loss of solvation may or may not outweigh the coulombic attraction¹⁵¹. Additionally, the complementarity in the size and shape of the CB[n] and its guest favours van der Waals interactions, which have a much stronger and significant impact on the binding affinity. However, it should be noted that due to the low polarizability of CB[n], interactions between hydrophobic guests and the bulk should slightly be more favourable as compared to those with the cavity, and the dispersion interactions between the guest and the cavity are relatively weak. It has thus been suggested that the driving force for the host-guest complexation is primarily the release of 'frustrated' water molecules from the CB[n] cavity, representing a non-classical enthalpic hydrophobic effect¹⁴⁴. This has been corroborated by the negative enthalpy of binding of some guests towards CB[7] as compared to CD's .

Binding entropy: The change in configurational entropy upon host-guest complexation does not depend on the charge of the guest and the rigidity of the CB[n], with constrained encapsulation of guests leading to higher binding affinities^{151,152}. While flexible guest molecules suffer from entropic penalties on complexation, more rigid guests display entropically favourable complexation. For these guest molecules, the total binding entropy, as measured by ITC, becomes less favourable with the addition of positively charged units. Therefore, the contribution of configurational entropy can be disregarded, and the only parameter that has a drastic effect on the binding affinity is the difference in solvation entropy caused by ejection of water molecules from CB[n] upon guest binding. This observation diverges from the common enthalpy-entropy compensation model seen in most supramolecular systems, where gains in binding enthalpies are compensated by losses in binding entropies leading to narrower binding affinities¹⁴⁷. This deviation is primarily observed in CB[7] systems resulting in broader range of binding affinities.

Thus, in summary the high binding affinities observed with CB[n] are majorly due to: (1) the ability of guests and their substituents, especially positively charged ones that are close to the CB[n] portal to return as many hydration water molecules as possible to the bulk upon binding. This process

is both enthalpically and entropically favourable, (2) the structural rigidity of the CB[n] and some select guest molecules, (3) the minimal loss of solvation energy upon encapsulation and (4) the favourable ion-dipole interactions between positively charged guest and the carbonyl rims of the CB[n]¹⁵³. Thus, the calculations for ultrahigh binding affinity guests show a high enthalpic gain over negligible and negative entropic effect making the molecular recognition process by the CB[n] thermodynamically favourable¹⁵⁴.

2.2.2. Properties of CB[n] for sensing applications

The members of the cucurbit[n]uril family have been extensively studied for their utility as chemical sensors. In such systems there exists a thermodynamic equilibrium between the bound and free states of CB[n] and the amount of complex formation depends on the concentration and affinity of CB[n] to the analyte of interest. Given that most organic analytes of biological interest such as proteins, peptides, hormones, DNA, drugs typically occur in aqueous media at the concentration range of mM to nM, CB[n] based sensors with affinities generally within this range are well-suited for the detection of these analytes. These CB[n] macrocycles possess unique properties that facilitate their use for sensing applications¹⁵⁵:

1. *High affinity hosts*: The CB[n] family exhibits high binding affinities that are typically in the range of $10^3 - 10^9$ M for various organic molecules. As discussed in detail earlier, this can be attributed to the release of water molecules from the confined hydrophobic cavity upon binding of different guest molecules and ensures their interaction with a variety of analytes.
2. *Charge selectivity*: Organic molecules that possess complementary size-wise fit into the cavities of the different members of the CB[n] family tend to have appreciable affinities in water. Exceptions are mainly observed in the case of guests with negatively charged groups, which are not stabilized by ion-dipole interactions with the carbonyl portals of CB[n]. Therefore, the macrocycles possess higher affinities for positively charged analytes as compared to the neutral ones, often by a factor of 10-100.
3. *Size selectivity*: The rigid structure of CB[n] allows for constrictive and size-selective binding of analytes, which is crucial for designing selective sensors.
4. *Fast Kinetics*: To facilitate the development of sensors for diagnostic applications, it is essential to adopt systems that have rapid response times to shorten assay periods and enable faster reaction monitoring for point-of-care applications. The rate of complex formation with CB[7] and CB[8] have been observed to be fast ($\sim 10^7 \text{ M}^{-1} \text{ s}^{-1}$), falling short by only 2-3 magnitudes with respect to the diffusion limit of $10^9 \text{ M}^{-1} \text{ s}^{-1}$. Although some tight complexes with CB[6] exhibit slower kinetic rates, using higher members of the homologous series can improve the kinetics.
5. *Wide applicability*: The members of the CB[n] family exhibit several other beneficial properties that are practically relevant to their use in sensor development. Firstly, the CB[n] are inert to a number of common chemicals, enabling their use in a wide range of reaction media, including medium at different

pH levels, and in the presence of cofactors/co-reagents. They are also redox-inert, making them suitable for the setup of redox-based sensing systems for redox-active analytes. Additionally, they are photochemically inert, allowing their use in sensing applications involving the presence of strong light sources and lasers. Importantly, CB[n]'s have been identified as biologically non-toxic and can be adapted for use in biologically relevant conditions.

6. *Functionalization with transducers*: For the design of an effective sensor, the CB[n] mediated complexation of an analyte has to be coupled to a transducer. Kim *et al.*, pioneered the functionalization of CB[n] with reactive functional groups to improve solubility and enable the synthesis of tailor-made CB[n] derivatives¹⁵⁶. This advancement aided in the covalent conjugation of optical transducers like fluorophores, to the spectroscopically dark CB[n]. Additionally, with non-modified CB[n], the use of spectrally active guest molecules as transducers by indicator displacement or associative binding assays (ABA) has widened the scope of utilizing CB[n] for development of CB[n]-based sensing strategies.

However, the nature of the coupling adopted for the sensor design is often limited by the inherent characteristics of each member of the CB[n] family. CB[5] has a smaller cavity size, limiting its ability to bind a narrower range of guest molecules. In contrast, CB[7] has a larger cavity size, making it capable of binding a wider variety of guests. It is also water-soluble and more amenable to chemical functionalization. On the other hand, CB[6] and CB[8] present challenges due to their lower solubility, which complicates functionalization efforts. Therefore, with regards to the size of the cavity CB[6] and CB[7] are preferred for developing IDA sensors, while CB[8] uniquely enables the design of ABA systems with 2:1 host-guest complexes¹⁵⁷.

2.2.3. Analyte binding properties of CB[n] homologues

While the general properties of CB[n] that make them suitable as recognition elements were covered in the previous section, the specific variations in these properties among different family members and how these variations influence the selection of analytes they can potentially interact with are discussed in detail below:

Binding properties of CB[5] : CB[5] is the smallest member of the cucurbit[n]uril family, with a portal diameter of 2.4 Å and a cavity volume of 82 Å³. Consequently, CB[5] primarily encapsulates gas molecules such as noble gases, methane, ethane and small molecules like methanol and acetonitrile. Crystallographic evidence also supports the encapsulation of chloride or nitrate ions. The carbonyl groups along the portals also allow CB[5] to bind cationic species, facilitating the formation of portal complexes with alkali, alkaline earth and ammonium cations, in addition to hexamethylenetetramine which acts as a lid on the carbonyl portals. Due to its inability to bind large organic molecules, CB[5] is well-suited for the development of size selective gas sensors. However, these CB[5] sensor systems still lack appropriate signal transduction mechanisms for effective sensor set-ups.

Binding properties of CB[6] :

The next member of the series, CB[6], is the most abundant homologue, featuring a portal diameter of 3.9 Å. It is known to bind both aliphatic and aromatic hydrocarbons along with small cyclic and polycyclic aromatic hydrocarbons. However, its insolubility in organic solvents and sparing solubility in water constrains its application in sensor designs. Despite this, it most commonly forms stable compounds with aliphatic amines, where strong electrostatic interactions with the carbonyl portals and the release of water molecules from the cavity stabilize the formed complexes. Thus, the most stable complex is formed for an optimal length of the aliphatic chain, such that the amine groups remain positioned in close proximity to the portals. The surrounding medium significantly influences the complex formation by CB[6]. Higher stability is observed in the solution phase as compared to the gas phase, due to the additional H-bonding with water molecules in solution, whereas in the gas phase, only ion-dipole interactions are predominant. Further, in the gas phase, the guest molecules undergo structural distortion to maximize the ion-dipole interactions with the portal, however this requirement is compensated in solution through H-bonding. Thus, CB[6] is best suited for binding small organic molecules or large organic species with suitable side chains.

Binding properties of CB[7] :

The third member of the family, CB[7], possesses a larger cavity that allows it to form inclusion complexes with bulkier guest molecules beyond the simple aliphatic chains and gas molecules typically encapsulated by CB[5] and CB[6]. With a portal diameter of 5.4 Å, cavity diameter of 7.3 Å and cavity volume of 279 Å³, CB[7] can encapsulate a broader range of guests while exhibiting highly favorable binding parameters. They are general binders for most organic compounds and display an appreciable affinity for the aromatic side chain residues of peptides and proteins (Tryptophan (Trp), Tyrosine (Tyr) and Phenylalanine (Phe)). The binding affinities within the CB[n] family are highest for CB[7] due to its excellent cavity-filling effects, which facilitate the expulsion of high-energy water molecules from the macrocyclic core. However, CB[7] is not preferred for guests that are too small, as this leads to incomplete water release from the cavity, or for very large guests that cause steric repulsion. Additionally, CB[7] has significantly higher water solubility (30 mM) compared to other family members, making it more suitable for developing cucurbit[n]uril-based sensors. Though not an ideal representation of the entire family, CB[7] also exhibits a good degree of biocompatibility as demonstrated in studies that have highlighted the binding of CB[7] to a range of biologically relevant molecules such as peptides, proteins, dyes, neurotransmitters at sub-nanomolar concentrations.

Binding properties of CB[8] :

CB[8] is 1.7 times larger than CB[7], with a cavity volume of 479 Å³, and its binding parameters parallel those of other members of the family. They display strong binding affinities towards bulky amphiphilic positively charge guest molecules, including natural and artificial steroids and drug molecules. Due to their larger cavity, CB[8] encapsulates guests with larger aliphatic chains by the formation of U-shaped

conformations, whereas in CB[7] a significant portion of the guest molecule remains protruding outside the macrocyclic cavity. Additionally, CB[8] possesses the unique ability to bind two guest molecules simultaneously within its cavity, resulting in the formation of homo (2:1) and hetero-ternary (1:1:1) complexes with appropriate guest molecules. These features make CB[8] particularly useful in designing complex sensor systems that require the detection and analysis of larger and more complex molecules

While the CB[n] family offers extensive and valuable properties for its role as a recognition element in the proposed sensor array, it remains crucial to evaluate new optical transducer elements in the context of developing an effective sensor design. These transducer elements must thus be capable of interacting with CB[n] receptors and providing appropriate transduction for fingerprint generation, facilitating the identification and discrimination of target analytes. Therefore, a group of conjugated triphenylamine derivatives have been considered as suitable optical transduction elements.

2.3. Triphenylamine derivatives: structure and properties

2.3.1. Structural properties

Vinyl-triphenylamine derivatives (TPA's) are a group of fluorogenic molecules recognized for their high two-photon absorption properties. They feature a central triphenylamine electron-rich donor core, with two or three conjugated linkers terminating in acceptor groups. The molecular geometry of central TPA is influenced by two opposing forces: the resonance stabilization of the π -electron system and the steric repulsion between protons of adjacent phenyl groups. While resonance stabilization promotes the delocalization of π -electrons, favoring a planar structure with D_{3h} symmetry, the steric repulsions are mitigated by tilting away from the molecular plane. This results in a propeller-like structure with C_3 symmetry, making the phenyl-substituted branches non-coplanar and forming a three-dimensional configuration. This unique structure is highly sought after for the design of π -conjugated systems with non-linear absorption properties¹⁵⁸. The central core of the TPA exhibits unique radical characteristics and significant steric hindrance. The lone pair on the central nitrogen atom contributes to the role of TPA core as a donor, while the addition of electron acceptor groups at the branches induces strong intramolecular charge transfer. This enables the formation of octupolar structures in the three-branched derivatives and V-shaped structures with a strong quadrupolar character in the two-branched derivatives¹⁵⁹ (Figure 2.4).

These derivatives are synthesized by functionalization of the TPA core by classical cross coupling (Heck, Sonogashira) or Wittig reactions, allowing the creation of versatile π -conjugated

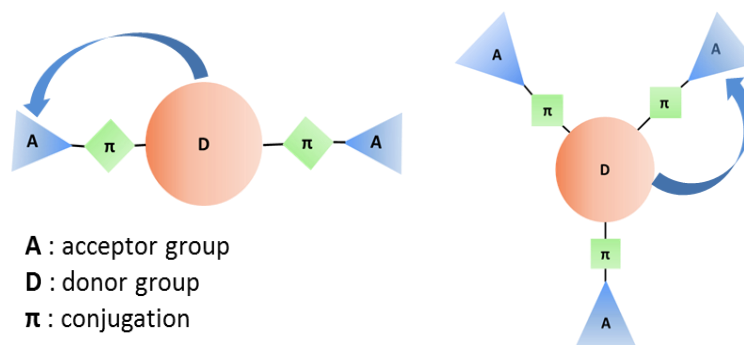


Figure 2.4: General structure of quadrupolar and octupolar structure with D- π -A branches based on triphenylamine donor core exhibiting intramolecular charge transfer transitions (adapted from Mantooth, S.M. *et al.*¹⁶⁰).

multi-branched systems with diverse optical properties. However, these synthesized derivatives with extended π -conjugated systems, are large in size and water-insoluble, which is often considered to be a significant drawback for their use in biological applications primarily as imaging probes. In this context, small cationic electron acceptors have been introduced at the terminals of the branches, which not only help in maintaining a relatively modest size but also improve overall water solubility¹⁶¹. The most commonly incorporated cationic acceptors include π -deficient heteroaromatic rings such as pyridine, quinoxaline, 1,3,5-triazine, benzimidazole and triazole moieties¹⁶².

2.3.2. Photophysical properties

The vinyl linkers that connects the donor to the acceptors, plays a crucial role in determining the photoluminescence properties of TPA derivatives by imparting a certain degree of conformational flexibility to the system. This flexibility influences how the molecule behaves upon photoexcitation, where the excited state can relax through two primary pathways: radiative processes, such as fluorescence emission, or through non-radiative mechanisms like internal molecular rotation. The latter typically occurs around the σ -bonds of the vinyl linkers, which are particularly sensitive to the surrounding microenvironment of the probe. The degree of intramolecular rotation around these bonds can significantly affect the optical properties of the TPA derivatives. By altering the microenvironment—such as changing the solvent viscosity or synthesizing more rigid molecular frameworks—the intramolecular rotation can be restricted. This phenomenon, known as restriction in intramolecular rotation (RIR), is critical as it directly modulates the optical properties of the molecule. Enhanced RIR often leads to increased fluorescence efficiency, as it reduces the non-radiative decay

pathways, allowing more energy to be released as light. Therefore, controlling the flexibility of the vinyl linker and the surrounding environment is essential for fine-tuning the photophysical behaviour of these compounds for specific applications.

2.3.3. Interactions with (bio)molecules

Interaction with DNA: The incorporation of cationic groups as terminal acceptors in the TPA derivatives not only enhances the solubility but also provides a functional motif that is well-suited for interacting with the grooves of DNA double helices, facilitating these derivatives to function as DNA-binding probes. The absorption of the TPA-based octupolar systems TP-1Py, TP-2Py and TP-3Py with one, two or three arms terminating in pyridinium moieties were investigated by Allain *et al.*, in the presence of duplex DNA. The study revealed that the interaction with DNA induces significant fluorescence enhancement in TP-2Py and TP-3Py, accompanied by red shifts in absorption maxima, clearly indicating the binding of these TPA derivatives to DNA. The observed DNA-induced fluorescence enhancement was attributed to restriction in intramolecular rotation (RIR) caused by the stacking or anchorage of the TPA molecules within the DNA grooves. This anchorage immobilizes the molecule, thereby reducing its internal flexibility and enhancing fluorescence. Additionally, the sequestration of the probe within the hydrophobic grooves of DNA, away from the polar aqueous environment, likely contributes to this fluorescence increase, as highly polar solvents such as water are known to significantly reduce quantum yield values.

Interestingly, it was also observed that the bis and tris derivatives exhibit similar binding affinities with the DNA duplex, suggesting that only two branches are buried within the DNA, while the third arm of the tris derivative remains outside the helical structure. This remaining arm undergoes rotational de-excitation, leading to reduced quantum yields for TP-3Py compared to TP-2Py. Based on the shape of the TPA derivatives, it was proposed that these molecules are likely to be involved in surface binding by inserting into the minor grooves of the DNA, which is typical for arc-shaped, nonplanar molecules. However, partial intercalation of the extra helical single branch of the tris derivative cannot be entirely ruled out¹⁶¹.

Interaction with proteins: Dumat *et al.* demonstrated that the introduction of variable terminal acceptor groups could regulate the affinity of the derivatives towards other biomolecules like proteins (human serum albumin). This interaction was accompanied by an enhancement in fluorescence quantum yield and a blue-shifted emission wavelength.¹⁵⁹ Gong *et al.*, reported a ~ 40- fold enhancement of their triphenylamine-based all-organic compounds upon addition of proteins like egg albumin and fetal bovine serum (FBS).¹⁶³ The rationale behind this behaviour was elucidated through numerical docking simulations, suggesting that this observation results from the TPA derivative being enclosed within the hydrophobic core of bovine serum albumin.

Interaction with macrocycles: More recently, attempts have been made to utilize host-guest complexation with different macrocyclic hosts like cyclodextrin and cucurbit[n]uril, to impose a structural confinement of fluorogenic guest molecules to suppress non-radiative relaxation and improve the luminescence quantum yield and the lifetime of the photoexcited states¹⁶⁴. This strategy was extended to the TPA derivatives and the secondary ammonium salts of a TPA derivative was probed for its ability to form inclusion complexes with CB[7], β -CD, and dibenzo-24-crown-8 by monitoring the emission response of the TPA derivative¹⁶⁵ (Figure 2.5a). Further, a 1-naphthylmethylamine terminated TP derivative has been reported for its characteristic inclusion complex formation with CB[7] and β -CD hosts. These complexes exhibit a rare dual emission property in solution at room temperature, with a lower and higher energy band arising from a locally excited state and an intramolecular charge transition respectively. The formation of two distinctly different inclusion complexes based on the different host shapes, electrostatic surface potentials and cavity polarities results in dissimilar luminescence properties¹⁶⁶ (Figure 2.5b).

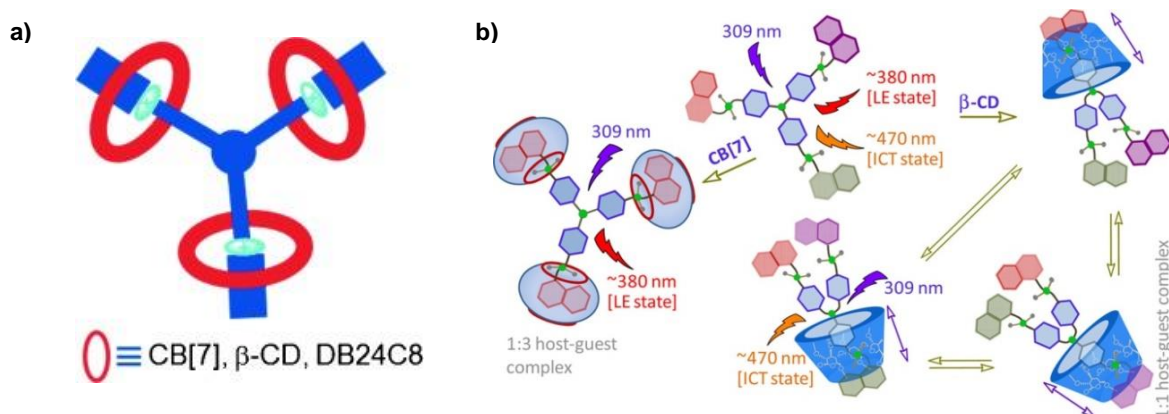


Figure 2.5: Schematic illustration of a) restricted conformational flexibility of TPA mediated by host-guest formation with hosts CB[7], β -CD, and dibenzo-24-crown-8 (adapted from Mandal, A.K. et al.¹⁶⁵), b) Tuning emission responses of a TPA derivative with a 1-naphthylmethylamine acceptor group by host-guest complexation with CB[7] and β -CD, resulting in an unusual dynamic inclusion phenomena (adapted from Gangopadhyay, M. et al.¹⁶⁶).

Thus, the integration of TPA derivatives as optical transducers with CB[n] receptors provides a comprehensive sensor design that could be further investigated and optimized to develop the proposed sensor array.

2.4. Results and Discussion: TPA-CB[n] fluorescent array-based sensor

2.4.1. Design strategy for construction of sensor array

Herein, our sensor array was designed initially by utilizing the macrocyclic host CB[n] as the recognition element and a library of TPA derivatives as the transduction elements.

TPA, featuring terminal positively charged acceptor groups, encapsulate within the hydrophobic cavity of CB[n] via host-guest interactions. This encapsulation provides the structural confinement needed to enhance radiative de-excitation by RIR, thereby promoting the function of the TPA as optical transduction elements. It enables the TPA to be engaged in competitive indicator displacement assays in the presence of analytes, further modulating their optical properties and generating a distinct optical response that is characteristic of the interacting analytes and suitable to design a sensor array.

While alternative dye molecules could have also been utilized for this purpose, the structure of the TPA derivatives was hypothesized to undergo partial encapsulation within the CB[n] cavity, thereby providing uncomplexed regions as additional binding sites for analytes. This feature leverages the tendency of TPA derivatives to interact with biomolecules like DNA or proteins and, hence, provides an orthogonal recognition scaffold for interacting analytes.

By exploiting this possible bimodal interaction of the TPA derivatives and their complex with CB[n], we hypothesised the possibility of bringing additional modes of interactions for the analytes while doubling the number of output channels for each sensing element of the sensor array (TPA and TPA+CB[n]), by retaining a limited number of sensing elements in the array (Figure 2.6).

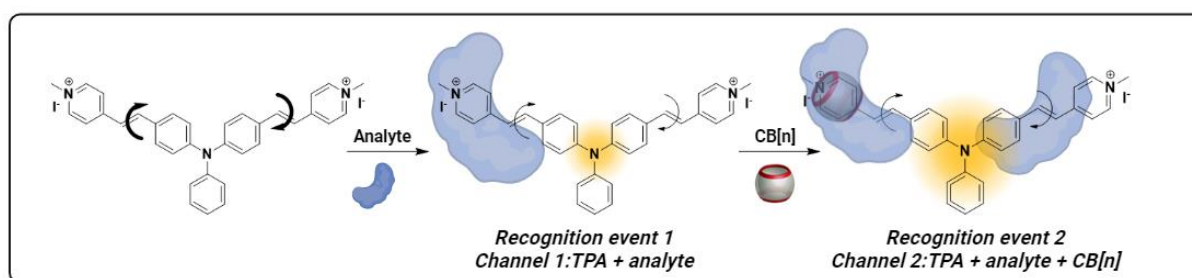
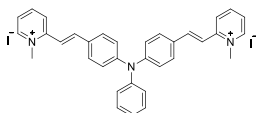
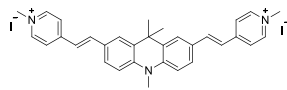
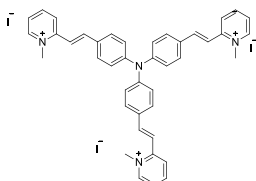
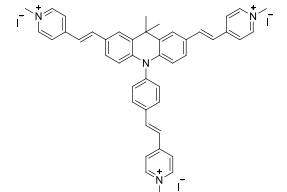
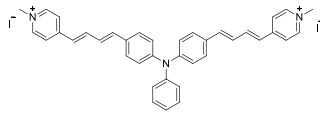
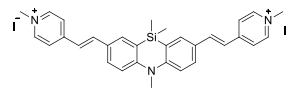
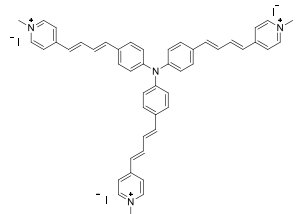
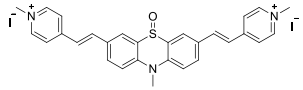


Figure 2.6: Schematic illustration of bimodal recognition of analytes using a macrocyclic fluorescence-based sensor array. Fluorescence emission of TPA is modulated sequentially by restriction in intramolecular rotation first by TPA alone and then by TPA-CB[n] host-guest complex.

2.4.2. Diversity of library of triphenylamine derivatives

The recognition elements of the proposed sensor array were selected from an available library of TPA derivatives synthesized by Dr. Delphine Naud-Martin from the team of Dr. Florence Mahuteau-Betzer (*Institut Curie*) (Table 2.1).

Table 2.1: Chemical structures and diversity of the library of TPA derivatives used for the development of fluorescence based sensor array.

Name	Structure	Name	Structure
TP_2Pyo		AcriPy	
TP_3Pyo		Acri_3Py	
DV_2Py		PhenazPy	
DV_3Py		SulfoxPy	

A variety of donor cores, including the fundamental triphenylamine backbone, along with the acridane, phenazasiline and phenothiazine derivatives were evaluated. The atoms added to the central core of these derivatives play an important role in determining their photophysical properties, as indicated by molecular modeling calculations conducted on a few quadrupolar members of the library. The study revealed that the major contribution to the charge transfer transition of these derivatives from the ground state (S_0) to the first excited state (S_1) comes from the HOMO to LUMO transition. While the electronic density of the HOMO is predominantly located on the donor core, the LUMO is located on the branches of the derivative. This remains consistent with the quadrupolar design of the A- π -D- π -A system. Since

the energy levels of the LUMO in these systems shows only slight variations, the energy difference (ΔE) of the HOMO-LUMO gap is primarily influenced by the energy of the HOMO. Although the shape of the HOMO is similar across these derivatives, the inductive effect of the atoms added to modify the central core significantly determines the ΔE and, consequently, the absorption wavelength of the corresponding derivative. For instance, the inductive effect of the electron-attracting groups of phenothiazine derivatives lowers the energy of the HOMO orbitals, while the electron-donating groups of triphenylamine, acridane and phenazasiline tend to increase their energy. This results in a larger HOMO-LUMO energy gap for the former, accompanied with a decrease in the maximum absorption wavelength and a corresponding decrease in the HOMO-LUMO energy gap for the latter, with an associated increase in the maximum absorption wavelength.

Furthermore, this variation in absorption wavelength inversely affects the fluorescence quantum yield, leading to corresponding changes in the brightness of the respective TPA derivative¹⁶⁷. This variation in photophysical properties essentially facilitates the introduction of diversity in the sensor units by the generation of differential optical outputs across a wide range of emission wavelengths. This, in turn, is known to improve the differentiating index of the array¹⁶⁸, potentially offering more information on sensor-analyte interactions, without the overlapping of generated optical outputs.

These central donor cores are coupled through vinyl linkers that vary in length, featuring single or double C=C bond to the terminal pyridinium acceptor moiety either at the ortho- or para- positions. The binding of the organic cationic group of pyridinium derivatives with the CB[n]'s has been well-studied¹⁶⁹. For instance in CB[6] complexes with alkyipyridinium cations, the alkyl chain is encapsulated within the cavity, while the charged nitrogen interacts with the carbonyl portals^{170,171}. Similarly, CB[7] forms inclusion complexes with the alkyl chain of 4-naphthyl-N-alkyipyridinium cations¹⁷² and one of the alkyl chains of dialkyl-4,4'-bipyridiniums (viologens) when the alkyl chain are longer¹⁷³. For derivatives with bulky substituents linked to the pyridinium ring N atom, CB[7] encapsulates only the pyridinium ring, whereas CB[8] fully encapsulates the large guest¹⁷⁴. Furthermore, CB[8] can bind a pair of stacked N-benzylpyridinium¹⁷⁵ or 4-aryl-N-methylpyridinium¹⁷⁶ derivative guests in its cavity. Studies on the complexes of such bulky pyridinium cations with CB[7] and CB[8] has led to the conclusion that the CB[n] hosts encapsulate these guests or shuttles along them, depending on the substitution of the bulky pyridinium moiety, the length of the alkyl chain of the guest and the cavity size of the host^{177,178}.

Additionally, the number of branches also contributes to the diversity of the TPA derivatives, as observed in the interaction with DNA. The host-guest encapsulation of the TPA with CB[n] could result in the encapsulation of all or just some branches of the bis and tris derivatives, leaving behind un-complexed branches that may contribute to additional interactions with analyte molecules and enhance the bimodal recognition properties of the sensor array system.

Thus, these structural properties of the selected TPA derivatives significantly contribute to their ability to function as effective guests for complexation with cucurbit[n]urils and play a crucial role in the development of an efficient fluorescence-based sensor array.

2.4.3. Characterization of TPA-CB[n] complexation

Then, we experimentally investigated the interaction of the library of TPA derivatives with the macrocyclic family of the CB[n]'s^{165,166,179,180}. The host-guest complexation between TPA and CB[n] was expected to be driven by hydrophobic interactions between the inner cavity of CB[n] and the aromatic moieties of TPA, as well as ion-dipole interactions between the carbonyl groups lining the CB[n] portals and the protonated amines of TPA. The TPA-CB[n] interaction was anticipated to constrain TPA and influence the process of electron-transfer in the excited states, resulting in tuneable photophysical properties, such as shift in absorption wavelength and enhancement of fluorescence quantum yields¹⁸¹. We, therefore, evaluated the photophysical characteristics of the library of TPA derivatives with three different members of the CB[n] family, CB[6], CB[7] and CB[8], using UV-visible and fluorescence spectroscopy.

2.4.3.1 Modulation of TPA absorbance properties by complexation with CB[n]

We firstly analysed the absorbance spectra of the eight free TPA (5 μM) derivatives followed by their complex with CB[n]'s (150 μM) scanned over a wavelength range of 400-650 nm in aqueous solution. A distinct absorption band was observed for the different TPA derivatives, which can be attributed to the triphenylamine-based $\pi\text{-}\pi^*$ transition, with high molar extinction coefficient around 50,000 $\text{M}^{-1} \text{cm}^{-1}$ and wavelength maxima between 430 to 490 nm. The absorption maxima remained unchanged in the presence of CB[6] (Figure 2.7), suggesting a lack host-guest complexation, possibly due to the larger size of the derivatives as compared to the size of the hydrophobic cavity of CB[6]. However, a significant red-shift in the absorption maxima wavelength was observed in the presence of CB[7] (~25-40 nm) and CB[8] (~60-70 nm), a hallmark indicating the formation of host-guest complexes (Figure 2.8, 2.9). This could be attributed to the strong electron absorbing effect of carbonyl in the CB[n] hosts, which can neutralize the positive charge of the pyridinium. This neutralization enhances the electron-withdrawing away from the triphenylamine core, inducing the spectral red-shift. Meanwhile, between CB[7] and CB[8], the larger cavity of CB[8] could potentially enable the encapsulation of two TPA derivatives within its cavity, leading to a stable aggregation of the guests to form a $\pi\text{-}\pi$ complex and resulting in a more pronounced red-shift. The variation in the absorbance properties for the library of TPA derivatives is provided in Table 2.2.

Thus, the distinct modulation in the absorbance properties of the TPA derivatives leads to two key conclusions. Firstly, the wavelength shift confirms the successful host-guest complexation of TPA with CB[7] and CB[8], while no such complexation is observed with CB[6], as indicated by the negligible wavelength shift in the absorbance spectra of the TPA's with CB[6]. Secondly, a significant wavelength shift was observed for the TPA complexes with CB[8], and this large shift was identified as an ideal feature for the development of a differential sensor array. This insight led to the development of a colorimetric sensing array that exploits the variation in the absorbance properties of selected TPA-CB[7] and TPA-CB[8] systems. The details of this colorimetric array will be discussed in Chapter 6 of this thesis.

Table 2.2: Shift in absorbance maxima of TPA derivatives upon complexation with CB[6], CB[7] and CB[8]

Name	$\lambda_{\text{max(abs)}}$ (TPA)	$\Delta \lambda_{\text{(abs)}}$ (TPA+CB[6]) nm	$\Delta \lambda_{\text{(abs)}}$ (TPA+CB[7]) nm	$\Delta \lambda_{\text{(abs)}}$ (TPA+CB[8]) nm
TP_2Pyo	460 nm	0	40	66
TP_3Pyo	452 nm	0	42	60
DV_2Py	488 nm	0	23	60
DV_3Py	486 nm	18	21	62
Acri_3Py	486 nm	0	40	60
PhenazPy	472 nm	0	42	67
AcriPy	486 nm	0	39	71
SulfoxPy	430 nm	8	27	44

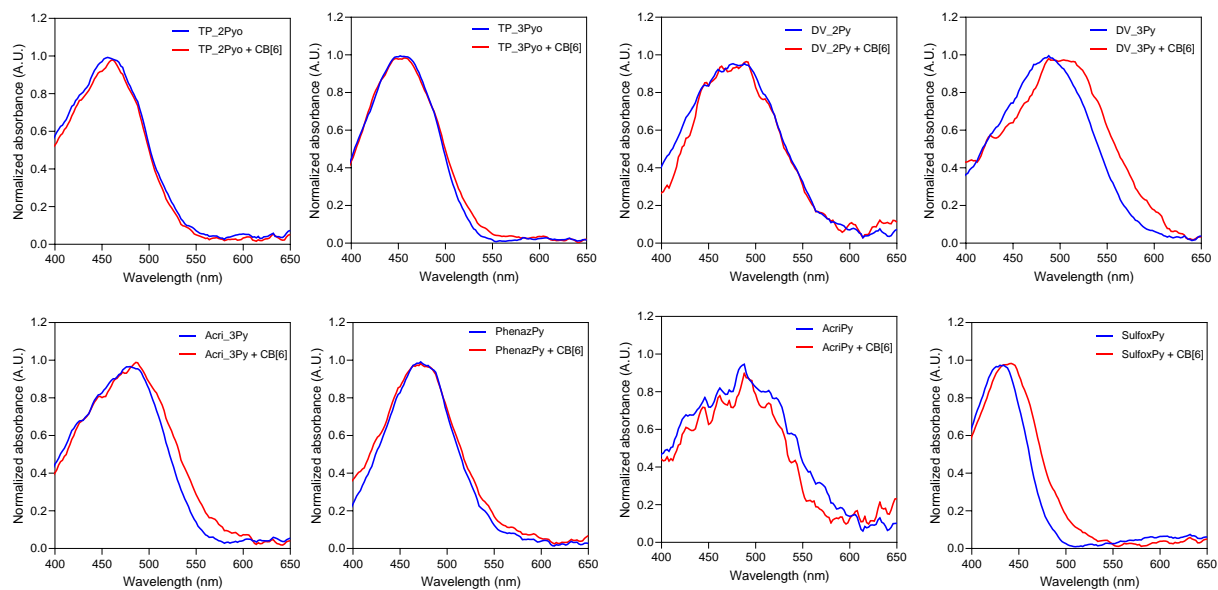


Figure 2.7: a) UV-visible absorbance spectra of library of TPA derivatives scanned over wavelength range of 400-650 nm. Blue trace corresponds to the absorbance spectra of free TPA derivative at (5 μM) while the red trace corresponds to the complex of the respective TPA derivative with CB[6] (150 μM).

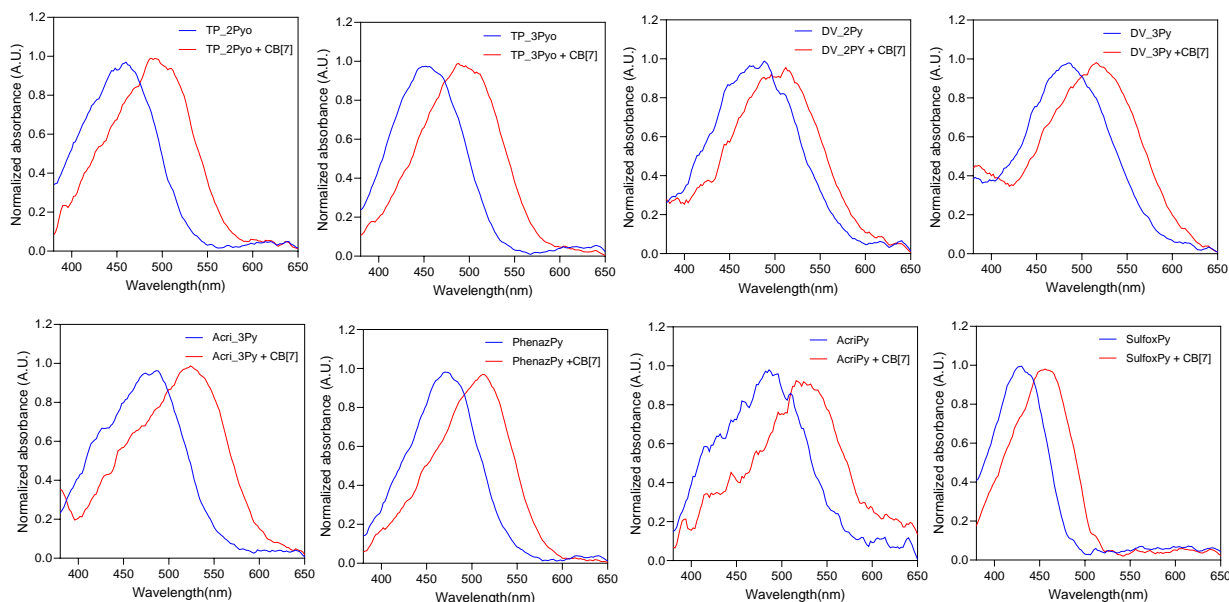


Figure 2.8: a) UV-visible absorbance spectra of library of TPA derivatives scanned over wavelength range of 400-650 nm. Blue trace corresponds to the absorbance spectra of free TPA derivative at (5 μM) while the red trace corresponds to the complex of the respective TPA derivative with CB[7] (150 μM).

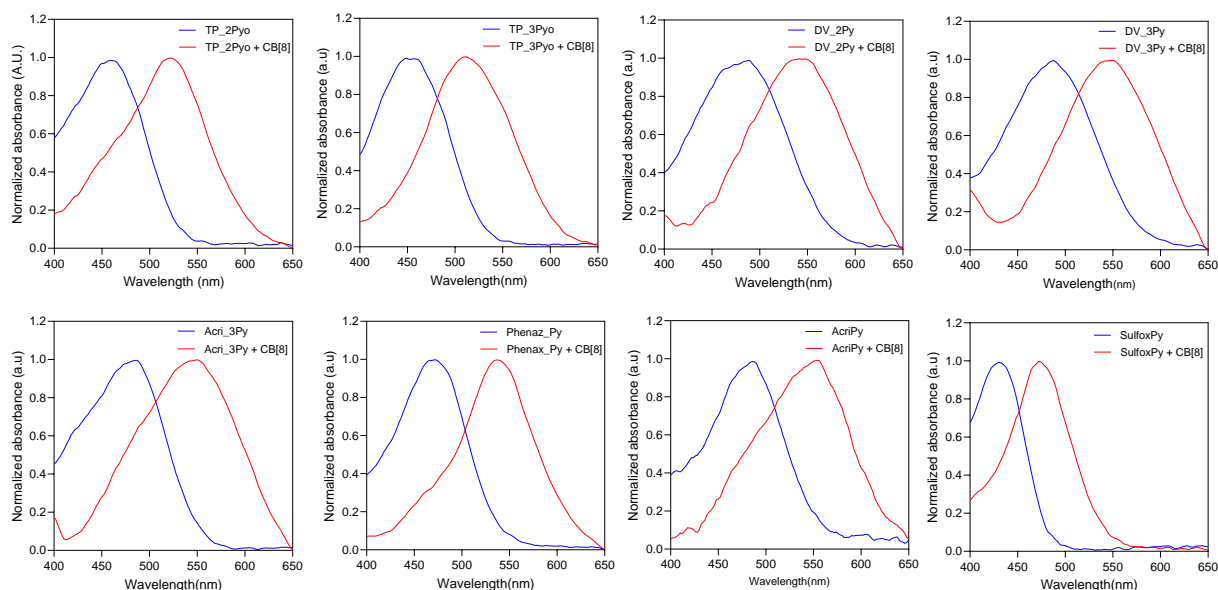


Figure 2.9: a) UV-visible absorbance spectra of library of TPA derivatives scanned over wavelength range of 400-650 nm. Blue trace corresponds to the absorbance spectra of free TPA derivative at (5 μM) while the red trace corresponds to the complex of the respective TPA derivative with CB[8] (150 μM).

2.4.3.2 Modulation of TPA fluorescence emission properties by complexation with CB[n]

We then proceeded to analyze the fluorescence emission spectra of the TPA library with the CB[n] hosts. To ensure a consistent comparison of fluorescence enhancement with the three CB[n]'s, and considering the observed subtle or pronounced bathochromic shift in the absorption wavelength induced by complexation, we selected the isosbestic wavelength between the absorption spectra of TPA and TPA-CB[n] complexes as the excitation wavelength for the fluorescence emission measurements. The fluorescence emission spectra of the TPA derivatives in the presence of the CB[n] hosts showed variable responses. CB[6] induced a negligible change in fluorescence intensity (Figure 2.10). A small increase or even a decrease was observed in the presence of CB[8] (Figure 2.12). This behaviour could be explained by the possibility that, initially, the addition of CB[8] to the solution of TPA leads to the formation of a 2:1 complex, accompanied with an increase in fluorescence emission. However, this enhancement reaches its maxima at a specific concentration of CB[8] beyond which the fluorescence emission is often decreased due to the prevalence of 1:1 complexes as opposed to 2:1 complexes when CB[8] is present in excess. This shift in complexation may trigger complex photophysical mechanisms that alter the fluorescence properties of the TPA derivatives resulting in the observed behaviour. In contrast, a significant fluorescence enhancement was monitored in the presence of CB[7] (~5-50 fold)

suggesting efficient encapsulation of TPA's within CB[7] for the formation of a 1:1 complex, leading to enhanced restriction in intermolecular rotation and higher radiative de-excitation (Figure 2.11). The superior fluorescence enhancement of TPA derivatives depicted in the investigation of the photophysical properties, coupled with the better water solubility criterion, established CB[7] as the optimal choice for the complexation of TPA's in the context of developing a fluorescence-based sensor array. The variation in the fluorescence properties for the library of TPA derivatives with CB[7] is provided in Table 2.3. This sensor array will be discussed in detail in the remainder of this chapter, as well as in Chapter 4 and Chapter 5 of this thesis.

Table 2.3: Shift in fluorescence emission maxima and extent of fluorescence enhancement of TPA derivatives upon complexation with CB[7].

Name	$\lambda_{\text{isosbestic}}$ (nm)	$\lambda_{\text{max(ems)}}$ (TPA) (nm)	$\lambda_{\text{max(ems)}}$ (TPA+CB[7]) (nm)	$\Delta \lambda_{\text{(ems)}}$	Fluorescence enhancement at $\lambda_{\text{max(ems)}}$ (TPA+CB[7])
TP_2Pyo	471	664	644	-20	21.5 fold
TP_3Pyo	472	660	658	-2	18.8 fold
DV_2Py	487	764	701	-63	7.5 fold
DV_3Py	497	714	704	-10	7.4 fold
Acrid_3Py	497	676	654	-22	3.4 fold
PhenazPy	486	660	640	-20	38.7 fold
AcridPy	497	680	665	-15	29.7 fold
SulfoxPy	444	590	552	-38	36.7 fold

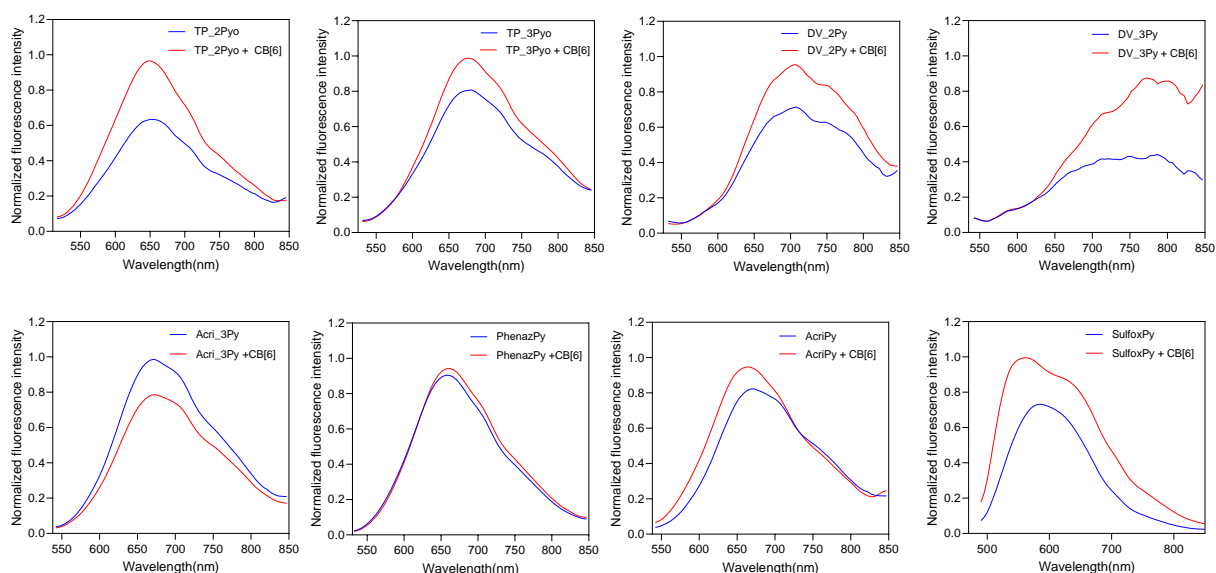


Figure 2.10: Normalized fluorescence intensity recorded at the emission maxima of the TPA derivatives at $5 \mu\text{M}$ in the presence of CB[6] at $150 \mu\text{M}$. Excitation wavelengths correspond to isosbestic point indicated in Table 2.3. Blue trace corresponds to the emission spectra of free TPA derivative while the red trace corresponds to the complex of the respective TPA derivative with CB[6].

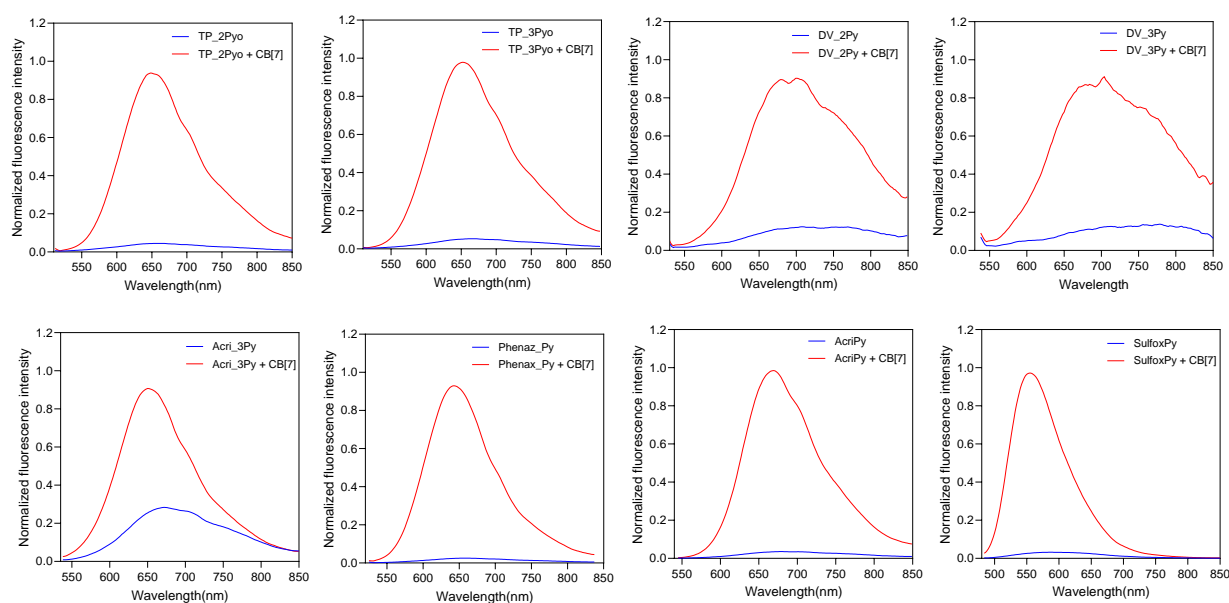


Figure 2.11: Normalized fluorescence intensity recorded at the emission maxima of the TPA derivative library at $5 \mu\text{M}$ in the presence of CB[7] at $150 \mu\text{M}$. Blue trace corresponds to the emission spectra of free TPA derivative while the red trace corresponds to the complex of the respective TPA derivative with CB[7].

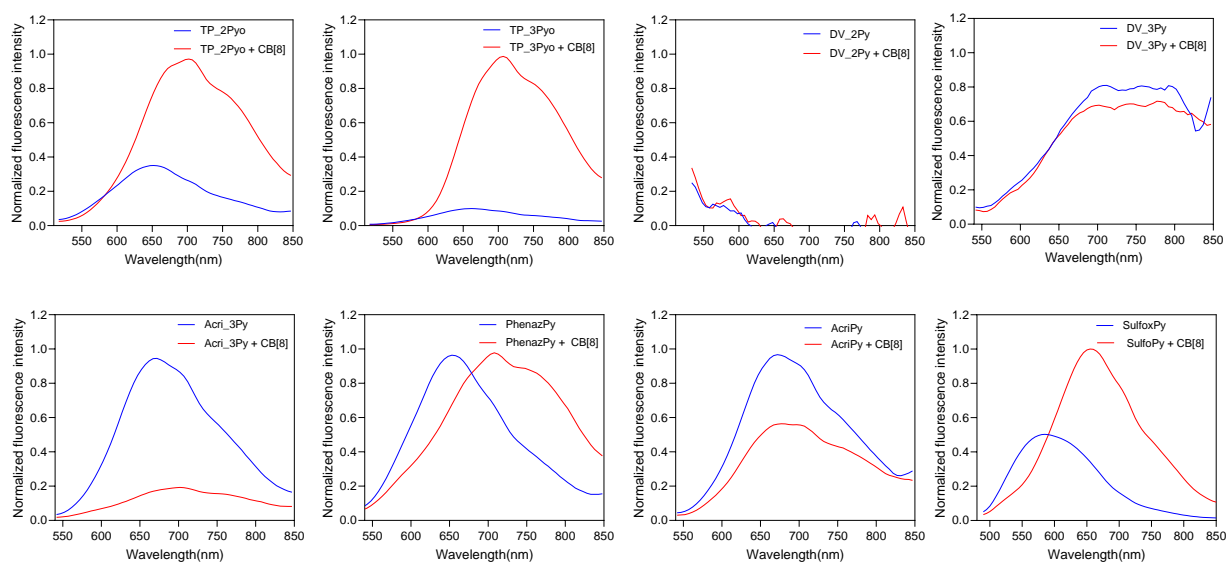


Figure 2.12: Normalized fluorescence intensity recorded at the emission maxima of the TPA derivative library at $5 \mu\text{M}$ in the presence of CB[8] at $150 \mu\text{M}$. Blue trace corresponds to the emission spectra of free TPA derivative while the red trace corresponds to the complex of the respective TPA derivative with CB[8].

2.4.4. Construction and characterization of a fluorescent based sensor array

Constructing an optimized sensor array requires careful consideration of factors like the dimension of the sensor array, the sensitivity of the output signals and the chemical diversity among the recognition elements. This ensures that the sensor array is equipped to interact with analytes by multiple binding modes to ensure identification and classification of even subtle variations among analytes in complex sensing conditions.

2.4.4.1. Selection of TPA-CB[7] sensing elements

After the initial optical characterizations, the most promising candidates were selected from the eight members of the TPA library. This selection was based on the degree of their fluorescence enhancement properties and the chemical structure of their acceptor/donor scaffolds.

The enhancement of the fluorescence intensity of each TPA derivative upon addition of CB[7] was monitored and the derivatives with the maximal fluorescence enhancement during host-guest complexation were selected (Figure 2.13b). This significant enhancement was expected to provide a higher signal-to-noise ratio and reduced background signal to account for the possible decrease in fluorescence in the presence of analytes with higher binding affinities that could engage in competitive displacement of the TPA derivatives.

The DV_2Py and DV_3Py derivatives, featuring two double bonds along their vinyl branches were not selected due to their limited fluorescence enhancement with CB[7]. This could be attributed to the greater ease of rotation around two double bonds as opposed to one, which reduces RIR and consequently, fluorescence enhancement. This narrowed the selection to the two remaining 3-arm derivatives, TP_3Pyo and Acri_3Py. Additionally, two 2-arm derivatives, PhenazPy and Sulfoxy, with the highest fluorescence enhancement were also selected. Notably, the 3-arm derivatives provided an ortho (TP_3Pyo) and a para (Acri_3Py) terminal methyl pyridinium derivative to the sensor array (Figure 2.13c), that most likely introduce diversity in the extent of interaction within the CB[7] cavity based on their different binding affinities and, consequently, influence the displacement in the presence of the analyte molecules.

The photophysical and CB[7] binding properties of this optimized sensor array was further investigated to establish the suitability of these TPA-CB[7] pairs for the development of the macrocyclic fluorescence based sensor array.

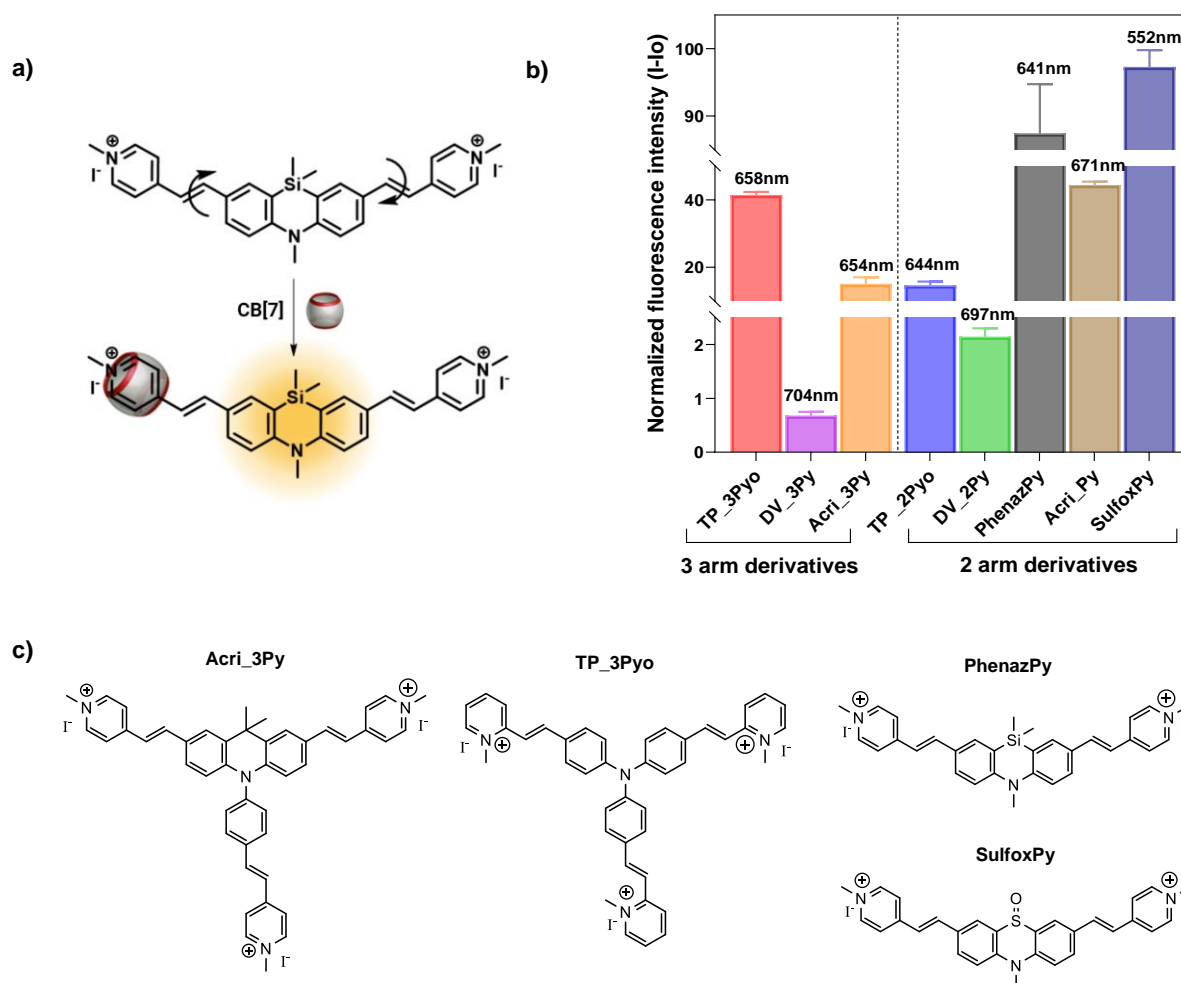


Figure 2.13: a) Schematic illustration of fluorescence enhancement resulting from the restriction in intramolecular rotation provided by the host-guest interaction of TPA derivatives with CB[7]. b) Normalized fluorescence intensity recorded at the emission maxima of the TPA derivative library at 5 μM in the presence of CB[7] at 150 μM . c) Chemical structures of the triphenylamine derivatives selected for the optimized sensor array.

2.4.4.2. Optical characterization of optimized sensor array

To further elucidate the differences in the nature and strength of the intermolecular forces driving the host-guest complexation, it is essential to determine the dissociation constants (K_d) of the TPA-CB[7] complexes. This information is crucial for understanding the range of binding affinities of various target analytes that can be considered to facilitate the design of efficient IDA systems.

We firstly examined the effect of addition of increasing concentration of CB[7] on the absorption of the four selected TPA derivatives. The main absorption peak for TP_3Pyo, Acri_3Py, PhenazPy and SulfoxPy was observed at 452, 486, 472 and 430 nm respectively. With the maximal addition of 250 μM of CB[7], the peak position of the absorbance maxima was shifted to 488, 524, 512 and 458 nm

with a $\Delta\lambda$ value of 36, 38, 40 and 28 nm accompanied with change in colour of the aqueous solution (Figure 2.14).

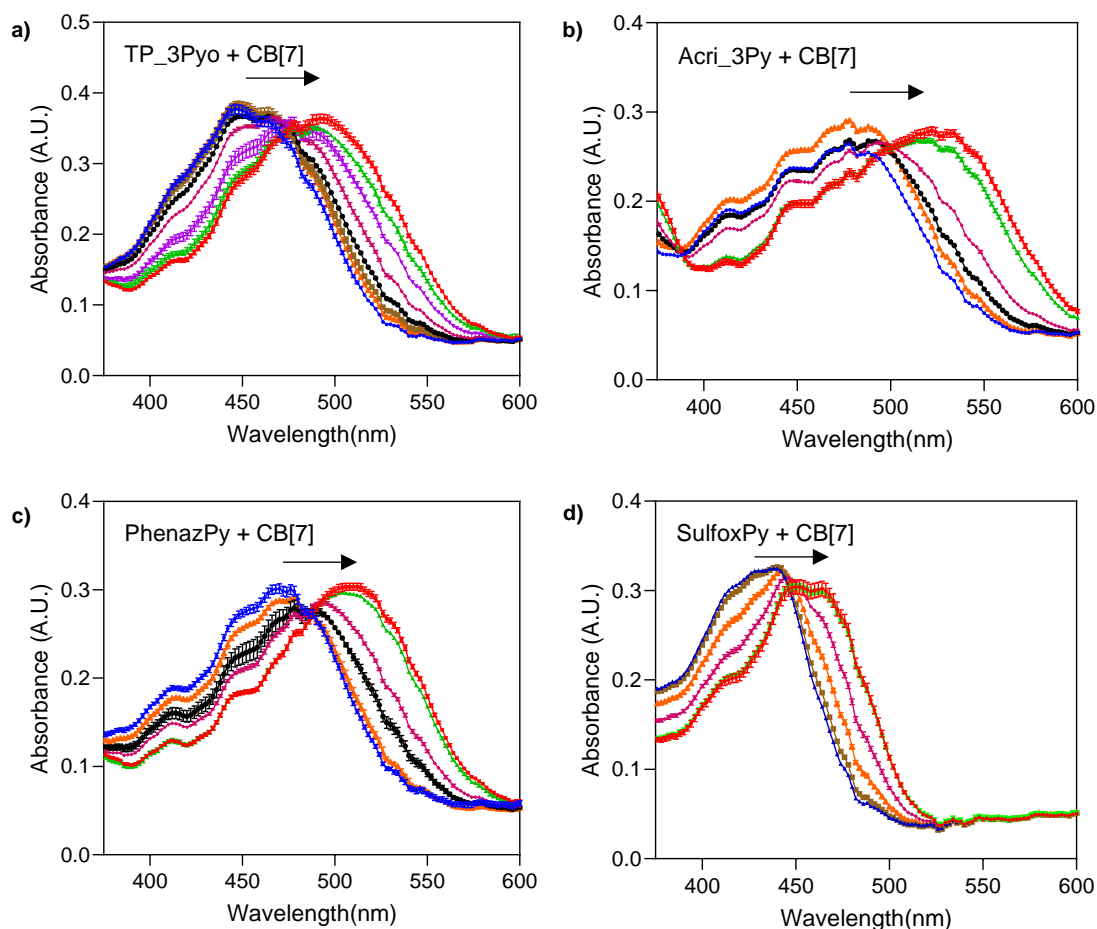


Figure 2.14: UV-visible absorbance spectra of a) TP_3Py, b) Acri_3Py, c) PhenazPy and d) SulfoxPy in the presence of increasing concentration of added CB[7] (serially diluted from 250 μM to 0.9765 μM). Blue trace corresponds to the absorbance spectra of free TPA derivative at (10 μM) while the red trace corresponds to the complex of the respective TPA derivative with CB[7] at maximum concentration (250 μM).

This was followed by the fluorescence titration of the derivatives in order to estimate a range of magnitude of dissociation constants for the corresponding TPA-CB[7] complexes formed. The global dissociation constant K_d (Figure 2.15, middle panel) can be extracted from the fluorescence plot and K_d of the TPA derivatives with CB[7] were determined to be 95.07 μM , 37.79 μM , 14.84 μM and 8.13 μM with 95% confidence intervals of [77.19 to 118.90], [21.03 to 72.41], [12.16 to 18.17] and [6.08 to 10.93], for TP_3Py, Acri_3Py, PhenazPy and SulfoxPy respectively.

We observed that the two-arm derivatives, PhenazPy and SulfoxPy have a slightly lower K_d as compared to the three-arm derivatives. In addition, the higher K_d of TP_3Py as compared to Acri_3Py

could be ascribed to the higher affinity of the terminal para substituted pyridinium in Acri_3Py as compared to the ortho substitution in TP_3Pyo. The K_d of the TPA-CB[7] complexes fall within the micromolar range, which indicates a sufficiently high affinity of the TPA for CB[7] as compared to other possible low binding interferents. This desired range, ensures an effective competition between TPA and potential bioanalytes, as expected in a typical IDA, with the equilibrium maintained at an ideal balance between a higher concentration of TPA-CB[7] complex in the absence of the analytes, while allowing for a possible displacement of TPA in the presence of analytes with higher affinity¹⁵⁵. We further compared these obtained K_d values to the values extracted from the EC50 determination, which represents the concentration of CB[7] required to complex half of the TPA binding sites, irrespective of the stoichiometry (Figure 2.15, far right panel) and obtained similar values. A comparison of the K_d values obtained from the fluorescence titrations of the TPA- CB[7] complexes are presented in Table 2.4.

Table 2.4: Dissociation constant and EC50 values determined for the host-guest complexation of TPA-CB[7] sensing elements .

	K_d (μM)	95% Confidence interval	EC50 (μM)	95% Confidence interval	R^2
TP_3Pyo	95.07	77.19 to 118.90	54.03	48.18 to 60.70	0.9945
Acri_3Py	37.79	21.03 to 72.41	30.52	22.32 to 42.13	0.9525
PhenazPy	14.84	12.16 to 18.17	11.22	9.539 to 13.20	0.9864
SulfoxPy	8.13	6.08 to 10.93	5.88	4.522 to 7.656	0.9608

Based on these experiments, the concentration of the TPA derivatives for the sensing was maintained at 5 μM to facilitate the fluorescence measurement variation in the presence of several analytes in multiple replicates to ensure efficient statistical analysis for analyte classification and discrimination. Fluorescence titrations with CB[7] at this TPA concentration revealed a saturation of fluorescence

enhancement around 125 μM , thus prompting the selection of a concentration of 100 μM for all subsequent analysis.

We further utilized the fluorescence titrations to analyse the stoichiometry of the TPA-CB[7] interactions using the continuous variation method. This method better known as the Job plot has been a common method to determine the stoichiometry of interactions during the process of host-guest complexation. However recent concerns from various groups have challenged the utility of this method in analysing supramolecular binding interactions, especially those involving the formation of more than one type of complex at equilibrium¹⁸²⁻¹⁸⁴. These challenges were evident in our results when applying this method to the TPA-CB[7] system. While theoretically, the formation of CB[7] complexes could reach a maximum of 2:1 or 3:1, based on the number of terminal pyridinium groups available in the multi-branched 2-arm and 3-arm systems respectively, we failed to obtain Job plots to support this, particularly for the 3-arm derivatives. Therefore, the Job plot can only be used for additional positive confirmation once the equilibrium association constants for both 1:1 (K_1) and 1:2 (K_2) host-guest complexes have been established by other titration experiments and it is not suitable for analysing more complex supramolecular systems like the TPA-CB[7] system. Thus, we have explored the utilization of other methodologies to establish the various parameters associated with the TPA-CB[7] host-guest binding interactions.

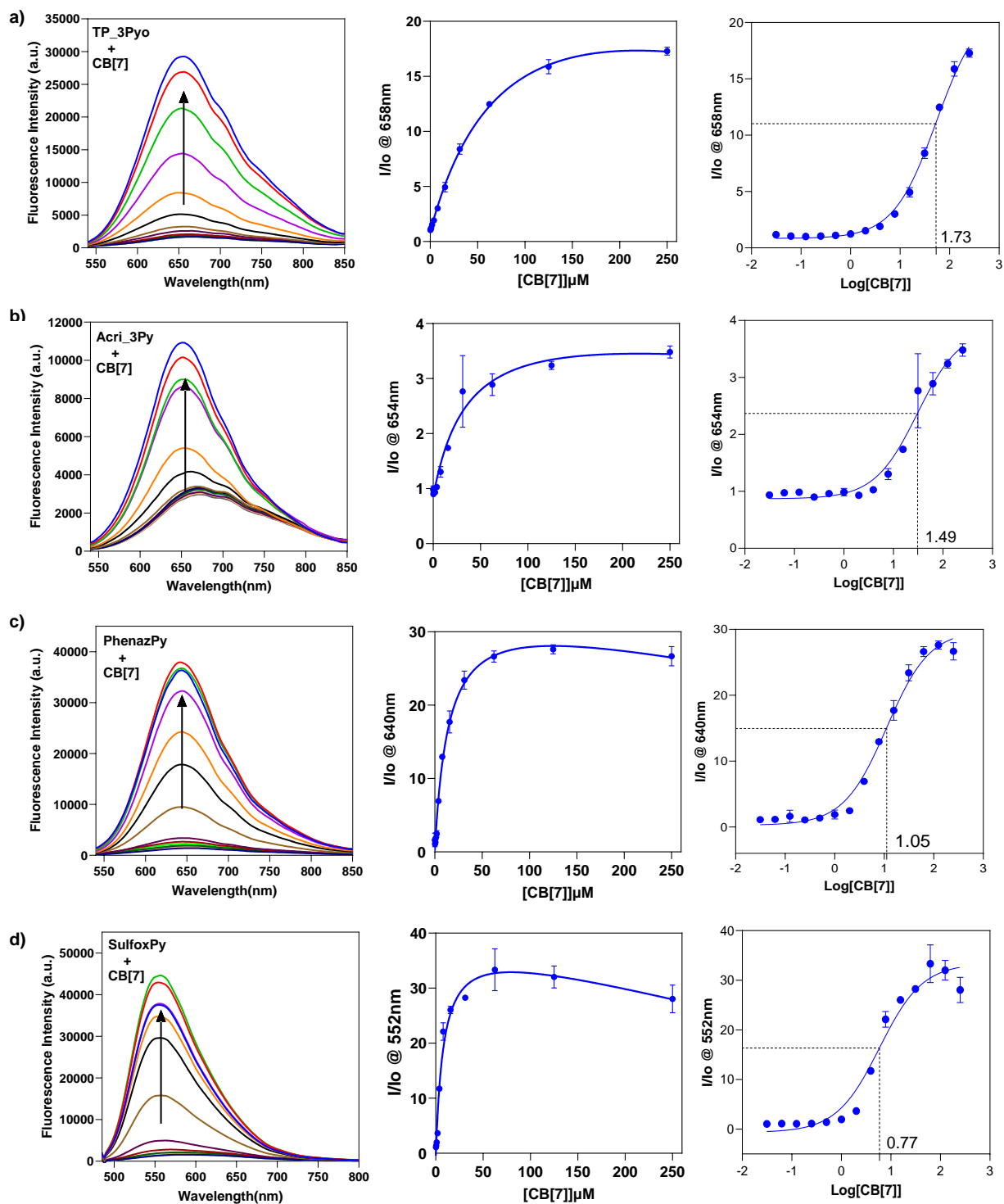


Figure 2.15: Fluorescence titration of a) TP_3Pyo, b) Acri_3Py, c) PhenazPy and d) SulfoxPy (5 μM) by CB[7] (250 – 3.96 μM serially diluted). Middle panel represents the corresponding binding isotherms obtained by plotting the fluorescence enhancement I/I_o versus concentration of CB[7] for determination of dissociation constant (K_d). Far right panel represents EC_{50} determination from full dose-response curves of the TPA derivatives.

2.4.4.3. Determination of binding parameters by Isothermal titration calorimetry (ITC)

Introduction

The process of host-guest binding is fundamentally a molecular association phenomenon, resulting in either the release or consumption of heat from the surrounding environment. These energetic variations can be measured and provide valuable insights into the characteristics of the interaction¹⁸⁵. Isothermal titration calorimetry (ITC) has emerged as a powerful technique that offers comprehensive thermodynamic¹⁸⁵ and even kinetic¹⁸⁶ profiling in a single experiment. ITC requires no chemical modifications, allowing for direct measurement of the binding constant, stoichiometry, and heat of reaction. Additionally, ITC provides indirect access to other thermodynamic parameters such as entropic binding contributions and Gibbs free energy¹⁸⁷. It is essentially suitable for determination of binding

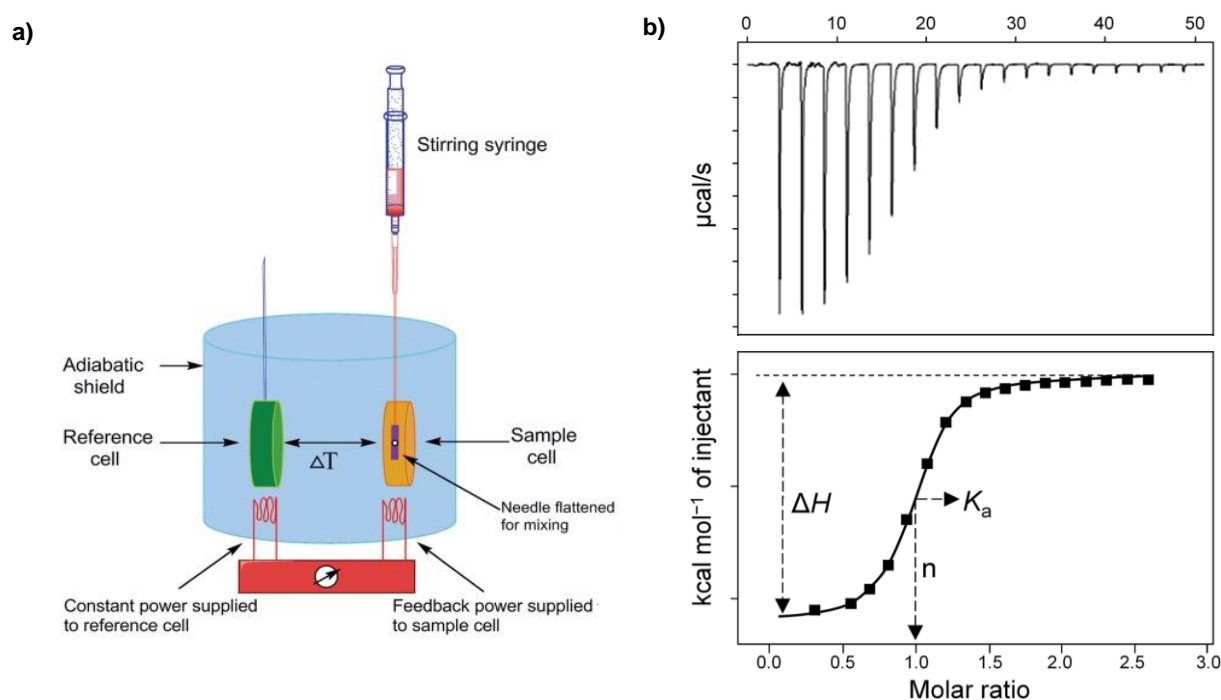


Figure 2.16: a) Schematic illustration of an isothermal titration calorimeter. b) Titration profile generated for an ITC experiment. The recorded power is plotted as a function of time, with each peak representing the heat generated upon injection of the ligand into the receptor solution in the calorimetric cell or vice-versa (upper panel). The titration curve is generated by the integration of peaks, depicting the heat of binding as a function of molar ratio (lower panel). Non-linear least-squares analysis allows determination of the parameters ΔH , n , and K_a (adapted from Song, C. et al.¹⁸⁸ and Callies, O. et al.¹⁸⁹).

affinities in the micromolar to nanomolar range¹⁹⁰. The general procedure of the experiment consists on the titration of ligand aliquots into a receptor solution, and the heat that is released (exothermic reaction) or taken up (endothermic reaction) is monitored in relation to a reference cell (Figure 2.16a). To further resolve cooperativity challenges, primarily observed in macromolecules, reverse titrations are also performed to produce reliable results¹⁹¹.

The output of the ITC experiment is represented as a plot of power required to maintain constant temperature as a function of time, with each peak indicating the thermal effects caused by the injection of the ligand to the receptor solution (or inversely). The area under each peak represents the heat energy changes upon injection. As the titration progresses and the binding site become more saturated, the heat variations decrease. When the titration reaches adequate saturation, the observed peaks with constant values will ultimately represent only the corresponding dilution heats, which are used for background analysis of the binding data (Figure 2.16b). The available raw data needs to be fitted into a suitable model that considers the nature of the ligand-receptor interaction. This fitting process is typically performed using the Wiseman isotherm (Equation 1). Wiseman *et al.*¹⁹² demonstrated that a parameter, c , defined as the product of the binding affinity K_a and the concentration of receptor or ligand $[cell]$ in the calorimetric cell, must fall within the range of 1 to 1000 for the reaction to be accurately characterized by ITC.

The variable n denotes the number of binding sites per molecule in the ITC cell device or the stoichiometry of binding interactions.

$$c = nK_a [Cell]_t \quad (1)$$

The integration of the generated peaks provides a titration curve from which various thermodynamic and kinetic properties can be derived. The amplitude of the peaks is proportional to the enthalpy of the interaction (ΔH), the inflection point indicates the corresponding stoichiometry (n), and the slope yields the binding affinity (K_a) or dissociation constant ($K_d = 1/K_a$) in case of sigmoidal titration curves (Figure 2.16b). Using the van't Hoff relationship, the Gibbs free energy can also be determined based on its link to K_a ($\Delta G = -RT \ln K_a$). Interpreting the calorimetric data reveals the nature of the physical phenomena driving the ligand-receptor interaction. Large favourable enthalpic contributions are primarily attributed to van der Waals interactions, hydrogen bonding, or electrostatic interactions, while unfavourable enthalpic contributions are associated with the desolvation of polar groups. Large entropy contributions indicate changes in the solvation of lipophilic and/or hydrophobic groups, which originate from the release of water molecules from the binding pocket upon complexation. Conversely, conformational changes that involve the loss of degrees of freedom are generally unfavourable in entropic terms¹⁹³. While conventional crystallography techniques enable the study of bimolecular interactions between receptors and ligands, a deeper analysis of the binding affinity among the species is elucidated from the thermodynamics of the association.

Therefore, ITC has emerged as a suitable tool to analyse host-guest complexation by directly measuring the binding enthalpy and the binding affinities in addition to the stoichiometry of the formed inclusion complexes.

ITC titrations of TPA derivatives and CB[7]

Thus, to determine these parameters and assess the stoichiometry of interaction we performed direct titrations by injecting CB[7] into the ITC device cell containing TPA derivatives in triplicates, establishing the binding enthalpy per mole of CB[7]. The raw ITC data was analysed by two or three site binding model i.e. the "ABBB three symmetric sites model", the "ABB two symmetric sites model", and the "AB heteroassociation model". The quality of the models employed for analysis is represented by curves calculated from average value of the three replicates in each thermogram at a confidence interval of $P=95\%$ (Figure 2.17). Initially, the TPA thermograms were analysed with 1:3 and 1:2 binding models.

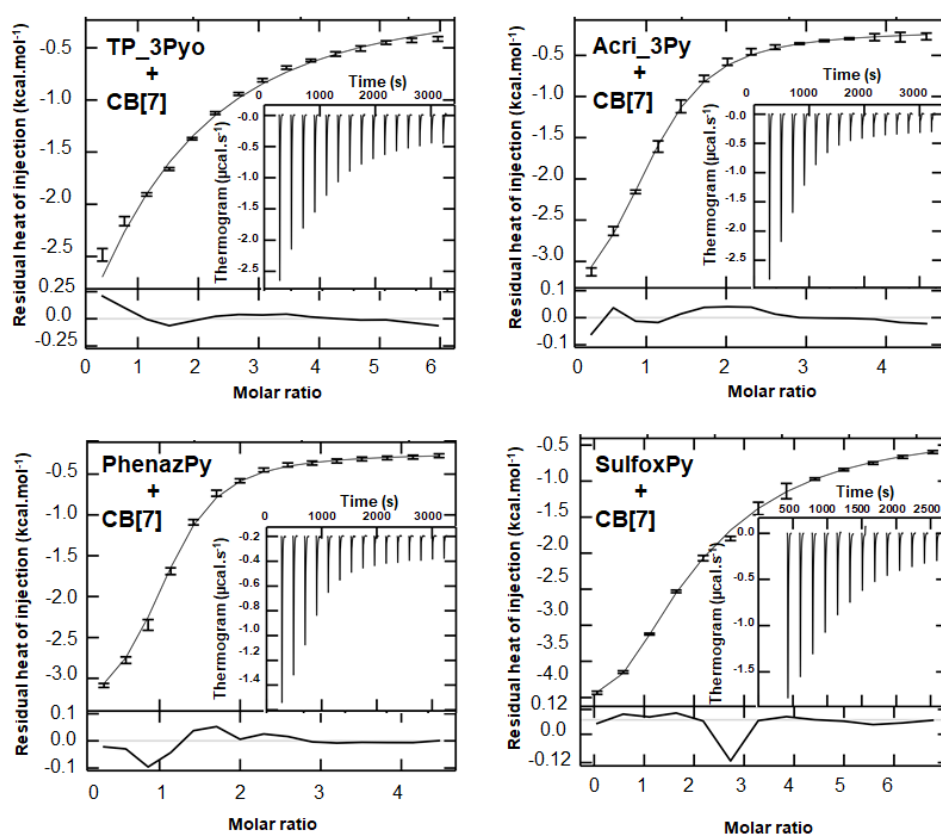


Figure 2.17: Isothermal titration calorimetry of TPA derivatives by CB[7] at 25°C and titrations curves calculated with one binding site (TP_3Pyo, Acri_3Py, PhenazPy) or two binding sites (SulfoxPy) from average values. Inset represents the raw thermogram.

However, cooperativity factors between the symmetric binding sites had to be refined to fit the data. This resulted in high negative cooperativity between the first and subsequent binding sites, which had very low affinities to be reliably titrated.

Consequently, the thermograms were analysed with a 1:1 binding model except for Sulfox-Py which had two titratable CB[7] binding sites. We hypothesized that the binding of CB[7] at the first site modified the electronic and structural properties of the TPA derivatives at the remaining sites. Indeed, the partial delocalization of positive charge from a terminal pyridinium group towards the core nitrogen atom can stabilize the first binding of CB[7] by electrostatic interactions with positively charged nitrogens of the pyridinium and of the central core. The first binding results in an asymmetric distribution of positive charges, which in turn prevents charge delocalisation from the subsequent binding sites. This hypothesis also implies a partial sp^2 hybridization of the central nitrogen and creation of a rigid molecular skeleton with a planar structure, which favours charge delocalization at all the terminal pyridinium moieties (Figure 2.18). Considering the complete TPA library, it is noteworthy to analyse that the tricyclic core are more prone to higher affinity for CB[7] (Table 2.5). Regarding TP_3Pyo, its lower affinity could result from the restriction due to the steric hindrance by the proximity of the ortho hydrogens on the phenyl_substituents of the central nitrogen, and due to unfavourable entropy of binding well characterized by ITC. Interestingly, the sulfoxide derivative can bind on both branches which may be due to the sulfoxide's ability to act as second mesomeric +M donor, allowing a second charge delocalization, once the first binding site is occupied.

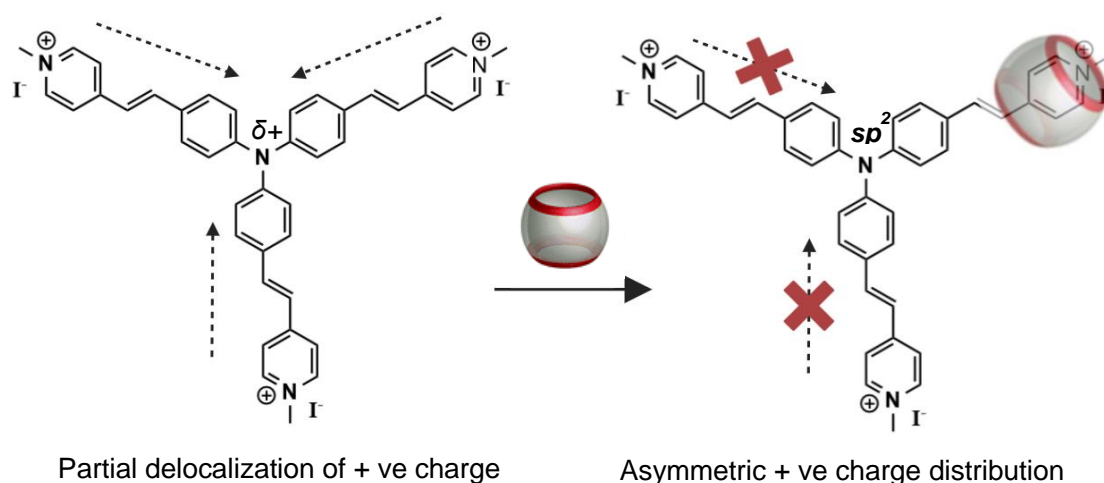


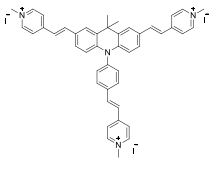
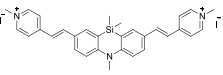
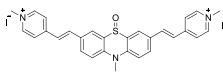
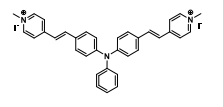
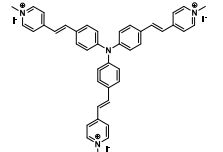
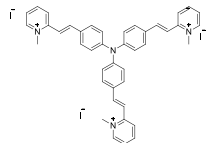
Figure 2.18: Schematic illustration of the hypothesized effect of changes in charge distribution on the electronic and structural properties of the TPA derivatives, resulting in highly negative cooperativity between the first and subsequent CB[7] binding events, as indicated by ITC titrations.

According to the measured thermodynamics of binding, all TPA derivatives exhibited negative cooperativity and the exhibited behaviour that can be classified broadly into three sets:

- a) Acridane and phenazasiline derivatives displayed high negative cooperativity, indicating that the second site was not significantly bound within the concentration range used. The binding was driven by both favorable enthalpy (exothermic reaction) and entropy of binding within the same intensity range, suggesting a decrease in the hydrophobic solvent-exposed surface on both guest and host upon binding.
- b) For the phenothiazine derivative, SulfoxPy, two sequential bindings were observed. Like the acridane and phenazasiline derivatives, both enthalpy and entropy were found to be favourable and of comparable magnitude for the first binding site. In the two-binding sites model, the anticipated statistical relationship of $Kd_2 = 4 \times Kd_1$ in the absence of cooperativity indicated negative cooperativity for SulfoxPy (K_{d1} and K_{d2} refer to the macroscopic dissociation constants for the first and the second binding respectively).
- c) The unmodified triphenylamine derivatives namely TP-2Py, TP-3Py (*structure presented in Table 2*) along with TP_3Pyo exhibited unfavourable entropy of binding, partially compensated by a more favourable enthalpy of binding compared to the other derivatives, resulting in sub-millimolar affinities. The low affinity of these compounds made quantitative comparison difficult due to the thermograms being recorded with very low “c” factors (Wiseman factor).

The unfavourable entropy variation during CB[7] binding, exceed the favourable entropy contribution due to the burial of the pyridinium derivative in the hydrophobic cavity of CB[7], which we interpret as a reduction of conformational space of the bound unmodified triphenylamine derivatives. These complementary results are in accordance with our hypothesis of a charge delocalization from pyridinium derivative toward the central nitrogen upon first binding, and structural constraints induced by its resulting partial sp^2 hybridization. CB[7] binding must be enhanced by the presence of two positively charged nitrogen in the vicinity of the partial negatively charged carbonyl functions on both sides of CB[7]. For the group of unmodified triphenylamine derivative, the unfavourable entropy of binding resulted from the transition of unconstrained structures in unbound state toward constrained and planar structures in bound state. For other derivatives (acridane, phenazasiline, phenothiazine) the flattening of the tricyclic core upon binding did not significantly reduce the conformational entropy, and the overall favourable entropy was dominated by the hydrophobic effect.

Table 2.5: Thermodynamics of triphenylamine derivatives binding with CB[7] measured by isothermal titration calorimetry. Numbers of titrations are indicated as 'n'. Uncertainties were calculated with 95% confidence.

Name (No. of titrations)	Structure	Kd ₁ (μM)	ΔH°_1 (kcal.mol^{-1})	$-\text{T}\Delta\text{S}^{\circ}_1$ (kcal.mol^{-1})	Kd ₂ (μM)	ΔH°_2 (kcal.mol^{-1})	$-\text{T}\Delta\text{S}^{\circ}_2$ (kcal.mol^{-1})
Acrid_3Py (n=2)		46.4 [31.7; 67.6]	-3.7 [-4.2; -3.3]	-2.2 [-2.7; -1.7]			
Phenaz_Py (n=3)		12.5 [9.5; 16.5]	-3.4 [-3.6; -3.2]	-3.3 [-3.6; -3.1]			
Sulfox_Py (n=3)		1.9 [0.8; 4.8]	-3.8 [-4.3; -3.6]	-4.0 [-4.5; -3.3]	118 [35; 204]	-2.3 [-3.8; 1.1]	-3.1 [-6.5; - 1.4]
TP_2Py ⁶⁴ (n=3)		422 [314; 565]	-11.7 [-14.9; - 9.5]	7.1 [4.9; 10.4]			
TP_3Py (n=3)		179 [162; 197]	-9.3 [-9.7; -9.0]	4.2 [3.9; 4.6]			
TP_3Pyo (n=4)		352 [287; 521]	-9.4 [-12.1; - 7.6]	4.7 [2.9; 7.4]			

2.4.5. Interaction of TPA and CB[7] with (bio)molecules

2.4.5.1 Interaction of TPA derivatives with protein

Building upon previous findings indicating that optical properties such as fluorescence emission of the TPA derivatives are modulated through interactions with proteins, we examined the variations in the fluorescence emission of the library of TPA derivatives in the presence of different commonly available proteins. Differential fluorescence responses of the TPA derivatives were monitored with seven proteins, namely BSA, catalase, lysozyme, chymotrypsin, amylase, peroxidase and pepsin in comparison to the free TPA in solution (Figure 2.19). The fluorescence emissions increased by 2 to 20 fold for protein such as catalase, or remained unaffected as in the case of lysozyme. A hypsochromic shift was also observed upon interaction with the proteins (max 20 nm). This could be attributed to the docking of the TPA derivatives on the protein surface and pockets through a combination of weak interactions. Consistent with prior reports, we hypothesized that the docking of the derivatives was most effective when it occurred at the hydrophobic patches of the proteins. This docking is expected to decrease hydrogen bonding of the nitrogen core with the surrounding water molecules, enhancing electron flow in the conjugation system and resulting in an increase of fluorescence. It also confirms that geometrical confinement hinders torsional motions of the TPA upon binding and enhances fluorescence.

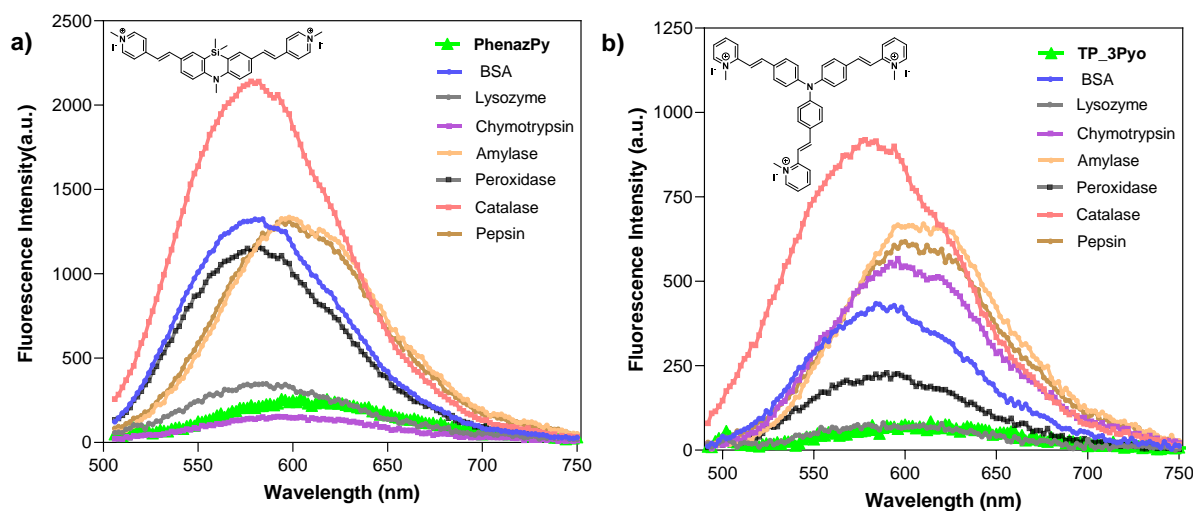


Figure 2.19: Fluorescence response patterns of a) two-arm PhenazPy and b) three-arm TP_3Pyo at 5 μM in the presence of a diverse range of proteins (catalase, BSA, lysozyme, chymotrypsin, amylase, peroxidase, pepsin at 0.625 mg/mL) monitored at excitation wavelength of 486 nm and 476 nm respectively.

To further highlight the importance of hydrophobic interactions between TPA and proteins, we investigated the impact of the number of surface exposed hydrophobic residues of proteins on the fluorescence variation of the TPA. On a set of selected protein crystal structures, PyMol (The PyMOL Molecular Graphics System, Version 4.6 Schrödinger, LLC software) was employed to visualize and count the solvent-accessible surface residues and among them, the hydrophobic residues. It was expected that these residues would have a higher likelihood of interacting with the hydrophobic group of the TPA.

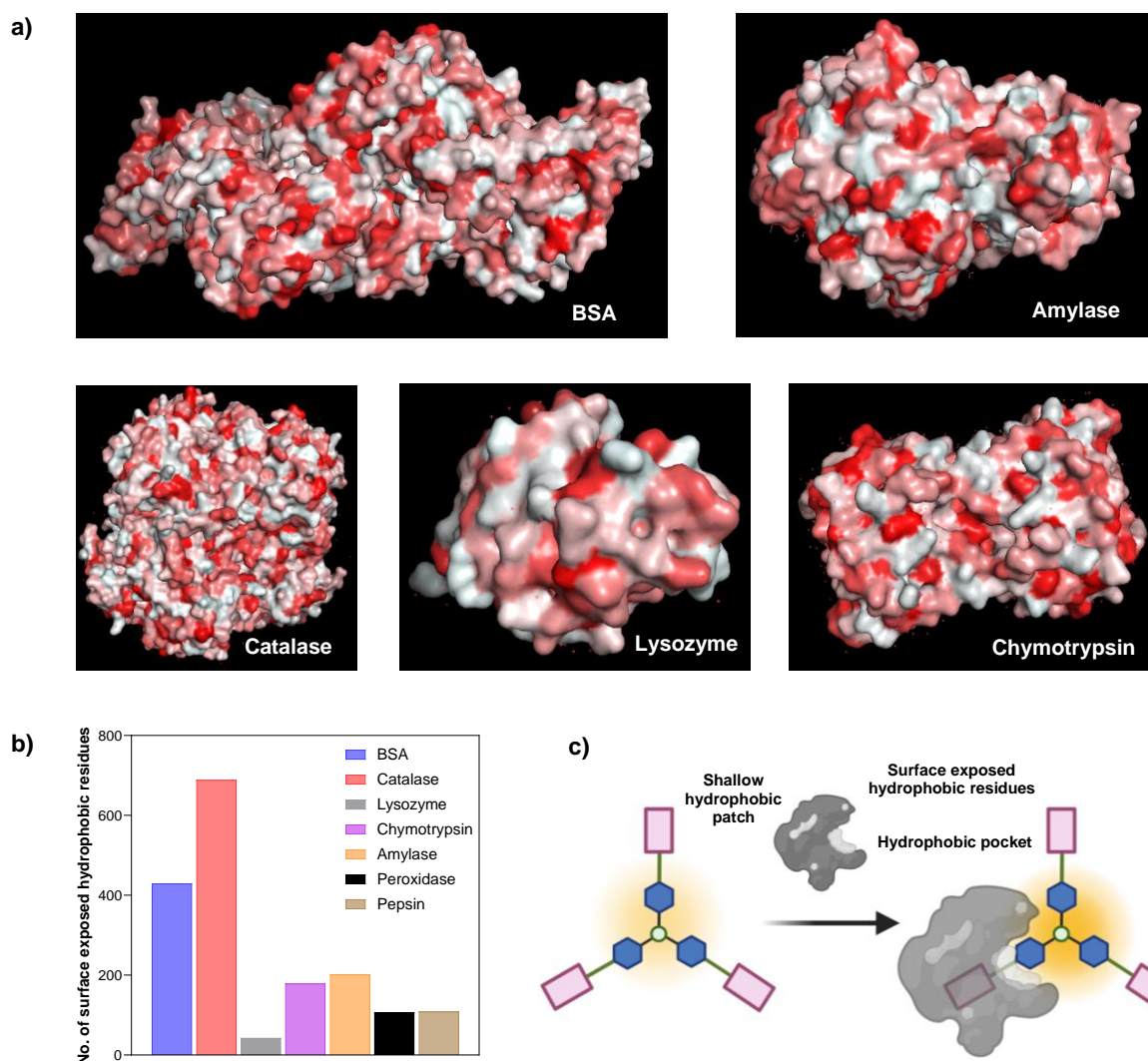


Figure 2.20: a) Visual analysis of surface hydrophobicity of proteins by PyMOL (*color_h.py*). Red: hydrophobic residues, white: hydrophilic residue. b) Evaluation of the number of surface exposed hydrophobic residues for different proteins generated using a PyMOL script (*findsurfaceresidues.py*) with a 2.5 Å cutoff, c) Schematic illustration of fluorescence modulation arising from the interaction of TPA derivatives with the hydrophobic patches or pockets of a protein.

Firstly, the hydrophobic regions of the proteins BSA, catalase, lysozyme, chymotrypsin, amylase, peroxidase and pepsin were visualized by using the PyMOL script (color_h.py) in accordance with the Eisenberg's scale of hydrophobicity (Figure 2.20a). The exposed surface residues of the crystal structure were then generated by using the PyMOL script (findsurfaceresidues.py) with a 2.5\AA^2 cutoff. The estimated numbers of surface exposed residues were grouped according to the nature of the amino acids (hydrophobic, polar, basic and acidic) for each protein to enable better quantitative analysis of the residues and their effect on interaction with the TPA derivatives. Catalase, with the highest hydrophobic residue count, increased the fluorescence of the TPA derivatives to a larger extent as compared to lysozyme or peroxidase, with the lowest hydrophobic residue count, which caused negligible or weak enhancement as compared to the free derivative in solution (Figure 2.20b). These results highlight the potential use of TPA derivatives as attractive scaffolds for the design of a fluorescent sensor array for protein recognition due to their radiative properties which are sensitive to the association with hydrophobic patches of proteins (Figure 2.20c) in addition to their optical transduction properties.

2.4.5.2 Interaction of TPA –CB[7] with biomolecules

CB[7] has been extensively studied for its ability to interact with wide range of guest molecules, with affinities ranging from micromolar to attomolar levels. It forms highly stable 1:1 complexes with simple three-dimensional molecules such as adamantanes and diadamantanes, bicyclooctanes and ferrocene derivatives (Figure 2.21). The diadamantane diammonium ion exhibits the highest binding at $K_a = 7.2 \times 10^{17} \text{ M}^{-1}$ in D_2O . Often these affinities approach or exceed the strongest non-covalent interaction found in nature, as in the case of the biotin–avidin pair. Additionally CB[7] also interacts with various classes of neutral and cationic species including biomolecules like amino acids, peptides and proteins.

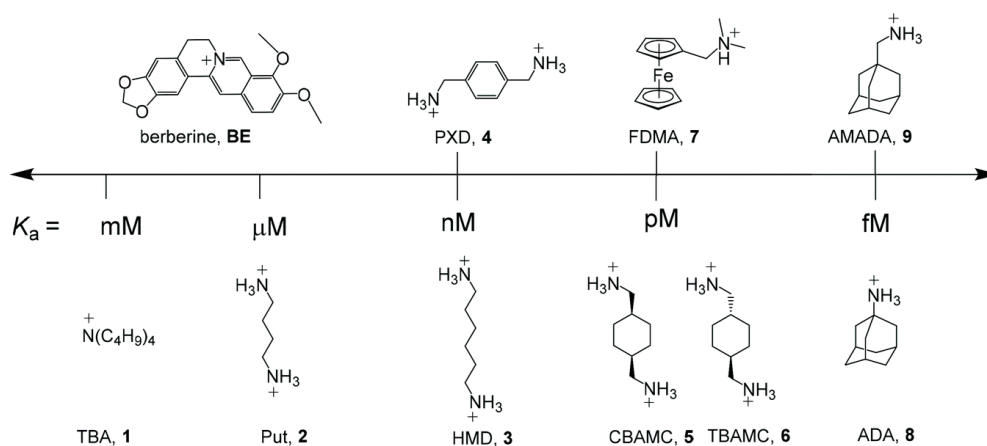


Figure 2.21: Structure of compounds displaying a range of binding affinities as indicated by the K_a values. (adapted from Alnajjar, M. A. *et al.*)¹⁹⁵

The reversible nature of the host-guest complexation and the variable binding affinities of the encapsulated guest molecules lead to the formation of different kind of complexed states:

- a) Guests with K_a in the nanomolar or higher range, form highly stable complexes, characterized by a maximum forward rate constant value (k_{on}) and extremely low backward rate constant value (k_{off});
- b) Guest with lower K_a in the associated state can be displaced by guest of higher binding affinities. This displacement can be either complete or partial, depending on the difference in their K_a values and the number of binding sites on the guest molecule.

Keeping these interactions in mind, we mechanistically expected that the interaction of analytes with TPA-CB[7] complexes may be driven by the interaction of analytes with TPA, the interaction of analytes with the outer surface of CB[7] and the competitive displacement of encapsulated TPA by analytes.

To validate these different modes of interactions, we considered different competitive analytes. First, it was observed that the fluorescence enhancement of the TPA's on complexation with CB[7] decreased drastically upon addition of the CB[7] high-affinity binder adamantylamine ADA.¹⁴⁵ The latter completely displaces TPA's from the CB[7] cavity, consequently reversing the inclusion-induced fluorescence. No residual fluorescence of TPA was observed, showing that ADA does not interact with TPA and acts only as a TPA competitor (Figure 2.22).

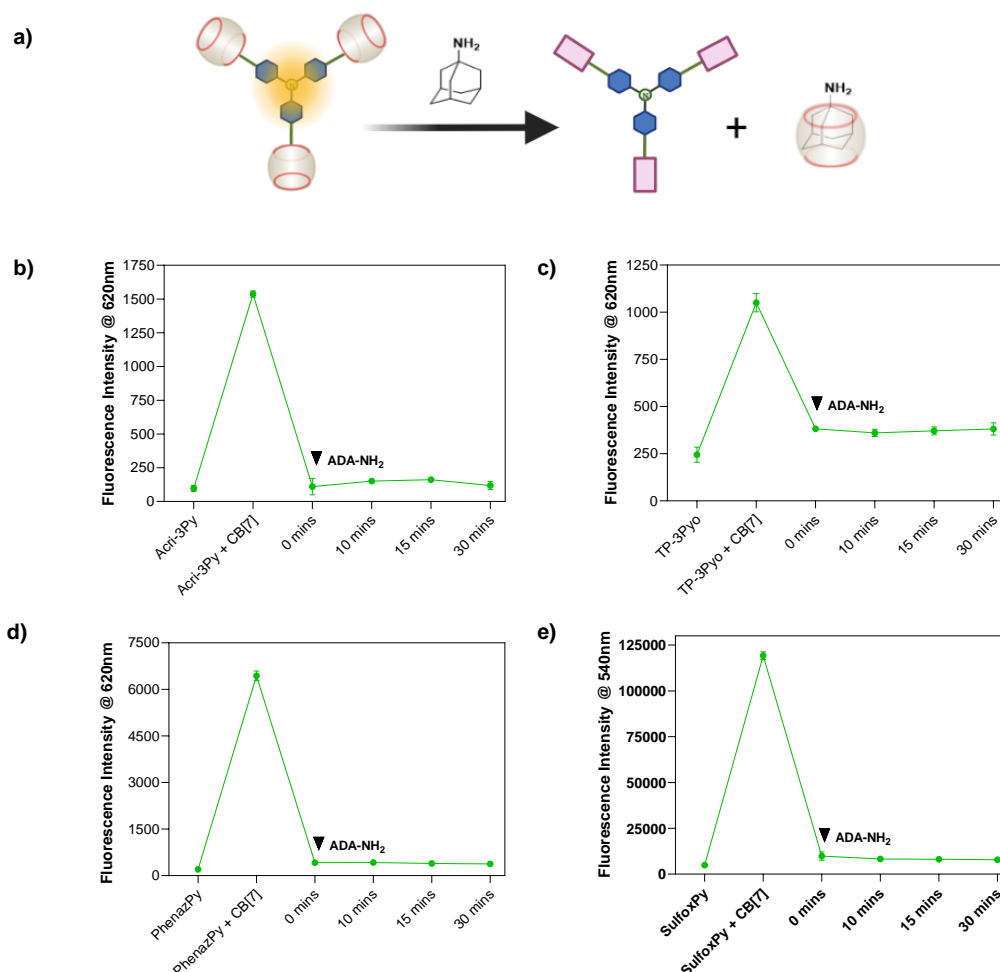


Figure 2.22: (a) Schematic illustration of decrease in fluorescence resulting from the displacement of TPA derivatives from their complex with CB[7] in the presence of high affinity binder adamantane (ADA). Fluorescence intensity variation upon formation of TPA-CB[7] complex by b) Acri_3Py, c) TP_3Pyo, d) PhenazPy e) SulfoxPy ($5 \mu\text{M}$) and CB[7] ($100 \mu\text{M}$) and corresponding changes in the presence of ADA ($250 \mu\text{M}$) monitored over 30mins at $\lambda_{em} = 620 \text{ nm}$ for Acri_3Py, TP_3Pyo, PhenazPy and 540 nm for SulfoxPy.

CB[7] is also known to encapsulate amino acids with diverse affinities. In particular, CB[7] has been noted for its selectivity toward aromatic amino acids: phenylalanine (Phe), tryptophan (Trp) and tyrosine (Tyr) at pH 7. Its highest affinity is observed with Phe, K_a ranging from 10^5 M^{-1} to 10^6 M^{-1} .¹⁹⁶ Considering this, we evaluated the extent and effect of interaction of amino acids on the fluorescence of TPA-CB[7] complexes. We subjected the TPA-CB[7] complexes to representative amino acids and observed a decrease in fluorescence upon interaction with the aromatic amino acid (Phe). No (or less) change was recorded in the case of the other amino acids like Arg, and Lys (Figure 2.23). Again, the free TPA derivatives were observed to have no significant alterations in their fluorescence intensity in the presence of any of the amino acids. This demonstrated the displacement of the TPA

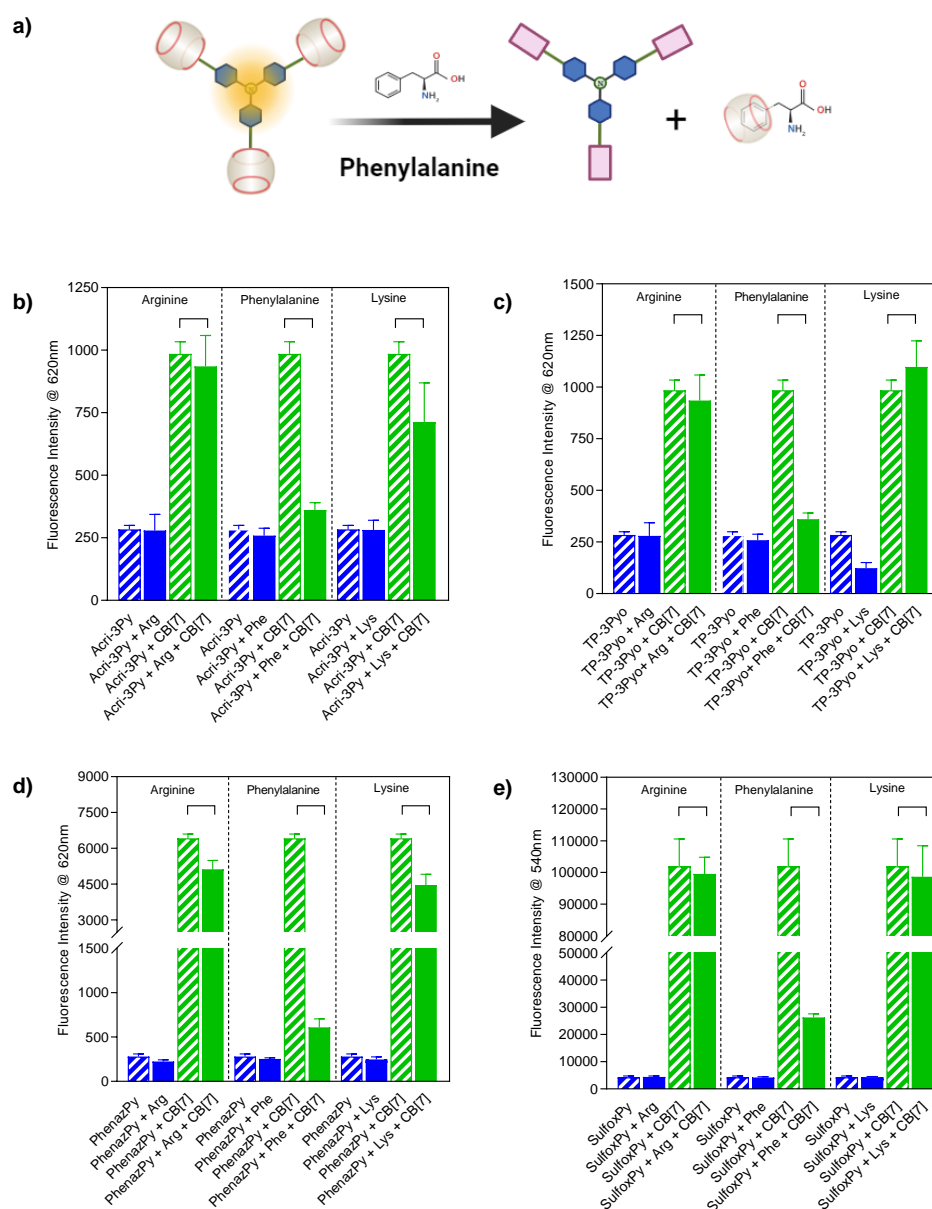


Figure 2.23: (a) Schematic illustration of variation in fluorescence resulting from the displacement of TPA derivatives from CB[7] in the presence of aromatic amino acid like phenylalanine. Fluorescence intensity variation upon formation of TPA.CB[7] complex by b) Acri_3Py, c) TP_3Pyo, d) PhenazPy e) SulfoxPy (5 μ M) and CB[7] (100 μ M) and corresponding changes in the presence of amino acids Arg, Phe, Lys at λ_{em} = 620 nm for Acri_3Py, TP_3Pyo, PhenazPy and 540 nm for SulfoxPy.

derivative from its complex with CB[7] in the presence of a stronger binder like Phe and confirmed the retention of the function of CB[7] as a recognition element even after complexation with TPA. These findings also suggest that while the larger protein moieties are more effective in interacting with the TPA's even in the absence of CB[7] to promote radiative de-excitation, the smaller size of the amino acids fails to generate the necessary structural confinement for the same effect. Furthermore, it is

observed that a complete displacement of TPA does not occur in the presence of amino acids, unlike with ADA. This suggests that only partial displacement occurs due to the lower binding affinity of the amino acids such as Phe towards CB[7] as compared to ADA, resulting in the retention of the fluorescence signal generated by the partially CB[7] complexed TPA derivatives.

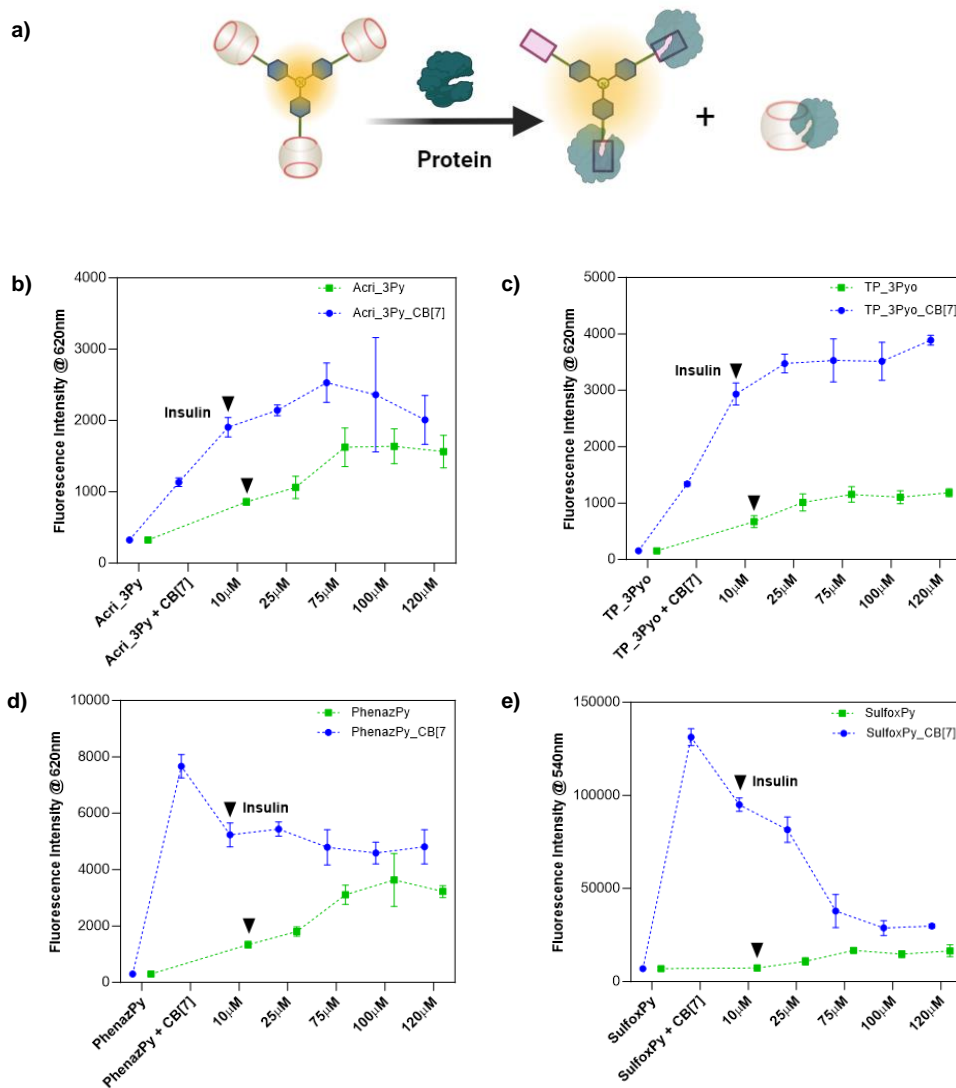


Figure 2.24: (a) Schematic illustration of displacement assay resulting from the interaction of TPA derivatives and their complex with CB[7] in the presence of protein insulin. Fluorescence intensity variation of TPA and TPA-CB[7] complexes of b) Acrid_3Py, c) TP_3Py, d) PhenazPy e) SulfoxPy (5 μM) with CB[7] (100 μM) in the presence of insulin (10, 25, 75, 100 and 120 μM) at λ_{em} = 620 nm for Acrid_3Py, TP_3Py, PhenazPy and 540 nm for SulfoxPy.

Similarly, the fluorescence response of the TPA-CB[7] complexes were monitored by a displacement assay in the presence of insulin. The binding of which has been previously documented by Urbach *et al.*, who reported the incorporation of the N-terminal Phe residue of insulin within the hydrophobic cavity of CB[7] with a binding constant (K_a) of $1.5 \times 10^6 \text{ M}^{-1}$.¹⁹⁷ Initially, we demonstrated that the free

TPA derivatives underwent a fluorescence enhancement upon interaction with increasing concentrations of insulin and further additions were stopped after reaching a plateau in the fluorescence intensity. This is indicative of the ability of TPA to interact with the surface exposed hydrophobic residues of the protein. The addition of insulin to the TPA-CB[7] complexes elicited varied responses for the 2-arm and 3-arm derivatives (Figure 2.24). An initial fluorescence enhancement followed by no further increase in intensity was observed for the 3-arm derivatives. However, for the 2-arm derivatives a significant decrease in fluorescence emission due to the competitive displacement of the TPA from the CB[7] by insulin similar to Phe was observed.

These biomolecular interactions with the TPA-CB[7] complexes validated the sustained functionality of CB[7] as a recognition element even after complex formation with the TPA derivatives. It has also demonstrated the diversity in binding interactions that CB[7], TPA and CB[7]/TPA offer, contributing to the sensor design of the developed fluorescence sensor array. These results establish the effectiveness of the proposed sensor design and support its potential for non-specific sensing of biomolecules.

2.5. Conclusions

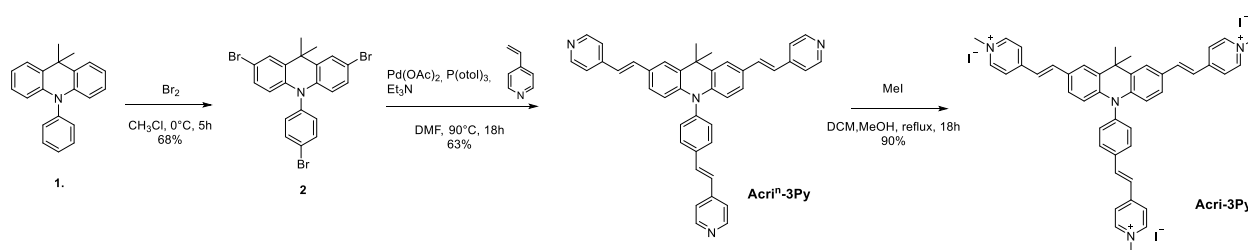
In summary, we have investigated the use of macrocyclic host-guest interactions to develop a fluorescence-based sensor array utilizing the conjugated system of triphenylamine derivatives and the members of the cucurbit[n]uril family. This chapter presents an extensive study of the photophysical and binding properties of the TPA derivatives and their modulation in combination with the various CB[n] hosts, ultimately establishing the suitability of CB[7] and CB[8] for the development of fluorescent and colorimetric sensor arrays, respectively. The TPA derivatives have been identified for their potential to function effectively as the transducers of the sensor array. This functionality is primarily attributed to the restriction in intramolecular rotation that occurs upon host-guest complexation, thereby modulating their photophysical properties. Moreover, these derivatives have been shown to interact with analytes such as proteins, thereby serving as orthogonal recognition scaffolds. This dual functionality facilitates the development of a bimodal recognition macrocyclic sensor array, effectively doubling the number of sensor channels while retaining a limited number of essential sensing elements. The fluorescence sensor array has been optimized by the selection of four specific TPA derivatives with the CB[7] hosts. The unique structural features and strong molecular recognition capabilities of CB[7] make it particularly effective in forming stable host-guest complexes with TPA derivatives. We have further characterized the binding properties of the TPA-CB[7] system and established the thermodynamic and kinetic parameters associated with the binding interactions using isothermal titration calorimetry (ITC), providing valuable insights into the binding affinities, stoichiometry, and enthalpic and entropic contribution of these interactions.

2.6. Materials and Methods

2.6.1. Synthesis and characterization spectra of unpublished TPA derivatives

****All synthesis were performed by Dr. Delphine Naud-Martin from the group of Dr. Florence Mahuteau-Betzer (Institut Curie)**

Synthesis of Acri_3Py



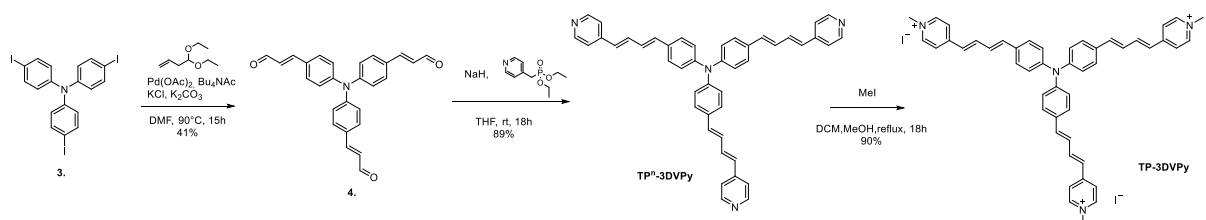
Compound 2 [880771-61-9]: To a stirred solution of acridane (**1**) (500 mg, 1.75 mmol, 1 eq.) in CHCl_3 (6 mL) cooled to 0°C , a solution of bromine (888 mg, 5.5 mmol, 3.15 eq.) in CHCl_3 (6 mL) was added drop wise. The resultant mixture was stirred at 0°C for 5 h. The reaction mixture was quenched by adding water carefully. The resulting organic layer was then extracted with dichloromethane. Organic phases were combined, dried and concentrated to afford the expected product as a white powder (620 mg, 1.18 mmol, $\eta=68\%$). $^1\text{H NMR}$ (300 MHz, CDCl_3) δ 7.77 (d, $J = 8.5$ Hz, 2H), 7.51 (d, $J = 2.0$ Hz, 2H), 7.17 (d, $J = 8.5$ Hz, 2H), 7.07 (dd, $J = 8.5, 2.0$ Hz, 2H), 6.12 (d, $J = 8.5$ Hz, 2H), 1.63 (s, 6H), $^{13}\text{C NMR}$ (75 MHz, CDCl_3) δ 139.6, 134.7, 132.8, 131.8, 129.5, 128.3, 122.8, 115.8, 113.8, 36.4, 31.0.

Acriⁿ-3Py: In a dry and degassed triethylamine/DMF (2:1, v/v, 5mL) mixture, $\text{Pd}(\text{OAc})_2$ (4.5 mg, 0.02 mmol, 0.1 eq.) and $\text{P}(o\text{-tol})_3$ (12 mg, 0.04 mmol, 0.2 eq.) were introduced and stirred 15 min. Mixture was degassed then compound (**2**) (105 mg, 0.2 mmol, 1 eq.) and 4-vinylpyridine (84 mg, 0.8 mmol, 4 eq.) were added. The mixture was stirred at 90°C under argon overnight. DCM was introduced and the organic layer was washed with water and saturated NaHCO_3 and dried over MgSO_4 . The crude was purified by flash chromatography on silica gel (DCM/MeOH, 100:0 to 95:5) to give **Acriⁿ-3Py** as an orange powder (75 mg, 0.12mmol, $\eta=63\%$). **MS(ESI+):** m/z 595.3 $[\text{M}+\text{H}]^+$, $^1\text{H NMR}$ (300 MHz, CDCl_3) δ 8.63 (d, $J=5.5\text{Hz}$, 2H), 8.53 (d, $J=5.0$, 4H), 7.83 (d, $J=8.0\text{Hz}$, 2H), 7.65 (s, 2H), 7.47 – 7.09 (m, 14H), 6.88 (d, $J=16.0\text{Hz}$, 2H), 6.34 (d, $J=8.5$ Hz, 2H), 1.80 (s, 6H), $^{13}\text{C NMR}$ (75 MHz, CDCl_3) δ 150.3, 150.0, 145.1, 144.2, 140.7, 136.6, 133.0, 131.9, 131.4, 130.3, 129.6, 129.2, 127.6, 125.4, 124.8, 123.1, 121.0, 120.6, 114.7, 36.1, 31.8.

Acri_3Py: A large excess of iodomethane (0.31 mL, 5.0 mmol, 100 eq.) was added to a solution of **Acriⁿ-3Py** (30 mg, 0.05 mmol, 1 eq.) in a MeOH/DCM (1:1) mixture (4.0 mL). The reaction was stirred at r. t. overnight. Et_2O was introduced to get a precipitate which was filtered to give **Acri_3Py** as a dark

red powder (43 mg, 0.042 mmol, $\eta=90\%$). **MS(ESI+)**: m/z 213.4 $[M-3I]^{3+}$, **1H NMR (300 MHz, DMSO- d_6)** δ 8.93 (d, $J = 6.5$ Hz, 2H), 8.79 (d, $J = 6.5$ Hz, 4H), 8.30 (d, $J = 6.0$ Hz, 2H), 8.25 – 8.07 (m, 7H), 8.01 - 7.97 (m, 4H), 7.70 (d, $J = 16.5$ Hz, 1H), 7.60 (d, $J = 8.0$ Hz, 2H), 7.51 – 7.35 (m, 4H), 6.31 (d, $J = 8.5$ Hz, 2H), 4.30 (s, 3H), 4.22 (s, 6H), 1.80 (s, 6H), **^{13}C NMR (75 MHz, DMSO- d_6)** δ 152.9, 152.1, 145.3, 144.8, 140.9, 140.9, 140.7, 139.2, 135.9, 131.3, 131.0, 130.4, 128.6, 127.6, 126.6, 124.9, 123.8, 122.8, 120.2, 114.7, 47.1, 46.7, 35.9, 32.3, 3.

Synthesis of TP-3DVPy

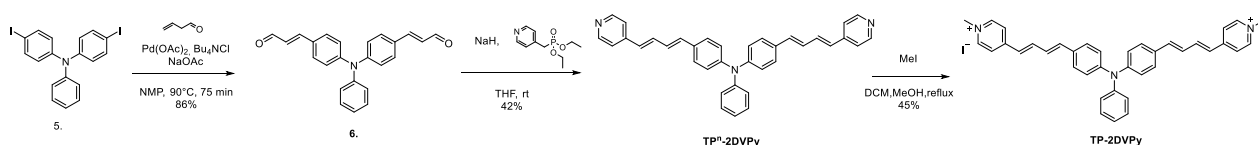


Compound 4: [925452-63-7] Under argon atmosphere, acrolein diethyl acetal (1.1 mL, 7.2 mmol, 9 eq.), tetrabutylammonium acetate (1.2 g, 4 mmol, 5 eq.), potassium carbonate (500 mg, 3.6 mmol, 4.5 eq.), potassium chloride (180 mg, 2.4 mmol, 3 eq.) and palladium diacetate (16 mg, 0.07 mmol, 0.09 eq.) were successively added to a solution of trisiodotriphenylamine (**3**) (500 mg, 0.8 mmol, 1eq.) in degassed dry DMF (10 mL). The reaction mixture was stirred for 15 h at 90°C. After being cooled, the reaction mixture was quenched with 10 mL of HCl 2N and extracted three times with CH_2Cl_2 . The combined organic fractions were washed with brine once and dried over $MgSO_4$. The crude was purified by flash chromatography on silica gel (Cyclohexane/ethyl acetate, 80:20 to 70:30) to give the expected product as a yellow powder (136 mg, 2.6 mmol, $\eta=41\%$). **MS(ESI+)**: m/z 408.4 $[M+H]^+$, **1H NMR (300 MHz, $CDCl_3$)** δ 9.66 (d, $J = 7.5$, 3H), 7.50 (d, $J = 8.5$, 6H), 7.42 (d, $J = 16.0$, 3H), 7.15 (d, $J = 8.5$, 6H), 6.63 (dd, $J = 7.6, 16.0$, 3H), **^{13}C NMR (75 MHz, $CDCl_3$)** δ 193.5, 151.7, 148.8, 130.1, 129.8, 127.6, 124.6.

TPⁿ-3DVPy: Under argon atmosphere, 4-pyridinylphosphonate (75.9 mg, 0.33 mmol, 3.3 eq.) and sodium hydride 60% in oil (52 mg, 0.74 mmol, 7.5 eq.) were added in dry THF (3mL). After 15 min stirring, trisaldehyde (**4**) (40 mg, 0.1 mmol, 1 eq.) in dry THF (2 mL) was added and the resulting mixture was stirred at room temperature for 18h in the dark. Water was introduced dropwise to quench the reaction and the resultant mixture was extracted three times with dichloromethane. Organic layer was washed twice with brine and dried over $MgSO_4$. After partial evaporation, addition of drops of Et_2O afforded **TPⁿ-3DVPy** as an orange powder (55 mg, 0.09 mmol, $\eta= 89\%$). **MS(ESI+)**: m/z 633.2 $[M+H]^+$, **1H NMR (300 MHz, $CDCl_3$)** δ 8.52 (d, $J = 4.0$ Hz, 6H), 7.36 (d, $J = 8.0$ Hz, 6H), 7.30 – 7.21 (m, 6H), 7.19 – 6.99 (m, 9H), 6.94 – 6.80 (m, 3H), 6.73 (d, $J = 15.0$ Hz, 3H), 6.54 (d, $J = 15.0$ Hz, 3H), **^{13}C NMR (75 MHz, $CDCl_3$)** δ 150.2, 147.0, 144.8, 135.2, 133.8, 131.9, 127.9, 127.2, 124.4, 120.6

TP-3DVPy: A large excess of iodomethane (0.31 mL, 5.0 mmol, 100 eq.) was added to a solution of **TPⁿ-3DVPy** (30 mg, 0.05 mmol) in a MeOH/DCM (1:1) mixture (4.0 mL). The reaction was stirred at r.t. overnight. Et₂O was introduced to get a precipitate which was filtered to give **TP-3DVPy** as a dark red powder (45 mg, 0.04 mmol, η = 90%). **MS(ESI+):** *m/z* 226.1 [M-3I]³⁺, **¹H NMR** (300 MHz, CDCl₃): δ 8.78 (d, *J* = 6.5 Hz, 6H), 8.11 (d, *J* = 6.5 Hz, 6H), 7.82 (dd, *J* = 10.5 Hz, *J* = 15.5 Hz, 3H), 7.61 (d, *J* = 7.5 Hz, 6H), 7.24-7.03 (m, 12H), 6.86 (d, *J* = 15.5 Hz, 3H), 4.21 (s, 9H), **¹³C NMR (75 MHz, DMSO-*d*₆)** δ 152.3, 146.9, 144.8, 142.1, 139.8, 131.4, 129.0, 127.1, 126.0, 124.2, 123.0, 46.8.

Synthesis of TP-2DVPy



Compound 6: Under argon atmosphere, acrolein (0.17 mL, 2.6 mmol, 4 eq.), tetrabutylammonium chloride (358 mg, 1.29 mmol, 2 eq.), potassium acetate (5.33 mg, 0.032 mmol, 0.05 eq.) and palladium diacetate (7 mg, 0.03 mmol, 0.05 eq.) were successively added to a solution of diiodotriphenylamine (**5**) (400 mg, 0.64 mmol, 1 eq.) in NMP (14 mL). The mixture was then heated at 90 °C for 75 min. After being cooled, the reaction mixture was poured into a half-concentrated aqueous solution of NaHCO₃ and extracted three times with CH₂Cl₂. The combined organic fractions were washed with brine once, dried over MgSO₄, and concentrated under reduced pressure. The resulting solution in NMP was directly loaded onto a column packed with silica gel and eluted with Et₂O-pentane (15:85, v/v) to give the expected product as a yellow powder (197 mg, 0.56 mmol, η = 86%). **MS(ESI+):** *m/z* 354.3 [M+H]⁺, **¹H NMR (300 MHz, CDCl₃)** δ 9.67 (d, *J* = 7.5 Hz, 2H), 7.46 (d, *J* = 8.5 Hz, 4H), 7.41 (d, *J* = 16.0 Hz, 2H), 7.35 (d, *J* = 8.0 Hz, 2H), 7.23 – 7.08 (m, 7H), 6.63 (dd, *J* = 16.0, 7.5 Hz, 2H), **¹³C NMR (75 MHz, CDCl₃)** δ 193.8, 152.2, 149.7, 146.0, 130.1, 130.0, 128.6, 127.1, 126.6, 125.6, 123.4.

TPⁿ-2DVPy: Under argon atmosphere, 4-pyridinylphosphonate (100 mg, 0.44 mmol, 2.2 eq.) and sodium hydride 60% in oil (69 mg, 1 mmol, 5 eq.) were added in dry THF (3 mL). After 15 min stirring, dialdehyde (**6**) (70 mg, 0.2 mmol, 1 eq.) in dry THF (3 mL) was added and resulting mixture was stirred at room temperature for 18h in the dark. Water was introduced dropwise to quench the reaction and the resultant mixture was extracted three times with dichloromethane. Organic layer was washed twice with brine and dried over MgSO₄, filtered and concentrated to dryness. The crude residue was purified by dissolution in a minimal amount of dichloromethane and precipitated with Et₂O afforded **TPⁿ-2DVPy** as an orange powder (62 mg, 0.12 mmol, η = 62%). **MS(ESI+):** *m/z* 504.4 [M+H]⁺, **¹H NMR (300 MHz, CDCl₃)** δ 8.53 (d, *J* = 5.5 Hz, 4H), 7.35 (d, *J* = 8.5 Hz, 4H), 7.29 (m, 6H), 7.18 – 7.03 (m, 9H), 6.92 – 6.81 (m, 2H), 6.73 (d, *J* = 15.5 Hz, 2H), 6.54 (d, *J* = 15.5 Hz, 2H), **¹³C NMR (75 MHz, CDCl₃)** δ 150.0, 147.5, 147.0, 145.0, 135.5, 134.1, 131.3, 129.6, 128.9, 127.8, 126.8, 125.3, 124.0, 123.8, 120.7.

TP-2DVPy: A large excess of iodomethane (0.37 mL, 6.0 mmol, 100 eq.) was added to a solution of **TPⁿ-2DVPy** (31 mg, 0.06 mmol) in a MeOH/DCM (1:1) mixture (4.0 mL). The reaction was stirred at r.t. overnight. Et₂O was introduced to get a precipitate which was filtered to give **TP-2DVPy** as a dark red powder (22 mg, 0.028 mmol, η= 45%). **MS(ESI+)**: m/z 266.9 [M-2I]²⁺, **¹H NMR (300 MHz, DMSO-d₆)** δ 8.78 (d, *J* = 6.0 Hz, 4H), 8.11 (d, *J* = 6.0 Hz, 4H), 7.94 – 7.74 (m, 2H), 7.59 (d, *J* = 7.5 Hz, 4H), 7.45 - 7.37 (m, 2H), 7.27 – 6.96 (m, 11H), 6.86 (d, *J* = 14.5 Hz, 2H), 4.22 (s, 6H), **¹³C NMR (75 MHz, DMSO-d₆)** δ 152.3, 147.5, 146.0, 144.0, 142.1, 140.0, 130.6, 129.9, 128.8, 126.6, 125.7, 125.6, 123.1, 123.0, 119.2, 46.7, **HRMS (ESI)** m/z [C₃₈H₃₅ N₃I]⁺ calculated: 660.1876, found: 660.1884.

****All relevant spectra are found in the Appendix 1 of the Thesis**

2.6.2. Protein visualization and determination of solvent exposed hydrophobic residues on protein surfaces

On Pymol (download xxx), run the following script

Script:

run findSurfaceResidues.py

fetch (protein PDB file), async=0

findSurfaceResidues doShow=1, cutoff=2.5

(Iterate to return names of residues and corresponding number)

iterate exposed_res_02, resn, resi

2.6.3. UV- Visible and fluorescence spectroscopy

Primary stocks of the TPA derivatives (5mM) were prepared through weighing of synthesized, purified solid product and dissolution in DMSO and were stored at -20°C. All working solutions of TPA were diluted in milliQ water. For the CB[n] derivatives, the sparingly soluble nature of CB[6] required the preparation of primary stock solutions (300μM) in 500mM KCl, while CB[7] (3mM) and CB[8] (300μM) were prepared in milliQ water and were stored at 4°C. All working solutions of CB[n] were diluted in milliQ water.

Absorbance spectra were recorded on a TECAN SPARK CYTO400 spectrophotometer using a Corning UV-transparent half area 96-well plate. As the instrument automatically adjusts the length of the optical path crossed to 1cm (according to Beer-Lambert's law), no manual corrections were required and the absorbance values recorded could therefore, in theory be found by any other measuring device under the same conditions. Spectra were recorded with a wavelength scan ranging from 400-650 nm with an incremental step size of 2 nm. All measurements were performed in triplicates. The absolute absorbance

of the TPA and TPA-CB[n] complexes were treated by background subtraction of control values and normalized for visualization on GraphPad Prism version 10.0 for Windows, GraphPad Software, www.graphpad.com.

Fluorescence emission spectra were recorded on the same instrument using Corning black polystyrene half-area 96-well plates. The measurements were made with an incremental step size of 4nm, excitation wavelength was set at the isosbestic point obtained from the absorbance spectra of TPA and TPA-CB[n] spectra and a suitable emission range was set ($\sim\lambda_{\text{exc}} + 40 \text{ nm}$) up till 850 nm to avoid any superposition between λ_{exc} and λ_{ems} .

Preparation of protein solutions and measurement of fluorescence spectra with TPA's

Protein stock solutions at 10mg/mL were prepared by dissolution of commercially available solid products (Details provided in Materials and Methods of Chapter 3) in 1X phosphate buffered saline (10X PBS –Sigma Aldrich). The interaction of proteins with the TPA derivatives was recorded by adding 40 μl of the TPA solutions at 5 μM to the microplate, followed by the addition of 2.5 μl of protein solutions to obtain a final concentration of 0.625mg/mL. The plate was incubated for 10 minutes and the fluorescence spectra were then recorded. The fluorescence intensity of the TPA with protein solution was plotted for visualization on GraphPad Prism version 10.0 for Windows, GraphPad Software, www.graphpad.com.

2.6.4. Isothermal Titration Calorimetry

Isothermal calorimetric titrations were conducted using the Microcal™ ITC200 instrument at 25°C. CB[7] concentrations were determined through weighing and dilution, while TPA derivative concentrations were obtained from triplicate absorbance measurements using the SpectraMax Quickdrop® spectrophotometer at the maximum absorbance wavelength. The typical titration scheme included an initial injection of 0.5 μL , followed by 15 main injections of 2.54 μL , with an equilibration time of 210s. The reference power was set at 5 $\mu\text{cal/s}$, and the stirring speed was maintained at 800 rpm. For each TPA, the raw ITC data were integrated in Nitpic and analysed globally in Sedphat^{198–200}. Two or three binding models were applied to analyze the data: the "ABBB three symmetric sites model", the "ABB two symmetric sites model", and the "AB heteroassociation model". Fixed stoichiometry (1:3, 1:2, or 1:1) required refining a concentration correction factor for the TPA, which was constrained to be between 0.8 and 1.2; the binding enthalpy was normalized by setting the CB[7] concentration correction to 1.000. Confidence intervals were estimated at the P=95% confidence level with F-statistics-based error surface contours implemented in SEDPHAT, using the optimized correction factors of the dyes. Finally, Gussi²⁰¹ was employed to produce illustrations of the thermograms.



Chapter 3

***Data management and statistical analysis for pattern recognition in
array-based sensing***

3.1. Introduction

Cross-reactive array-based sensors employ a collection of low specificity receptors to interact with a wide range of analytes present in complex mixtures generating unique fingerprints. Regardless of the type of signal outputs generating these fingerprints, this sensing approach produces a large amount of raw data that is not easy to manage and interpret by simple visual inspection. Analysis using basic calibration techniques such as simple linear regression for every single analyte does not provide useful information from the raw data. Thus, chemometric methods of data analysis have been established to alleviate this challenge of data management and analysis³⁷.

Chemometrics refers to the chemical discipline that utilizes mathematical and statistical approaches that employ formal logic to extract information from chemical systems by data-driven means. The characteristics of the chemometric approach can be better understood when compared to classical methods of data extraction in chemical systems. The classical approach focuses on understanding the impact of various factors, identifying which ones are dominant and which are negligible. In contrast, the chemometric approach does not require an understanding of these effects; instead, it aims for objectives such as prediction, pattern recognition, and classification. The classical approach is reductionist, examining one factor at a time and isolating effects as much as possible. On the other hand, the chemometric approach employs multivariate methods, considering all variables simultaneously and fitting the model to the data. When constructing a model to fit the data, the conclusions should align with the information contained within the data. This differs significantly from the classical approach, where the model is based on theory and data is used to validate the model. While these two approaches are complementary, they cannot be substituted by each other. Consider blood tests of healthy and ill patients. If one feature is examined at a time, all features might fall within the healthy limits, yet a patient could still be ill. Conversely, some healthy patients might exhibit extreme values. By analysing a large number of patients and considering all features simultaneously, healthy and ill patients can typically be distinguished using multivariate chemometric methods²⁰².

Hence the chemometric method offers unique data insights that are otherwise inaccessible and involves two main ideas that are crucial for analysing data generated by cross-reactive sensor arrays: prediction and pattern recognition.

- Prediction refers to making a declaration on the basis of observation, experience, or scientific reason. It involves not only temporal processes but also other aspects, for example, the prediction of the toxicity of a compound based on similar compounds. Even without a causal model, predictions using black box models can be valuable.
- Pattern recognition involves identifying patterns in the data. A pattern is a natural or chance configuration, of traits, tendencies, or other observable characteristics of data. Although the recognition of patterns may happen automatically, the process is difficult to define. In

chemometrics, patterns are often simplified to clusters (groupings), separators, thresholds and outliers.

In addition to chemometric methodologies, which primarily focus on the statistical analysis and interpretation of fingerprint data generated by sensor arrays, the initial preparation, organization, and management of the substantial volumes of raw data produced by these arrays are critical steps in the overall process of array-based sensing systems. Streamlining this data management process is not only vital for the effective adoption of sensor arrays in various applications but also needs to be developed in a highly efficient manner to enable robust sensing. This ensures that the system can handle large datasets reliably, allowing for scalability and adaptability to different sensing environments and challenges.

Thus, this chapter will provide a detailed discussion on the streamlined methodology employed for data acquisition, management and analysis used to evaluate the optical fingerprints generated by the TPA-CB[n] sensor arrays. By elaborating each of these stages, we aim to highlight the robustness and efficacy of the developed methodology, demonstrating its crucial role in ensuring accurate and reliable analysis of the optical fingerprints. This discussion will highlight the technical aspects while also illustrating how these methodologies contribute to the broader objectives of this study, such as the development of diagnostic tools for multifactorial disease.

3.2. From data management to statistical treatment of fingerprints

The process from recording sensor-analyte interactions to classifying analyte fingerprint output signals involves two primary steps: data management and chemometric analysis. While data management appears to be only a single step process, transitioning from signal measurement to feature extraction for statistical analysis, is a highly challenging task that encompasses data acquisition, organisation and pre-processing. Therefore, this step lays the foundational groundwork for the high-throughput operation of the sensor array. The chemometric analysis on the other hand involves a series of steps involved in the process of pattern recognition, including dimensionality reduction, exploratory data analysis, classification, and clustering. Ultimately, together, these processes lead to the development of a classification model that accurately describes the data and the nature of the sensor array-analyte behaviour for the specific application being investigated. This workflow is tailored to the nature of data output collected from the transducing elements. An overview of this process is represented in Figure 3.1.

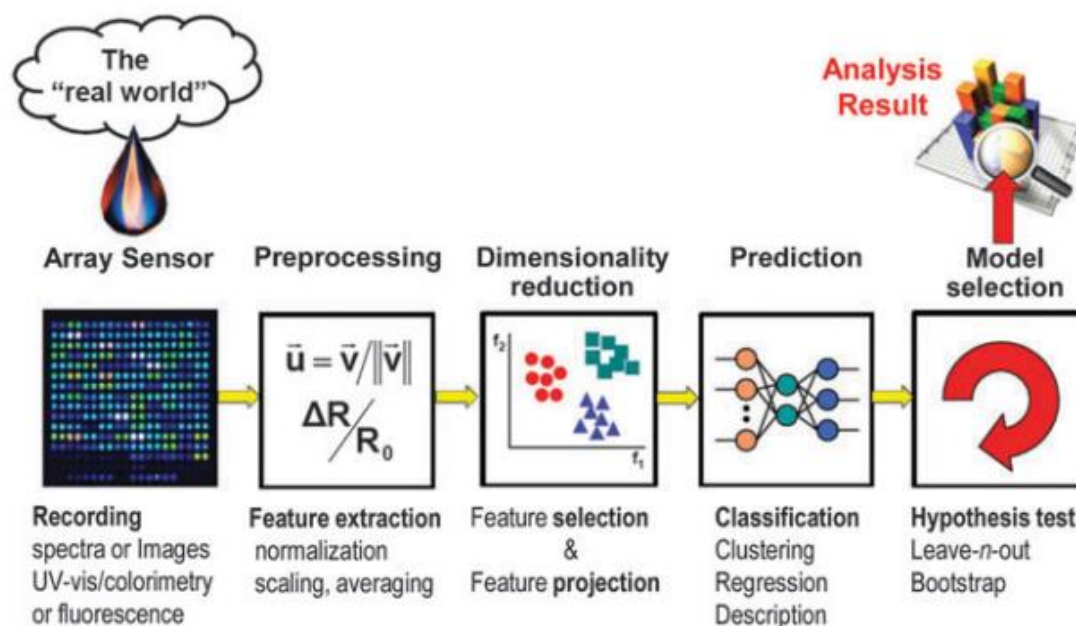


Figure 3.1: Schematic illustration depicting the workflow adopted in chemometric analysis of array-based fingerprinting data (adapted from Anzenbacher, P. et al.⁴²).

3.2.1. Design of sensing experiments

The first step of developing the workflow for analysing the data generated by the TPA-CB[n] fluorescence sensor array was to design a consistent experimental protocol to ensure that the nature and structure of the data output was compatible with the development of an automated pipeline using R coding.

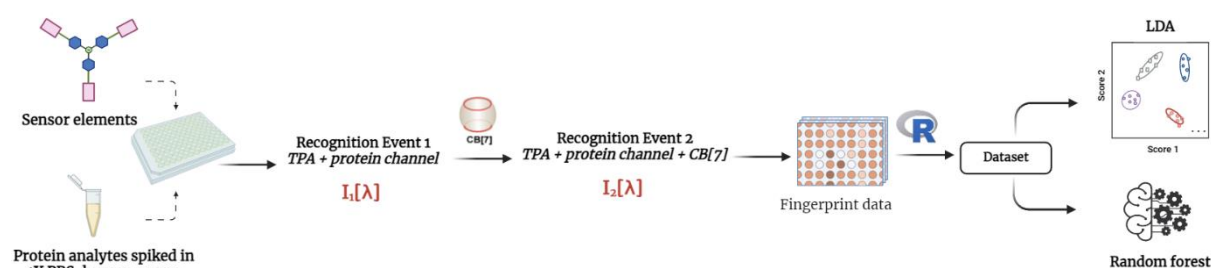


Figure 3.2: Schematic illustration of protocol established for sensing and data treatment of obtained fluorescence fingerprint for discrimination of proteins by TPA-CB[7] sensor array.

Initially, an experimental plan was formulated to capture the bimodal recognition events of the TPA-CB[7] sensor in a time-efficient and effective manner. We recorded the complete emission spectra of

all the TPA's ($5\ \mu\text{M}$, $40\ \mu\text{l}$) in the array after adding the analyte spiked in physiological buffer or serum ($2.5\ \mu\text{l}$) in six replicates within a 96-well half-area microplate. This provides the fingerprint corresponding to the first recognition event of [TPA + analyte channel] as $(I_1[\lambda_i])$. Subsequently, CB[7] was added to the same wells before recording the fluorescence spectra once again to obtain the fingerprint corresponding to the second recognition event [TPA + analyte + CB[7] channel] as $(I_2[\lambda_i])$ (Figure 3.2). This protocol was established after evaluating the effects of different sequences of CB[7] and analyte addition to the microwells. We observed that adding CB[7] either before or after the analyte (proteins- BSA, lysozyme, peroxidase and chymotrypsin) did not impact the nature of the fluorescence response of TPA (PhenazPy) recorded (Figure 3.3). Therefore, to efficiently measure both recognition events in the same well, we decided to add the CB[7] after recording the first event.

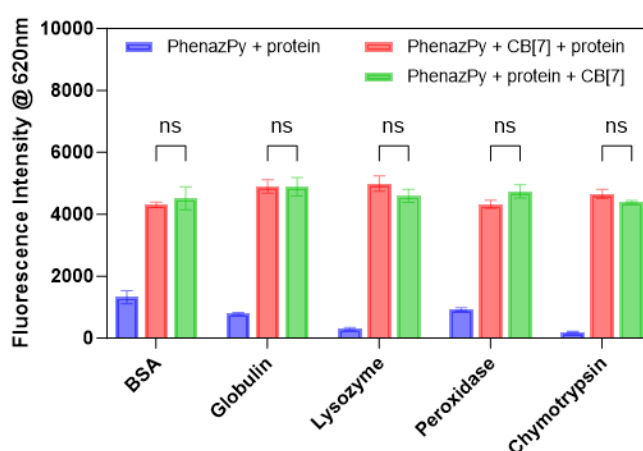


Figure 3.3: Fluorescence intensity variations of TPA derivative PhenazPy ($5\ \mu\text{M}$) in the presence of proteins BSA, globulin, lysozyme, peroxidase and chymotrypsin ($5\ \text{mg/ml}$) and CB[7] ($100\ \mu\text{M}$) with different sequential additions (TP + Protein + CB[7]) and (TP + CB[7] + Protein), showing no impact on the fluorescence response recorded at $\lambda_{\text{ems}}=620\ \text{nm}$.

- Then, having established the protocol to strictly follow for the measurement, we focused on designing the layout of the 96-well microplates to keep them consistent for the development of the R code. As seen in Figure 3.4, the edges of the 96-well plates were not considered for any measurements to avoid the 'edge effect' arising during multiple fluorescence measurements due to increased evaporation rate at the circumferential wells as compared to the centrally located wells on the plate. Each of the plates were assigned for a single TPA derivative, with the vertical columns corresponding to the six replicates for nine analyte being tested (indicated as analyte 1, 2, 3... etc.). For more than nine analytes, two plates were utilized for each TPA derivative.

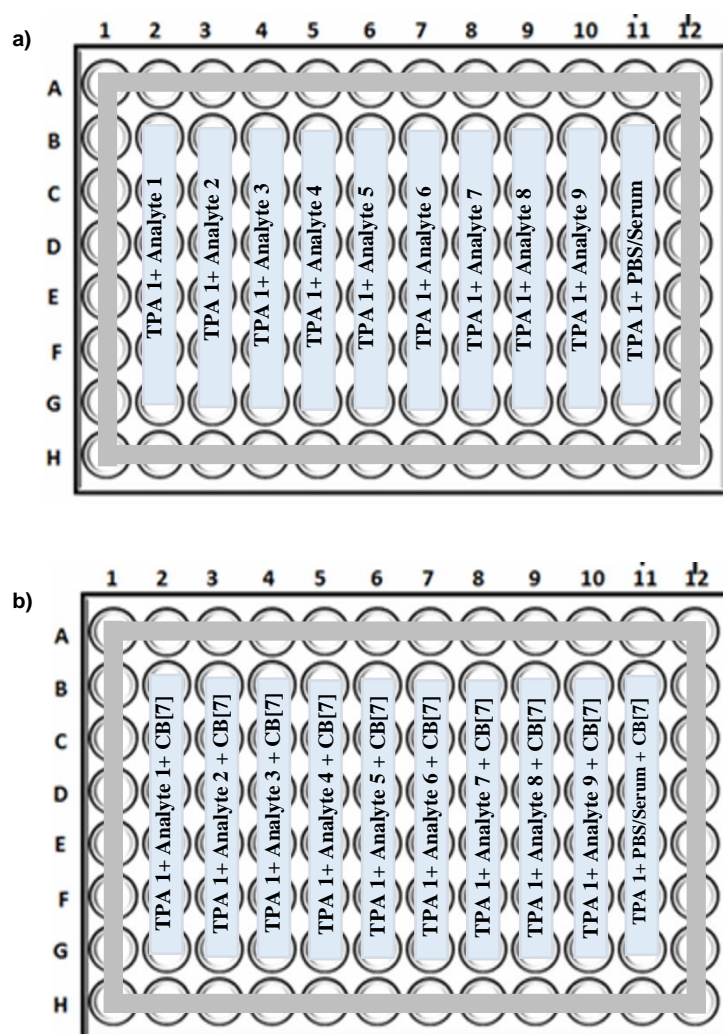


Figure 3.4: 96-well plate layout design adopted for sensing experiments: a) first recognition event corresponds to the first sensing channel [TPA + analyte], b) second recognition event corresponding to the second sensing channel [TPA + protein + CB[7]] measured in the same well so as to maintain consistency for the development of a streamlined data management and statistical analysis workflow.

- While these parameters were optimized to enable the automation of the statistical analysis, it is important to also consider the experimental setup to avoid any design flaws that might introduce artificial trends during the analysis. Ensuring identical conditions and uniform parameters for the sensing of each analyte is crucial.
- When testing clinical samples, it is also essential to include both diseased and healthy patient samples on the same plate and not physically separate them onto two different plates. This approach helps avoid any potential experimental bias during the development of the classification model.
- It is also crucial to understand the difference between technical and experimental replicates to estimate the benefits of each type and prevent false discrimination patterns and incorrect

conclusions. Technical replicates involve data derived from the same stock solutions, helping to evaluate pipetting accuracy and the homogeneity of analyte stock solutions. In contrast, experimental replicates require the entire experiment to be repeated from start to finish, ensuring that discrimination patterns are not based on irrelevant features such as the type of experimental microplate used or the storage conditions of stock solutions.

3.2.2. Management of generated fluorescence response data

3.2.2.1. Data extraction from generated raw optical fluorescence fingerprint

The fluorescence responses recorded for the two recognition events of the optical fingerprinting experiments were arranged, from the raw data, into a response matrix containing columns corresponding to the number of features detected (number of wavelengths at which the fluorescence emission intensity were measured for each TPA and TPA-CB[7] pair) and the number of rows corresponding to the number of recorded replicates for each of the different analytes (Figure 3.5). This extraction generates a collection of such response matrices corresponding to each of the four TPA derivatives and their complexes with CB[7]. The four TPA's have wavelengths recorded from 492 to 750 nm (130 λ), 518 to 750 nm (117 λ), 506 to 750 nm (123 λ), and 464 to 750 nm (144 λ), for TP_3Pyo, Acri_3Py, PhenazPy and SulfoxPy respectively with a wavelength step of 2 nm, adding up to a total of 1028 features for the dataset. An extra column with the description of the analytes is added for listing the classifier description for analysis by supervised statistical analysis such as the LDA.

3.2.2.2. Pre-processing of response data

The generated optical response data in some instances can be used in its raw form i.e. unprocessed and with absolute values. However, if the data is pre-treated, careful consideration is required to determine the appropriate treatment that is needed for the extracted data. Multiple pre-processing methods should be evaluated before the final application to ensure the best-suited one for the data. Often incorrectly selected methods lead to complications during the development of the final classification models, making it difficult to distinguish whether classifier features are due to pre-processing or the nature of the sensor array. Several methods such as relative scaling, background subtraction, signal averaging, linearization, auto-scaling and range scaling have been adopted in

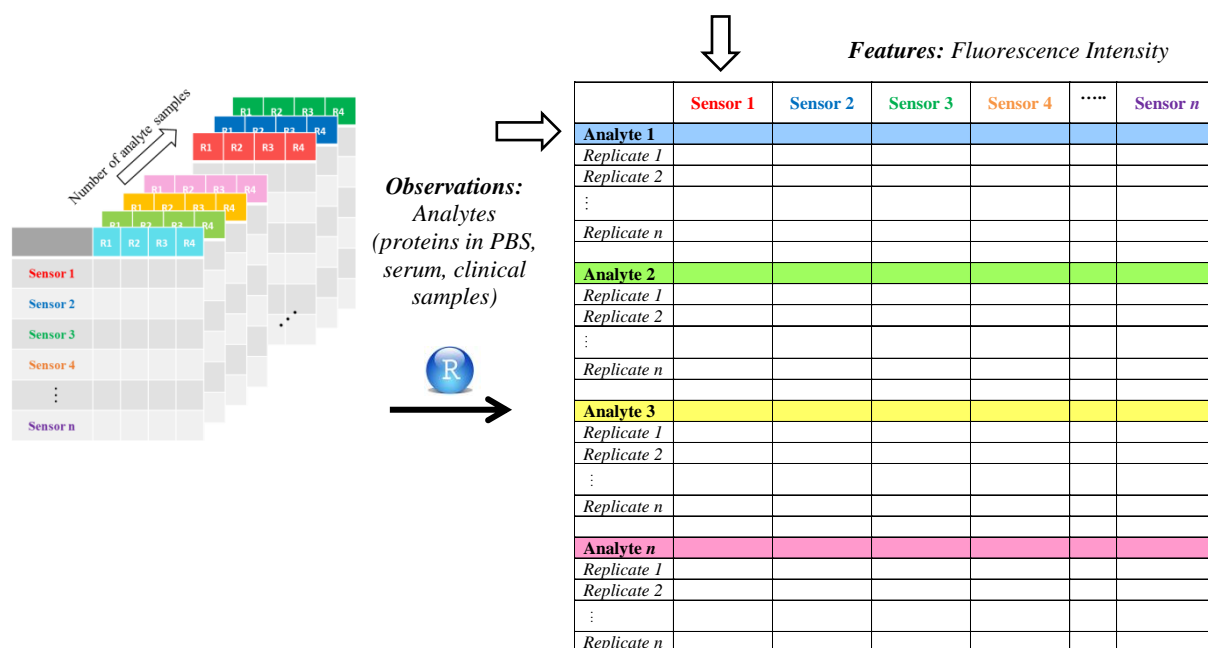


Figure 3.5: Schematic representation of response matrix extracted from the fluorescence fingerprint data generated by the TPA-CB[7] sensor array.

literature. For the TPA-CB[7] sensor array, pre-processing aimed to improve overall sensing resolution by normalizing through subtraction of the mean fluorescence intensity of controls, which corresponds to the fluorescence emission intensity of TPA and TPA-CB[7] complex upon the addition of matrix like PBS or human serum, without the spiked analyte. Based on the number of wavelengths recorded for the four TPA and TPA-CB[n] complexes we obtained a final pre-processed data matrix with dimension of $[N \times 1028]$ with $N = \text{number of analytes} \times \text{number of replicates for each analyte}$. This kind of a normalization approach is often employed for qualitative sensing applications, since some of these methods result in the elimination of the concentration dependence of the sensor response intensity. Thus, this approach is relatively less suitable for quantitative predictions. However, while this normalization approach mitigates the impact of concentration, it does not fully eliminate the potential variability introduced by the excitation light used during the measurement of fluorescence emission from the TPA-CB[n] sensing elements. A more effective strategy would thus involve considering the ratio of the fluorescence emission change in the presence of analytes.

The number of replicate measurements is also crucial, requiring a balanced compromise between the size of the array and the amount of information generated from multiple replicates. Nowadays, limiting replicates is less critical as a time-saving measure due to the advent of high-throughput automated analytical platforms that have reduced the time required for preparation and readout of arrays.

3.2.3. Multivariate statistical analysis of fingerprint response data

The pre-processed data obtained from the data management workflow is formulated in a structured format, ready for in-depth analysis through the application of various pattern recognition methods. In this section, the different approaches employed for the identification, classification, and discrimination of the tested analytes will be discussed in details. Additionally, a detailed examination of the corresponding general R code functions utilized in the analysis will be provided and each method will be explored in terms of its application, effectiveness, and how it contributes to the overall analysis, offering a thorough understanding of the techniques used to interpret the sensor array data in the following chapters.

3.2.3.1. Data sampling methods

The development of pattern recognition models involves three key steps: model selection, training and validation. The selection of an appropriate model is based on the nature of dataset and the classification problem at hand. Once selected, the model trains to identify patterns within the data, often involving feature engineering and adjustments to enhance performance of the model. Finally, the accuracy of the model is validated by testing its ability to perform on unseen data to ensure the reliability of the model. This evaluation of model accuracy is assessed by two fundamental techniques: cross-validation and bootstrapping, which are data sampling techniques that generate subsets of the pre-processed data for the training and testing of the developed pattern recognition model.

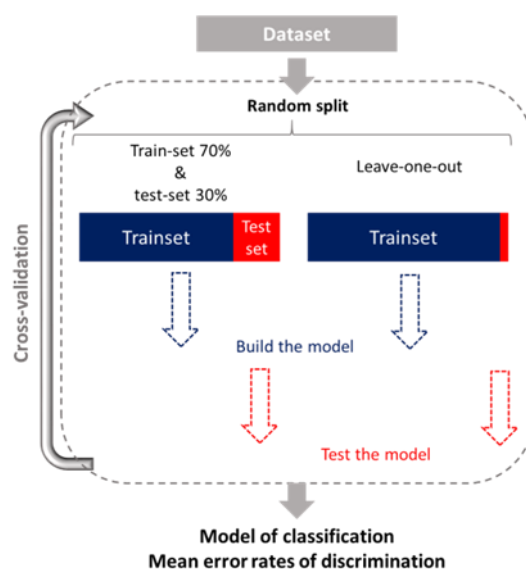


Figure 3.6: Schematic representation of the development and cross validation of classification model using the hold-out and leave-one-out cross validation methodologies.

Cross-validation: This technique involves the partitioning of the pre-processed dataset into subsets. The model is trained on one subset, known as the training set, and evaluated on the remaining subsets, known as test sets. This process is repeated multiple times, and the prediction results of the model are averaged to estimate its overall performance. The commonly used cross-validation methods that were adopted for analysing pre-processed TPA-CB[n] datasets are (Figure 3.6):

- *Hold-out cross-validation:* This is the simplest and most common cross-validation approach, where the dataset is randomly split to assign 70% to a training set and the remaining 30% is assigned to the test set. The model is firstly trained using the training set followed by the validation of the model using the test set. This process is repeated multiple times and the results are averaged
- *Leave-one-out cross validation:* In this approach only a single training set is analysed and one observation is removed at a time, and the model is recalculated using the remaining samples. The removed sample is then used to test the class prediction by the developed model. The most common approach is the Leave-One-Out (LOO) or jackknife cross-validation, where a single observation is left out at a time²⁰³. This process is repeated until all the observations in the dataset are left out and classified. Alternatively, for large datasets, v -folds cross-validation is used, where v -parts of the dataset is taken out and the rest is used to calculate the model. This process repeats v times, and the classification accuracies are averaged.

Bootstrapping: This technique involves resampling the data with replacement, meaning that some data points may appear multiple times in the resampled dataset, while others may not be included at all. The model is trained on this resampled dataset and then evaluated on the remainder of the dataset. The samples that are left out during the resampling are called the out-of-bag (OOB) samples and the error generated by the model in making predictions on the OOB samples is referred to as the OOB error rate. This resampling process is repeated numerous times to ensure a comprehensive evaluation, and the results from these multiple iterations are averaged to provide a more accurate and reliable estimate of the model's overall performance.

The application of these sampling techniques, along with their role in the development of pattern recognition models, will be highlighted in the discussion of each multivariate technique in the following sections of this chapter.

3.2.3.2. Random Forest Algorithm

There has been a gradual and emerging adoption of new pattern recognition algorithms that have been adopted for the analysis of chemical sensor arrays, like the random forest (RF) algorithms. RF is an ensemble learning method, where a set of classifiers are set-up and the aggregated prediction of the classifiers is used to identify the most popular result. RF can be imagined as a congregation of

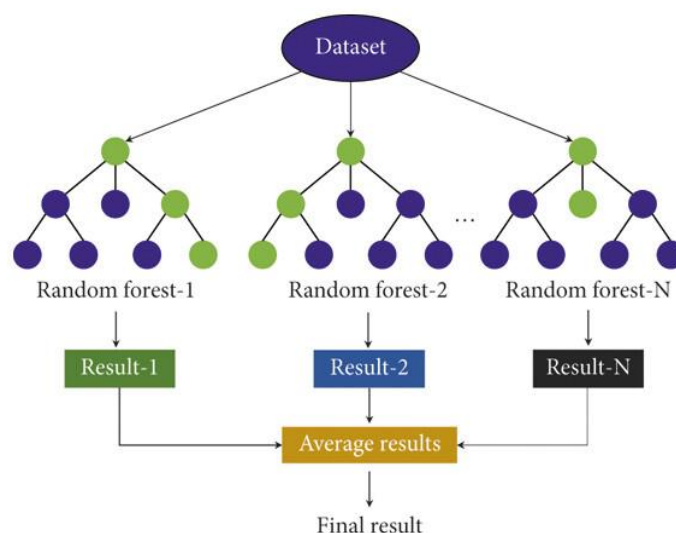


Figure 3.7: Schematic representation of random forest algorithm workflow for classification (adapted from Adyasha, M. et al.).

decision trees and each tree is created from a bootstrap sampling method. This method is robust and reduces variance especially in noisy datasets, while avoiding over fitting of data points. This approach although complex enables fast training, strong generalization ability, and adaptability to unbalanced datasets and multi-classification problems that are common among array-based systems²⁰⁴.

We firstly applied the random forest algorithm to our processed response datasets. The steps involved in the development of the model are described below²⁰⁵:

- **Random observation selection:** Every tree in the random forest is trained on approximately 2/3rd (63.2%) of the total training data. Observations are selected randomly with replacement from the original data and are set as the training dataset for the growth of each decision tree.
- **Random feature selection:** Few of the predictor features (m) are selected in random out of all the features and the best features from m is used to create a split at the node to proceed with the decision tree. The value of m is kept constant throughout the process.
- **Calculation of misclassification rate:** For each decision tree created, the remaining dataset (36.8%) is used to calculate the misclassification rate – ‘out of bag (OOB) error rate’. Aggregated errors from

all the decision trees are then utilized to determine the overall OOB error rate for the classification of the sensor array analytes. For example, if 200 trees are used, one observation on average will be out of the bag for about $0.37 \times 200 = 74$ trees

- ***Classification of test dataset:*** Each tree generates a classification of the type of analyte on the leftover test dataset. This classification is considered over all the trees in the forest and the major vote is considered as the prediction on the test set. For example if 500 trees are fitted with the model, a case is out of the bag in 200 of them: 160 trees could classify this as analyte 1 and 40 trees could classify them as analyte 2, in this case the maximum vote is predicted as analyte 1 for this observation with a probability of $160 \div 200 = 0.8$.

These steps are fundamental to the random forest algorithm and ensure that the model benefits from both the robustness of bootstrapping and the reduction of over fitting through feature randomness and aggregation.

Function:

The R package utilized for the application of the random forest algorithm to the pre-processed data is `randomForest`

The general function for the classification algorithm is:

```
# Load the library
library(randomForest)

# Train the random forest model
model <- randomForest(target ~ ., data=training_data, ntree=500,
mtry=2, nodesize=1, maxnodes=NULL, importance=TRUE)

# Predict on the test set
predictions <- predict(model, newdata=test_data)

# Evaluate the model for classification by generation of confusion matrix
conf_matrix <- table(predictions, test_data$target)
print(conf_matrix)

# Calculate accuracy of classification
accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)
print(paste("Accuracy: ", accuracy))
```

This function has a number of hyperparameters that need to be modified to achieve an efficient classification:

`ntree`: Number of trees to grow, with a default value of 500

`mtry`: Number of variables randomly sampled at each split, with the default being the square root of the number of features for the classification model

`nodesize`: Minimum size of terminal nodes, with a default of 1 for the classification model

`maxnodes`: Maximum number of terminal nodes. If not specified, trees are grown to the maximum number possible

`importance`: Each decision tree in the forest has its own out-of-bag (OOB) observation that was not used during its construction. This OOB sample is utilized to evaluate the importance of a feature. Initially, the prediction accuracy on the OOB sample is assessed. Then, the values of the feature of interest are randomly shuffled, while other features remain unchanged. The change in prediction accuracy due to this shuffling is measured, and the mean decrease in accuracy across all trees is reported as the feature's importance. This importance measure reflects how much the removal of the feature impacts the accuracy of the model. If a feature has minimal predictive power, shuffling may inadvertently improve accuracy due to random noise, potentially resulting in slightly negative importance scores, which are effectively treated as zero importance²⁰⁶(Figure 3.8c).

Results of analysis by random forest model:

The output of the random forest model, as depicted in Figure 3.8a, illustrates the hyperparameters used, including the number of trees and the number of variables considered at each split during the tree-building process. The results also display the OOB error rate, which reflects the errors made in predictions by the model on out-of-bag samples. Additionally, a confusion matrix is provided to evaluate the performance of the model by comparing the labels predicted by the model (e.g. healthy or unhealthy) against the actual labels. To determine the optimal number of decision trees for accurate prediction by the classification model, it is essential to stabilize the plot of the error rates of the model in comparison to the number of trees generated (Figure 3.8b). As expected, fewer trees in the model correlate with higher error rate. However, with an increase in the number of trees, the error rate eventually stabilizes. A lower and stable OOB error rate suggests a well-performing model. Thus, a balance must be struck between utilizing numerous trees to attain a stable prediction and using fewer trees to achieve efficiency.

Thus, upon applying the random forest algorithm, we generated the classification error rate for the entire pre-processed dataset [N×1024]. This required the recording of the full emission spectra of the TPA and TPA-CB[7] sensing elements, thus we reduced the number of recorded wavelengths to ten per TPA and TPA-CB[7] pair with a dimension of [N×80], and subsequently to five wavelengths each with a dimension of [N×40]. At each stage, the random forest algorithm was applied, and variations in the overall classification error rate of the analytes across different media were documented. The hyperparameters were optimized for all the discrimination experiments with `ntree=20,000` and `mtry=0`. For the final dataset [N×40], which included features corresponding

```

a) > Model<-randomForest(hd~.,data=data.imputed,proximity=TRUE)
> Model

call:
  randomForest(formula = hd ~ ., data = data.imputed, proximity = TRUE)
    Type of random forest: classification
    Number of trees: 500
    No. of variables tried at each split: 3

    OOB estimate of error rate: 16.83%
Confusion matrix:
      healthy unhealthy class.error
healthy   142      22  0.1341463
unhealthy  29     110  0.2086331
>

```

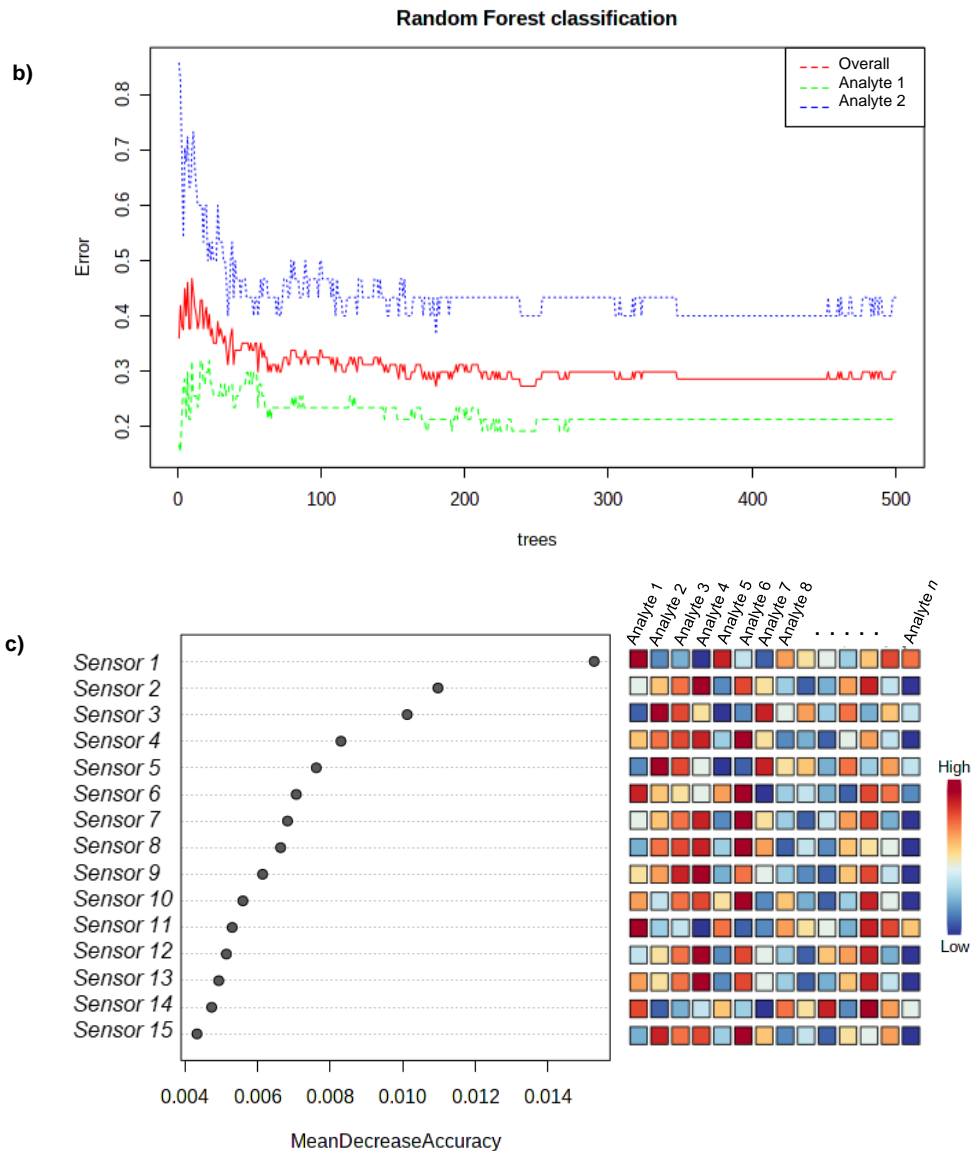


Figure 3.8: Representation of results and plots obtained from random forest algorithm. a) Output results generated by the classification model on R. b) Stabilization plot of the OOB error rate with the increase in the number of trees. c) Feature importance plot depicting the contribution of the features with high importance to the classification by the random forest model, where the features are ranked by their contributions to classification accuracy (MeanDecreaseAccuracy)²⁰⁷.

to five wavelengths for each TPA and TPA-CB[7] pair, the `importance` function was used to identify the top 16 features or wavelengths that most significantly contributed to the accurate classification of target analytes under the given conditions to generate a dataset with dimensions of [N×16]. This refined dataset was then exported for further analysis using other multivariate statistical methods, as discussed further.

3.2.3.3. Linear discriminant analysis (LDA)

The datasets generated from the processing by the random forest analysis were analysed by LDA analysis. This analysis was performed by both the streamlined R code and the software Systat 13, Version 13.2.01.

The LDA analysis enables the dimensionality reduction of the fingerprint data obtained from the sensor array. The analysis generates a new lower dimensional space described by the canonical discriminant factors or canonical functions which best describe the likeness and differences between the target analytes. The first discriminant function is the linear combination of variables that best distinguishes among the groups. The second discriminant function, orthogonal to the first, is the next best combination, and so forth. The first discriminant function can be expressed as:

$$L_{DF} = c + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

where a_1 through a_n are discriminant coefficients, x_1 through x_n are discriminating variables, and c is a constant. The coefficients for each variable in each discriminant function indicate the contribution of that variable to the discrimination between groups; the larger the standardized coefficient, the greater the contribution. The maximum number of functions generated is equal to the total number of observations minus one or the number of features whichever is smaller²⁰⁸.

Function:

The R package utilized for the application of the random forest algorithm to the pre-processed data is MASS

The general function for the supervised classification LDA model utilizing the train and test validation is:

```
# Load the library
library(Mass)

# Load a generic placeholder for describing the datasets
lda_train_test <- function(train_data, train_labels, test_data,
test_labels) {

# Fit the LDA model by grouping the dataset by the labels given to the different analytes to
enable classification and discrimination by the applied model
lda_model <- lda(train_data, grouping = train_labels)
```

```

# Predict the class labels for the test data
  predictions <- predict(lda_model, test_data)$class

# Calculate the accuracy of the model
  accuracy <- mean(predictions == test_labels)

# Return the accuracy
  return(accuracy)
}

# Call the function
accuracy <- lda_train_test(train_data, train_labels, test_data,
test_labels)
print(paste("Accuracy:", accuracy))

```

The `lda_train_test` function takes four arguments:

```

train_data: Training data as a data frame or matrix.
train_labels: Corresponding labels for the training data.
test_data: Test data as a data frame or matrix.
test_labels: Corresponding labels for the test data.

```

The general function for the supervised classification LDA model utilizing the LOOCV validation is:

```

# Load the library
library(Mass)

# Load a generic placeholder for describing the datasets
lda_loocv <- function(data, labels) {

# Perform LDA with Leave-One-Out Cross-Validation (CV=True)
  model <- lda(data, grouping = as.factor(labels), CV = TRUE)

# Compute the classification error rate
  error_rate <- mean(model$class != labels)

# Print the model summary
  print(model)
}

```

Results of analysis by LDA:

The dataset was also analyzed using Systat software with leave-one-out cross-validation. This software like other typical statistical programs automatically creates a two-dimensional plot using the canonical functions or factors (LD1 and LD2) that most contribute to the model's discrimination. While this usually produces a good plot, it may not always be the optimal visual representation of the LDA analysis. In such cases, other canonical functions generated by the analysis should be considered, and alternative plots, like a three-dimensional plot, should be created to observe the distinct clusters of different analytes. Therefore, it is important to carefully consider all canonical functions to visually represent the discriminated data in the most effective way possible. Each plot is represented by canonical functions or factors along the axes, along with the percentage that each function contributes to the overall classification of multiple analytes. This percentage is derived from the eigenvalue generated by the model, which reflects the relative ability of that specific function to capture the most differences among the various analyte classes (Figure 3.9a-b)²⁰⁹. The clustered analyte data is also displayed with confidence ellipses (95%), which help in identifying how close each sample is to the centroid. The centroid serves as the reference point for calculations, ensuring that the distances between different analyte clusters are maximized, while the distances between replicates within the same cluster are minimized. Additionally, the LDA provides a confusion matrix that shows accurate classifications on the main diagonal and misclassifications on the off-diagonal of the matrix, allowing easy identification of analytes that are commonly misclassified (Figure 3.9d)²¹⁰.

The efficiency of this methodology has therefore made LDA the most commonly used supervised technique in recent sensor array applications. The above two approaches have primarily been adopted to streamline the data management and statistical analysis of the fluorescence based TPA-CB[7] sensor array. The associated results of the discrimination study with this pipeline with different analyte sensing will be further discussed in Chapter 4 of this thesis.

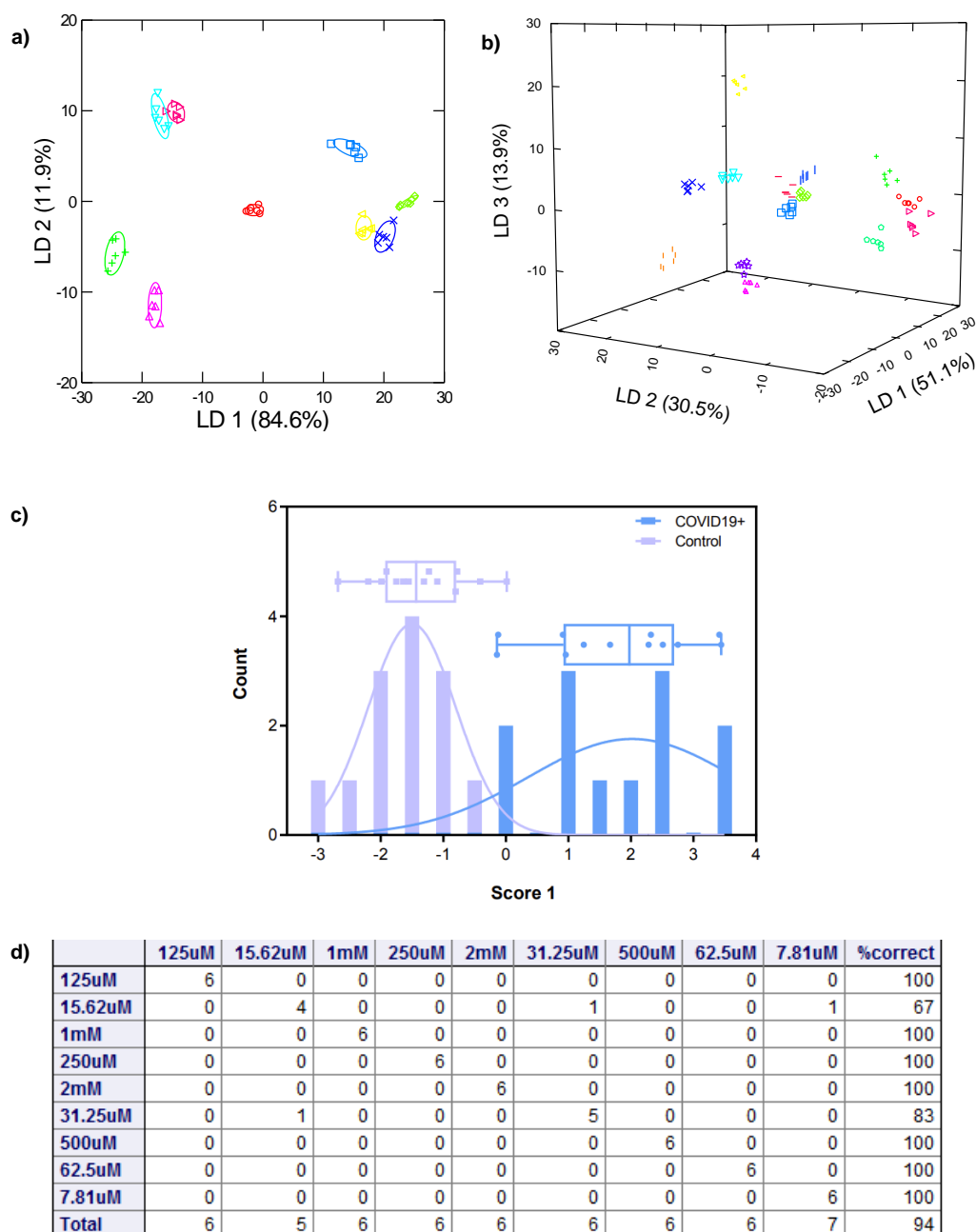


Figure 3.9: Different representations of canonical score plots generated by Systat software include: a) A two-dimensional plot showcasing canonical factors 1 and 2, highlighting their contributions to the classification model, with distinct clusters for different analytes and a 95% confidence interval. b) Three-dimensional plot of the canonical factors 1, 2 and 3 for better visualisation of the non-overlapping clusters of the analytes c) The LDA analysis for a two-class classification model is illustrated as a histogram with normal distributions fitted to the averaged replicates of each data class, while a box plot shows the range of canonical scores obtained from the LDA. d) Jackknifed classification matrix demonstrating the accuracy of the classification model in identifying different analyte classes.

3.2.3.4. Receiver Operating Characteristic curves (ROC)

The receiver operating characteristic curve (ROC) analysis is a statistical tool that is used in medicine to assess the accuracy of diagnostic tests. In general, a diagnostic classification test typically yields a binary outcome indicating whether the tested patient sample is non-diseased ($D_x=0$) or diseased ($D_x=1$). The classification accuracy is estimated by firstly measuring the disease status of each patient without error and comparing it to the gold standard (GS), which is the true status of the patient that is available from clinical follow-up, surgical verifications or autopsy. However, the fundamental measures of diagnostic accuracy are sensitivity (or true positive rate; TPR) - which is the probability that the test correctly classifies a diseased subject as positive and specificity (or true negative rate; TNR) - which is the probability that the test correctly classifies a non-diseased subject as negative. These parameters are represented in a confusion matrix or decision matrix (Figure 3.10a), where the four possible outcomes of a binary classification are represented as true positive, false positive, true negative and false negative.

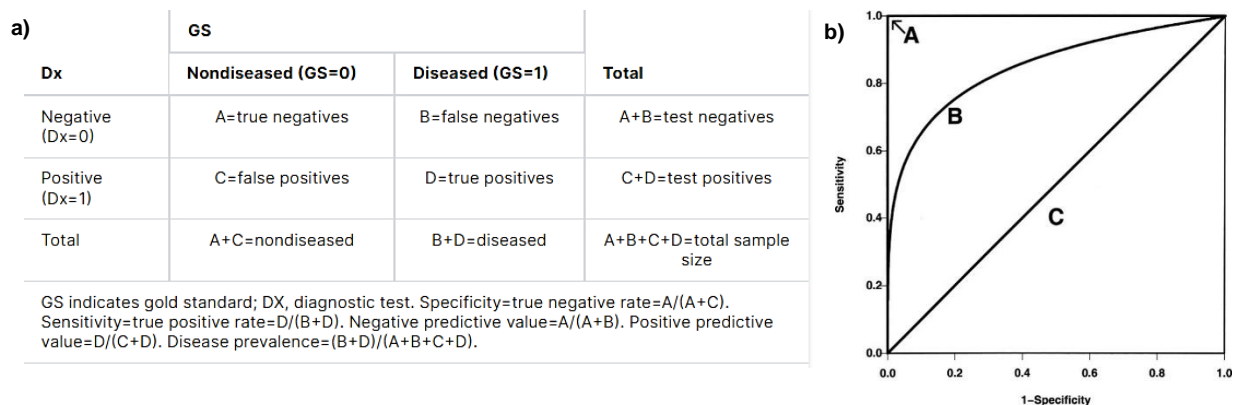


Figure 3.10: a) Decision matrix or confusion matrix generated from the diagnostic test results and the gold standard, to facilitate the plotting of the Receiver Operating Characteristic curve to determine diagnostic accuracy of a given diagnostic test. b) Three hypothetical ROC curves representing the diagnostic accuracy of the gold standard (line A; $AUC=1$) on the upper and left axes in the unit square, a typical ROC curve (curve B; $AUC=0.85$), and a diagonal line corresponding to random chance (line C; $AUC=0.5$). As diagnostic test accuracy improves, the ROC curve moves toward curve A (adapted from Zou, K. H. et al.²¹¹).

The ROC curve is generated by plotting the sensitivity (sensitivity = True positive / (false negative + true positive)) on the y axis against (1- specificity) on the x axis (specificity = true negative / (true negative + false positive)). The 45° diagonal line connecting the coordinates (0,0) to (1,1) corresponds to the ROC curve indicating the possibility of a random chance of the diagnostic test making a right classification, while the ROC curve of the gold standard is the line connecting the coordinates (0,0) to (0,1) and (0,1) to (1,1).

In general, the ROC curves lie between these two extreme cases and the area under the curve (AUC) represents the average diagnostic accuracy across the total set of subjects that have been tested. The AUC takes value from 0 to 1, with 0 indicating a perfectly inaccurate test and a value of 1 indicating a perfectly accurate diagnostic test. In general, an AUC of 0.5 suggests no discrimination (i.e., ability to diagnose patients with and without the disease or condition based on the test), 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding²¹². With improvement in the diagnostic ability of a test the ROC curve moves to curve A as indicated in Figure 3.10b, while the AUC approaches 1. The AUC is often presented with a 95% confidence interval due to variability in the sample data which does not have fixed values but is rather influenced by statistical errors. This interval provides a range of plausible values for the AUC, accounting for statistical errors. For a test to be statistically significant, the lower bound of the 95% confidence interval for the AUC must be greater than 0.5²¹³.

We have also adopted this statistical approach to evaluate the diagnostic capability of the TPA-CB[7] sensor array in a proof-of-concept study for diagnosing the disease model of preeclampsia. To achieve this, LDA analysis assigns a single score to each patient sample, which is subsequently utilized to construct the ROC curve using the Graphpad Prism software (version 10.0). The area under the ROC curve (AUROC) is then calculated to evaluate the clinical significance of the diagnostic test using the TPA-CB[7] sensor array.

3.2.3.5. Principal Component Analysis (PCA)

We have also streamlined the development of another statistical approach (PCA-LDA routine) for the analysis of the TPA-CB[n] colorimetric array that will be discussed in chapter 5 of this thesis. An overview of the workflow will be discussed in this section.

PCA is an unsupervised statistical treatment used for the dimensionality reduction of fingerprint data generated from a sensor array, in such a way that the maximal variance of the data is retained in the dataset. It rotates and combines the original dataset by generating orthogonal eigenvectors for new orthogonal axes that result in the apparent shifting of the data points to be centered around the origin of each axis. This is the main advantage of the PCA, as it is often not necessary that the most important features of a dataset are those with the greatest variance, and the combination of variables that yields the most variance-explaining vector is not always obvious. In large datasets, such as those with spectroscopically obtained fingerprints, there is often significant redundancy among features, mostly wavelengths. PCA eliminates this redundancy by identifying collinear features and representing them on a single orthogonal axis. Essentially, PCA finds the axes that best fit the n dimensional data space and projects them into a simpler space. The contributions of each sensing element to the PCs are represented as vectors on specific plots known as ‘loading plots’, from which the ones with lowest

contributions/loading of similar and highly correlated loadings can be identified and removed without affecting the overall discrimination of the array.

Thus, this approach was used in the development of the LDA-PCA routine. An initial LDA and PCA analysis is performed on the fingerprint generated from the TPA-CB[n] colorimetric based sensor array. While the LDA analysis provides an initial percentage of classification error, the PCA analysis identifies the sensors that contribute least to discrimination, based on the dimensions of the analysis. Using this information, the features that contribute least to the discrimination are excluded. A subsequent LDA analysis is then conducted without these excluded features to determine if the percentage of classification error decreases. This iterative process continues until no further improvement in discrimination is observed. This facilitated the optimization of the array into a more efficient and experimentally convenient array with the highest accuracy of discrimination and reduced dimensions, enabling efficient data treatment time (Figure 3.11).

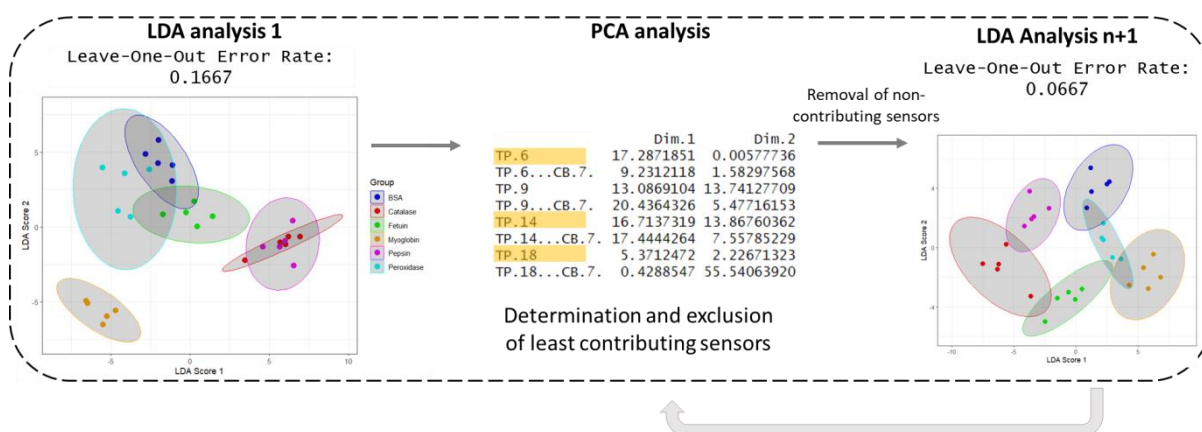


Figure 3.11: Representation of the LDA-PCA routine adopted for the analysis of the colorimetric TPA-CB[n] sensor array indicating the improvement in the LOOCV error rate of LDA discrimination by excluding non-contributing features identified by PCA.

The R package utilized for the application of the to the pre-processed data is

FactoMineR and factoextra.

The general function for the unsupervised classification by PCA is:

```
# Load the library
library(FactoMineR)
library(factoextra)

# Load a generic placeholder for describing the dataset and allows for optional scaling
perform_pca <- function(data, scale_data = TRUE, graph = TRUE) {

# Perform PCA and display the PCA graphs
pca_result <- PCA(data, scale.unit = scale_data, graph = graph)
```

```
# The variance explained by each PC is extracted
explained_variance <- pca_result$eig

# Return a list containing the PCA results and explained variance
list(
  pca_result = pca_result,
  explained_variance = explained_variance
)
}
```

This LDA-PCA routine allowed for the reduction of a 33-channel colorimetric sensor array to an 18-channel array while maintaining complete and accurate discrimination of the tested analytes. Thus, these two techniques are most effective when used in conjunction to optimize the analysis of the sensor array.

3.3. Conclusions

The analysis of the high-dimensional data generated during array-based sensing necessitates the development of a precise and efficient methodology for processing and statistical analysis. To address this, we have created an automated process for data management and analysis using R. Initially, we optimized the experimental design and data acquisition protocols for both fluorescence and colorimetric TPA-CB[n] sensor arrays. This included the measurement of appropriate controls and their integration into the data management and preparation process. Following this, the generated raw data was meticulously extracted and processed to enhance the resolution of the output signal. Multivariate statistical analysis has been applied to the pre-processed data to identify the most distinctive features of the sensor array that best discriminate the different analytes. By leveraging these identified classifiers, we implemented pattern-recognition algorithms such as random forest and linear discriminant analysis (LDA). These algorithms facilitate the streamlined identification, classification, and discrimination of analytes by the TPA-CB[n] sensor array across different sensing applications under different experimental conditions.



Chapter 4

***Optical fingerprinting for diagnostic applications: Proteins and
disease models***

4.1. Introduction

Proteins form vital constituents of the human body, serving as crucial components in various physiological processes. They play an integral role in cellular physiology, being meticulously regulated and affected during diseased conditions. As such, proteins are an important class of biomolecular targets that can be utilized as invaluable biomarkers for diagnosing and monitoring pathological conditions, as well as evaluating the efficacy of treatments. The development of advanced protein sensing technologies is therefore a promising route toward more accurate and comprehensive protein analysis and disease diagnostics. Such advancements have great potential in revolutionizing the emerging fields of precision and personalized medicine, enabling more tailored and effective healthcare solutions²¹⁴.

However, detecting proteins poses a greater challenge compared to other biomarker species. Unlike nucleic acids, proteins cannot be directly amplified or targeted using base-pair complementarity. To achieve the necessary specificity and sensitivity, protein detection assays generally rely on surface capture of target molecules by affinity reagents such as in ELISA, which are time-consuming, expensive and may also lead to false-positive results. This approach also falls short when dealing with diseases for which no specific biomarkers have been identified, as well as for the early diagnosis of diseases that are highly multifactorial and linked to changes in multiple biomarkers in body fluids. Developing multiple antibody-based affinity capture probes to target these biomarkers is a challenging task, and these probes may often lack the sensitivity required to detect very low concentration changes that are observed in the early stages of a diseased state²¹⁵.

Array-based profiling offers an attractive complement to such specific biomarker-focused strategies. The selective interactions of the sensor elements with biomolecules present in body fluids facilitate diagnosis even in the absence of specific disease biomarkers. These arrays can non-invasively capture holistic differences in body fluids by generating cross-reactive signals for multiple analytes at once, effectively capturing the compositional diversity of biomolecules encoded in unique fingerprint signatures, rather than being limited to specific target molecules. This approach enhances the accuracy for early disease diagnostics²¹⁶. Effective global fingerprinting of human serum with these arrays necessitates training them to recognize a wide range of structurally similar analytes, such as amino acids, carbohydrates, and proteins, and ultimately healthy or diseased state.

Therefore, with these aspects in mind, we propose to train the developed TPA-CB[7] sensor array in the identification and classification of proteins in both simple and complex matrix, followed by the fingerprinting of artificial urine and human serum sample to establish a proof-of-concept for the application of this system for non-invasive diagnosis of diseases from body fluids.

4.2. Results and Discussion: Protein fingerprinting

To evaluate the ability of the sensor array to interact with analytes by the displacement of TPA encapsulated within CB[7], we assessed the fluorescence modulations of the four selected TPA's and their corresponding TPA-CB[7] complexes with a few model proteins. The chemical and topological features on the surface of proteins provide essential information about their identity, as these features depend on the solvent-exposed amino acid residues²¹⁷. CB[7], as a recognition scaffold combines multiple orthogonal binding interactions (encapsulation within hydrophobic cavity, H-bonding, electrostatic, dipolar interaction) to provide a complex target probing mechanism. This is achieved via primary recognition by direct interaction with amino acids such as Phe, Tyr, and Trp, paired with secondary effects arising from the three-dimensional geometric arrangement of adjacent residues that participate in secondary interactions^{196,197,218,219}. This mechanism results in varying binding affinities of different proteins with CB[7], leading to differential displacement of encapsulated TPA, inducing variations in their fluorescence and resulting in the development of unique fingerprint responses. As detailed in Chapter 3, this generated fingerprint response will be evaluated using two primary statistical methods: the random forest algorithm, which will produce an error rate of protein classification, and LDA analysis, which will offer a two or three-dimensional plot showing the percentage of accuracy in discriminating the different protein analytes.

4.2.1. Sensitivity of TPA-CB[7] sensor array toward proteins

To this end, we initially evaluated the fluorescence response of the sensor array across a range of concentrations, from 200 mg/ml serially diluted to 0 mg/ml, using representative protein analytes – BSA and trypsin spiked in 1X PBS. Considering that the total protein concentration in serum is approximately 60-80 mg/mL, with 50% being albumins and 40% globulins²²⁰ and the remaining 10% corresponding to other proteins, this initial test was performed across a broader concentration range to determine the sensitivity of the assay, i.e. the concentration range over which the TPA-CB[7] sensor array exhibited efficient discrimination ability. This experiment also enabled us to select a suitable and consistent analyte concentration for all future discrimination experiments. We firstly recorded the complete emission spectra for each TPA (5 μ M, 40 μ L) in the array after the addition of the various concentrations of the two proteins (2.5 μ L), in six replicates in a microplate. Subsequently, CB[7] (100 μ M, 10 μ L) was added to each well, and the fluorescence spectra were recorded again. The fluorescence emission intensity (I) at few specific wavelengths for each of the TPA-CB[7] sensing elements were selected for further statistical analysis (Figure 4.1a). These included fluorescence intensity at wavelengths close to the emission maxima wavelength, at 600, 620 nm for TP-3Pyo and AcriPy; 590, 620 nm for PhenazPy; and 540 nm for SulfoxPy, under two conditions: (TPA + protein) and (TPA + protein + CB[7]). This approach generated a 14 λ channel-based unique fluorescence fingerprint for each

concentration. An example of fingerprint with BSA is represented by the heatmap in Figure 4.1b. While the heatmap presents a visual representation of the fluorescence fingerprint, it does not provide suitable information regarding the extent of discrimination by the TPA-CB[7] sensor array between the different concentrations being tested. Thus, this fluorescence fingerprint was analyzed using a straightforward LDA analysis based on the absolute values of fluorescence emission intensities measured. From the canonical score plot of the first two discriminant factors obtained from the dimensionality reduction of the fluorescence response fingerprint, two linear ranges were extracted for BSA, one from 0 to 12.5 mg/mL (Figure 4.2a-c) and one from 25 to 200 mg/mL and a single linear range was extracted from 0.39 to 12 mg/mL for trypsin (Figure 4.2d-f). We used these results with BSA and trypsin to choose a low concentration of 0.625 mg/mL for all protein samples in subsequent discrimination studies. This concentration was chosen to ensure that the sensor array could detect low protein concentration modulation, facilitating its use for diagnostic purposes where protein variations in the secretome might undergo only subtle changes amidst a background of numerous other interfering proteins during the early stages of a disease state.

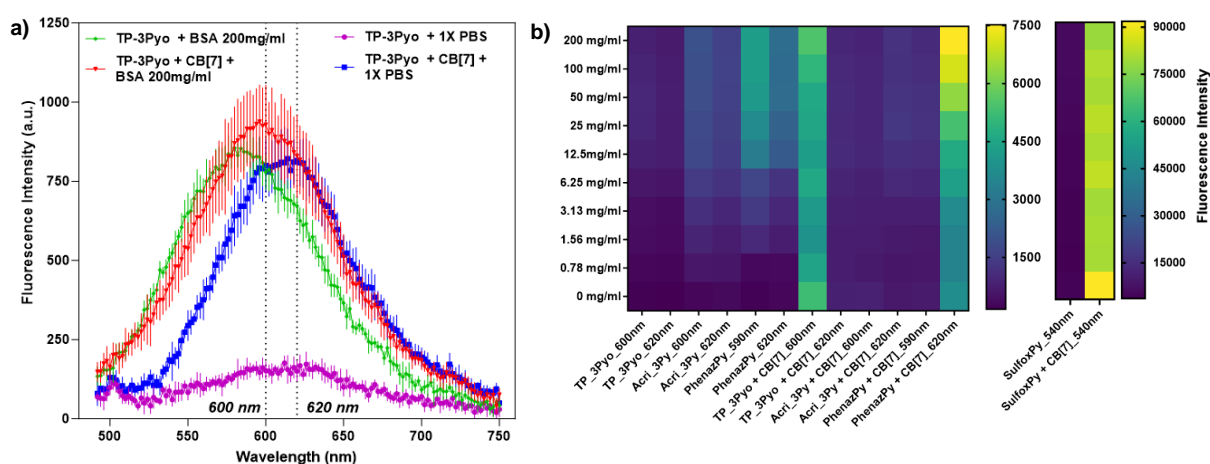


Figure 4.1: a) Fluorescence emission spectra of TP_3Pyo and corresponding complex with CB[7] in the presence of protein BSA at a concentration of 200 mg/ml (green and red resp.) and 0 mg/ml as represented by controls with the addition of 1X PBS alone (purple and blue resp.). The dotted lines represent the wavelengths (600 and 620 nm) chosen for the generation of the fluorescence fingerprint of the sensor array. b) Heatmap representing the fluorescence fingerprint generated by the TPA-CB[7] sensor array with different concentrations of BSA.

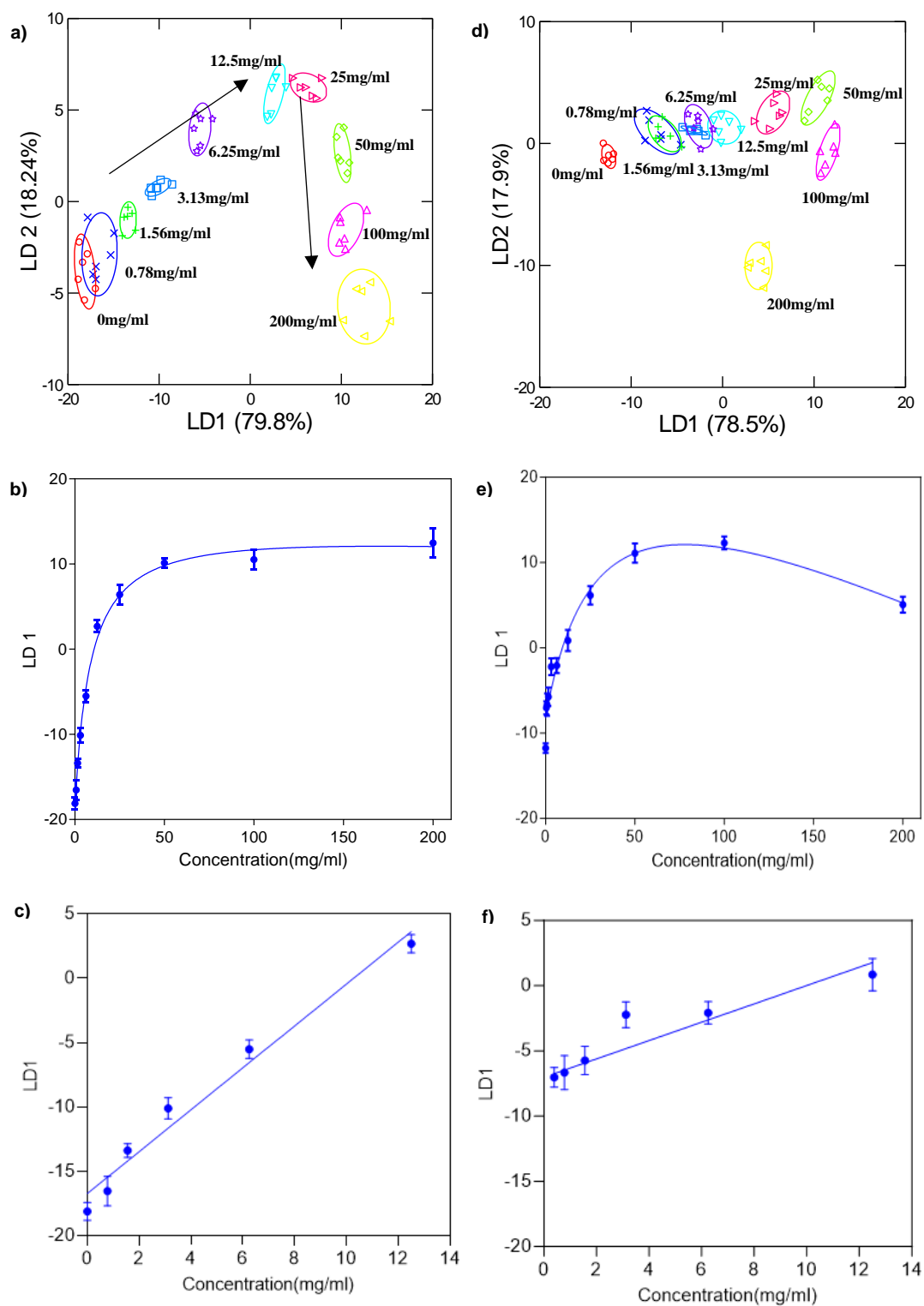


Figure 4.2: Canonical score plot of the first and second factors of fluorescence fingerprint obtained with TPA-CB[7] sensor array against varying concentrations of proteins a) BSA, d) trypsin (200mg/ml serially diluted to 0mg/ml). b) Plot of first discriminant factor versus concentration of b) BSA and e) trypsin along with the corresponding linear detection range for c) BSA and f) trypsin with the TPA-CB[7] sensor array.

4.2.2. Discrimination of proteins in physiological buffer

a. Fluorescence modulation of sensor array with proteins

In this section, we aimed to monitor the fluorescence response patterns generated by the four different TPAs and their corresponding TPA-CB[7] complexes with a few model proteins spiked in phosphate-buffered saline (PBS). This evaluation was intended to assess if the sensor array was capable of generating differential interactions with various protein analytes.

Thus, the fluorescence emission spectra of a three-arm derivative (Acri_3Py) and a two-arm derivative (PhenazPy), along with their respective CB[7] complexes, were measured alone and in the presence of few proteins (here catalase and peroxidase spiked in PBS, 0.625 mg/mL) to analyse their interactions.

- As shown in Figure 4.3a-b, the expected increase in the fluorescence emission of the TPA derivatives upon complexation with CB[7] was observed due to restriction in intramolecular rotation, with approximately 4-fold and 40-fold increases for Acri_3Py and PhenazPy, respectively.
- The addition of the proteins to the TPA derivatives resulted in variations in their fluorescence properties. This can be attributed to the interaction of the TPA derivatives with the hydrophobic solvent exposed residues of the proteins. The extent of the variation was dependent on the specific protein as seen in Figure 4.3c. While peroxidase enhanced the fluorescence of both derivatives to the same extent, catalase resulted in a greater enhancement with PhenazPy as compared to Acri_3Py.
- Furthermore, it was observed that the CB[7] complex of the TPA derivatives exhibited a decrease in fluorescence emission with both the proteins (Figure 4.3a-b). This behaviour indicates the competitive displacement of the TPA derivative from the cavity of CB[7] by the proteins. Furthermore, the emission maxima wavelength observed after the displacement remain nearly identical to those seen when proteins interact with TPA alone, suggesting that residual interactions between the proteins and the TPA derivatives persist even after the displacement.
- The fluorescence decrease in the Acri_3Py-CB[7] complex was noticeably more pronounced with peroxidase than catalase, indicating the potentially higher binding affinity of peroxidase to CB[7] as compared to catalase, leading to a larger extent of displacement of TPA from the CB[7] cavity.
- A significant hypsochromic shift was also observed upon addition of proteins to the TPA or TPA-CB[7] complex. This is primarily due to the binding of the TPA derivatives along the hydrophobic regions of the proteins. This binding shields the TPA derivatives from the surrounding water molecules, leading to a reduction in polarity that can be correlated to the blue-shift of the emission wavelength.¹⁵⁹

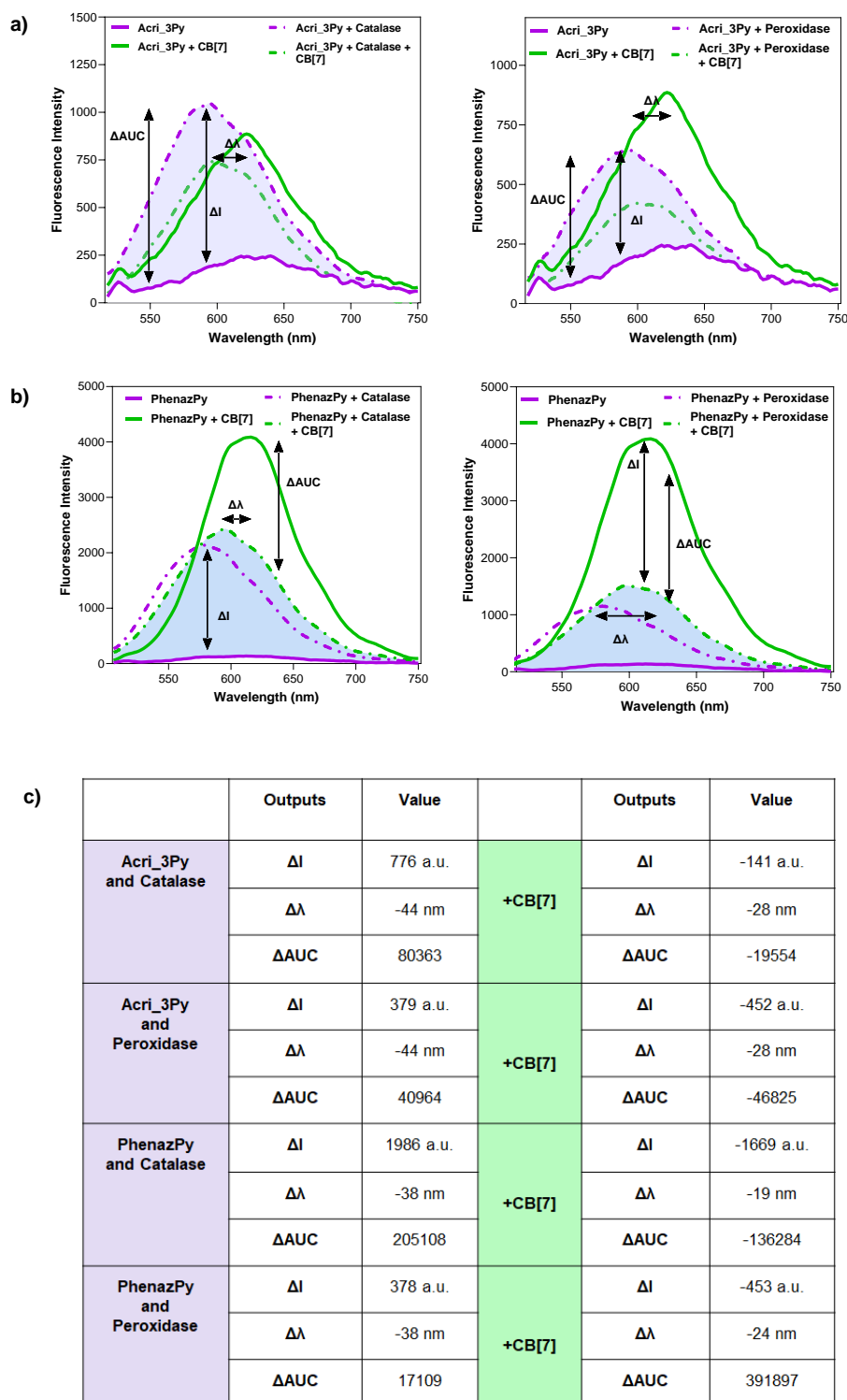


Figure 4.3: Modulations in the fluorescence emission spectra of a) Acrid_3Py and b) PhenazPy (5 μM) and their complex with CB[7] (100 μM) in the presence of the proteins catalase and peroxidase (0.625 mg/mL). c) Output parameters (ΔI : change in fluorescence intensity; $\Delta\lambda$: shift in wavelength of emission maxima; ΔAUC : change in area under the fluorescence emission curve) extracted from the fluorescence emission spectra of Acrid_3Py and PhenazPy.

Thus, from these results we could infer that the interaction of each protein with TPA and TPA-CB[7] modifies the associated complexation equilibrium and creates specific modulations in the photophysical properties of the TPA derivatives. These alterations in the optical properties result in unique fluorescence patterns or fingerprints which are characteristic of each of the tested protein analytes. A representation of this kind of differential fluorescence fingerprint generated by the TPA-CB[7] sensor array with proteins - catalase, peroxidase and lysozyme is shown in Figure 4.4. While the whole fluorescence emission spectra of the TPA-CB[7] sensing elements were registered in the presence of the proteins, we only used at this stage the fluorescence intensity variation ΔI measured at $\lambda_{\text{ems}} = 620\text{nm}$ for Acri_3Py, TP_3Pyo, PhenazPy and 540nm for SulfoxPy respectively, as discussed in section 4.2.1. The ΔI is calculated such that $\Delta I = I - I_0$, with I corresponding to the fluorescence intensity of TPA or TPA-CB[7] in the presence of protein and I_0 corresponding to the fluorescence intensity of TPA in the presence of PBS only. The combined ΔI variations generated by the four TPA derivatives has been considered to obtain a heatmap representing the fluorescence fingerprint of each of the tested proteins (Figure 4.4b).

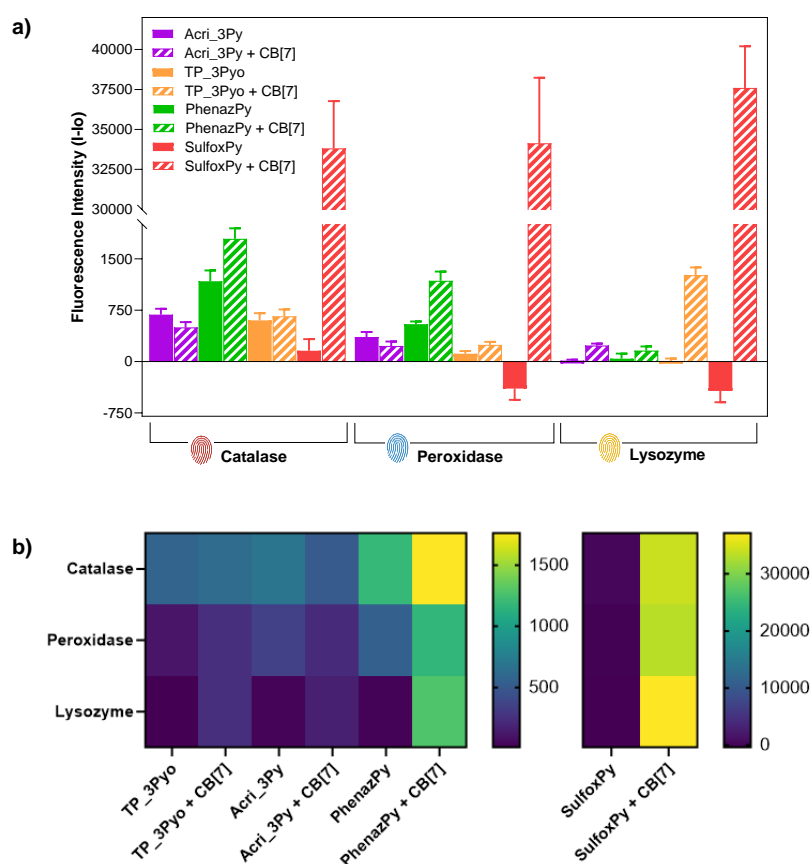


Figure 4.4: a) Fluorescence emission intensities of proteins catalase, peroxidase and lysozyme with four TPA derivatives and their corresponding complexes with CB[7] measured at $\lambda_{\text{ems}} = 620\text{nm}$ for Acri_3Py, TP_3Pyo, PhenazPy and 540nm for SulfoxPy respectively. b) Heatmap depicting the unique fluorescence fingerprint of each of the proteins as generated by the TPA-CB[7] sensor array.

b. Discrimination of a set of proteins by sensor array

Since, we observed differential fluorescence responses for various proteins with the different TPA's, we further assessed the discrimination performance of the developed sensor array. This involved evaluating its ability in discriminating a wider range of proteins by the generation, reduction and statistical analysis of their fluorescence fingerprints.

We spiked 14 biologically relevant proteins with a wide range of molecular weights and isoelectric points (Materials and methods), at a concentration of 0.625 mg/mL in 1X PBS and examined the fluorescence response of each element of the sensor array (Figure 4.5). Again, we firstly recorded the complete emission spectra of all the TPA's in the array in the presence of the 14 proteins ($I_1[\lambda_i]$) in six replicates in a microplate. Subsequently, CB[7] was added to the wells before recording the fluorescence spectra once again ($I_2[\lambda_i]$). This protocol was designed to measure the two recognition events effectively (TPA + Proteins and TPA + CB[7] + Protein) in a timely manner, so as to obtain maximum replicates to ensure robust statistical analysis. Thus, the background correction of the fluorescence intensity for the determination of ΔI was done by considering I_0 corresponding to the fluorescence intensity of TPA or TPA-CB[7] upon addition of PBS without any spiked protein. This approach generates a large fluorescence based dataset and to effectively handle the large volume of data at this stage, a pipeline of R codes has been established to automatically organise and format this raw fluorescence fingerprint data. In a first attempt, since we retained the whole registered fluorescence emission spectra, the streamlined data treatment generates here, a dataset with dimensions of $[84 \times 1028]$ (6 replicates of 14 proteins \times number of wavelengths for each sensor element (whole spectra), which was then used for subsequent analysis.

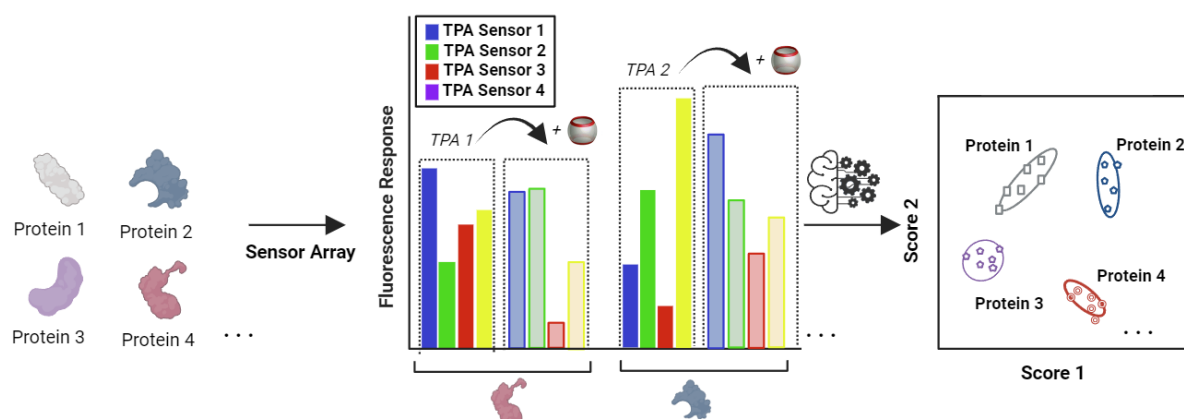


Figure 4.5: Schematic illustration of protein sensing by TPA and TPA+ CB[7] fluorescence sensor array by bimodal recognition interactions.

Statistical analysis: Complete fluorescence fingerprint dataset

The initially generated dataset with dimensions of $[84 \times 1028]$ has been evaluated by a random forest algorithm, which provides a discrimination error rate of 2.4% among the 14 proteins, corresponding to 2 confusions among the 84 predictions made for the 14 proteins (6 replicates \times 14 proteins).

The registration of the complete emission spectra during the course of the above fluorescence experiments enabled the comparison of the discrimination rates obtained by taking into consideration

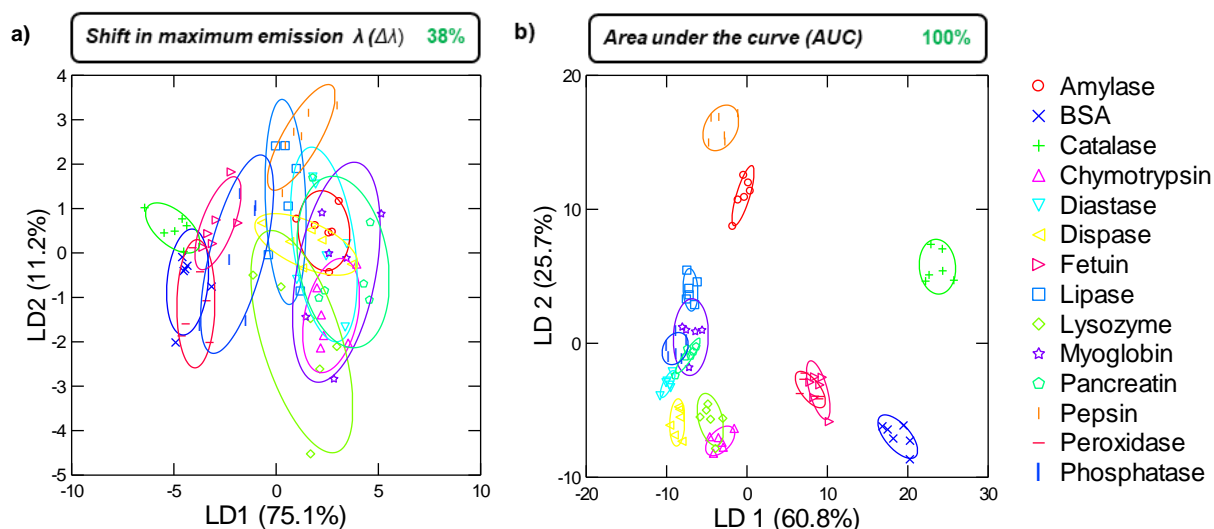


Figure 4.6: Canonical score plot of the first and second factors of fingerprint response obtained based on a) shift in maximum wavelength of emission ($\Delta\lambda$) and b) area under fluorescence emission spectra of the TPA-CB[7] sensor array elements in the presence of 14 proteins spiked in 1X PBS.

other variable output parameters such as shift in the maximum emission wavelength ($\Delta\lambda$) or area under the curve (AUC). While these outputs reduce the dataset dimensions from $[84 \times 1028]$ to $[84 \times 8]$, considering two outputs for each TPA/CB[7] pair, they still require the time-consuming measurement of the complete emission spectra to extract these parameters. We analysed the discrimination accuracy of the 14 proteins by using either $\Delta\lambda$ or AUC as different outputs. While the analysis with AUC allows a complete discrimination with an accuracy of 100% over the 14 proteins (Figure 4.6b), using only $\Delta\lambda$ was found to be non-discriminant with an accuracy of 38% (Figure 4.6a).

Statistical analysis: Reduced fluorescence fingerprint dataset

Therefore, instead of relying on these output parameters, we systematically reduced the number of features of the dataset based on the fluorescence intensity variations, to enable faster fluorescence

measurements with multiple replicates and improve the resolution of the output signal by reduction in noise generated by features that are non-discriminant towards the protein analytes.

To achieve this systematically, we adopted two main steps:

Increase the step between wavelengths to register the fluorescence emission spectra: When a complete emission spectrum is recorded, for instance, for SulfoxPy, the spectrum is registered between 464 to 750 nm with a step of 2 nm, generating 144 measurements per spectral scan. Similarly, for TP_3Pyo, Acri_3Py, and PhenazPy, there are 130, 117, and 123 measurements per spectral scan, respectively. To reduce the dataset, we initially decided to increase the step between each consecutive wavelength by selecting only five wavelengths for each spectrum. This would reduce the time of measurement and resulted in a reduced dataset with dimension of $[84 \times 40]$. It was observed that by reducing the size of the original dataset to this dimension there was no effect on the discrimination accuracy of the 14 proteins spiked in PBS.

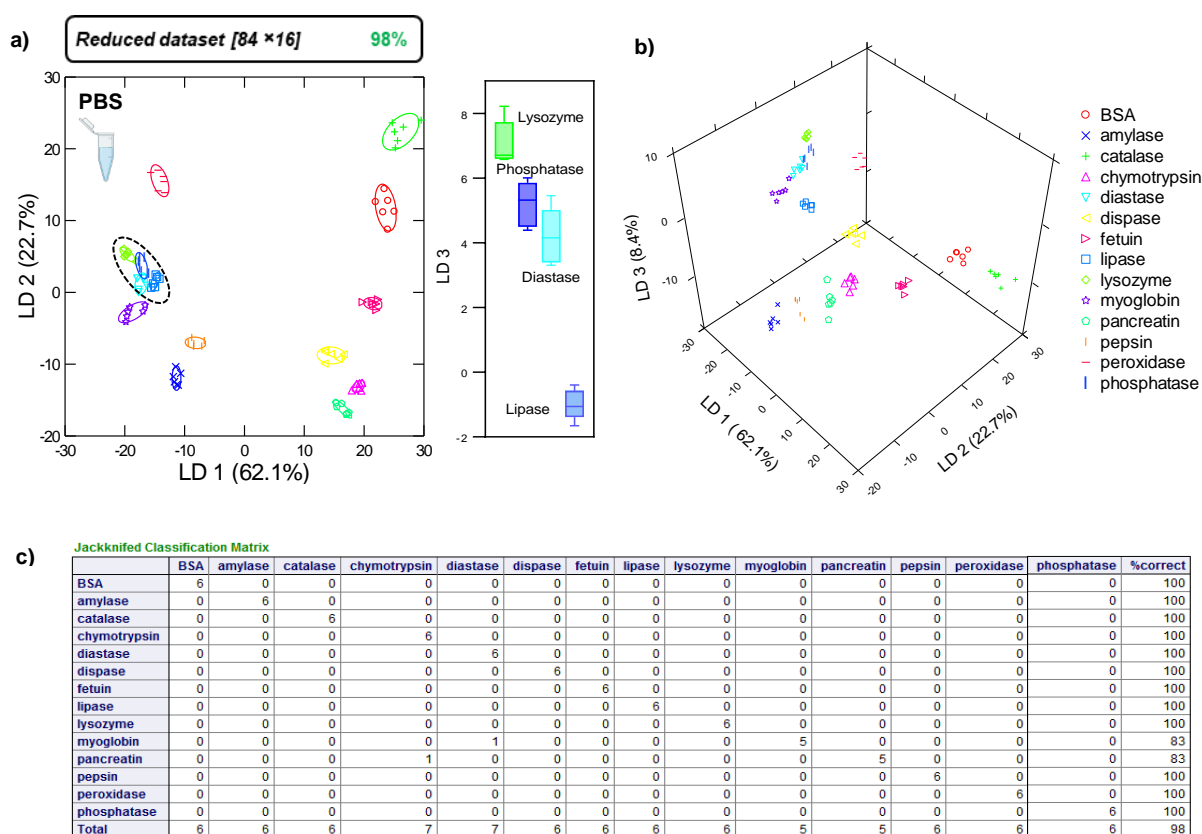


Figure 4.7: a) Canonical score plot of the first and second factors of fluorescence response pattern calculated by LDA for the identification of 14 proteins in 1X PBS with 16 λ channels with adjoining box plot highlighting the circled region to depict clearer separation between overlapped classification clusters. b) Three-dimensional canonical score plot of the first three discriminant factors of the corresponding reduced fluorescent spectral fingerprint $[84 \times 16]$. c) LDA classification accuracy represented by the Jackknifed classification matrix for the reduced dataset.

Selection of wavelengths contributing most significantly towards discrimination: From this reduced dataset with the selected 40 wavelengths, the 16 most important wavelengths for the discrimination were identified by the ‘importance’ function of the randomForest package on R as discussed in Chapter 3. The random forest analysis using this reduced dataset [84 × 16] provided a good discrimination rate of 97.6% in PBS. This dataset was also evaluated statistically by LDA. The leave-one-out cross-validation was evaluated and the plot of the first and second determinant factors, with a 95% confidence ellipse, discriminated the proteins spiked in PBS with an accuracy of 98% (Figure 4.7a). The circled region indicates an apparent overlap among four protein clusters, however the box plot alongside the canonical score plot clearly represents the extent of separation among these proteins against their corresponding LDA 3 scores. Further clarification in the extent of discrimination of the 14 proteins can be visualised with the three-dimensional plot of the LD 1, LD 2 and LD 3 as indicated in Figure 4.7b.

These results clearly demonstrate the sensor array's ability to discriminate diverse protein analytes at low concentrations in a physiological buffer. However, to enhance its utility, it is essential to evaluate the array's performance in more complex matrix systems, especially in the presence of other interfering biomolecules. By assessing the sensor array under these conditions, we can better understand its robustness and expand its applicability to real-world sensing challenges.

4.2.3. Discrimination of proteins in complex matrix: serum and treated serum

a. Protein discrimination in complete human serum by sensor array

To address the utility of the system for clinical and diagnostic applications, we tested the TPA-CB[7] sensor array with the same set of proteins spiked in more complex media as compared to simple PBS. Specifically, we spiked the 14 proteins into complete human serum, which comprises a diverse array of biomolecules including proteins, antibodies, lipids, nutrients, metabolic substances, and more. The fluorescence fingerprints of these proteins were obtained by following the same streamlined methodology used in the previous discrimination experiments conducted in PBS.

Statistical analysis: Complete fluorescence fingerprint dataset

Following the same protocol used for discrimination in PBS, we spiked the proteins into complete human serum to achieve a final concentration of 0.625 mg/mL upon addition to the TPA derivatives. Fluorescence emission spectra for the two recognition events were recorded for each protein in six replicates, resulting in a fluorescence fingerprint dataset with dimensions of [84 × 1028]. This dataset was initially analysed using the random forest algorithm, which accurately discriminated the proteins

with an error rate of 3.6%. As anticipated, there was a slight decrease in discrimination accuracy compared to PBS, which can be attributed to the complexity of the serum matrix and the notably high concentration of albumin and immunoglobulins (IgG), which together constitute approximately 80-90% of the total serum protein concentration.

Statistical analysis: Reduced fluorescence fingerprint dataset

We further processed this data using the streamlined R code to generate a reduced dataset with dimensions of $[84 \times 40]$, based on the selection of five wavelengths for each TPA-CB[7] sensor pair. It was observed that this reduction in dataset size had a limited negative impact on the discrimination error rate, which increased to 7.1% when analysed using the random forest algorithm. This indicates that while the reduced dataset still provided reliable results, some degree of accuracy was compromised compared to the full dataset. We then identified the 16 wavelengths that contributed most significantly to the discrimination of the $[84 \times 40]$ reduced dataset by the 'importance' function with the random forest algorithm. The analysis of this further reduced dataset revealed an increased

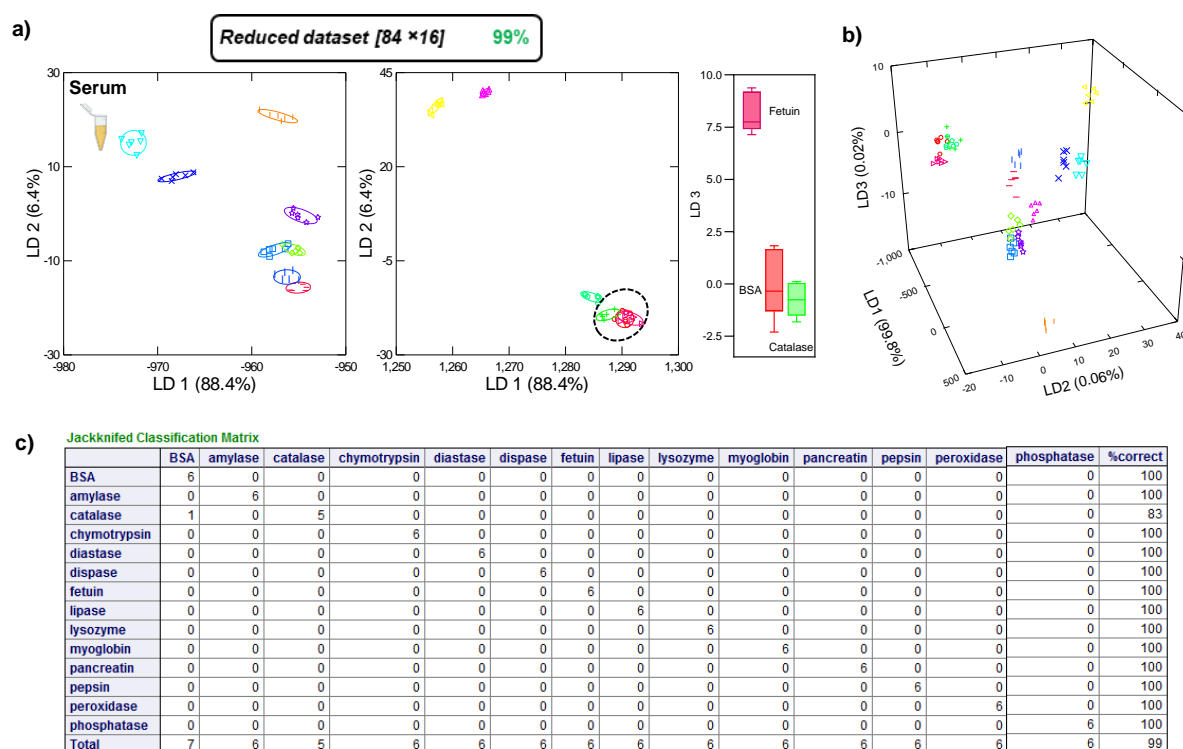


Figure 4.8: a) Canonical score plot of the first and second factors of fluorescence response pattern calculated by LDA for the identification of 14 proteins in complete human serum with 16λ channels with adjoining box plot highlighting the circled region to depict clearer separation between overlapped classification clusters. b) Three-dimensional canonical score plot of the first three discriminant factors of the corresponding reduced fluorescent spectral fingerprint $[84 \times 16]$. c) LDA classification accuracy represented by the Jackknifed classification matrix for the reduced dataset.

discrimination error rate of 8.3%. This further indicated that the reduction in the number of features in the dataset compromised the accuracy in discrimination, particularly in the already challenging matrix of complete human serum. This dataset was also evaluated statistically by LDA. The leave-one-out cross-validation of the reduced dataset [84 × 16] was evaluated and the plot of the first and second determinant factors, with a 95% confidence ellipse discriminated the proteins spiked in serum with an accuracy of 99% (Figure 4.8a). Further clarification in the extent of discrimination of the 14 proteins can be visualised with the three-dimensional plot of the LD 1, LD 2 and LD 3 as indicated in Figure 4.8b. The accuracy of discrimination between the two methods, random forest and LDA, differs and cannot be directly compared. This difference arises from the distinct approaches these methods employ. LDA is less flexible in handling complex data because it requires a normal data distribution and assumes an equal covariance matrix. In contrast, random forest does not require any hypothesis about data distribution and is less sensitive to noise. Therefore, while comparisons can be made for LDA and random forest individually under different conditions, they cannot be compared directly with each other.

b. Protein discrimination in treated human serum by sensor array

The presence of interfering biomolecules in complete human serum may have contributed to the decreased accuracy in discrimination of the spiked proteins. Consequently, we anticipated that the removal of excess proteins would enable better discrimination of the spiked protein analytes, which are present at a much lower concentration (0.625 mg/mL). To achieve this, we depleted the serum samples of Human Serum Albumin (HSA) and IgG using the High Select HSA/Immunoglobulin Depletion Resin, which is optimized to remove over 95% of both IgG and albumin. The proteins were then spiked into this pre-treated serum and evaluated using the TPA-CB[7] sensor array.

Statistical analysis: Complete fluorescence fingerprint dataset

The fluorescence fingerprint generated from the array was firstly analysed by the developed random forest algorithm. A recovery in the discrimination error rate to 2.4% was obtained in comparison to the discrimination in complete human serum.

Statistical analysis: Reduced fluorescence fingerprint dataset

We further reduced the dataset of the protein fingerprints and observed that decreasing the dataset size to [84 × 40] did not negatively impact the discrimination rate in the case of treated serum. In fact, an improvement in the discrimination rate to 1.4% was observed. These 40 wavelengths were further reduced using the ‘importance’ function, resulting in a dataset with dimensions of [84 × 16]. Random forest analysis of this further reduced dataset achieved perfect discrimination with a 0% error

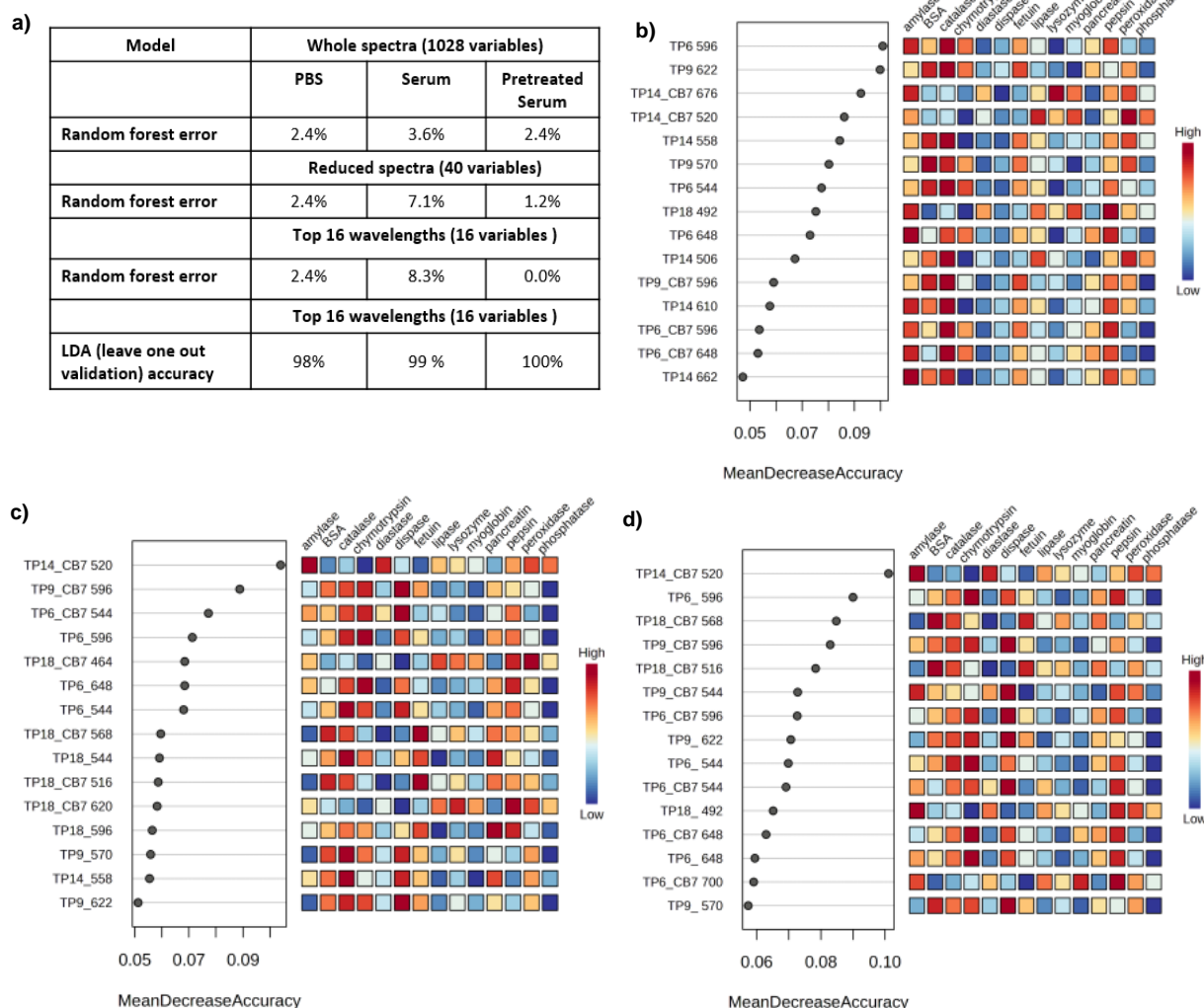


Figure 4.9: a) Discrimination accuracy determined by the streamlined data treatment and analysis methodology adopted in this study for the discrimination of 14 proteins spiked in different matrix. Representation of feature ranking based on contribution to the overall classification accuracy. The features have been obtained with the application of random forest model on the reduced fluorescence spectral dataset of top sixteen wavelengths [84 X 16] for b) PBS, c) human serum and d) treated serum. The figure also illustrates the contribution of these features individually towards classification of the 14 analytes (TP 6 = TP-3Pyo, TP 9 = Acri-3Py, TP 14 = PhenazPy, TP 18 = SulfoxPy)²⁰⁷.

rate. Additionally, leave-one-out cross-validation of the reduced dataset using LDA yielded a discrimination accuracy of 100% for the 14 proteins spiked in depleted serum. While serum depletion enhances protein discrimination accuracy by removing excess proteins, this method has limitations, particularly in diagnostic applications of the sensor array. The process of depleting individual samples from a large cohort can be time-consuming and may introduce errors into the experimental protocol, potentially compromising reliability and reproducibility.

Thus, the developed TPA-CB[7] sensor array demonstrated remarkable potential for distinguishing a variety of proteins at low concentrations in both simple media, such as PBS, and more complex media, like human serum, containing high concentrations of interfering proteins such as HSA and IgG. The discrimination accuracies across these different conditions and an overview of the results from the streamlined data treatment and statistical analysis by random forest and LDA are illustrated in Figure 4.9a. Furthermore, the selected wavelengths in the reduced dataset with dimensions of $[84 \times 16]$ and the extent of their contributions to the discrimination of each of the 14 protein analytes are shown for each of the tested matrix; PBS, complete serum and treated serum Figure 4.9b-d. Among the three groups, 44% of the wavelengths from the 16-wavelength reduced dataset were shared by all three groups, while another 44% were common to two out of the three groups.

These findings emphasize the underlying selectivity mechanism of the array-based system, highlighting its effectiveness in sensing multiple analytes with a limited number of sensor elements. This capability allows the system to detect overall changes in the composition of various complex mixtures and generate distinct fluorescence fingerprints, even in the presence of interfering substances.

4.3. Fingerprinting of disease models: Phenylketonuria

The biomolecular composition of body fluids reflects the onset, severity and progression of diseased states. Thus, detecting subtle alterations in biomolecular components, such as amino acids and proteins, holds great potential for the design and development of diagnostic methodologies. The cross-reactive nature of the chemical nose strategy can thus be a suitable choice to rapidly sense global proteomic changes and profile complex body fluid matrix like blood or urine. To demonstrate the capability of the TPA-CB[7] sensor array for such clinical and diagnostic applications, we tested the sensor array for the discrimination of disease models by profiling body fluid mimics such as artificial urine diluents and true clinical samples such as human serum collected from diseased/ healthy patients. The proof-of-concept of these diagnostic approaches will be discussed in the following sections of this chapter.

4.3.1. Phenylketonuria: State of the art diagnosis

Phenylalanine (Phe) is an essential amino acid acquired through the consumption of protein-rich foods or dietary supplements. It is crucial for the biosynthesis of other amino acids and the maintenance of the structure and function of various proteins and enzymes in the human body. Most dietary Phe is converted into tyrosine (Tyr), which is then converted into several other products, including neurotransmitters such as dopamine, serotonin, epinephrine and norepinephrine. These neurotransmitters play essential roles in coordination, movement, mood regulation and cognitive functions²²¹.

The autosomal recessive metabolic disorder of phenylalanine metabolism is called Phenylketonuria (PKU), and is also known as phenylalanine hydroxylase (PAH) deficiency. This disorder is linked to a defective mutation in the gene that encodes for the enzyme phenylalanine hydroxylase, which converts phenylalanine (Phe) to tyrosine through hydroxylation. PKU leads to the toxic accumulation of Phe in body fluids, which can cross the blood-brain barrier and also be converted into phenylketones such as phenylpyruvate, phenylacetate, and phenylacetate, which are excreted in the urine. If not diagnosed and treated early during the process of neonatal screening, this elevated Phe can cause intellectual disability, behavioural and psychiatric problems, microcephaly, motor deficits, eczematous rash, autism, seizures, and developmental problems^{222,223}. The severity of PKU depends on the degree of enzyme activity, ranging from a complete absence to a milder form with some enzyme functionality but still unusually high Phe levels.

The conventional treatment of PKU primarily focuses on dietary restriction to limit the consumption of excess Phe-rich foods. During this treatment, assessing patient blood Phe levels is crucial for evaluating the effectiveness of the dietary control^{224–226}. Most screening methods for monitoring the Phe levels use blood plasma, including enzymatic commercial assay kits, tandem mass spectrometry and high-pressure liquid chromatography (HPLC). Although these methods provide high sensitivity and accuracy, they require laborious and time consuming sample preparation, skilled technicians and highly sophisticated and expensive testing apparatus/reagents. These laboratory tests often cannot be carried out in resource-limited situations for real-time monitoring. Therefore developing more accessible testing methodologies under various settings is critical²²⁷. Additionally, the analysis of Phe in urine is preferred over blood, since urine-testing is non-invasive and has a higher concentration of Phe than blood²²⁸. Consequently, there is a need for improved methods that are simple, inexpensive and capable of offering rapid determination of phenylalanine in human urine. In the last few years fluorescence and colorimetric-based sensors^{227,229,230} have played a significant role in the monitoring of PKU and digital-imaging with paper based techniques having evolved as a cost-effective approach.

Weiss *et al.* were pioneers in developing a dehydrogenase-based biosensor to monitor Phe levels in human urine without requiring additional reagents. The sensor consisted of a carbon paste electrode incorporating NAD, phenylalanine dehydrogenase (PDH), uricase, and 3,4-dihydroxybenzaldehyde (3,4-DHB), all thoroughly mixed into the paste. In this setup, 3,4-DHB served as a base-stable electron mediator. This mediator interacts with the electrode surface, generating two redox species that catalytically oxidize NADH. This biosensor holds a distinct advantage as it can be effortlessly regenerated through polishing, owing to its preparation with carbon paste. The reported limit of detection (LOD) with this system for Phe was determined to be 0.5 mM²³¹. Further, Sun *et al.* developed hybrid nanoflowers using phenylalanine ammonia-lyase (PAL), an enzyme substitute for PAH, to provide a semi-quantitative measure of Phe concentration in urine samples. The PAL@Ca₃(PO₄)₂ hybrid nanoflowers (PAL@NF) were prepared by combining PAL and Ca²⁺. This

paper-based biosensor demonstrated the ability to detect Phe concentration in urine samples with good linearity in the range of 60 to 2400 μM , achieving a response time of approximately 10 min²³².

We anticipated that the global compositional changes in urine under such diseased conditions would provide an ideal scenario for the utilizing the cross-reactive nature of the chemical nose strategy to rapidly sense these variations, facilitating the development of an easy diagnostic strategy. Particularly, the competitive displacement of TPA's by amino acids from their CB[7] complexes, as described earlier would be a suitable system to facilitate the sensing of Phe in the complexity of the urine matrix of this disease model. Therefore, we used a non-biological diluent that mimics human urine and spiked it with varying concentration of Phe to develop a proof-of-concept for the utilization of the TPA-CB[7] sensor array in the diagnostic sensing of PKU.

4.3.2. Setup of PKU sample cohort using artificial urine diluent

To utilize an array-based system for diagnostic application, it is crucial to train the sensing elements to identify and classify analytes within a highly heterogeneous sample pool. This step is essential because the classification accuracy for both known and unknown samples indicates the effectiveness of the algorithm and whether the system requires further training or additional sensing elements. Achieving algorithms with accurate predictive power can be challenging, as they require a large number of samples to create an effective training set²⁸. A large dataset allows the model to more accurately identify underlying patterns, leading to better generalization when encountering unseen new data. It also ensures statistical reliability concerning class means and variances, and captures variability, making the developed model robust enough to handle different scenarios effectively. In classification tasks, such as disease diagnosis, it is essential that each class is adequately represented by a large sample set to provide a balanced representation and avoid the risk of overfitting, where the model learns the noise in the dataset rather than the actual class signal. It further ensures proper division of data into training and testing sets, which are essential for tuning the mode and assessing overall performance of the model. However, to acquire such a large cohort of clinical samples that are meticulously collected and recorded with their corresponding clinical and biological patient data and disease outcomes, has made the process of accessing these cohorts highly challenging.

Thus, to address this issue and proceed with a proof-of-concept study demonstrating the disease diagnostic capability of the TPA-CB[7] array, we have mimicked the disease model of PKU as it is clinically presented in patient urine samples. We used a non-biological diluent that mimics human urine, spiked with different concentrations of L-phenylalanine comparable to healthy levels, mild and severe forms of PKU. The Sigmatrix Urine diluent (Sigma Aldrich) is primarily a buffer solution containing calcium chloride, magnesium chloride, potassium chloride, sodium chloride, sodium phosphate, sodium sulfate, urea and creatinine with sodium azide as a preservative. By using this diluent instead of human urine, we envisioned the set-up of a large cohort of samples that could enable the training of the sensor

array towards the future possible detection of PKU from patient urine samples. We spiked the artificial urine with various concentrations of Phe (2 mM serially diluted to 7.8 μ M). This range covers the threshold values for the disease, where urinary Phe levels are maintained below 3 mM during dietary-therapy treatment and usually close to the typical reference range of 61-314 μ M in physiological conditions^{168,230,233}.

4.3.3. Optical fingerprinting of L-phenylalanine in artificial urine

We anticipated that the previously described competitive displacement of TPA's by amino acids from their CB[7] complexes would facilitate the sensing of Phe in the context of the phenylketonuria disease model. Since we previously observed minimal fluorescence modulation of the TPA's in the presence of amino acids (Chapter 2), we generated fluorescence fingerprints of the artificial urine samples spiked with various Phe concentrations (9 concentrations \times 6 replicates for each concentration) using only the four TPA-CB[7] complexes of sensor array and did not take into consideration the sensing elements corresponding to the TPA channels alone.

Statistical analysis: Complete fluorescence fingerprint dataset

The fluorescence fingerprints of the different Phe concentrations were obtained by firstly recording the complete emission spectra of the four TPA-CB[7] on addition of the urine diluent as ($I_0[\lambda_i]$) in six replicates. It is important to note that the spectra closely resemble those recorded in PBS, with fluorescence emission intensities falling within the same range. Following this, Phe spiked artificial urine samples were added, before recording the fluorescence spectra once again ($I_1[\lambda_i]$), including control samples of the artificial urine without Phe. The ΔI was calculated for all the six replicates of each concentration, at each wavelength, such that $\Delta I[\lambda_i] = I_1[\lambda_i] - I_0[\lambda_i]$. This generated a fluorescence fingerprint dataset with dimensions of [54 \times 514] (6 replicates of 9 Phe concentrations \times number of wavelengths for each sensor element). This dataset was firstly analysed by the random forest algorithm to obtain a discrimination error rate of 7.4%.

Statistical analysis: Reduced fluorescence fingerprint dataset

Similar to the previous analysis, we reduced the size of the dataset. To achieve this, we first increased the step between each consecutive wavelength, selecting only five wavelengths for each spectrum. This would reduce the time of measurement and resulted in a reduced dataset with dimension of [54 \times 40]. However, it was observed that this reduction led to a significant increase in the discrimination error rate, doubling it to 14.8%. Consequently, reducing the number of features compromised the accuracy of discrimination between different Phe concentrations, indicating that using the complete dataset is more advisable for this study. To refine the dataset further, we identified the 16 most significant wavelengths among the spectra contributing to the maximum discrimination of Phe

using the ‘importance’ function. This generates a dataset with dimensions of $[54 \times 16]$. Applying the random forest algorithm to this refined dataset resulted in an error rate of 7.4%, which was comparable to that obtained with the complete dataset.

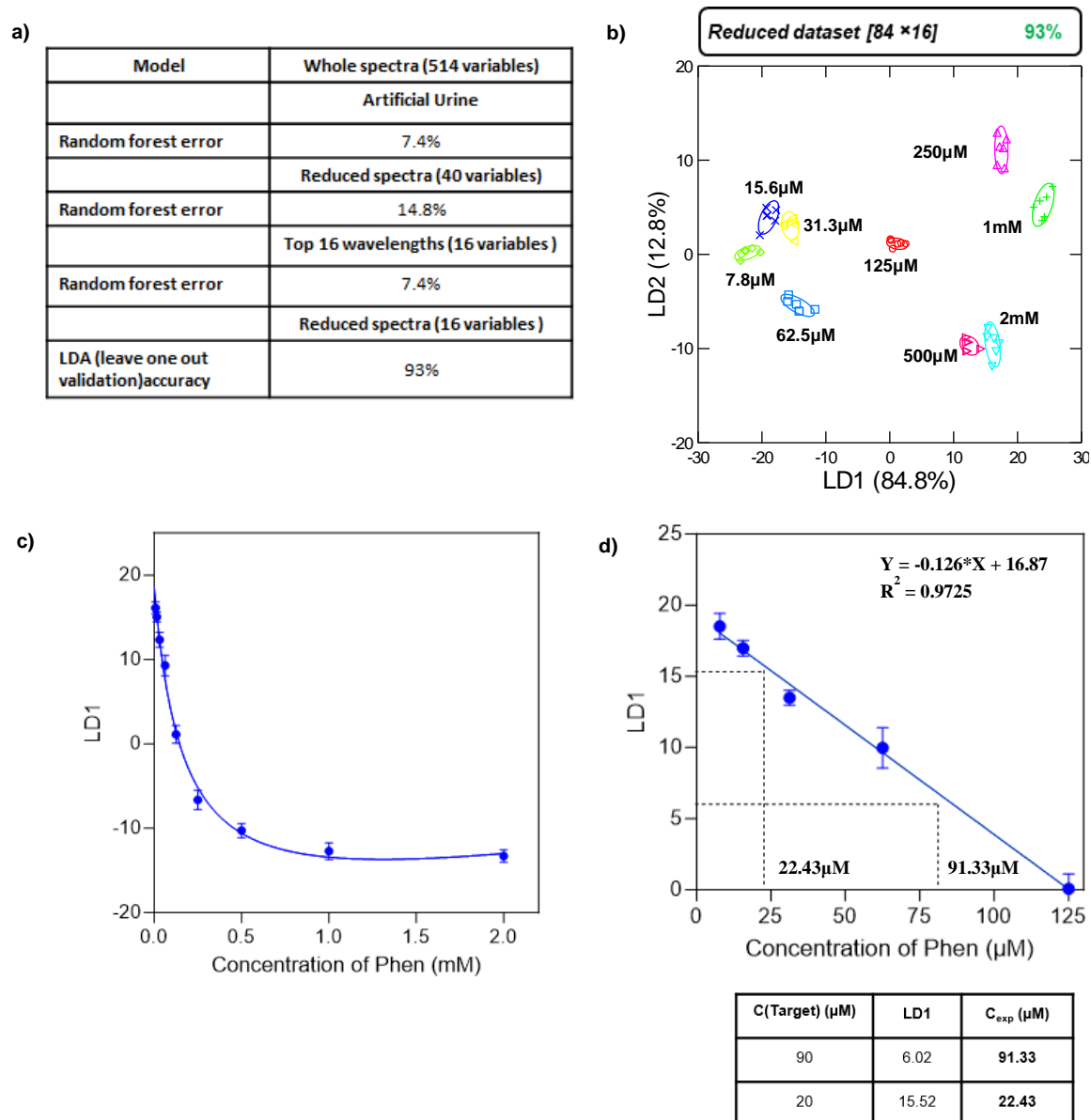


Figure 4.10: Clinical applications of the TPA-CB[7] sensor array: a) Discrimination accuracy determined by the streamlined data treatment and analysis methodology adopted in this study for the discrimination of phenylalanine spiked in artificial urine. b) Canonical score plot for the first and second factors of fluorescence response patterns calculated by LDA with 16λ channel. c) Plot of first discriminant factor versus concentration of Phe tested d) Linear range of variation of LD1 with Phe concentration (7.8 μM to 125 μM), extrapolation of LDA scores to determine concentration of unknown solutions with table indicating the identification of unknown concentrations from linear fit.

This reduced dataset [54×16] was also evaluated statistically by LDA. The leave-one-out cross-validation of the reduced dataset was evaluated and the plot of the first and second determinant factors, with a 95% confidence ellipse discriminated the different Phe concentration spiked in artificial urine with an accuracy of 93% (Figure 4.10a-b). We further evaluated the linear range across which the TPA-CB[7] showed efficient discrimination ability by analysing the linear variation of the first determinant LDA score with Phe concentration. It appears that a good linear fit with $r^2 = 0.97$ covers the range of concentration between $7.8 \mu\text{M}$ and $125 \mu\text{M}$ (Figure 4.10d). To evaluate the reliability of the sensor to detect Phe concentrations, we further utilized this linear fit to identify two unknown Phe solutions ($90 \mu\text{M}$ and $20 \mu\text{M}$). The extrapolation of their respective LDA scores from the linear fit accurately predicted the concentrations as shown in the table in Figure 4.10d, (lower panel).

The results presented here show that the proposed array-based sensing has a dynamic range for the detection of Phe, spanning from the disease threshold range to very low micromolar concentrations. Therefore, this approach could be explored for the direct detection of PKU in patient urine samples, offering a less invasive, faster, and simpler alternative to standard blood-based methods for monitoring the control of the disease by dietary restrictions.

4.4. Fingerprinting of disease model: Preeclampsia

4.4.1. Preeclampsia: State of the art diagnosis

Preeclampsia (PE) is a multifactorial pregnancy related disorder characterized by hypertension and proteinuria after 20 weeks of gestation^{234–236}. It affects 2 to 7% of pregnancies and is associated with maternal and perinatal mortality and morbidity, with both short- and long-term consequences for mother and child. PE can lead to serious and unpredictable adverse outcomes and is one of the major causes of extreme prematurity and maternal deaths globally. Due to the lack of specific biomarkers or adequate combination of biomarkers, predicting pregnancy outcomes after diagnosis is highly challenging. This limitation restricts the panel of therapeutic strategies available to clinicians and often results in unexpected emergency situations. No curative treatment exists, and although symptomatic management has improved, preeclampsia remains one of the top five causes of maternal deaths in developed countries.

The etiology of preeclampsia remains unclear, as no single causal factor links all theories due to its multifactorial and polymorph nature. It is classically associated with abnormal remodeling of the utero-placental vascularization but can also result from placental senescence, activation of coagulation in the intervillous space, infections, and fetal hydrops²³⁷. Preeclampsia involves significant quantitative and qualitative biological changes in maternal blood, with the placenta releasing high amounts of proteins^{238–240}, lipids²⁴¹ and nucleic acids. Additionally, it is also associated with protein misfolding and

oxidation of proteins and lipids. The presentation of the disease is highly variable. It is typically diagnosed by the presentation of hypertension and proteinuria after 20 weeks of gestation. Even with conservative management, the mother and fetus may still face severe complications such as eclampsia, HELLP (Hemolysis, Elevated Liver enzymes, Low Platelet count) syndrome, kidney failure, stroke for the mother; and prematurity, lung problems for the baby. The time from initial diagnosis to these complications and the severity of the disease can vary significantly between patients.

The follow-up of preeclamptic patients is complicated by the difficulty in distinguishing between women at low risk of perinatal complications (for which ambulatory surveillance is sufficient and avoiding unnecessary intervention) from those at high risk requiring more intensive management and close follow-up. Maternal factors used as severity criteria by some international clinical practice guidelines^{242,243} are not uniform and do not reliably identify women at high risk of developing maternal complications^{244,245}. To date, no biomarker profile satisfactorily predicts preeclampsia progression and outcomes²⁴⁶. Hence, classical methods targeting specific biomarkers, typically antibody-based assays, are not well suited to address the complexity of PE. Instead, holistic and untargeted approaches that rely on the global analysis of the serum, that contains thousands of different proteins, may provide richer data on the various physiological, dysfunctional or pathological states of the patient. These methods can detect subtle changes in the balance between several biomolecules (proteins, hormones, ions) in the serum, potentially offering more comprehensive insights into the condition.

We therefore tested the developed TPA-CB[7] sensor array as a non-specific serum-based diagnostic strategy for the identification and discrimination of a cohort of serum samples collected from preeclamptic and healthy patients. We hypothesized that the selective interactions of the sensor array, which allows the discrimination of proteins within complex matrices like serum, would also generate unique fluorescence fingerprints for each patient sample. These fingerprints could then be correlated with the multifactorial nature of the diseased state, reflecting the global proteomic variations present in the serum of PE patients as compared to healthy patient samples.

4.4.2. Fingerprint of PE samples by four TPA with CB[7]

We conducted a proof-of-concept study to demonstrate the application of the TPA-CB[7] sensor array as a diagnostic tool for non-specifically identifying the diseased preeclampsia (PE) state in human serum samples. The sensor was tested using serum from five PE patients and twelve non-PE patients, all being a subset of the APHERESE cohort (NCT03188900). Detailed clinical information for these selected samples can be found in the Materials and Methods section of this Chapter.

The samples were classified based on the levels of two specific biomarkers: soluble fms-like tyrosine kinase 1 (sFlt-1) and placental growth factor (PlGF). In preeclampsia, the circulating maternal serum levels of these biomarkers show characteristic changes, with sFlt-1 levels typically increased and PlGF

levels decreased. Currently, a high sFlt-1 to PlGF ratio is associated with an increased risk of preeclampsia and is considered a better predictor of risk than either biomarker alone²⁴⁷. Thus, based on the sFlt-1 to PlGF ratio provided in the clinical details, we selected 17 serum samples with relevant sFlt-1 to PlGF ratio for testing with the TPA-CB[7] sensor array (PE samples ratio : 200-300; non-PE samples ratio: 0-15). This selection ensured that the samples represented both the PE and non-PE states efficiently, enabling a comprehensive evaluation of the sensor array's diagnostic capability.

a. Preeclampsia discrimination in patient serum samples by sensor array

Statistical analysis: Complete fluorescence fingerprint dataset

We firstly recorded the complete emission spectra of all the TPA's (5 μ M, 40 μ L) in the array after the addition of the serum samples of the PE and healthy patients (2.5 μ L), in three replicates in a microplate ($I_1[\lambda_i]$). Subsequently, CB[7] (100 μ M, 10 μ L) was added to each well, and the fluorescence spectra were recorded again ($I_2[\lambda_i]$). This generated a complete fluorescence dataset with dimensions of $[51 \times 824]$ (3 replicates of 17 patient serum samples \times number of wavelengths for each sensor element), where each of the replicates of the 17 samples are considered as individual data points for classification by the two-class classification model. This dataset was firstly analysed by the random forest algorithm to obtain a discrimination error rate of 3.9%.

Statistical analysis: Reduced fluorescence fingerprint dataset

We further reduced the dataset by increasing the step between each consecutive wavelength, selecting only four wavelengths for each spectrum. This resulted in a reduced dataset with dimension of $[51 \times 32]$. However, it was observed that this reduction led to no change in the discrimination error rate, of 3.9% as analysed by the random forest algorithm. This indicated that the reduced dataset retained all the features that were essential to make an accurate discrimination between the different patient serum samples. To refine the dataset further, we identified the 16 most significant wavelengths contributing to the maximum discrimination of the patient samples using the 'importance' function to generate a dataset with dimensions of $[51 \times 16]$. Applying the random forest algorithm to this refined dataset resulted in an error rate of 5.9%, indicating that the reduction in the number of features has almost no effect on the discrimination accuracy. (Figure 4.11a). This reduced dataset was also analysed using Linear Discriminant Analysis. A plot of the first and second determinant factors displayed the discrimination of the 17 serum samples, achieving a classification accuracy of 96% (Figure 4.11b).

We further analysed the accuracy and sensitivity in the performance of our sensor system in prediction of PE through the Receiver Operator Characteristic (ROC) curve analysis (Figure 4.11c) and the area under the ROC curve (AUROC). While an AUCROC ≥ 0.75 is generally considered to be clinically relevant for a diagnostic test, our ROC yielded an AUC=1, signifying the efficient ability of the sensor array to distinguish healthy patients from those with preeclampsia.

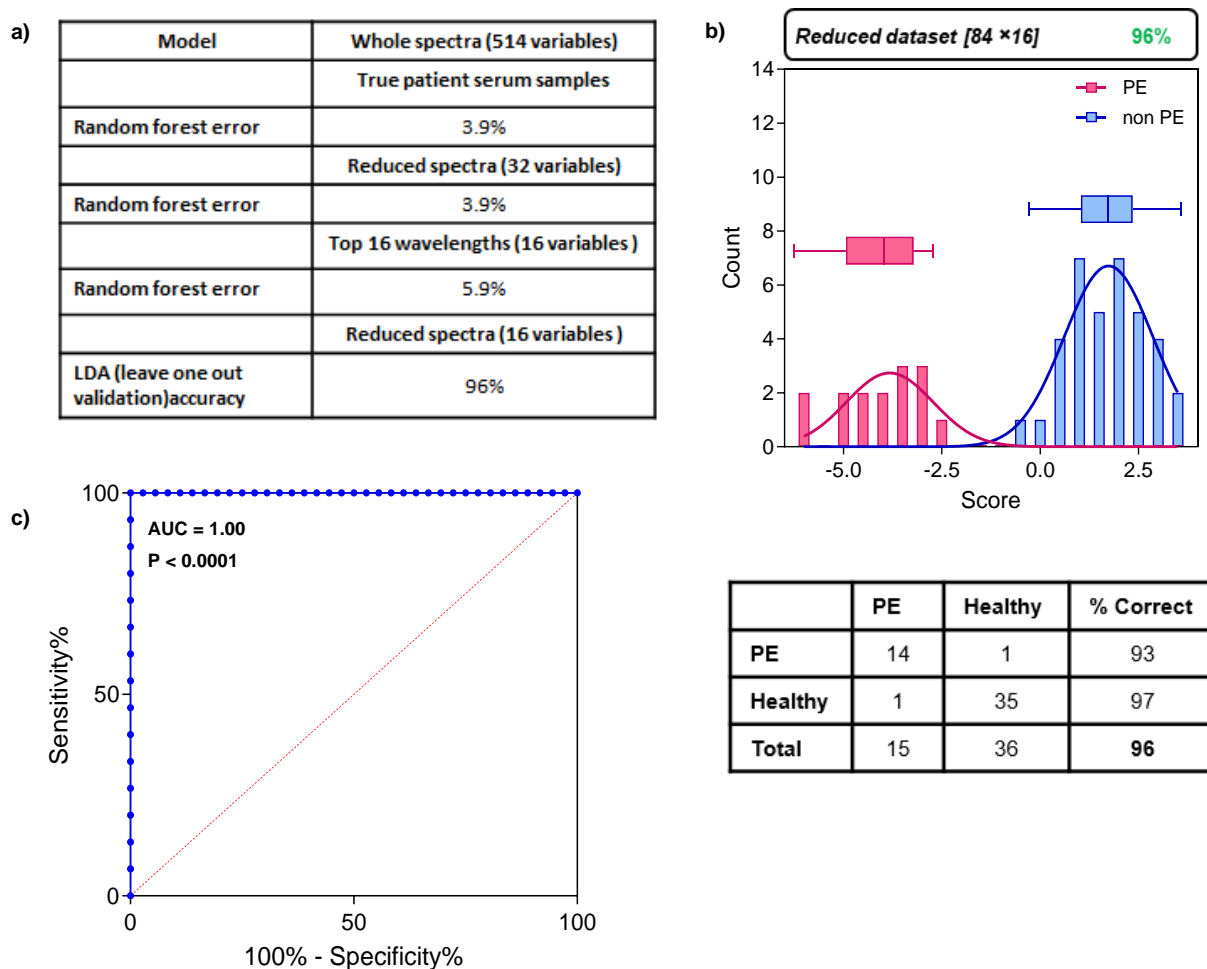


Figure 4.11: Clinical applications of the TPA-CB[7] sensor array: a) Discrimination accuracy determined by the streamlined data treatment and analysis methodology adopted in this study for the discrimination of PE disease in patient serum samples. b) Histogram marked with a normal distribution fitted to complete data analyzed by LDA with three replicates for PE and healthy serum samples and a box plot of combined samples in each group. c) ROC analysis of the two-group sensing by TPA-CB[7] sensor with corresponding AUC=1 with corresponding jackknifed classification matrix for the discrimination of 17 serum samples in triplicates depicts a classification accuracy of 96%.

b. Preeclampsia discrimination in mis-assigned patient serum samples by sensor array

To validate the robustness and eliminate possible bias in the treatment of the fluorescence fingerprint data by the developed random forest algorithm and LDA classification model, we deliberately partially mis-assigned serum samples, by labeling PE samples as non-PE and vice versa. This resulted in 60% mis-assignment of the samples, which was then used to analyse the variation in the results of the statistical analysis.

Statistical analysis:

We firstly analysed the complete fluorescence dataset with dimensions of $[51 \times 824]$ with random forest algorithm to obtain a discrimination error rate of 47.1%. The reduction in the dataset to $[51 \times 32]$ does not improve the discrimination error rate (error rate of 43.1%). To refine the dataset further, we identified the 16 most significant wavelengths using the ‘importance’ function to generate a dataset with dimensions of $[51 \times 16]$. Applying the random forest algorithm to this refined dataset resulted in an error rate of 33.3% (Figure 4.12a). While an improvement was observed, the error rate

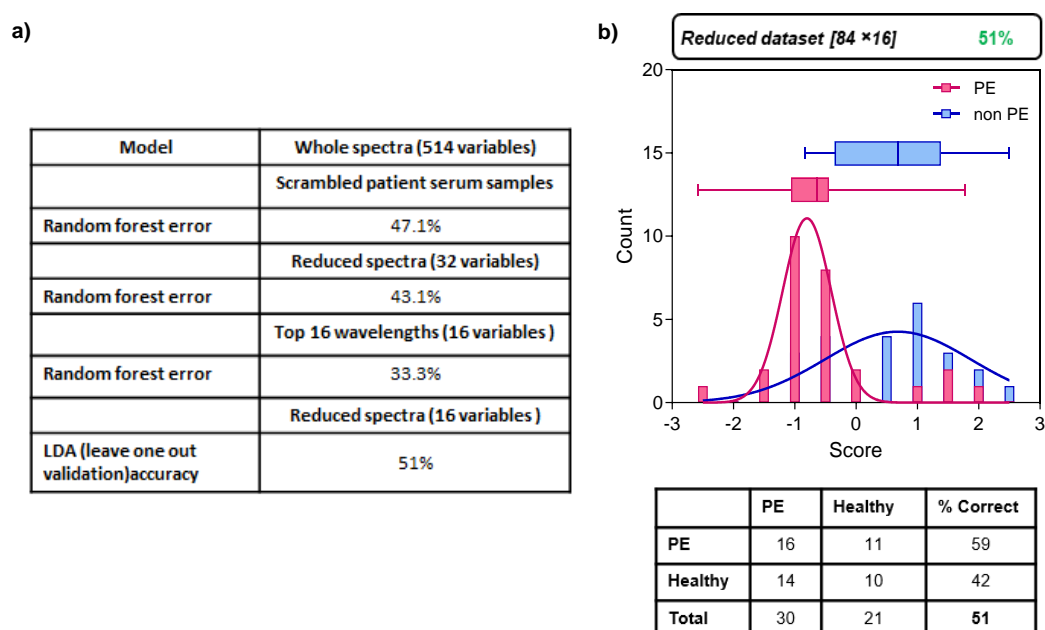


Figure 4.12: Clinical applications of the TPA-CB[7] sensor array: a) Discrimination accuracy determined by the streamlined data treatment and analysis methodology adopted in this study for the discrimination of PE disease in mis-assigned patient serum samples. b) Histogram marked with a normal distribution fitted to complete data analyzed by LDA with three replicates for mis-assigned PE and healthy serum samples and a box plot of combined samples in each group. Jackknifed classification matrix for the discrimination of 17 serum samples depicts a classification accuracy of 51%.

still remained too high to make an accurate discrimination among the different samples. The reduced dataset was also analysed using LDA, which produced a plot of the first and second determinant factors displaying the discrimination of the 17 serum samples with a classification accuracy of 51% (Figure 4.12b). This decrease in accuracy due to sample mis-assignment highlighted the robustness of the original model developed with our TPA-CB[7] sensor array, confirming its capability to effectively discriminate and classify serum samples from PE and healthy patients.

While this initial proof-of-concept analysis indicates promising potential for utilizing this sensor in diagnosing preeclampsia (PE), the current sample size used in the study is inadequate to thoroughly validate the predictive power of the sensor array. The small sample cohort studied limits the ability to generalize the findings and assess the robustness of the diagnostic performance. Thus, access to the entire cohort established for the study is crucial, but administrative limitations have made this difficult and time-consuming. Therefore, to more accurately establish the diagnostic capabilities of the sensor array, it is essential to conduct testing with a substantially larger and more diverse clinical sample cohort along with the use of high-throughput analytical platforms for efficient sensing. This expanded testing will provide a more definitive assessment of the potential of the TPA-CB[7] array for clinical application in diagnosing PE.

4.5. Conclusion

In conclusion, we have demonstrated the effectiveness of the developed TPA-CB[7] sensor array in optically fingerprinting bioanalytes. By leveraging the molecular recognition properties of the CB[7] macrocycle and TPA derivatives, along with the fluorescence properties of TPA, we have showcased the ability of the sensor array to discriminate a variety of proteins as model analytes spiked in both simple and complex matrices like human serum. Additionally, the sensor system exhibited a strong capability to discriminate pathological diseased states, such as phenylketonuria and preeclampsia, in both artificial urine and human serum. The treatment and statistical analysis of the generated raw data were streamlined using R coding, ensuring consistent and effective analysis. These overall results validate the potential utility of the array in fingerprinting actual clinical specimens and highlights its selectivity toward the global proteomic alterations associated with multifactorial diseases such as preeclampsia in the complex milieu of body fluids.

4.6. Materials and Methods

4.6.1. Physiochemical properties of protein used in the discrimination study

Proteins	MW (kDa)	pI	Commercial Source
Peroxidase from horseradish	44.0	7.2	Sigma Aldrich, 77332
Pepsin from porcine gastric mucosa	35.0	1.0	Sigma Aldrich, P7000
Diastase (from fungi)	-	-	Sigma Aldrich, 1036040050
α -amylase from hog pancreas	51-54	form I: 5.95 form II: 5.25	Sigma Aldrich, 10080
Lipase from <i>Aspergillus niger</i>	45.0	-	Sigma Aldrich, 62301
Lysozyme from chicken egg white	14.6	11.0	Hampton Research, HR7-108
Myoglobin from equine skeletal muscle	17.6	7.3; 6.8	Sigma Aldrich, M0630
Phosphatase, Acid from porcine	69.0	-	Sigma Aldrich, P3752
Pancreatin from porcine pancreas	-	-	Sigma Aldrich, P1750
Catalase, from bovine liver	250.0	5.4	Sigma Aldrich, C1345
α -Chymotrypsin from bovine pancreas	25.0	8.75	Sigma Aldrich, C4129
Bovine serum albumin	66.3	4.8	Sigma Aldrich, A4503
Fetuin, from fetal calf serum	48.4	3.3	Sigma Aldrich, F3004
Dispase II (neutral protease, grade II)	-	-	Roche, 11760200

4.6.2. Protocol for discrimination of proteins spiked in 1X PBS, serum and depleted serum

For the sensing experiments with the 14 proteins, 40 μ l of 5 μ M solution of the four TPA derivatives in water were taken in six replicates in a 96-well, half area microplate. To this solution, 2.5 μ l of 10mg/mL solution of the respective proteins spiked either in 1X PBS, serum or depleted serum were added to the wells. The plate was incubated for 10 minutes and the fluorescence spectras ($I_1[\lambda]$) were recorded after initial shaking of the plate for 10 seconds in the microplate reader at the excitation wavelengths indicated in Table S1. After this measurement, 10 μ l of 500 μ M solution of CB[7] in water was added to the wells and the plates were incubated for 10 minutes before the fluorescence spectras were measured again ($I_2[\lambda]$). Control spectras were recorded with the addition of 1X PBS, serum or depleted serum in the absence of spiked protein to the TPA derivatives, followed by the addition of CB[7]. Similar protocols were adopted for experiments with the disease models.

4.6.3. Depletion of human serum

The Thermo Fisher Scientific High Select HSA/immunoglobulin depletion resin (Cat. No. A36367) was utilized for the process of pre-treatment of human serum samples used for the discrimination of the 14 protein analytes. This depletion step involves the process of reducing the albumin and antibody components that are known to be abundantly present in human plasma samples, mainly IgG and albumin. The kit has been optimized for the removal of >95% removal of both these components by highly specific immobilized anti-HSA and anti-immunoglobulin antibodies (IgG, IgA, IgM, IgD, and IgE).

- The provided resin slurry stored at 4 $^{\circ}$ C filled into the spin column and equilibrated to room temperature
- Centrifuged the column at 500 rpm for 1 minute to remove the excess buffer of the slurry from the column
- Then, 120 μ l of the human serum sample was added directly to the resin in the column and gently end-over-end mixed several times with occasional vortexing at room temperature for 10 minutes to homogenous the mixture
- The spin column was then placed into a 15mL collection tube and centrifuged at 1000 \times g for 2 minutes
- The collected filtrate contained the depleted serum sample which was stored at -20 $^{\circ}$ C and spiked with proteins to be used for the sensing experiments

4.6.4. Preeclampsia: Clinical data of patient samples from APHERESE cohort

Table S3: Library of serum samples from the APHERESE cohort used for discrimination studies with TPA-CB[7] sensor array

N° inclusion	GROUP	DOB	Age at inclusion	Date of collection	PIGF	sFlt1	Ratio sFlt1&PIGF
001-0009-DH	Healthy	5/1/1983	29 SA + 3	6/20/2017	783	2679	3.421456
001-0038-RZ	Healthy	12/2/1990	31 SA + 3	8/8/2017	788	974	1
001-0055-PM	Healthy	5/11/1985	27 SA + 6	8/31/2017	7197	906	0
001-0151-LI	Healthy	6/8/1983	28 SA + 4	4/16/2018	671	1513	2
001-0152-SA	Healthy	4/21/1977	30 SA + 2	4/17/2018	560	1274	2
001-0159-OM	Healthy	1/4/1988	29 SA	4/26/2018	727	2115	3
001-0170-IN	Healthy	5/10/1984	30 SA + 1	5/15/2018	1781	2043	1
001-0197-BA	Healthy	11/7/1985	28 SA	6/26/2018	559	4400	8
001-0199-TM	Healthy	9/15/1984	29 SA + 4	7/2/2018	300	1045	3
001-0232-SA	Healthy	11/28/1985	30SA+5	11/14/2018	127	1887	15
001-0233-NC	Healthy	7/26/1983	30SA+6	11/14/2018	535	3856	7
001-0113-FN	Healthy	9/5/1983	30 SA + 6	1/16/2018	1067	1559	1
001-0096-TS	PE	9/18/1986	30 SA + 3	12/19/2017	41.6	16611	399
001-0134-SN	PE	6/6/1989	30 SA + 3	2/11/2018	26.1	9448	362
001-0226-CT	PE	11/20/1985	31SA+6	10/18/2018	40.1	7649	191
001-0227-PI	PE	8/8/1978	31SA	10/24/2018	33.7	10228	304
001-0241-KV	PE	10/30/1977	27SA+6	12/19/2018	38.8	10182	262

4.6.5. Ethical Declaration and Biological Samples

The local ethics committee (Comité de Protection des Personnes Ile de France 3) approved the studies of the human serum sample from the APHERESE case/control prospective study (<http://www.clinicaltrials.gov> unique identifier: NCT03188900). All patients gave written consent and non-opposition form for participation in these studies and to the use of their serums for research purposes. After centrifugation of the blood ($1,000 \times g$ for 15 min at 4°C), supernatant was transferred into a clean polypropylene tube, aliquoted and stored at -80°C . The serum samples were used after one year of conservation in accordance with French law. Serum total proteins and total HCG were measured using the Cobas analyzer (Roche Diagnostic, Meylan, France).



Chapter 5

***Optimization of analytical platform for throughput analysis: use of
droplet microfluidic platform and pipetting robot***

5.1. Introduction

The application of sensor arrays in diagnostic settings requires rigorous evaluation to ensure their effectiveness in distinguishing among large cohorts of clinical samples. Such evaluations demand substantial volumes of both the sensing elements and analytes to create sufficient replicate data, which are crucial for thorough analysis and for validating the diagnostic capabilities of the sensor array. To address these requirements, it is important to develop a sophisticated analytical system that is not only automated and miniaturized but also capable of high-throughput analysis. This system must handle the complexities associated with the analysis of large sample sets and the substantial dimensions inherent to sensor arrays. To fulfil this need, we have engineered a droplet-based microfluidic device tailored specifically for the TPA-CB[n] sensor array, focusing on its optimization for efficient performance. Additionally, we have also investigated the integration of a pipetting robot to refine the experimental design, aiming to minimize variability and ensure uniform sampling conditions. This approach enhances reproducibility in the fingerprint generation process by the sensor array, ensuring consistent and reliable results. By combining such analytically advanced technologies with the TPA-CB[n] sensing system, we aim to significantly improve the precision and scalability of the sensor array for applications such as diagnostic testing.

5.2. Results and Discussion: Droplet-based microfluidic platform

Microfluidics over the years has enabled high-throughput and inexpensive analysis through parallel liquid handling and reaction miniaturization. An emerging subcategory of microfluidics is droplet-microfluidics. Unlike continuous flow systems, droplet-based systems focus on creating discrete volumes with the use of immiscible phases. It provides an alternative approach for generation of large scale and parallel chemical or biological reactions without increasing the device size or complexity²⁴⁸. This approach involves the compartmentalization of aqueous solutions as droplets in oil in a controlled manner and at high frequency (Hz to kHz).

The droplet-based microfluidics approach offers several significant benefits. Reactions occur in very small sample volumes which can be reduced from milliliters and microliters to nanoliters and femtoliters, while reducing reaction times to a few seconds or lesser. This method minimizes analyte dilution within droplets, providing a stable microenvironment and delivering highly reliable information for long-term monitoring. It facilitates high-throughput generation and detection of numerous droplet reactors, enabling large-scale screening of samples and reaction conditions, which is particularly useful in applications like directed evolution and drug discovery. The flexibility in droplet manipulation allows for automation of multistep reactions, including generation, merging, splitting, and sorting. The possibility to aliquot samples and reagents into independent compartments prevents undesired cross-

contaminations, evaporation, and non-specific adsorption onto channel surfaces. Overall, this approach is cost-effective, as a single device can generate millions of droplets, significantly reducing research costs²⁴⁹.

To evaluate and optimize the TPA-CB[7] chemical nose sensor array for high-throughput screening of analyte samples, a droplet microfluidic platform was engineered and implemented together with the team of Dr. Stephanie Descroix (*Institut Curie*) in order to develop a miniaturized version of the sensor array with the aim to make hundreds of measurements simultaneously with rapid analysis and detection.

5.2.1. Experimental set-up of droplet microfluidic platform

The platform is established based on the ‘plug format’ previously developed by the group, with minor modifications and adjustments²⁵⁰. In this set-up generated droplets are kept separated by oil in a confined capillary. It is composed of different parts and is fully controlled by a computer station to allow automated droplet experiments. The different parts and setup of the developed platform are shown in Figure 5.1.

The platform includes a pipetting arm-robot (Rotaxys, Cetoni, Germany, Figure 5.1a-b, No.1) coupled with one high-precision syringe pump (Nemesys, Cetoni, Germany) mounting a 250 μL syringe (SGE Analytical Science, UK) (Figure 5.1a-b, No.3). The motorized pipettor arm is connected to the syringe by a circular polytetrafluoroethylene (PTFE) tubing (0.3 mm inner diameter x 0.6 mm outer diameter, 113 cm long) (BohlanderTM PTFE tubing, Sigma Aldrich, USA). This material is an ideal choice for various microfluidic applications where precision, chemical compatibility and reliability are essential, due to its chemical resistance, non-stick properties, thermal stability, flexibility, durability, and semi-transparency. The droplets are generated by pipetting alternatively the water and oil phase solutions from a 384-wells microplate (Nonbinding surface (NBS)TM, Corning, USA). The NBS of the plate prevents undesirable non-specific interactions between the plate surface and the sensing elements or analytes being examined, thereby preserving their integrity and concentration. This is especially crucial when working with low volumes of valuable samples. The carrier oil is fluorinated (FC-40, 3MTMFluorinertTM, Belgium) with 2% w/w of a fluorinated surfactant (1H, 1H, 2H, 2H-perfluoro-1-decanol (PFD), Fluorochem, UK). The microplate is placed on a home-made holder placed below the motorized pipetting arm (Figure 5.1a-b, No.1). The holder could be moved in X, Y, Z directions by a manual stage (Thorlabs Inc., USA). Once the pipetting PTFE tubing is aligned with the reference well, the QmixElements software (QmixElements V5.0, Cetoni, Germany) controlling the motion of the pipettor’s arm and the syringe could be operated. Thus, the PTFE tubing is moved from one well to another, pipetting in each one the desired volume of sample and inserting between each water droplet an oil plug. The QmixElements software also allows the possibility to program different protocols of confined droplets with determined order and desired number of iterations.

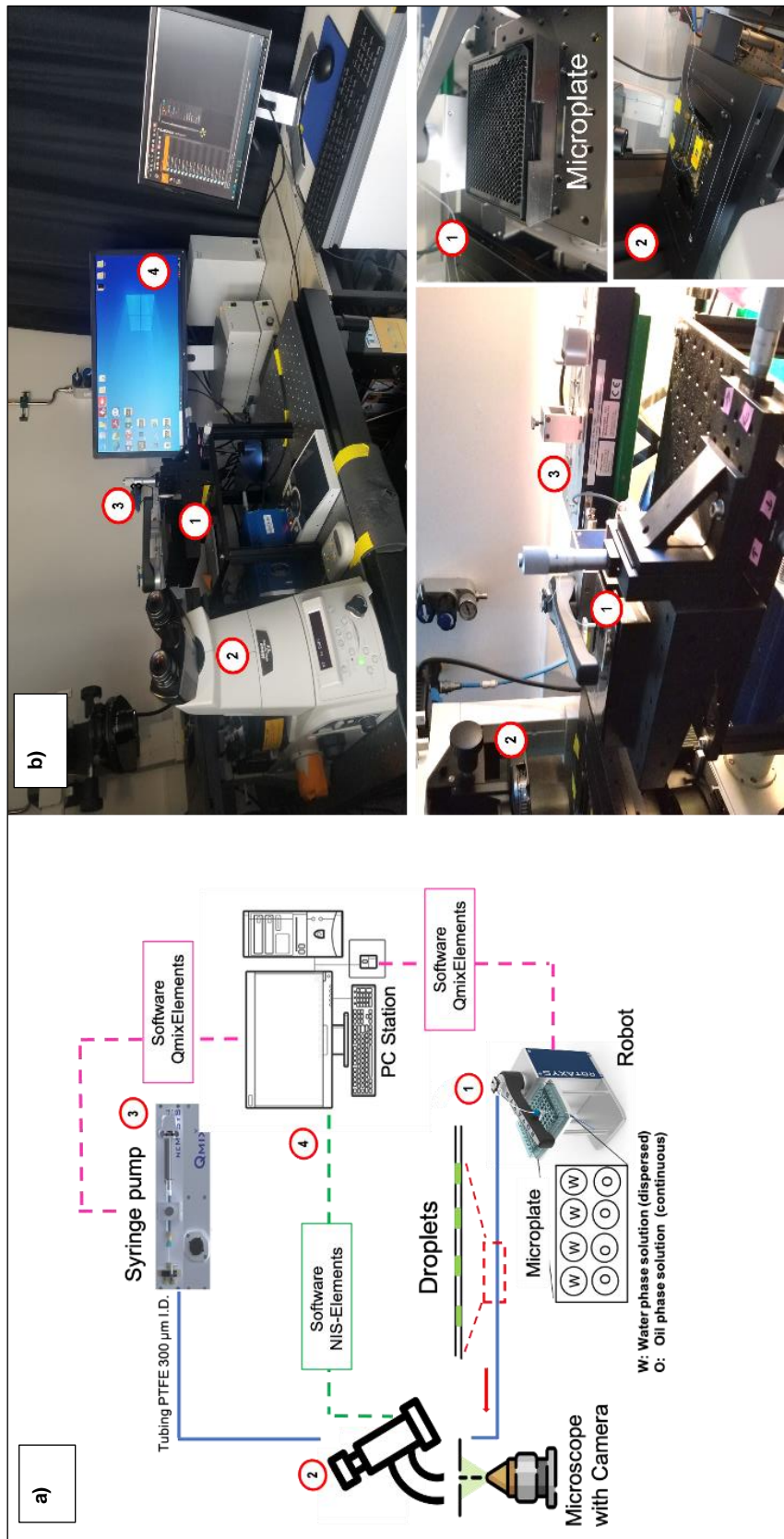


Figure 5.1: a) Schematic representation of experimental setup of the platform and b) pictures of the developed platform for droplet microfluidic based chemical sensing. The droplets are generated by pipetting alternatively the water phase solution and oil from the microplate (1). The droplets are moved in front of the microscope equipped with a camera to record the fluorescence intensities (2). A syringe pump (3) allowed the moving of the droplets in the PTFE tubing prefilled with oil phase. The droplet platform was fully controlled by a PC station (4). Fluorescence analysis and pipetting sequences can be programmed by using the NIS-Elements and QmixElements software, respectively.

An optical epifluorescence microscope (Model Nikon Eclipse *ti* with 4X objective, Nikon, Japan) was used for fluorescence measurements. It is equipped with an optical fiber coupled to a mercury lamp (Nikon Intensilight C-HGFIE, Nikon, Japan) and a CoolSNAP™ HQ2 CCD camera (Teledyne Photometrics, Tucson, AZ, USA). The image processing is performed by the NIS Elements AR software (NIS Elements V5.11.02, Nikon, Japan) (Figure 5.1a-b, No.4). The acquisition frame rate is varied between 1fps and 10fps according with the applied flow rate, in order to enable visualization of the flowing droplets. In addition, the microscope is equipped with optical filter cubes (Bandpass filter cubes, Nikon, Japan) that are chosen to match the required spectral excitation and emission characteristics.

5.2.2. Optimization of the microfluidic platform for the TPA-CB[n] working model

The platform was designed with the goal of integrating multiple functionalities, including droplet generation, merging, and online optical detection, all specifically optimized for the unique requirements of our TP-CB[n] sensor array. The following section provides a detailed description of each of these functionalities.

Droplet generation:

The advantage of droplet microfluidic systems lies in their ability to produce uniform droplets and particles, making precise control over droplet size, shape, and monodispersity critically important. Aqueous droplets in a capillary channel of oil are generated by the use of specific channel geometries or by programmed pipetting. The use of a fluorinated carrier oil phase (FC-40) in combination with a fluorinated surfactant (PFD) at an optimized concentration assures complete wetting of the fluorinated tubing preventing any cross-contamination during the pipetting. The fluorinated oil also reduces the risk of molecule diffusion from the aqueous phase droplet into the continuous oil phase because the aqueous droplet is surrounded by a thin lubrication film of the continuous phase. This combination further helps in obtaining stable droplets and avoiding non-specific adsorption of the sensing elements or analytes at the water-oil surface.

The aqueous droplet containing the TPA-CB[7] sensor elements or the analyte sample are generated as confined droplets with a volume of 100 to 200 nl at a flow rate of 0.2 $\mu\text{L/s}$ in fluorinated oil containing 2% PFD, by the pipettor arm from solutions in a 384-well microplate. These confined droplets provide a lot of flexibility in the development of a pipetting protocol, since the order, volume and content of the droplets as well as the oil gap separating two consecutive droplets could be customised by adjusting the pipetting sequence. The black 384-well microplate with V-bottom shape was chosen to store the solutions during the experiments instead of the traditional 96-flat wells plate commonly used with this

kind of motorized pipettors to perform the assay with reduced volume of sample. The 384-wells microplate allows working volumes between 10-40 μL rather than the standard 75-200 μL for the 96-wells microplate, resulting in a 5 to 7.5-fold volume reduction of the sample required to start an experiment. By programming the inter-well distance and the Z-well dimension as a step of the droplet generation sequence, the 384-wells plate was configured to be compatible with the pipettor robot. Considering the volume capacity of a single well and the volume of the droplets, it is possible to generate from a single solution of 40 μL of the sensing element or analyte, about 400 droplets of 100 nL or 200 droplets of 200 nL which would be ideal for the set-up of the sensing experiments on this platform.

Droplet merging:

Containment within the generated droplets provides an effective method to keep droplets containing the sensing elements and analytes separate, facilitating their controlled merging to allow effective interaction between the TPA-CB[7] sensors and the different analytes, thereby preventing the generation of unreliable data. Controlled coalescence of droplets is thus crucial for the automated applicability of the microfluidic platform in evaluating the TPA-CB[7] sensor array.

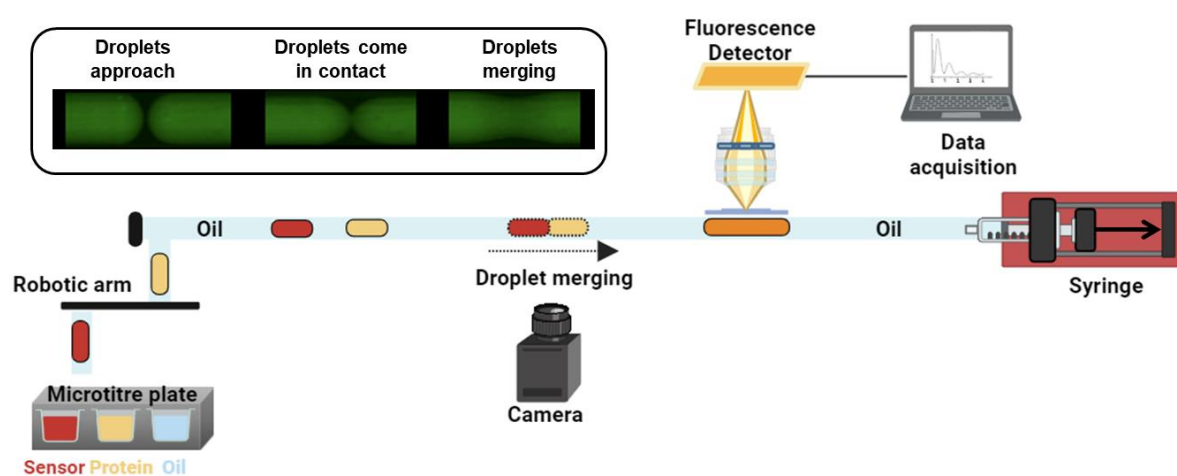


Figure 5.2: Schematic illustration of droplet based microfluidic platform. The analyte and sensor droplets are generated by alternatively pipetting the water phase solution and oil from a microplate and a passive merging/coalescence strategy is adopted for the interaction of sensor elements of the array with analytes. The inset shows the merging of the droplets as monitored by recording their fluorescence intensity.

Droplet merging can be achieved through either active or passive processes. After generating individual confined droplets of the sensing elements and analytes in a deterministic order (discussed in Section 5.2.3), we adopted a passive merging approach within the capillary, utilizing the difference in the

velocity of droplets with varying surface interfacial tensions depending on their content (Figure 5.2). This method is a simpler alternative to active merging strategies with no needs of external equipments or channels with specific geometric features. As previously reported, when a confined droplet is transported by a carrier fluid, its speed depends on the interfacial tension (γ) between the two liquids with lower interfacial tension (γ) resulting in higher speeds²⁵⁰. This effect was exploited to induce droplet contact and thus their merging after a travelled distance as shown in the sequence of images in Figure 5.2.

Optical detection: The droplet microfluidic platform is also incorporated with an optical microscopic set-up for the detection of the fluorescence of the sensor element inside the different droplets. After the generation and merging of the droplets, they are transported along the PTFE tubing of the capillary towards the microscope for fluorescence measurements. These measurements are performed through a detection window or region of interest (ROI) ($50 \times 50 \mu\text{m}^2$) defined at the center of the tubing with the aim to record the highest fluorescence intensity emitted by the droplet (Figure 5.3). In this way, it is possible to continuously record the emitted fluorescence signal in the ROI which is then collected by the camera and processed by the Imaging NIS Elements AR software (NIS Elements V5).

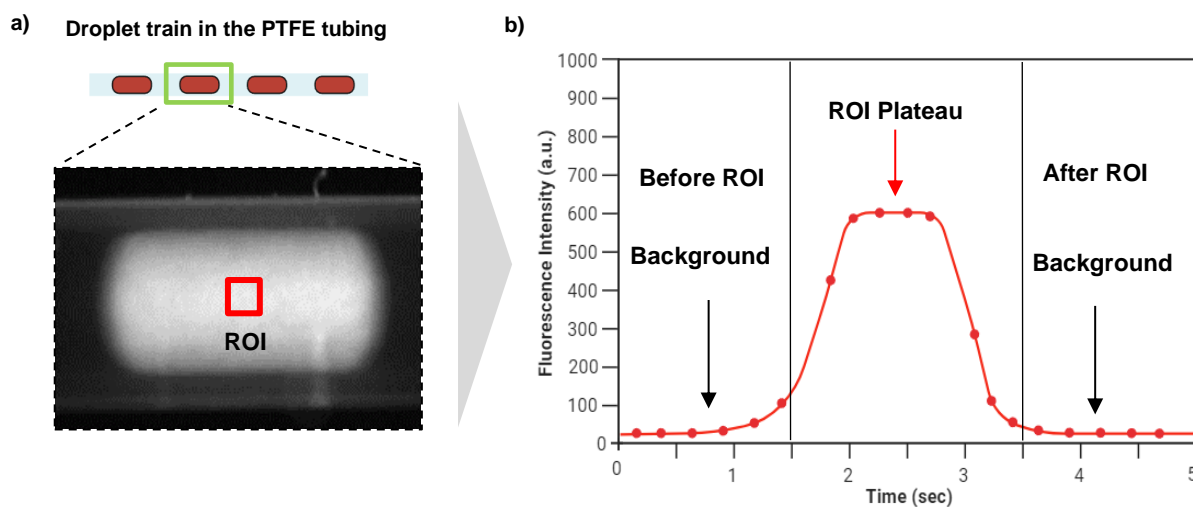


Figure 5.3: a) Representative image of generated fluorescent droplet. The red square represents the region of interest (ROI) where fluorescence is measured. b) Plot of the fluorescence intensity as a function of the time obtained upon image processing. When the droplet crosses the ROI a peak is generated. The plateau of the peak represents the highest fluorescence intensity emitted by the droplet at several times with a given wavelength range.

11.02, Nikon, Japan). The flow rate of the droplets passing in front of the camera was evaluated at 0.02, 0.05 and 0.1 $\mu\text{L/s}$ to ensure the effective acquisition of images. Afterwards, the fluorescence intensity was recorded at an optimized rate of 0.02 $\mu\text{L/s}$ as a function of the acquisition time at the ROI. Initially, a constant background value is measured; however when the droplet crosses the ROI a peak is observed

as shown in Figure 5.3b. The maximum of the peak, which is usually observed as a plateau, was considered as the fluorescence intensity of the droplet and is detected within a specific range of wavelengths determined by the cutoff of the emission bandpass filter cube. The acquired data were then exported in excel format for further analysis. Since large data files were generated after the image processi and the analysis was time-consuming, a dedicated MATLAB program was developed to facilitate this task (Materials & Methods).

5.2.3. Optimization of protein-sensing element interactions

With the successful initial optimization of various parameters of the droplet microfluidic platform, we conducted preliminary droplet tests using the four TPA derivatives and their complexes with CB[7] to evaluate the variation in their fluorescence response upon interaction with a model protein, BSA.

- **Design of droplet merging experiment with protein analyte (BSA):**

We designed an experimental plan to replicate the discrimination experiments previously performed on the microplate reader onto the droplet microfluidic platform. We programmed individual droplets of the sensing elements and analytes, each with a volume of 100 nL, to merge and form larger droplets of 200 nL, resulting in a half dilution of the initial concentrations. Thus, to achieve a final concentration of 5 μM of TPA, 100 μM of CB[7] and 600 mg/mL of BSA in the final merged droplets, we prepared a primary stock concentration of 10 μM , 200 μM and 1200 mg/mL of TPA, CB[7] and BSA respectively, to be added to dedicated wells of the 384-well microplate.

To facilitate effective merging and to monitor bimodal recognition events, we pre-prepared the TPA-CB[7] complex in the microplate, thus avoiding the need for merging of three droplets (TPA, CB[7], and BSA) to generate the second recognition event (TPA + CB[7] + protein). Using the QmixElements software, we generated sequences of consecutive droplet merging of TPA and TPA + CB[7] with BSA droplets spiked in 1X PBS. The fluorescence of the droplets was measured by using the FITC filter cube with an excitation wavelength of 480 nm and an adjoining band range of 30 nm. The experiments were repeated several times to ensure reproducibility in the merging approach. We also evaluated the effect of the oil plug volume (8-50 nL) and the flow rate (0.2 to 0.6 $\mu\text{L/s}$) of the continuous phase on the merging efficiency. During the initial tests, we achieved good reproducibility of droplet merging at a flow rate of 0.5 $\mu\text{L/s}$ with a 10 nL oil plug. These parameters were adopted for all future experiments. Additionally, we tested the effect of the sequence of merging of the sensing elements and analytes on the variation in fluorescence intensity. We found that the sequence (TPA \rightarrow BSA or BSA \rightarrow TPA) did not impact the measured output (Figure 5.4). However, we observed that the possible high interfacial surface tension of the BSA droplet reduced its velocity, and using a sequence where the BSA droplet trailed the sensing element droplet did not achieve 100% consistency in merging for long trains of

droplets with multiple merges corresponding to the multiple replicates. Thus, we determined that adding the analyte droplet in front of the sensing element enabled more efficient merging.

We also observed that while the fluorescence of AcriPy, TP₃Py, and PhenazPy was easily detectable at a final concentration of 5 μM and in the complexed form at a final concentration of 100 μM CB[7], emission signal saturation was observed for SulfoxPy at the same concentration. Consequently, all further experiments with SulfoxPy were conducted at a final concentration of 2 μM for the derivatives and 40 μM for CB[7] in the SulfoxPy + CB[7] complex.

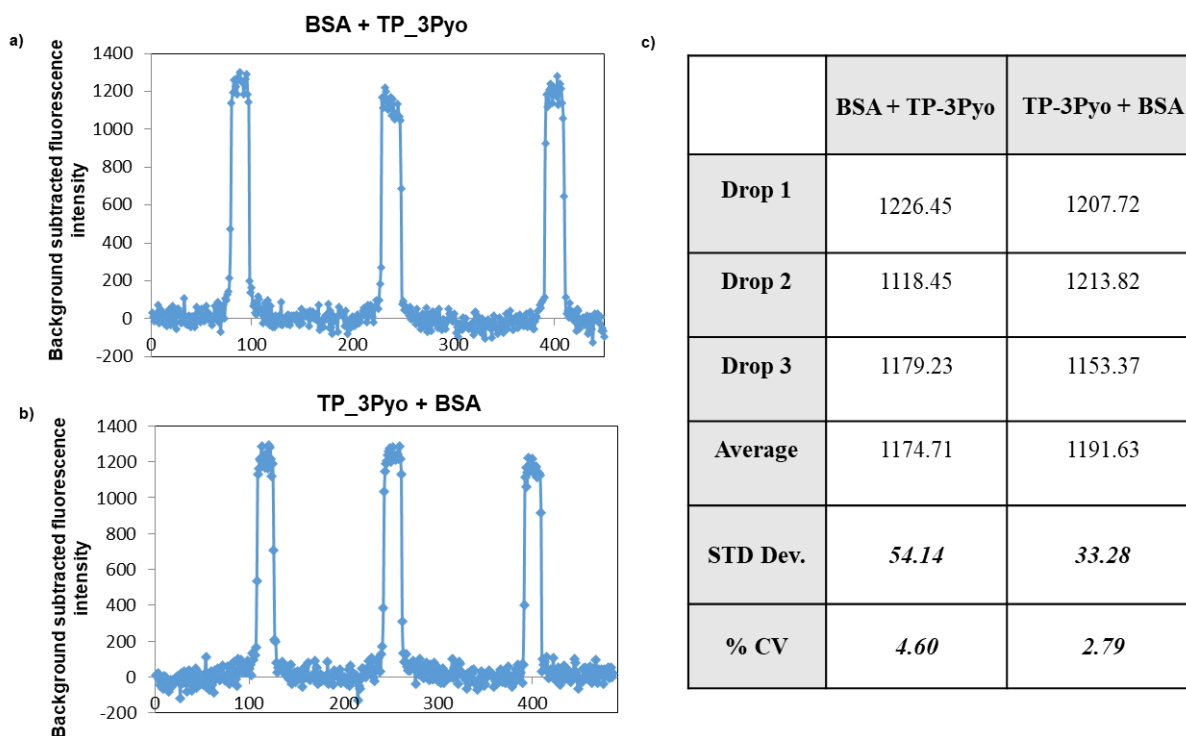


Figure 5.4: Fluorescence response of TP-3Pyo monitored in the presence of BSA in a train of three consecutive merged droplets with variation in sequence a) BSA trailed by TP-3Pyo droplet, b) TP-3Pyo trailed by BSA droplet. c) Statistical analysis indicating the standard deviation and associated percentage of coefficient of variation calculated with the three droplets considered as individual replicates.

- **Droplet merging with PhenazPy + CB[7] in the presence of BSA**

We tested several controls to assess the fluorescence variation of TPA and CB[7] in the presence of BSA, as well as to monitor the effect of merging on the dilution of the sensing element droplets and the corresponding impact on their fluorescence emission. While these test experiments were performed for all the TPA derivatives of the sensor array, representative results with PhenazPy and CB[7] are discussed below:

Merging of PhenazPy and CB[7] with BSA:

The droplets (100 nL) of PhenazPy and complex of PhenazPy + CB[7] were merged with the droplets of BSA and the fluorescence emission detected with the microfluidic platform were compared with the results previously obtained on the 96-well microplate. As seen in Figure 5.5, similar variations in fluorescence intensities were observed indicating no change occurs in the sensing process at lower volumes of the TPA derivatives with CB[7] and protein analytes.

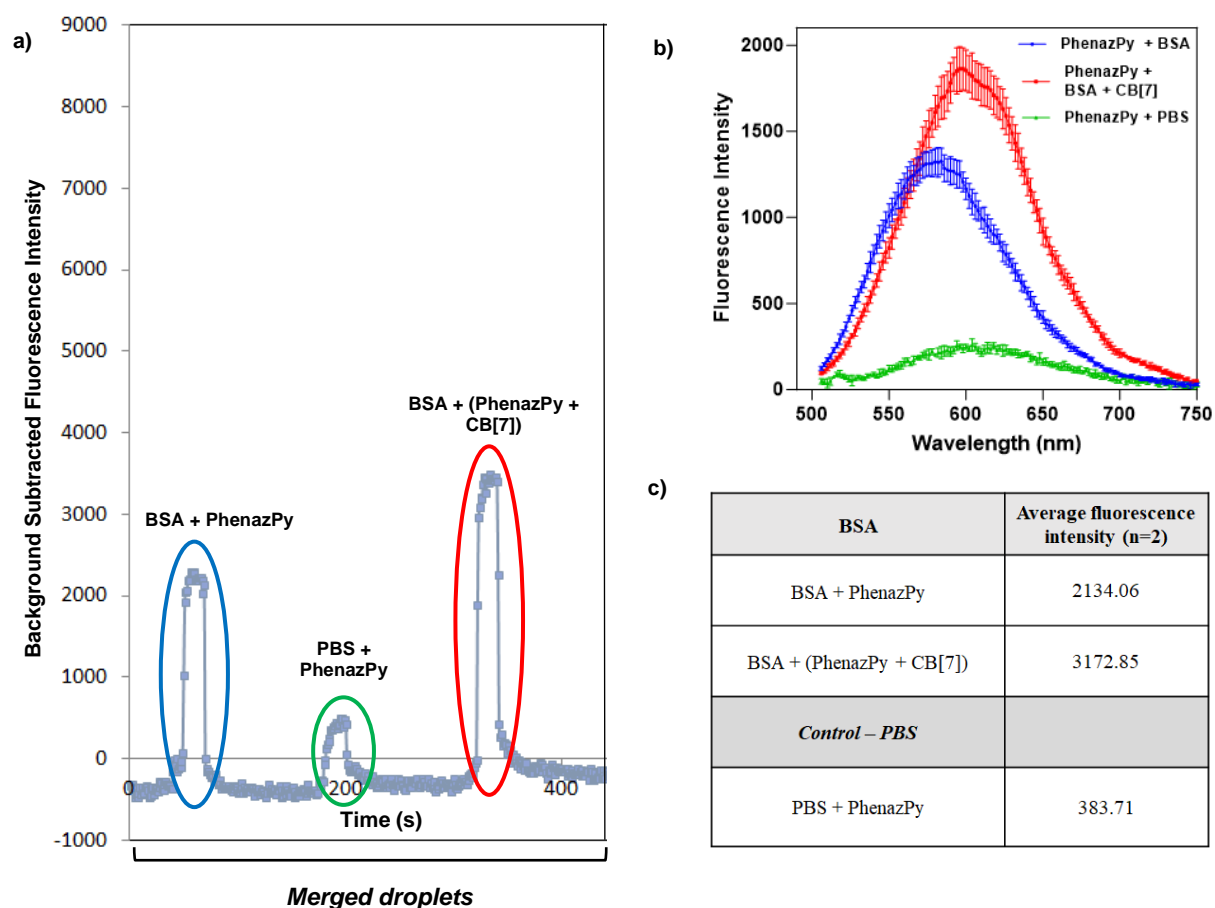


Figure 5.5: a) Fluorescence variation of droplets of PhenazPy and PhenazPy + CB[7] upon merging with analyte BSA. b) Fluorescence variation of PhenazPy and PhenazPy + CB[7] recorded on a 96-well microplate compared to that obtained on the microfluidic platform as indicated by the blue, green and red circles. c) Background subtracted fluorescence intensities of the PhenazPy and PhenazPy + CB[7] droplets merged with BSA as measured on the microfluidic platform.

Variation of fluorescence emission with change in concentration of TPA:

The test experiments with the proteins were programmed such that individual droplets of the sensing elements and analytes at doubled concentration, each with a volume of 100 nL, merge and form larger droplets of 200 nL, resulting in a half dilution of the initial concentrations. We thereby monitored controls to establish the variation in the fluorescence of the PhenazPy at these concentrations, by using test droplets of PhenazPy at 5 μM and 10 μM at a volume of 200 nL. This was also monitored by using test droplets of TPA + CB[7] at a concentration of (10 + 200) μM and (5 + 100) μM . As expected we observed a decrease in fluorescence signals with corresponding decrease in concentration. Furthermore, these results with the merged droplets of TPA + CB[7] emphasize the minimal interaction of the TPA derivatives with the microplate as well as the capillary tubing, preserving sample integrity and concentration (Figure 5.6a, corresponding peaks represented by double sided arrow).

Comparison of fluorescence of merged and pre-mixed TPA-CB[7] complex droplets:

To avoid the complexities involved in triplet droplet merging to monitor the second recognition event of the TPA + CB[7] interaction with proteins, we pre-mixed the TPA-CB[7] solution to form the host-guest complex in the microplate at a doubled concentration of (10 + 200 μM) prior to the droplet generation step. We then merged the pipetted droplet of this solution (100 nL) with the analyte BSA droplet (100 nL) to dilute the TPA+ CB[7] and obtain the final required concentration.

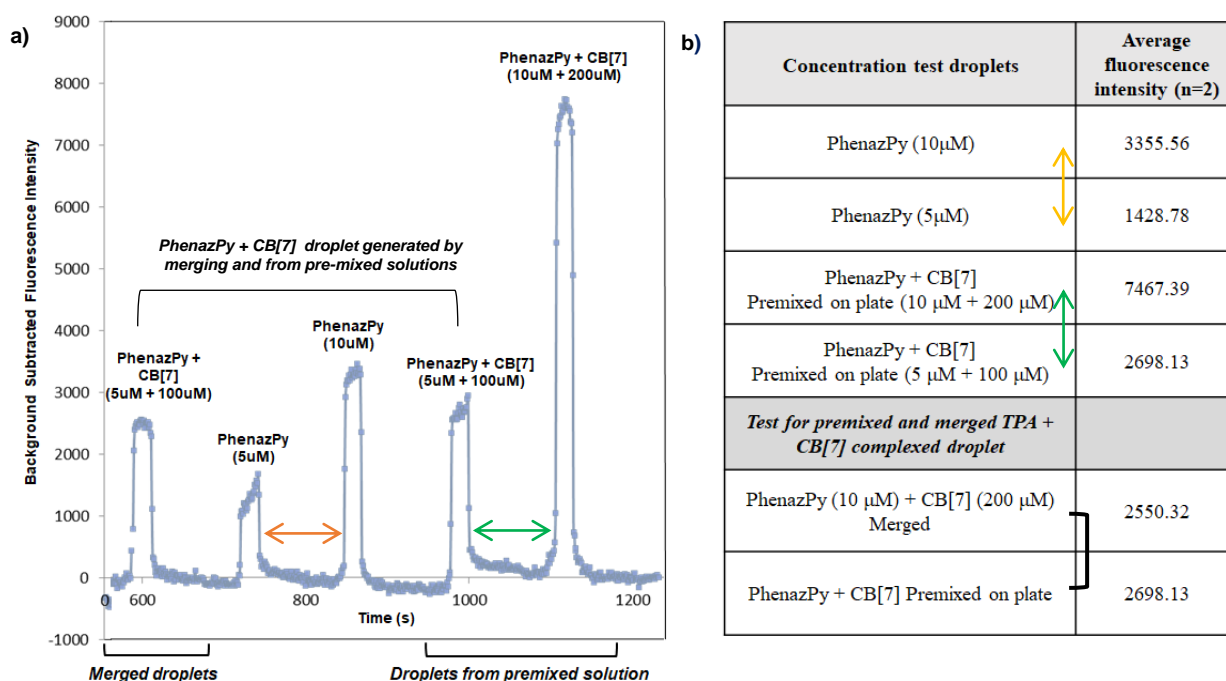


Figure 5.6: a) Variation of fluorescence intensities with concentration and pre-mixing of the PhenazPy and PhenazPy + CB[7] complex as determined by generated control droplets. b) Background subtracted fluorescence intensities of the PhenazPy and PhenazPy + CB[7] control droplets.

To ensure that this pre-mixing did not affect the detected fluorescence signal, we monitored the fluorescence variation between the TPA + CB[7] droplets formed by merging individual TPA and CB[7] droplets and a single droplet directly pipetted from the pre-mixed host-guest complex solution in the microplate. We observed that the recorded fluorescence intensity was consistent in both cases, with no significant variations. As a result, all subsequent experiments for measuring the second recognition event were conducted using the TPA+CB[7] pre-mixed on the microplate (Figure 5.6a, corresponding peaks compared by a box bracket).

5.2.4. Discrimination of proteins on droplet microfluidic platform

Discrimination of proteins in PBS:

To demonstrate the potential of this developed droplet microfluidic approach, we used the TPA-CB[7] sensor array in the droplet format to discriminate 6 proteins spiked in 1X PBS- catalase, peroxidase, fetuin, myoglobin, BSA and pepsin. The droplets are generated in a specific sequence to establish consistent merging of the individual droplets. By pipetting from the 384-microwell plate, the droplet of protein was generated first followed by the droplet of the sensing element of the array (TPA or the premix complex of TPA + CB[7]). This sequence was repeated multiple times creating a train of droplets corresponding to 20 replicates for the interaction of each sensing element with each tested protein analyte (Figure 5.7a).

The fluorescence emission signal was then analysed by the MATLAB software to obtain the background subtracted fluorescence fingerprint for each analyte. This fingerprint was analysed by LDA using the Systat software. The canonical score plots of the first and second discriminant factors that were produced by the LDA analysis using leave-one-out cross validation, displayed the clear discrimination of the six proteins with 95% classification accuracy (Figure 5.7b). Thus, the microfluidic platform enabled the consistent merging of the long train of droplets effectively with no complications even in the presence of multiple proteins.

Discrimination of proteins in complete human serum:

To further test the capability of the microfluidic system in discriminating more complex matrices, we proceeded to test the ability of the system to differentiate the 6 proteins spiked in complete human serum with no modifications. The droplets were generated in the same manner as with the proteins spiked in PBS, with the protein droplet preceding the droplet of the sensing element, which is either the TPA or the pre-mixed TPA+CB[7] complex. This sequence was repeated to create

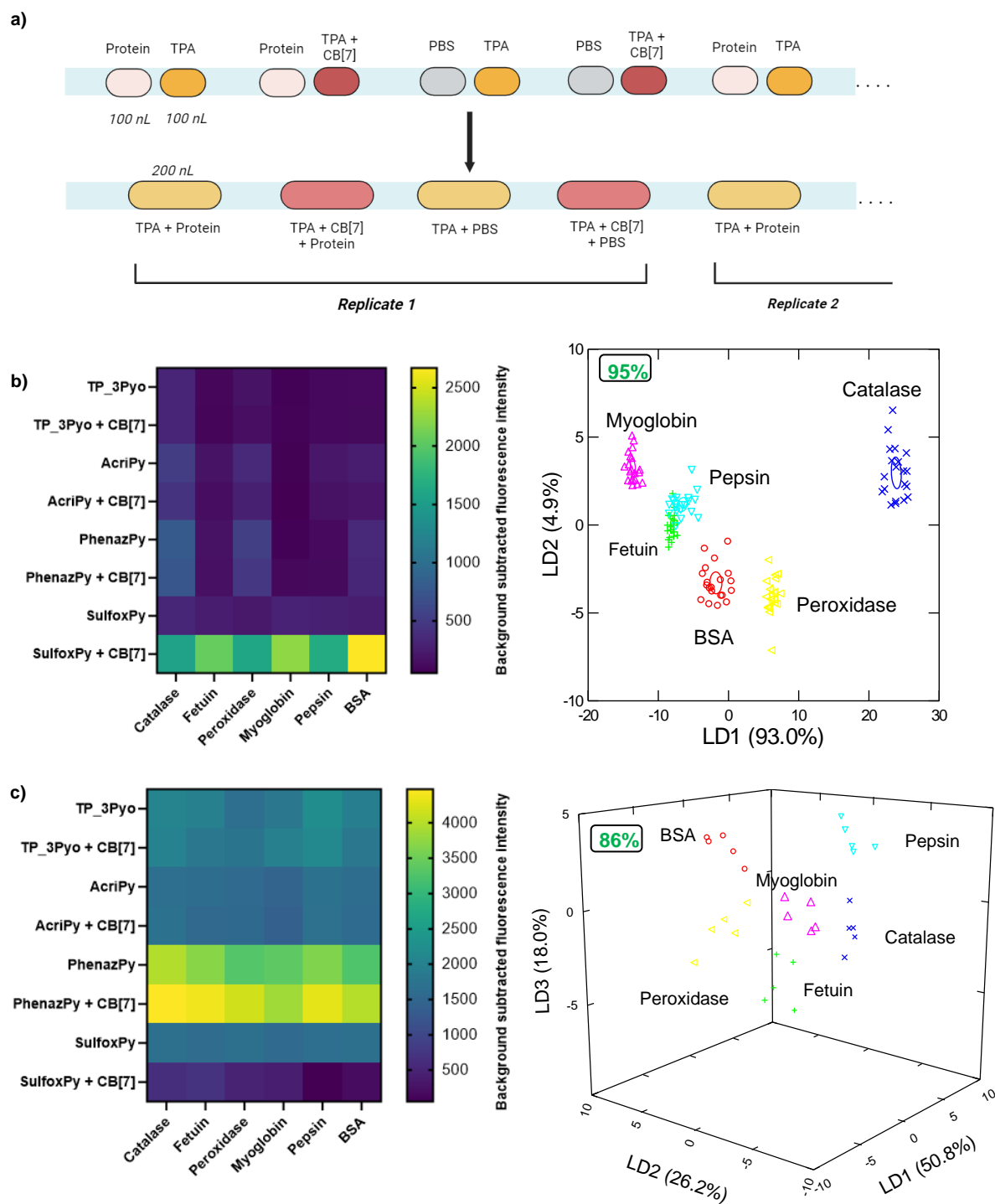


Figure 5.7: a) Pipetting sequence of droplet generation and merging to develop a train of multiple replicates, with protein analyte droplets preceding the sensing element droplets to facilitate merging. b) Canonical score plot for the first and second factors of fluorescence response patterns calculated by LDA for the identification of 6 proteins in 1X PBS based on fluorescence fingerprint generated on droplet microfluidic platform. c) Three-dimensional canonical score plot of fluorescence response patterns calculated by LDA for the identification of 6 proteins spiked in human serum generated on the platform.

a train of droplets corresponding to 5 replicates for the interaction of each sensing element and the tested protein. The fluorescence emission signals were analysed by the MATLAB software, and the obtained fluorescence fingerprints were then subjected to LDA analysis using the Systat software. The generated canonical score plot of the first and second significant factors displayed the discrimination of the 6 proteins with 87% classification accuracy (Figure 5.7c).

While the platform facilitated the generation of multiple replicates, a decrease in discrimination accuracy of the proteins was observed. This could be attributed to the manner of acquisition of fluorescence emission signals using this platform. Specifically, the fluorescence maximum observed as the ROI plateau is detected within a specific range of wavelengths as determined by the filter cutoff of the selected bandpass filter cube. Nevertheless, we have been able to successfully demonstrate the utility of the droplet based microfluidic platform to assess analytes with reduced volumes (~5-7.5 fold reduction from 96-well plate) and increased number of technical replicates (20-30 replicates). The establishment of this platform is a first crucial step for advancing the application of such chemical nose sensors for application in diagnostics accompanied by advanced statistical analysis.

5.2.5. Development of microfluidic chip for improvement of droplet merging strategy

Testing the droplet microfluidic platform with various protein analytes revealed that the passive merging strategy exhibited different merging behaviors based on the nature of the proteins and the sequence of droplet generation. This variation is due to the inherent interfacial tension of each protein in droplet form, which influences the droplet's velocity. This issue resulted in making the process challenging to reproduce consistently over time and space, which is a key requirement for any automatable and high-throughput analytical platform.

We therefore proposed to develop a microfluidic chip to be integrated into the droplet platform, designed to facilitate consistent active mechanical droplet merging within the channel of the chip.

Design strategy and fabrication of microfluidic chip:

The proposed microfluidic chip was primarily designed to facilitate the entry and control the confinement of droplets, thereby forcing their merging within a channel located in the central region of the chip Figure 5.8a. The chip was intended to be incorporated along the capillary tube, before the path leading to the camera used for fluorescence detection. Upon entering the chip, the droplet velocity was expected to decrease, promoting systematic and automated merging—a critical step for transitioning to a microfluidic-based automated analytical system.

The chip features a central channel formed by the strategic arrangement of solid pillars that create barriers to guide the entering droplets along the central channel. The pillars are aligned along the outer

edge of the chip at a suitable distance to prevent droplets from entering and getting retained along the edges and ensuring that they are directed exclusively into the central channel (Figure 5.8a). The chips were fabricated by 3D-printing using a biocompatible resin material DS 3000, which allows for the creation of multiple designs by varying the number, shape, and size of these pillar. This resulted in different patterns on the chip, with variation in the central channel thickness and shape as shown in Figure 5.8b. For each design, channel thicknesses of 120 μm and 240 μm were produced. Additionally, the outlet of the chip was designed with different structures, ranging from angular to bulb-shaped, to prevent the pinching and splitting of the merged droplets as they exit the central channel and re-enter the capillary. The outlet channels of the chip were designed to fit the dimensions of the capillary tubing, which were inserted into the openings and secured with adhesive glue around the edges of the tubing. The assembly was then cured at 80°C for 15 minutes before finalizing the microfluidic setup.

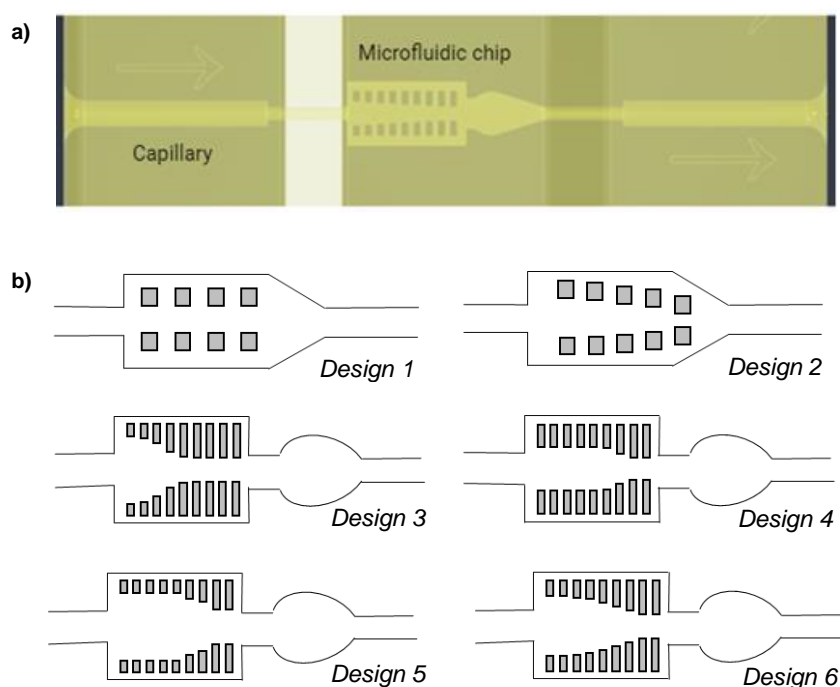


Figure 5.8: a) Schematic illustration of the microfluidic chip integrated with the droplet microfluidic capillary, featuring a central channel formed by different pillar arrangements on either side of the channel. b) Different designs of fabricated microfluidic chips with variation in the shape and thickness of the central channel to facilitate droplet constriction and merging.

Droplet merging with fabricated microfluidic chip:

The robustness, efficiency, and reproducibility of droplet merging across different chip designs were first assessed using food colouring solutions to enhance the visibility of the merging process within the chip. A series of 100 nL droplets were sequentially generated by the pipetting robot through the QmixElements software. Various parameters were optimized with these test droplets, including the flow

rate of the droplets and the volume of the oil plug between merging droplets, as well as between consecutive merged droplets. It was found that *design 5*, which featured a central channel with a thickness of 120 μm , achieved the highest merging efficiency. This was observed at a flow rate of 0.3 $\mu\text{L}/\text{sec}$, with an oil plug volume of 1.8 μL between merging droplets and 10 μL between consecutive merged droplets. These settings were effective in preventing the occurrence of badly controlled droplet merging within the chip (Figure 5.9).



Figure 5.9: Designed microfluidic chip (*design 5*) that facilitates merging by constriction of two smaller dye solution droplets (100 nL) into a single larger droplet (200 nL), which then moves into the capillary for detection by the camera.

However, it was noted that in some instances, droplets became trapped around the pillars, necessitating high-pressure flushing of oil to remove them. To address this issue, the microfluidic chip channels were silanized with a 5% solution of (tridecafluoro-1,1,2,2-tetrahydrooctyl)trichlorosilane to enhance their hydrophobicity. The silane solution, prepared in mineral oil, was flushed through the channel and was incubated for 10 minutes before being expelled with air. The channels were then dried at 80°C for 15 minutes. This silanization treatment significantly reduced droplet retention around the pillars, and all subsequent chips were treated using this protocol.

After optimizing these parameters using the test droplets we proceeded to evaluate the chips with droplets containing TPA derivatives and PBS to assess their performance under conditions more relevant for the intended application of the sensor array. While *design 5* (Figure 5.8b) of the microfluidic chip successfully facilitated droplet merging, even with the TPA derivatives, it failed to generate consistent and reproducible merging for long trains of droplets. We encountered several recurring issues throughout the testing process such as:

- **Droplet splitting:** As droplets entered the channel of the microfluidic chip, they often split, likely due to pressure build-up caused by constrictions in the channel. These constrictions were probably the result of blockages from residual glue or resin material.
- **Air bubble formation:** During droplet entry into the channel, air bubbles frequently formed. This was attributed to an improper seal of the capillary tubing at the channel inlet, which compromised the integrity of the setup.

- *Retention of silane material:* Residual silane material along the channel walls occasionally led to blockages or caused droplets to split, further disrupting the merging process.
- *Oil leakage:* A significant challenge was the leakage of oil from the chip's plug during cleaning and reuse. This issue was mainly due to inadequate sealing and ineffective curing of the glue used to secure the capillary tubing.

Improvement and further optimization of microfluidic chip:

These identified issues of the droplet system require further investigation and optimization to minimize sample and reagent consumption while maintaining droplet stability and detection robustness on this platform. While the microfluidic platform provides a promising way forward for the testing of large-scale cohorts using the TPA-CB[7] sensor array, further fine-tuning and refinement of the chip design and preparation protocols is necessary to make the process more robust and efficient to facilitate automation without the need for constant monitoring of the droplet merging.

A possible approach to addressing these challenges could involve altering the fabrication material of the microfluidic chip. By transitioning to polymer-based materials, the need for glue to attach the capillary tubing could be eliminated. Instead, the tubing could be inserted flexibly into the polymer, providing a more secure and adaptable connection. This change would allow for easier re-silanization and enable the chip to be reused multiple times without complications.

Furthermore, adopting polymer-based materials could help resolve several recurring problems. It could reduce the incidence of droplet splitting at the channel entry by minimizing the chances of blockages and pressure build-up. Additionally, achieving a more secure seal would decrease the formation of air bubbles, thereby improving the chip's overall reliability. This material change could greatly enhance the performance and durability of the microfluidic chip, thus making it more efficient and consistent for sensing applications.

Thus, given the challenges encountered with the droplet microfluidic system, we subsequently explored additional approaches to improve the throughput and efficiency of the TPA-CB[7] sensor array. This led us to investigate the use of a pipetting robot platform, which will be discussed in detail in the following section.

5.3. Pipetting robot platform

To enhance the automation and high-throughput capabilities of the TPA-CB[7] sensor array, we explored the integration of a pipetting robot into the developed experimental workflow. We employed the Gilson Pipetmax robot in collaboration with the team of Dr. Florence Mahuteau-Betzer (*Institut Curie*) to investigate this approach, which offers precise and reliable liquid sample handling. This advanced automation tool allows for the use of a 384-well plate format, significantly increasing the

throughput of the sensing protocol by allowing the rapid and simultaneous processing of multiple analytes on a single microplate with lesser volumes. By leveraging this technology, we aimed to streamline the experimental workflow, reducing the volume of sensing elements and analytes needed, and ultimately validating the sensor array with greater efficiency and accuracy.

5.3.1. Protein Discrimination using the TPA-CB[7] sensor array with pipetting robot platform

To validate the use of a pipetting robot for efficient processing and high-throughput sensing by improving the accuracy, reproducibility, and consistency of a large number of processed samples we firstly proceeded to the discrimination of proteins on this automated platform.

Experimental design and plate layout for pipetting robot protocol:

We carefully designed a plate layout allowing the development of an automated program for the Gilson Pipetmax pipetting robot. The plate layout was carefully designed to align with the robotic requirements, ensuring efficient and error-free plate preparation (Appendix 6). The robot operates by transferring solutions of the analytes or sensing elements from a 96-well plate into alternating rows of the final 384-well plate. This design, therefore, allows for the testing of only 12 analytes at a time, corresponding to the 12 columns of the 96-well plate, within a single program. To maximize efficiency, we divided the 384-well plate into two halves, each containing 12 columns. This setup facilitated the simultaneous testing of 11 analytes along with a PBS control, using two TPA-CB[7] sensing elements, enabling more streamlined and organized testing. The TPA derivatives were used at a final concentration of 5 μM with a volume of 20 μL per well, CB[7] at 100 μM with a volume of 10 μL per well, and proteins at a final concentration of 0.625 mg/mL with a volume of 2.5 μL , maintaining the parameters from previous experiments. A single 384-well plate was processed by the platform in approximately 35-40 minutes. Following this, the fluorescence emission spectra of the sensing elements were measured to generate unique fingerprints for the selected protein analytes. The complete emission spectra were recorded for all TPA-CB[7] sensing elements, covering the ranges of 495 to 739 nm (62 λ), 520 to 740 nm (56 λ), 510 to 740 nm (59 λ) and 467 to 739 nm (69 λ) with a step of 4 nm for TP_3Pyo, Acri_3Py, PhenazPy and SulfoxPy respectively. Consequently, the fluorescence fingerprint dataset was then processed using the streamlined data preparation, treatment, and statistical analysis strategy discussed in Chapter 3 of this thesis.

Statistical analysis: Complete and reduced fluorescence fingerprint dataset

The streamlined treatment generates a dataset with dimensions of [88 \times 492] (8 replicates of 11 proteins \times number of wavelengths for each sensor element) for subsequent analysis. This dataset was

evaluated using a random forest algorithm, which yielded a discrimination error rate of 0% among the 11 proteins.

We systematically reduced the number of features of the dataset to [88 × 50] by selecting fluorescence intensities at intervals of every ten wavelengths recorded for each TPA-CB[7] pair in the sensor array. This reduction did not affect the discrimination accuracy of the 11 proteins as determined by the random forest algorithm. This dataset was also evaluated statistically using LDA, where the leave-one-out cross validation plot of the first and second determinant factors, with a 95% confidence ellipse discriminated the proteins spiked in PBS with an accuracy of 100%. The discrimination was performed in two experimental replicates to further validate the efficiency of this approach, achieving a discrimination accuracy of 100% in both trials individually (Figure 5.10a). The canonical score plots for the first trial is shown in Figure 5.10b. The LDA plot displays highly compact clustering of the protein replicates within each cluster, emphasizing the minimal spread and close proximity of the identical data points. This indicates a high efficiency of reproducibility by eliminating anomalies due to pipetting errors and maintaining consistency across all replicates.

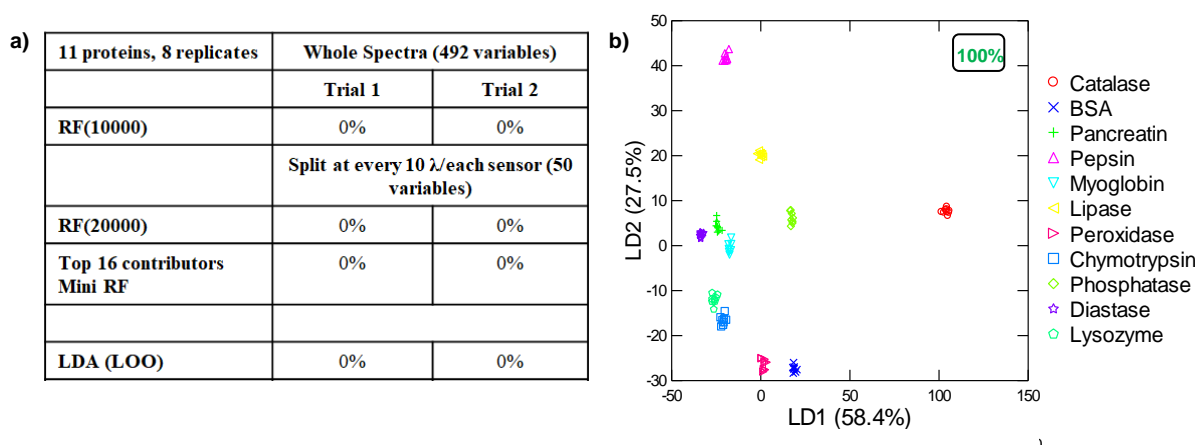


Figure 5.10: a) Table of discrimination rates generated by random forest and LDA algorithm for the two experimental replicates. b) Canonical score plot for the first and second factors of fluorescence response patterns calculated by LDA for the identification of 11 proteins in 1X PBS with 50 λ channel with the dataset generated from trial 1.

Cross-validation of developed LDA model:

To ensure the robustness and accuracy of the developed LDA model, we performed cross-validation using datasets from the two independent trials. In the first step, we used the initial dataset to train the LDA model, focusing on correctly identifying the 11 different protein analytes based on their unique fluorescence emission spectra. Once the model was trained, we proceeded to the validation phase, where the second dataset, which had not been used in the training process, was introduced as a test set. This step was critical in assessing the ability of the model to generalize and accurately predict

the identity of protein analytes in new, unseen data. By comparing the predictions made by the LDA model against the actual labels in the test dataset, we could evaluate the performance of the developed model. The results of this cross-validation process were promising; with the LDA model achieving a prediction error of just 5.11%. This low error rate indicates that the model is highly effective in distinguishing between the 11 protein analytes, demonstrating both the reliability of the LDA approach and the quality of the dataset used for training and testing.

Therefore, the pipetting robot proved to be effective and time-saving approach as compared to the manual pipetting protocol. Each sensing element (TPA, TPA + CB[7]) with the 11 proteins and PBS control could be prepared on a microplate in 35-40 minutes, while manual pipetting requires significantly more time. This efficiency allows for the testing of a large cohort of samples in a single day, unlike the extended duration needed for manual pipetting. Thus, this approach could pave the way for high-throughput sensing using the TPA-CB[7] sensor array, significantly enhancing the efficiency and scalability of such systems for applications in diagnostic testing.

5.4. Conclusion

In summary, the development of the droplet microfluidic platform and the implementation of the pipetting robot have collectively advanced the efficiency and scalability of sensing using the TPA-CB[7] sensor array. The microfluidic platform enabled precise control over droplet generation and merging, facilitating the utility of reduced volumes for the systematic interaction of sensing elements with analytes. Despite challenges in achieving reproducible droplet merging, the integration of a microfluidic chip and optimization of experimental parameters has laid a strong foundation for future improvements.

The pipetting robot on the other hand demonstrated remarkable accuracy, reproducibility, and time-efficiency in the preparation of 384-well plates containing the sensing elements and analytes. Transitioning from a 96-well to a 384-well plate system, the robot enabled high-throughput analysis with minimal pipetting errors and high consistency, allowing larger number of analytes to be processed in significantly less time as compared to manual pipetting. Collectively, these approaches highlight the promise of automated, miniaturized systems in enhancing high-throughput analysis of array-based sensors, especially offering promising pathways for future applications in clinical and research settings.

5.5. Materials and Methods

5.5.1. Analysis of fluorescence fingerprint generated by the droplet microfluidic platform

MATLAB software based ‘Fluo Droplets Analyzer – v0.5.1’ developed by Giacomo Gropplero, (*Institut Curie*) was utilized to analyse the fluorescence emission signals generated upon droplet merging between the sensing elements and protein analytes. This software parses output excel sheets generated by the image processing NIS Elements AR software to provide the background subtracted fluorescence intensity of each droplet plateau which collectively forms the fluorescence fingerprint for each tested analyte.

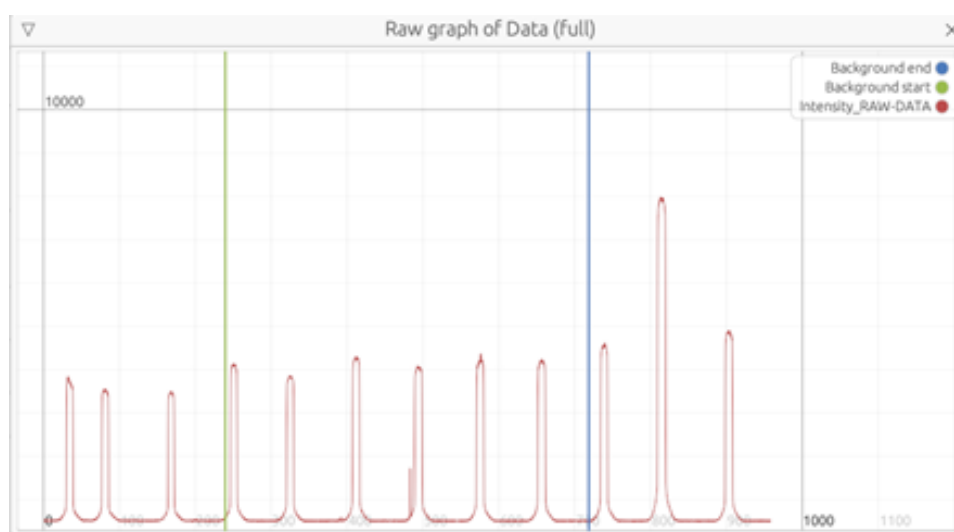
The user-interface of the software is shown below:



1. The excel output from the NIKON NIS elements software is uploaded
2. The excel sheet is pre-processed for the analysis and is presented as a graph of the raw data on the user screen.

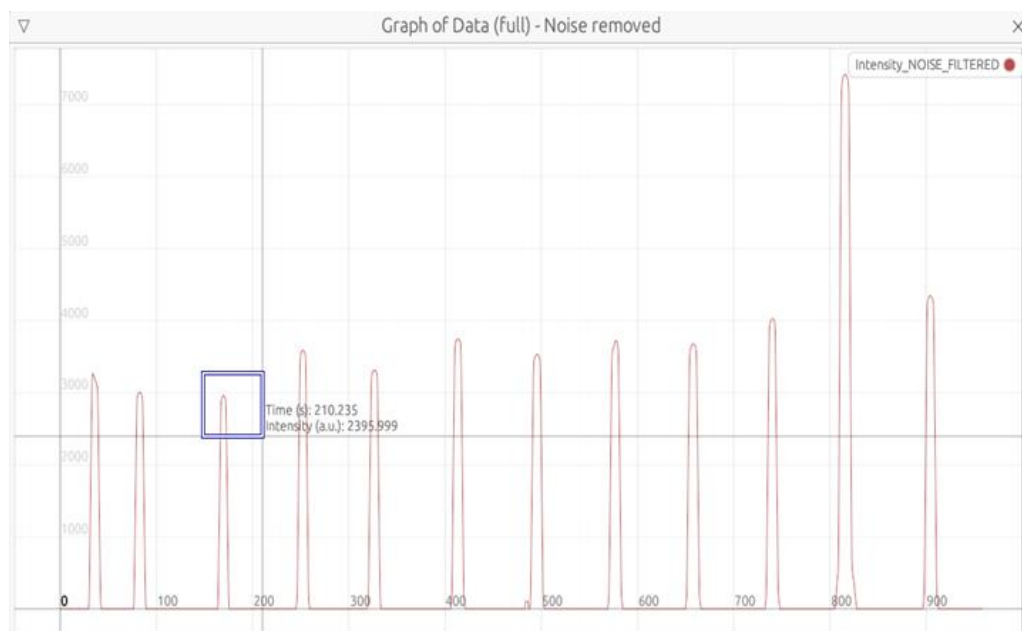


3. The dropdown indicates the choice for the x axis in the format of [m:s] for time, while intensity is selected as the y axis. The parameters are selected and a raw graph is generated as shown below



4. The background filtering range is selected by using the 'background end' and 'background start' lines shown in blue and green respectively.
5. The 'background noise threshold' is varied to adjust the extent of baseline removal that is required so as to detect all the necessary peaks.

- The signal to noise ratio can be adjusted at the 'noise filtering' tab to obtain relevant values at the plateau, a general range from 10-25 is maintained to prevent making the peaks too sharp for determination of the peak intensity.



- The 'peak detection' tab generates the different peaks and their corresponding intensity values averaged across the plateau. In some cases the peak intensity is below the threshold, this can be adjusted by the 'minimum prominence' tab to identify all the necessary peaks.
- The output graph and corresponding peak intensities can then be saved as a 'csv' file for further statistical analysis.





Chapter 6

**Host-guest based colorimetric sensor array for discrimination of
pharmaceutical compounds**

6.1. Introduction

In recent years, colorimetric sensor arrays have emerged as powerful and versatile tools in the field of chemical and biological sensing, gaining considerable attention for their broad range of applications. A primary benefit of colorimetric sensor arrays is their simplicity—these systems are designed to produce easily interpretable colour changes in response to specific analytes. This simplicity translates into a cost-effective solution for complex detection tasks, as they do not require expensive or sophisticated instrumentation. Additionally, they offer rapid discrimination, which is particularly valuable in time-sensitive situations, with detection of optical signals achievable through the use of a basic UV–visible absorption spectrophotometer or even by visual inspection. This flexibility makes colorimetric sensor arrays accessible to a diverse range of users, from specialized laboratories to field environments where advanced equipment may be unavailable. As a result, colorimetric sensor arrays have been employed in various applications, such as detecting explosives²⁵¹, volatile organic compounds (VOCs)³⁷, environmental monitoring^{252,253}, and for healthcare diagnostics^{254,255}.

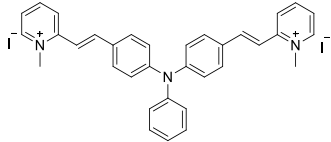
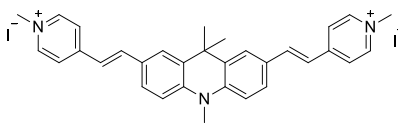
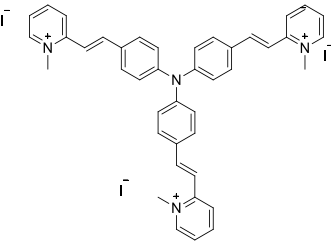
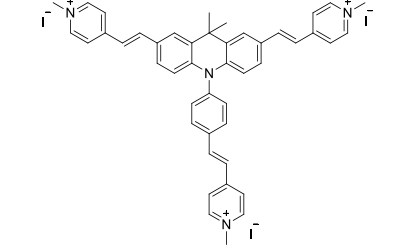
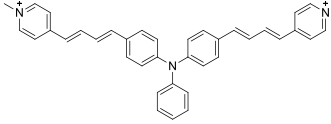
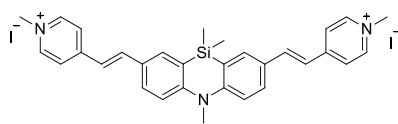
With these advancements in mind, our objective was to harness the TPA-CB[n] system for the development of a colorimetric sensing platform. The optical properties of TPA derivatives when involved in host-guest complexation with various members of the CB[n] family were investigated in Chapter 2 of this thesis. While the fluorescence modulations of the system have been extensively studied for creating a fluorescence-based sensor array, the variations in absorbance properties of TPA derivatives upon host-guest complexation with CB[n] will be investigated in this chapter. These absorbance variations which are a distinctive feature of host-guest complexation, offer a straightforward and effective optical transduction mechanism with potential utility for the detection of analytes in a time-efficient manner through an indicator displacement approach. In this study, a comprehensive library of TPA derivatives were analysed with different members of the CB[n] family, CB[6], CB[7] and CB[8]. The most promising TPA-CB[n] combinations, identified by their wavelength shifts upon host-guest complexation, were chosen to construct the colorimetric sensor array. This assembled sensor array was designed to discriminate analytes that interact with CB[n], especially those capable of displacing TPA from its TPA-CB[n] complex, resulting in a modification of the optical properties of TPA and the generation of unique colorimetric fingerprints for effective array-based sensing of the target analytes. Additionally, a streamlined LDA-PCA routine was developed for the efficient statistical analysis of the generated optical fingerprints. We applied this array to discriminate a specific class of pharmaceutical compounds that are designed to interact with the cavity of CB[n], based on their chemical structure. This newly developed array thereby provides a potential tool for addressing issues related to drug counterfeiting and enhancing the reliability of pharmaceutical quality control.

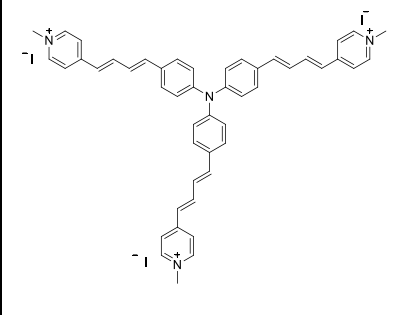
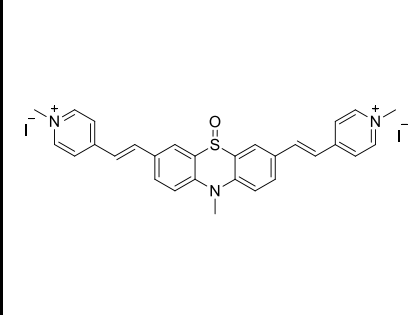
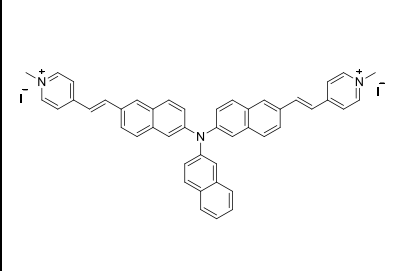
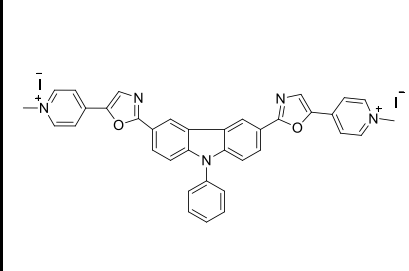
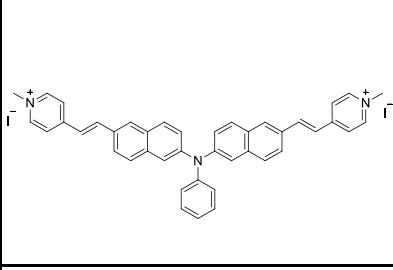
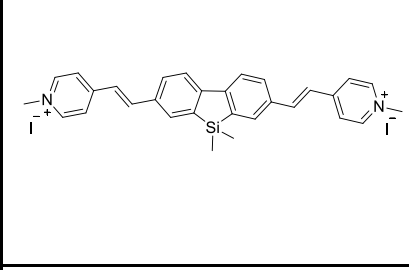
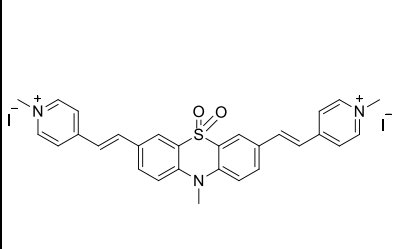
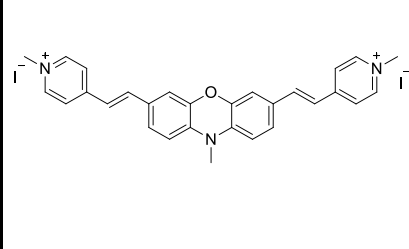
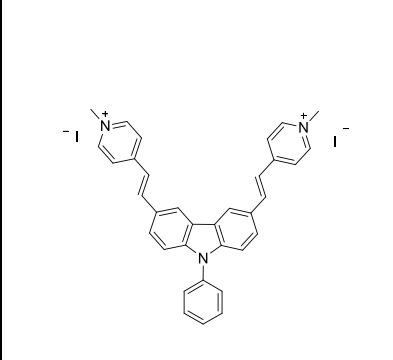
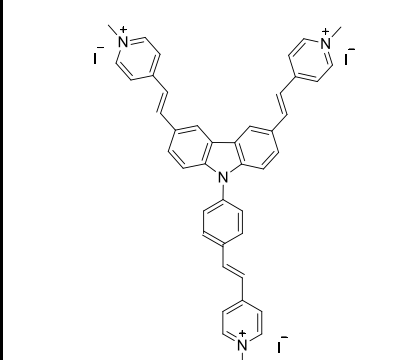
6.2. Results and Discussion: TPA-CB[n] colorimetric sensor array

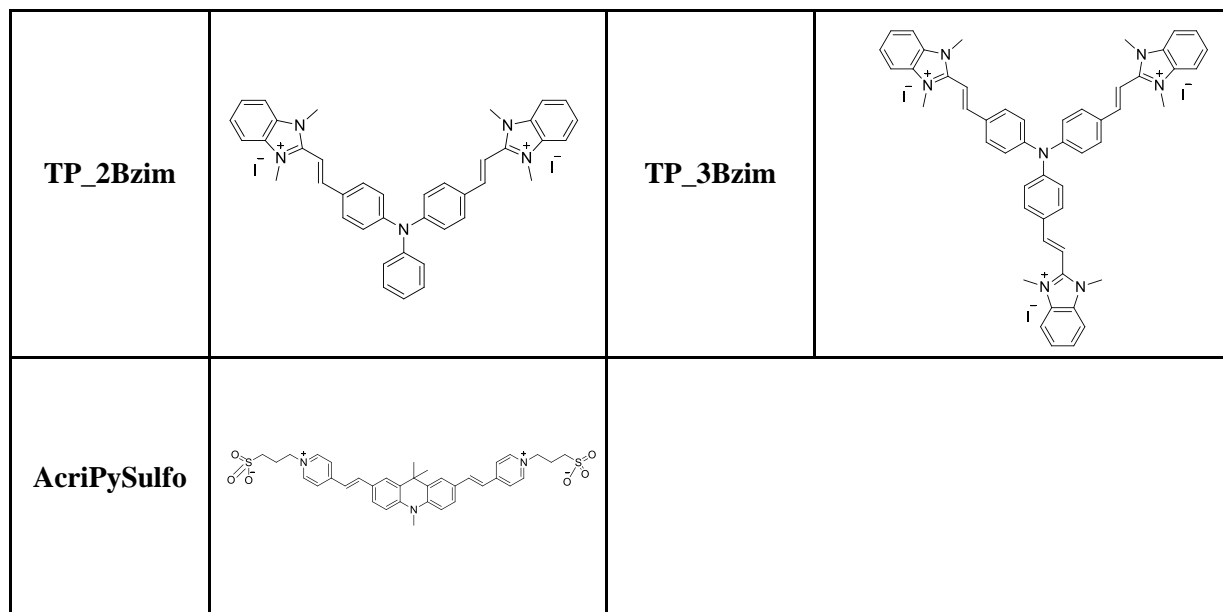
6.2.1. Characterization of TPA-CB[n] complexes

Building on the optical characterization results from an initial library of 8 TPA derivatives described earlier, we aimed to explore a more comprehensive library of 19 TPA derivatives with Carlos Gonzalez Galindo (M2 internship) in collaboration with the team of Dr. Florence Mahuteau-Betzer (*Institut Curie*). The structures of these additional derivatives are provided in Table 6.1 with fourteen 2-arm derivatives and five 3-arm derivatives. In addition to the previously discussed TPA derivatives, the remaining compounds feature a range of different donor cores and terminal acceptor groups. These include analogs with a rigid N-phenylcarbazole-based donor core, such as Cbz_2Py and Cbz_3Py, which have vinyl linkers that end in pyridinium acceptor groups. Another derivative, Cbz-2ox5Py, substitutes the original double bond of the TPA structure with a 2,5-diaryloxazole group. Moreover, TPA analogs with pyridinium units replaced by a more extended π -deficient heterocyclic N-benzimidazolium moiety specifically, TP_2Bzim and TP_3Bzim have also been considered. Additionally, the library has been further expanded to include 2-arm derivatives featuring donor moieties such as 9-silafluorene, phenothiazine, and phenoxazine.

Table 6.1: Chemical structures and diversity of the library of TPA derivatives used for the development of colorimetric sensor array

Name	Structure	Name	Structure
TP_2Pyo		AcriPy	
TP_3Pyo		Acri_3Py	
DV_2Py		PhenazPy	

DV_3Py		SulfoxPy	
TN2Py		Cbz_2ox5Py	
DN_2Py		SiFluoL	
SulfonPy		PhenoxPy	
Cbz_2Py		Cbz_3Py	



We firstly titrated the different TPA derivatives at 5 μM with increasing concentrations of CB[6], CB[7] and CB[8] up to 150 μM and evaluated the variations in the absorbance spectra registered between a wavelength range of 400 to 650 nm. As noted earlier, limited variations in the UV-visible spectra of the TPA derivatives upon complexation with CB[6] was observed. More prominent red-shift in wavelength was observed with CB[7] and CB[8], attributed to the electron absorbing effect of carbonyl in the CB[n] hosts, which can neutralize the positive charge of the pyridinium. This neutralization enhances the electron-withdrawal away from the triphenylamine core,

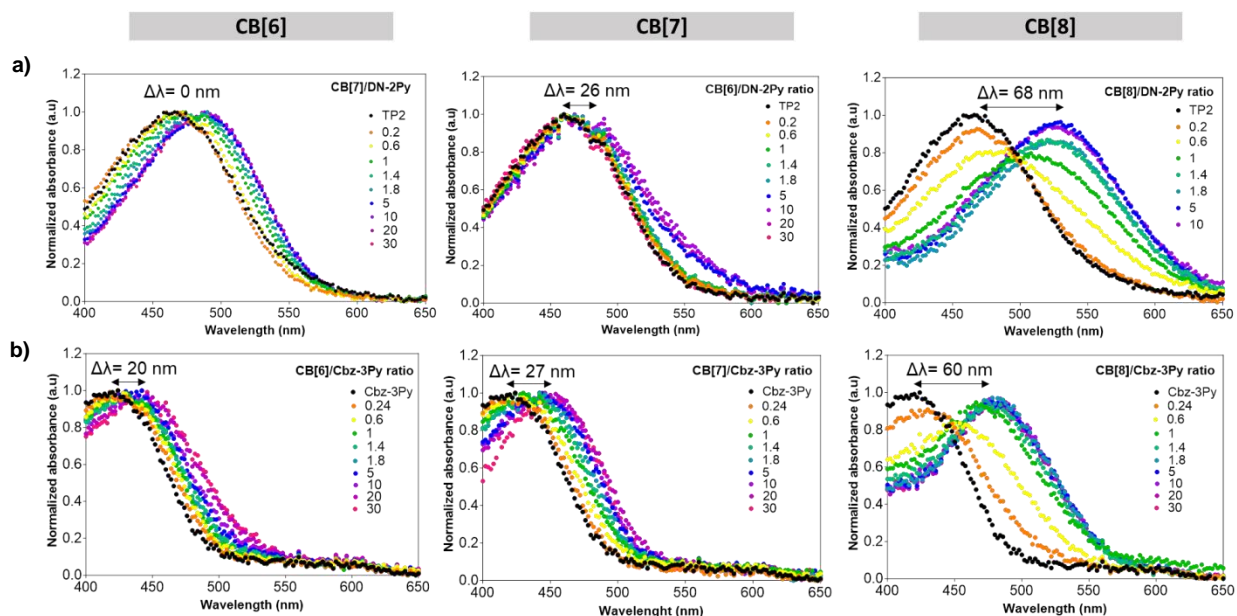


Figure 6.1: Normalized absorbance spectra of a) 2-arm (DN_2Py) and b) 3-arm (Cbz_3Py) TPA derivatives at 5 μM , recorded on titration with increasing concentration of CB[6], CB[7] and CB[8] in water.

inducing the spectral red-shift. This behavior is represented with a 2-arm and a 3-arm derivative, where the maximum shift in λ_{\max} among all the TPA-CB[n] pairs of 60 nm was observed for Cbz-3Py upon complexation with CB[8] (Figure 6.1a-b). Based on these initial results, we also observed that the absorbance titrations did not indicate the formation of different intermediate TPA-CB[n] complexes based on the TPA to CB[n] ratio used, but rather suggested the formation of a single and stable complex at the saturation point of absorbance. This finding is supported by previous literature reports, where NMR titrations of TPA derivatives with CB[8] showed no effect of increasing concentration on the stable host-guest complex formed¹⁸¹. Thus, from the absorbance titrations, we established the optimal CB[n]/TPA ratio for the 2-arm and 3-arm derivatives to be 5:1, with TPA at 5 μM and CB[n] at 25 μM . This ratio was utilized for all future experiments and provided an effective compromise between using minimal amount of sensor, while achieving significant absorbance and maximal λ_{\max} shift.

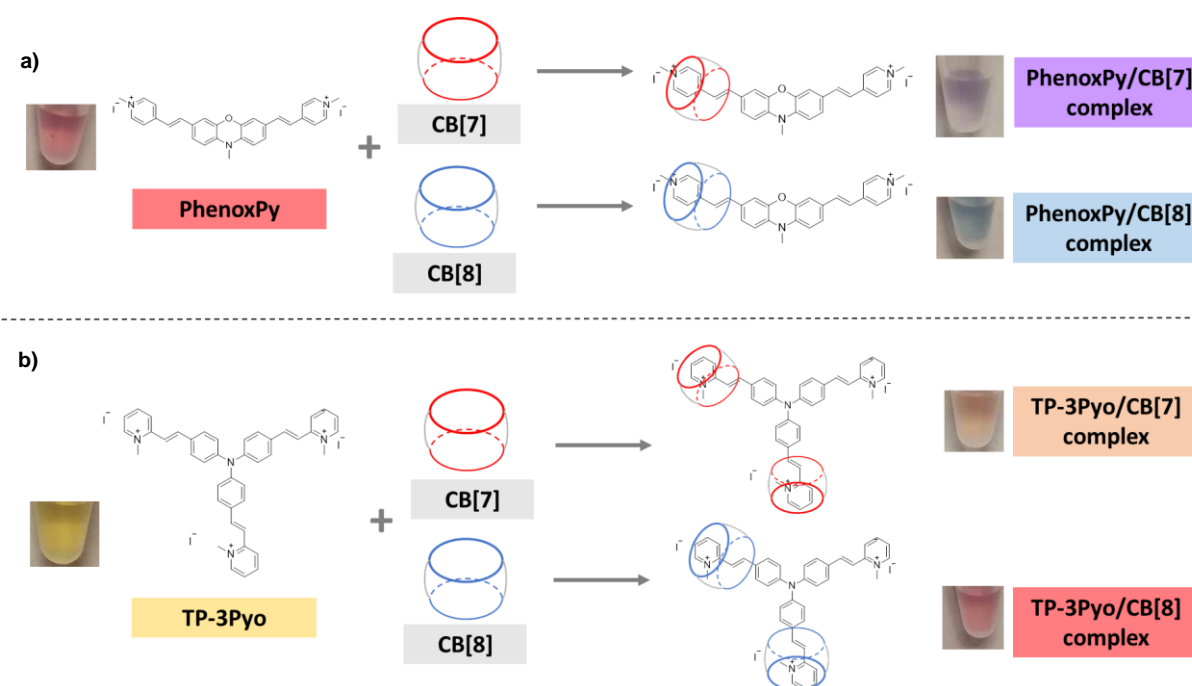


Figure 6.2: Visual comparison of absorbance variation of 2-arm (PhenoxPy) and 3-arm derivative (TP_3Pyo) upon complexation with CB[7] or CB[8] indicating a larger bathochromic shift for the CB[8] complex as compared to the CB[7] complex.

6.2.1.1. Selection of TPA-CB[n] sensing elements for colorimetric sensor array

To select and assemble the final TPA-CB[n] sensing elements of the colorimetric sensor array, we evaluated from the previous experiments the wavelength shift of the TPA derivatives with the CB[n]

($n=6, 7, 8$) at a concentration of $150 \mu\text{M}$. The goal was to identify the TPA-CB[n] pairs that exhibited the largest bathochromic shifts (Figure 6.2), under the premise that a greater wavelength shift would ensure greater sensitivity in detecting analytes capable of partially or completely displacing TPA from the CB[n] cavity. To evaluate this the spectrum recorded at the highest CB[n] concentration ($150 \mu\text{M}$) was chosen to ensure that the λ_{max} was captured at complete complexation of TPA-CB[n] (Table 6.2). From the spectra we observed negligible bathochromic shift for the complexes of TPA with CB[6], thus we focused on investigating the absorbance properties of the 19 derivatives primarily in complexation with the larger members, CB[7] and CB[8], and did not proceed with CB[6] for the development of the sensor array. The final TPA-CB[n] sensing elements were chosen primarily based on their bathochromic shifts, while also ensuring diversity in the variable sizes presented by the CB[7] and CB[8] hosts. This diversity is essential for facilitating differential interactions and variable binding affinities of the same analyte with different sensing elements, resulting in the generation of unique colorimetric fingerprints. Consequently, 11 TPA-CB[n] complexes were chosen to form the sensor array, with five CB[7] and 6 CB[8] complexes (Table 6.2, highlighted cells). Although a larger number of complexes could have been selected, increasing the number of sensing elements in the array would result in longer experimental time and more complex data management. Thus, this selection balanced the need for sensitivity and specificity with practical considerations of experimental efficiency.

Table 6.2: $\Delta\lambda_{\text{max}}$ shift determined from UV-visible absorbance spectra of 19 TPA derivatives (with alternative sample naming) complexed with CB[6-8].

TPA derivative	λ_{max} (TPA alone) (nm)	$\Delta\lambda_{\text{max}}$ (CB[6]/TPA) (nm)	$\Delta\lambda_{\text{max}}$ (CB[7]/TPA) (nm)	$\Delta\lambda_{\text{max}}$ (CB[8]/TPA) (nm)
TN2Py (TP1)	461	12	0	54
DN_2Py (TP2)	458	0	26	68
Cbz_2Py (TP3)	431	25	25	65
Cbz_3Py (TP4)	420	20	27	60
TP_2Pyo (TP5)	457	0	40	66
TP_3Pyo (TP6)	452	0	42	60
DV_2Py (TP7)	484	0	23	60
DV_3Py (TP8)	488	18	21	62
Acri_3Py (TP9)	485	0	40	60
Cbz_2ox5Py (TP10)	398	0	24	50
SiFluoL (TP11)	409	12	26	57

TP2Bzim (TP12)	433	0	10	0
TP3Bzim (TP13)	434	24	10	0
PhenazPy (TP14)	471	0	42	67
AcriPySulfo (TP15)	487	0	20	32
AcriPy (TP16)	488	0	39	71
PhenoxPy (TP17)	530	34	24	75
SulfoxPy (TP18)	430	8	27	44
SulfonPy (TP19)	423	4	28	39

6.2.1.2. Characterization of binding affinity of sensor array

UV-visible absorbance titrations of the selected TPA–CB[n] combinations were further conducted to assess the binding affinity of their complexation with CB[7]/CB[8]. The absorbance spectra were monitored at two distinct wavelengths for each TPA–CB[n] combination, the λ_{\max} for TPA derivative alone and the λ_{\max} for the TPA–CB[n] complex. Two representative TPA derivatives, TP_2Pyo and TP_3Pyo, titrated with CB[8] and CB[7], are shown in Figure 6.3. The absorbance at 500 nm for TP_3Pyo and 524 nm for TP_2Pyo, corresponding to the λ_{\max} of the TPA–CB[n] complex, was first plotted against the concentration of CB[7]/CB[8]. The data were fitted using a one-site binding, non-linear regression model, yielding global dissociation constants (K_d) of 13.85 μM and 4.01 μM respectively (Figure 6.3).

This approach was applied to all the 11 sensing elements of the sensor array and an overview of the results are presented in Table 6.3. The calculated binding affinities ranged from 10^4 M^{-1} to 10^5 M^{-1} for the 11 sensing elements of the array. This range supports the strategic design of the sensor array for interactions with potential analytes, as commonly seen in a typical IDA. This range of binding affinities also significantly contributes to the dynamic range of the sensor array since larger the difference between the lowest and highest affinity TPA guests with CB[n], the wider the range of analytes that can be detected by the array using the IDA approach.

It was observed that the TPA derivatives selected with CB[8] have a lower K_d value compared to those selected with CB[7], indicating a slightly higher binding affinity of TPA towards CB[8]. Among the 2-arm derivatives with CB[8], TP_2Pyo with a terminal ortho-pyridinium group exhibited a slightly higher K_d value compared to the TPA derivative with a para-pyridinium group. A similar trend was noted with the 3-arm derivative complexes with CB[7], where TP_3Pyo exhibited a slightly higher K_d value compared to Acri_3Py. This pattern was also observed in their respective 2-arm derivatives (TP_2Pyo and AcriPy).

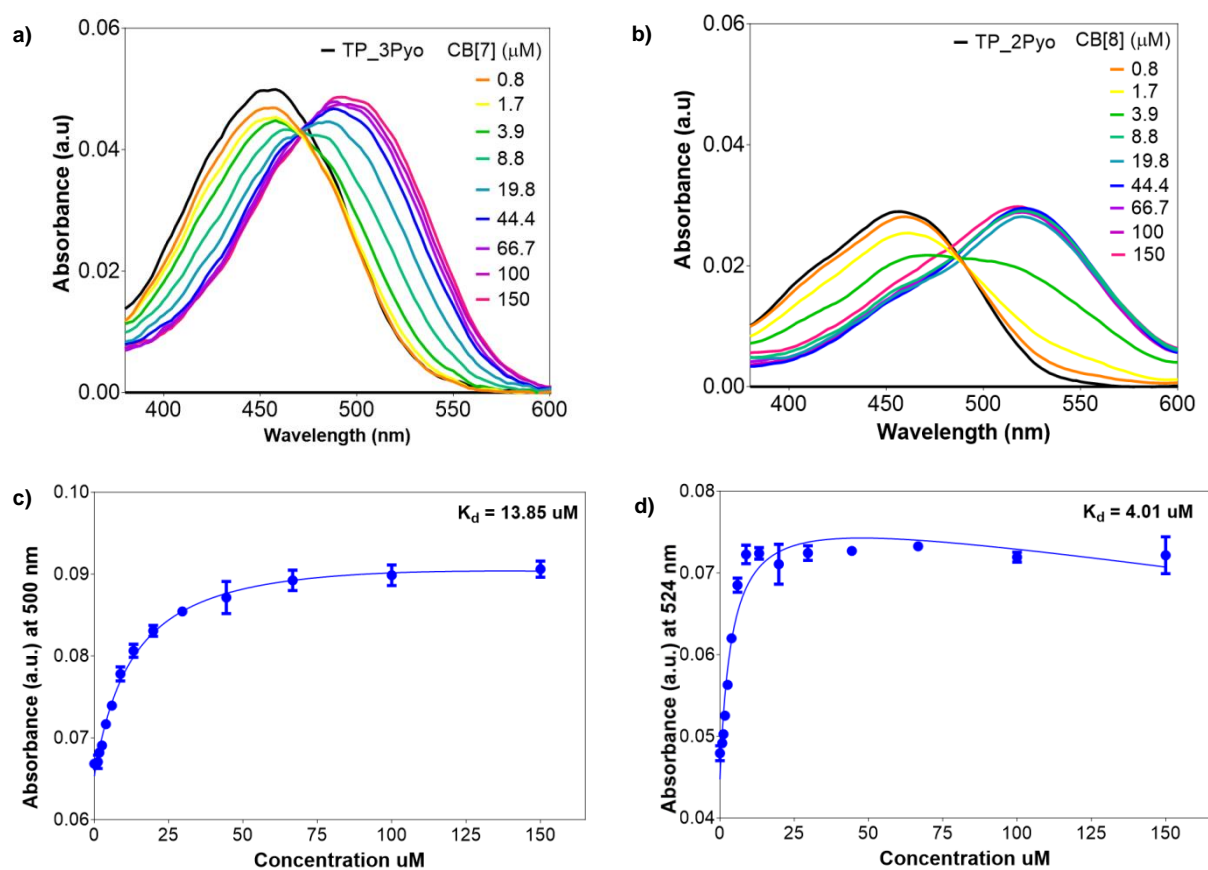


Figure 6.3: UV-visible titration of a) TP_3Pyo and b) TP_2Pyo (5 μM) in aqueous solution by CB[7] and CB[8] respectively (0.8 to 150 μM). The binding curves were obtained by plotting the variation in the absorbance for c) TP_3Pyo and d) TP_2Pyo with increasing concentration of CB[7] and CB[8] added. The absorbance was monitored at the λ_{max} at TPA + CB[n] with triplicate measurements for determination of the dissociation constant (K_d).

Table 6.3: Binding affinity parameters for the 11 sensing elements of the colorimetric sensor array determined by UV-visible titrations monitored at (λ_{max}) at TPA-CB[n]

CB[7]	Wavelength (λ_{max}) at TPA-CB[n] (nm)	K_d (μM)	K_a (M^{-1})
TP_2Pyo	500	9.96	1.00×10^5
TP_3Pyo	500	13.85	7.22×10^4
Acri_3Py	520	10.21	9.79×10^4
PhenazPy	516	10.69	9.35×10^4
AcriPy	520	7.09	1.41×10^5
CB[8]	Wavelength (λ_{max}) at TPA-CB[n] (nm)	K_d (μM)	K_a (M^{-1})
DN-2Py	530	2.39	4.18×10^5
Cbz-2Py	490	2.05	4.88×10^5
TP_2Pyo	524	4.01	2.49×10^5
PhenazPy	540	2.43	4.12×10^5
AcriPy	560	3.34	2.99×10^5
PhenoxPy	620	2.91	3.44×10^5

6.2.2. Optical fingerprinting by TPA-CB[n] sensor array

To assess the optical fingerprinting potential of the optimized TPA-CB[n] colorimetric sensor array, we conducted a preliminary test to firstly determine the concentration of analyte required to maximally displace the TPA derivatives from their complex with CB[n]. We analysed the variations in the UV-visible spectra of DN_2Py and TP_2Pyo with CB[8] in the presence of varying concentrations of adamantanol, a high-affinity small molecule binder of CB[8] ($K_a = 2.3 \times 10^{10} \text{ M}^{-1}$)⁶⁷.

The concentrations of the TPA derivatives and CB[8] were fixed at 5 μM and 25 μM , respectively, while the concentration of adamantanol was varied between 10 μM and 100 μM . The results showed that TP_2Pyo exhibited complete displacement from CB[8] at lower concentrations of adamantanol, whereas DN_2Py required a much higher concentration of adamantanol—beyond 100 μM —for complete displacement (Figure 6.4). This behaviour is also consistent with the K_d values determined in the previous section, where DN_2Py demonstrated a higher binding affinity to CB[8] as compared to TP_2Pyo. Thus, based on this experiment, a concentration of 100 μM was selected as the optimal analyte concentration for TPA displacement in all subsequent experiments. This concentration was chosen to ensure that the system remains within the detection range for most biomolecules.

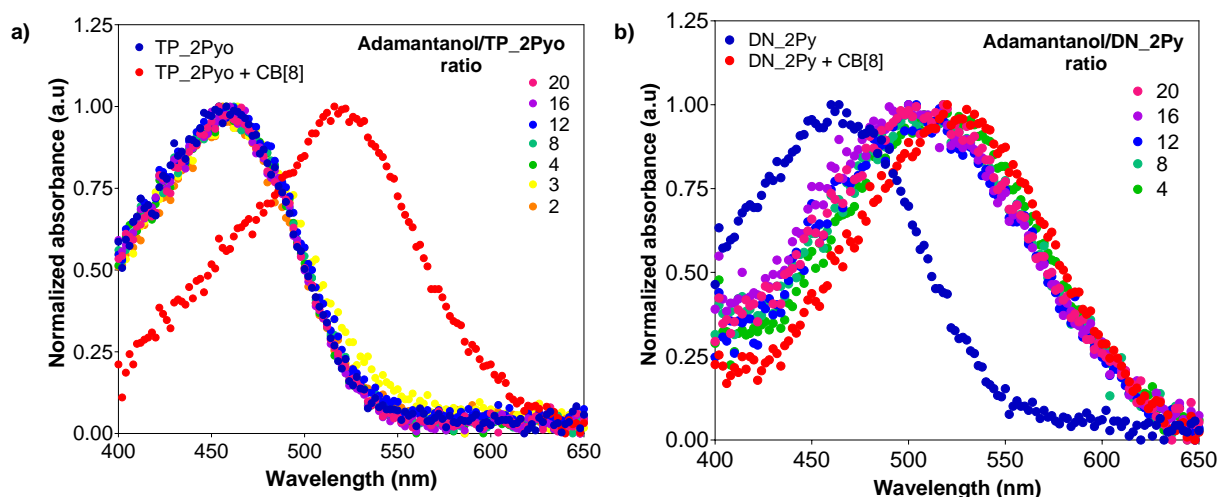



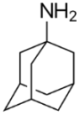
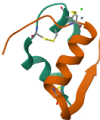
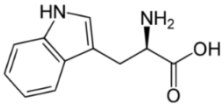
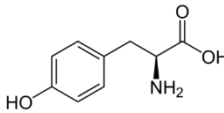
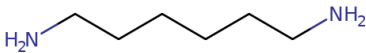
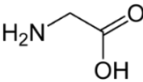
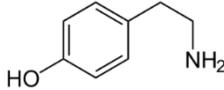
Figure 6.4: UV-visible absorbance titration of a) TP_2Pyo and b) DN_2Py ($5 \mu\text{M}$) complex with CB[8] ($100 \mu\text{M}$) upon addition of adamantanol ($10 - 100 \mu\text{M}$), with blue indicating the absorbance spectra of TPA derivative alone and red indicating the saturated complex of TPA-CB[8] derivatives.

6.2.2.1. Interaction of colorimetric sensor array with small molecules

The sensing ability of the sensor array was then evaluated using a group of eight molecules, including amino acid, protein, and small molecules known to form host-guest complexes with CB[n]s with a range of affinity constant from 10^{10} to 10^4 M^{-1} . These molecules were chosen due to the hydrophobic nature of their aromatic groups or the presence of positive charges, both of which enhance their interaction with the CB[n] hosts. A comprehensive list of the selected molecules, along with their corresponding association constants with CB[7], is provided in Table 6.4.

We proceeded by adding the TPA derivatives and CB[n] to a UV-transparent half-area 96-well plate, ensuring that the final concentrations were maintained at $5 \mu\text{M}$ and $25 \mu\text{M}$, respectively. After this, the initial absorbance spectra were recorded. Following this, the analytes were introduced to the wells in three replicates to achieve a final concentration of $100 \mu\text{M}$, and the absorbance spectra were measured again. The shift in the maximum absorbance wavelength was monitored for each analyte, generating a unique colorimetric fingerprint with dimensions of $[24 \times 11]$ (3 replicates for each of the 8 analytes \times 11 sensing elements, with 5 using CB[7] and 6 using CB[8]). In the heatmap depicting the colorimetric fingerprint (Figure 6.5a), it is clear that some sensing elements play a more prominent role in discrimination, as shown by the wider range of color changes across the 8 analytes. In contrast, other elements contribute far less. The smallest and largest wavelength shifts are associated with glycine (purple) and adamantanol (yellow), respectively, which can be linked to their corresponding association constants. To identify the non-contributing sensing elements and develop an efficient discrimination model, this data was further analyzed using the LDA-PCA method, as described in Chapter 3 of this thesis.

Table 6.4: List of analytes tested with the TPA-CB[n] colorimetric sensor array including amino acid, protein and small molecules and their associated K_a with CB[7]

Analyte	Structure	K_a
Adamantanol		$2.3 \times 10^{10} \text{ M}^{-1}$
Amantadine		$1.0 \times 10^{10} \text{ M}^{-1}$
Insulin		$1.5 \times 10^6 \text{ M}^{-1}$
Tryptophan		$3.7 \times 10^5 \text{ M}^{-1}$
Tyrosine		$2.7 \times 10^5 \text{ M}^{-1}$
Hexamethylenediamine		$8.9 \times 10^4 \text{ M}^{-1}$
Glycine		10^2 M^{-1}
Tyramine		-

Statistical data analysis of colorimetric fingerprint data

First, the obtained fingerprint dataset was analysed using LDA with leave-one-out cross-validation, yielding a discrimination error rate of 21%. Next, Principal Component Analysis (PCA) was employed to assess the contribution and quality of each sensing element in the sensor array towards the discrimination process. PCA revealed that only six sensing elements were necessary to effectively distinguish the 8 analytes: PhenazPy + CB[7], DN_2Py + CB[8], TP_2Pyo + CB[8], PhenazPy + CB[8], AcriPy + CB[8], and PhenoxPy + CB[8] as indicated in the scree plot in Figure 6.5b. The PCA analysis involves first computing the eigenvalues for each principal component, reflecting the amount of variance in the data captured by each component. These eigenvalues are then expressed as a percentage of the total variance, represented by the sum of the eigenvalues of all the principal components. This percentage is visualized in a linear scree plot, where the x-axis represents the different components, and

the y-axis shows the percentage of variance explained by each component. This plot therefore helps in the determination of the optimal number of principal components required for effective discrimination, typically identified by the point where the plot exhibits a distinct change in slope. Using this insight, the original fingerprint dataset was reduced to a dimension of $[24 \times 6]$ by considering only the sensing elements that contribute to the first two dimensions of the scree plot. Upon reanalysis with LDA, the discrimination error rate was significantly reduced to 4.2% (Figure 6.5c). This outcome demonstrated the effectiveness of the LDA-PCA routine in optimizing the combination of sensing elements for discriminating the analytes, while eliminating non-contributing features that added noise and hindered accurate discrimination. Notably, it was observed that the TPA-CB[8] sensing elements (5 TPA-CB[8] pairs) contributed to a larger extent towards the discrimination as compared to CB[7] (1 TPA-CB[7] pair). These overall results demonstrated the capability of the sensor array to generate distinct optical colorimetric fingerprints, which can effectively discriminate between analytes that have the potential to bind with CB[n] hosts.

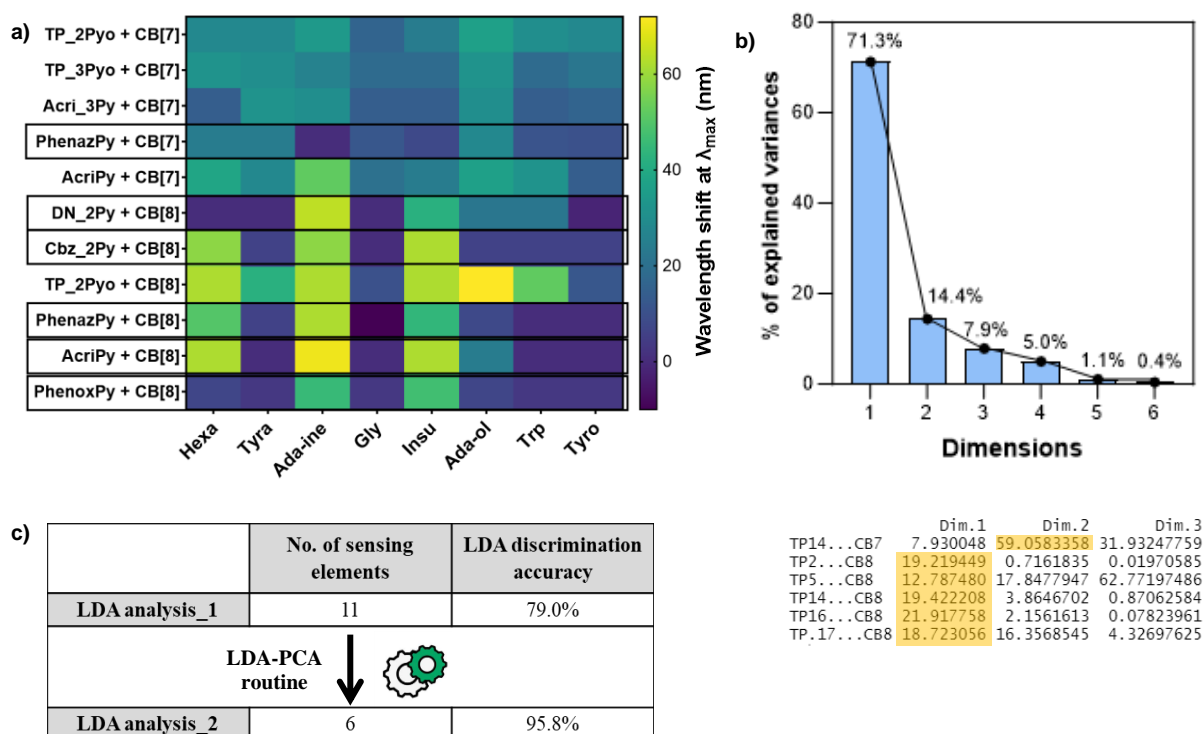


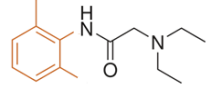
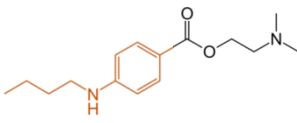
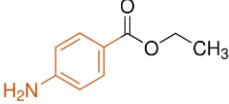
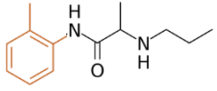
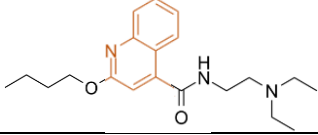
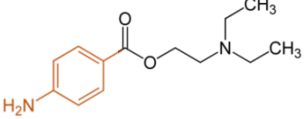
Figure 6.5: a) Heatmap illustrating the colorimetric fingerprint generated by the TPA-CB[n] sensor array with 11 sensing elements in response to various analytes (small molecules, amino acid, protein). b) PCA analysis showing the contribution of each sensing element toward the discrimination of the analytes in a scree plot. c) Table depicting improvement in discrimination accuracy achieved through LDA by removing non-contributing sensing elements identified by PCA, resulting in a final accuracy of 95.8% using 6 sensing elements. The final six optimized and contributing sensing elements have been highlighted on the heatmap.

6.2.2.2. Interaction of colorimetric sensor array with pharmaceutical molecules

Pharmaceuticals encompass a diverse array of molecules, including organic compounds, peptides, proteins, and nucleic acids, each with specific pharmacological mechanisms and therapeutic effects. As the pharmaceutical industry continues to expand, so does the counterfeit drug industry. Beyond the intellectual property concerns related to patent and trademark infringement, counterfeit medicines pose significant risks to both the economy and public health. In countries with weak regulatory systems, the absence of a reliable and quality-controlled medicine supply creates an environment where falsified and substandard drugs can thrive. These counterfeit drugs can lead to poisoning, untreated diseases, disease progression, and in severe cases, death^{256–258}. Since 1970, the World Health Organization (WHO) has emphasized the principle of essential medicines as a cornerstone for an effective and equitable healthcare system. However, within this list²⁵⁹, the group of anaesthetic medicines²⁶⁰ are often overlooked, resulting in supply shortages and increased instances of counterfeiting, largely due to the high cost of standard medications. To combat and detect falsified and substandard anaesthetic drugs, various quantitative techniques have been employed, such as HPLC, Raman spectroscopy, near-infrared spectroscopy, fluorescence spectroscopy, and mass spectrometry^{261–263}. While these methods are generally effective in identifying counterfeit and substandard drugs, they often require complex sample preparation, and the high cost of the equipments limits their accessibility in low- and middle-income countries. Additionally, these techniques are not well-suited for high-throughput screening²⁶³. The chemical nose strategy has emerged as an alternative strategy in the field of drugs counterfeit²⁶⁴ and has been employed for the successful discrimination of non-steroidal anti-inflammatory drugs²⁶⁵, carboxylate drugs²⁶⁶ and antimalarial tablets²⁶⁷.

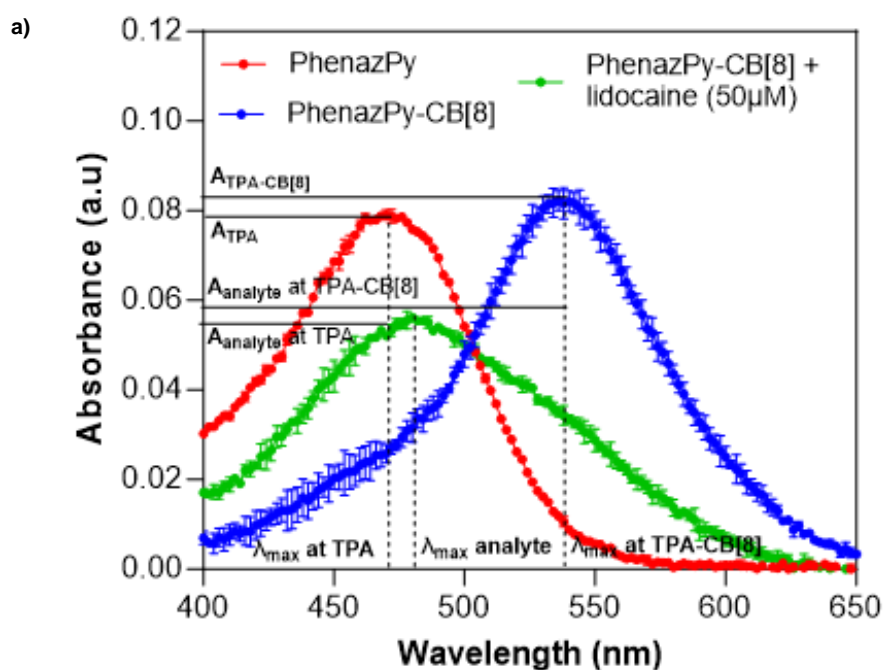
We proposed to use the developed colorimetric assay to identify and discriminate a group of pharmaceutical anaesthetic molecules bearing chemical structural properties such as aromatic rings and protonated amine groups. These features were expected to facilitate interactions with the TPA-CB[n] system to produce distinct colorimetric fingerprint outputs that allow for the identification and classification of these compounds. We hypothesized that by training the sensor array towards these anaesthetic molecules, we could extend the potential use of this system towards counterfeit drug detection applications. For this purpose, we selected a group of six commercially purchased pharmaceutical molecules, detailed in Table 6.5. These molecules were chosen due to their easy availability and their documented encapsulation within CB[n] hosts, as reported in literature.

Table 6.5: Structure and binding properties of anaesthetic molecules tested with TPA-CB[n] colorimetric sensor array

Anaesthetic molecule	Chemical structure	K_a with CB[7] ²³
Lidocaine		-
Tetracaine		$1.5 \times 10^4 \text{ M}^{-1}$
Ethyl-p-aminobenzoate		$2.2 \times 10^4 \text{ M}^{-1}$
Prilocaine		$2.6 \times 10^4 \text{ M}^{-1}$
Dibucaine		$1.8 \times 10^5 \text{ M}^{-1}$
Procaine		$3.5 \times 10^4 \text{ M}^{-1}$

We firstly measured the absorbance spectra of the 11 TPA-CB[n] pairs added to a UV-transparent half-area 96-well plate with final concentrations maintained at 5 μM and 25 μM , respectively. The initial absorbance spectra were recorded [$A_{\text{TPA-CB}[n]}$]. Following this, the six anaesthetic molecules spiked in 1X PBS were added to the well in eight replicates to achieve a final concentration of 100 μM , and the absorbance spectra were measured again [A_{analyte}]. From these measurements we extracted three main output sensing channels for each TPA-CB[n] pair (Figure 6.6):

- i) Absorbance variation at the maximum absorbance wavelength for the TPA/CB[7-8] complex after analyte addition [$\text{Abs}(\text{analyte})_{\lambda(\text{TPA-CB}[8])} - \text{Abs}_{\lambda(\text{TPA-CB}[n])}$], abbreviated as **Abs CB[n]-TPA (λ)**
- ii) Absorbance variation at the maximum absorbance wavelength for TPA after analyte addition [$\text{Abs}(\text{analyte})_{\lambda(\text{TPA})} - \text{Abs}_{\lambda(\text{TPA})}$] abbreviated as **Abs 2 CB[n]-TPA (λ)**
- iii) Maximum absorbance wavelength shift of the TPA/CB[7-8] complex after analyte addition [$(\lambda_{\text{max}})_{\text{analyte}} - (\lambda_{\text{max}})_{\text{TPA-CB}[n]}$] abbreviated as **λ CB[n]-TPA**



b)

<i>Parameters measured</i>			
	λ_{max}	<i>Absorbance (a.u.)</i>	
<i>TPA</i>	470 nm	470 nm	0.078
<i>TPA-CB[8]</i>	540 nm	540 nm	0.080
<i>Analyte</i>	480 nm	470nm	0.054
		540nm	0.058
<i>Sensing channels</i>		<i>Fingerprint Output</i>	
Channel 1 [Abs(analyte) $_{\lambda(TPA-CB[8])}$ - Abs $_{\lambda(TPA-CB[n])}$]		-0.022 a.u.	
Channel 2 [Abs(analyte) $_{\lambda(TPA)}$ - Abs $_{\lambda(TPA)}$]		-0.024 a.u.	
Channel 3 [(λ_{max}) $_{analyte}$ - (λ_{max}) $_{TPA-CB[n]}$]		-60 nm (480 - 540 nm)	

Figure 6.6: a) Representative UV-visible spectra of PhenazPy – CB[8] sensing element in the presence of anaesthetic molecule lidocaine. b) Three-sensor channels are extracted from each TPA-CB[n] sensing element pair as indicated by the parameters measured in the plot ($[A_{analyte} - A_{TPA-CB[n]}]$, $[A_{analyte} - A_{TPA}]$ and $[(\lambda_{max})_{analyte} - (\lambda_{max})_{TPA-CB[n]}]$).

Thus, with 3 channels considered for each of the 11 sensing elements, a total of 33 sensing channels were obtained from the sensor array. A total of 22 channels correspond to the absorbance variations while the remaining 11 channels correspond to the wavelength shift registered in the presence of the analytes. This colorimetric fingerprint of the analytes generates a dataset with dimension of $[48 \times 33]$ (8 replicates for each of the 6 analytes \times 3 channels for each of the 11 TPA-CB[n] sensing elements). This fingerprint data is represented in a heat map in Figure 6.7 corresponding to the two different sets of signal outputs extracted to obtain the colorimetric fingerprint.

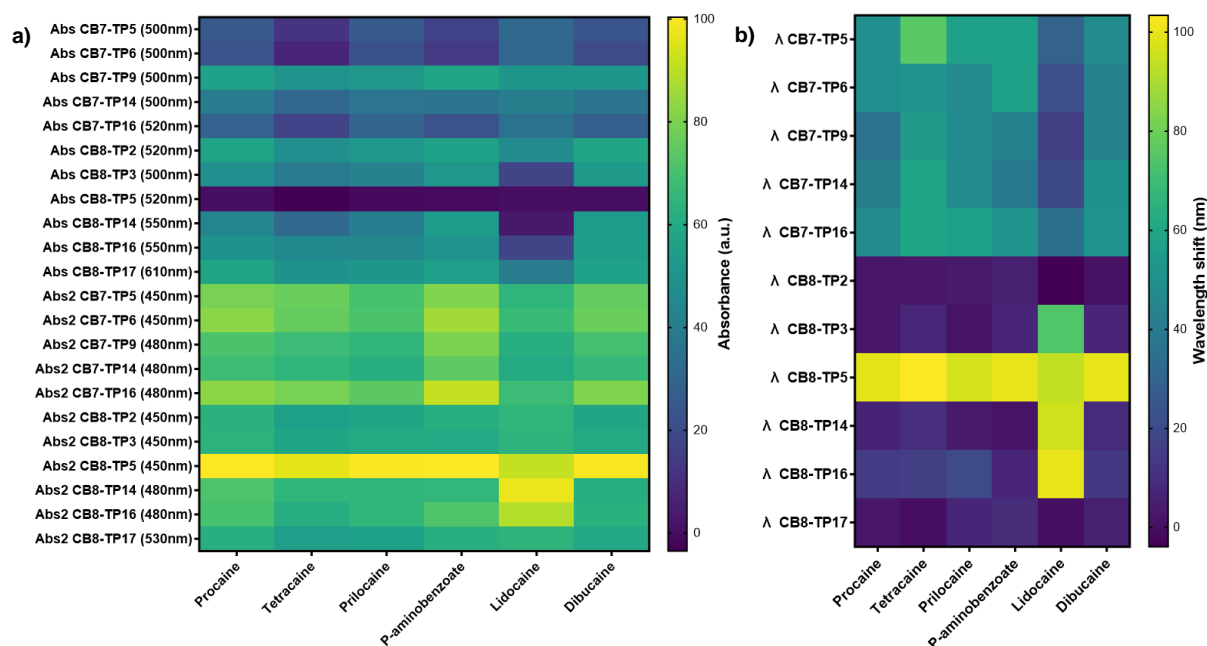


Figure 6.7: Heatmap illustrating the colorimetric fingerprint generated by a) absorbance variations (2 channels \times 11), where $Abs\ CB[n]-TPA(\lambda) = [Abs(\text{analyte})_{\lambda(TPA-CB[8])} - Abs_{\lambda(TPA-CB[n])}]$ and $Abs\ 2\ CB[n]-TPA(\lambda) = [Abs(\text{analyte})_{\lambda(TPA)} - Abs_{\lambda(TPA)}]$ and b) λ_{max} variations of the 11 TPA-CB[n] sensing elements (1 channel \times 11) in response to the six anaesthetic molecules.

Statistical data analysis of colorimetric fingerprint data

The generated dataset was analysed by the LDA-PCA routine that has been streamlined for the developed colorimetric sensor array. The dataset was firstly analysed LDA with leave-one-out cross-validation, yielding a discrimination error rate of 3.5%. However, it is possible to observe from the heatmap that some sensor channels yielded redundant responses, either by producing similar outputs for different analytes or by generating the same response for different sensor channels (e.g., Abs 2 CB[8]-TP2 and Abs 2 CB[8]-TP3). This redundancy hindered accurate discrimination. To address this, PCA was utilized to identify and remove the non-contributing sensing elements. The scree plot generated from the PCA indicated that only 18 sensing elements were necessary for effective

discrimination of the anaesthetic molecules. Of these, 8 channels corresponded to wavelength shifts, while 10 channels reflected absorbance variations in response to the analytes. Consequently, the original dataset was refined to a dimension of $[48 \times 18]$. When reanalysed using LDA, the refined dataset yielded a discrimination error rate of 0%, achieving perfect discrimination of the analytes.

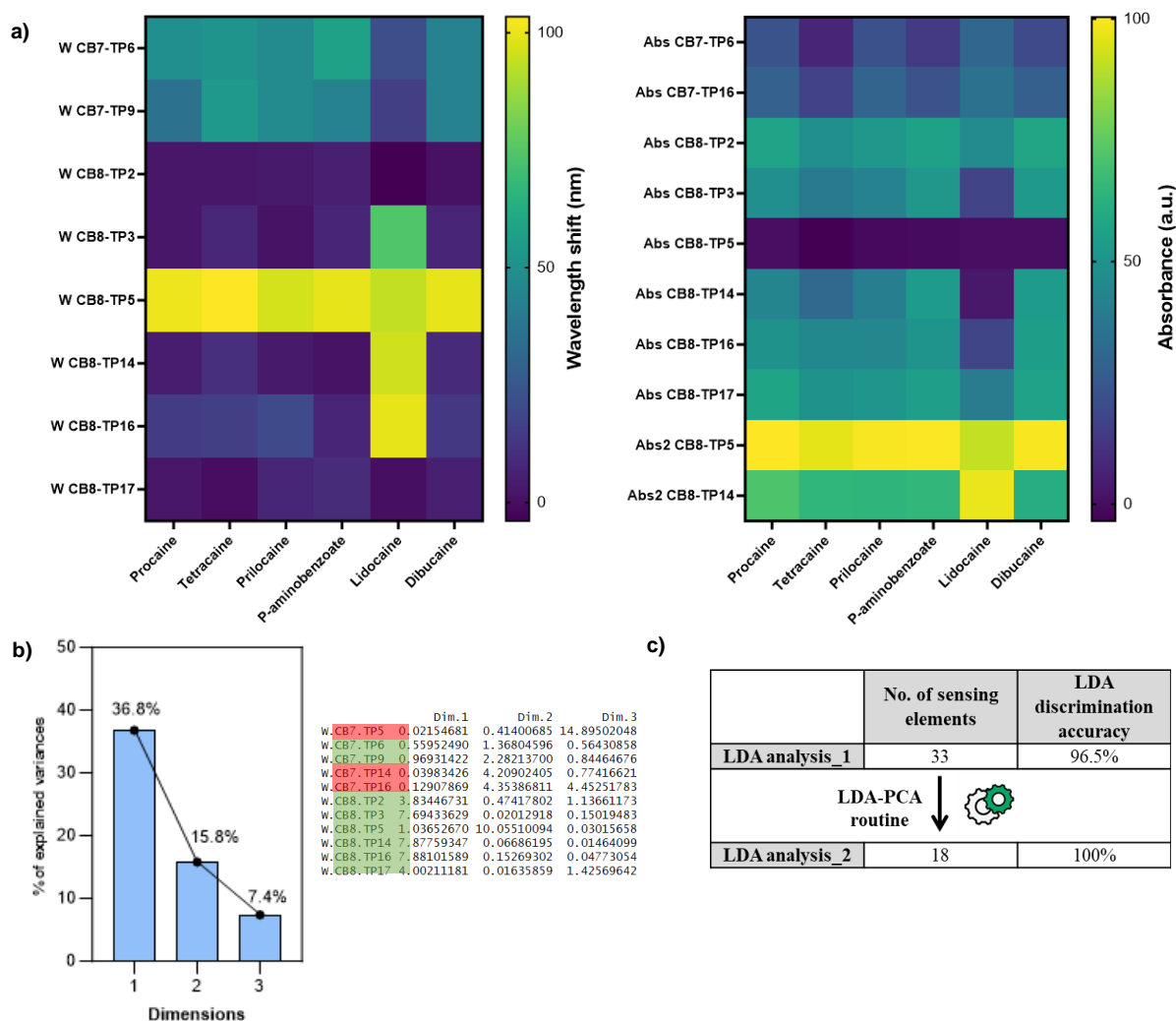


Figure 6.8: a) Heatmap illustrating the colorimetric fingerprint of the six anaesthetic molecules generated by the reduced 18 channel TPA-CB[n] sensor array with 8 channels corresponding to the λ_{max} variations and 10 channels corresponding to the absorbance variations. b) PCA analysis showing the contribution of each sensing element toward the discrimination of the analytes in a scree plot for the reduction of 33-channel sensor array. c) Table depicting improvement in discrimination accuracy achieved through LDA by removing non-contributing sensing elements identified by PCA, resulting in a final accuracy of 100% using 18 sensing elements.

This successful application of the TPA-CB[n] sensor array marks a promising step towards developing a sensor array capable of detecting counterfeit anaesthetic drugs. With further refinements the array would allow discrimination of similar products from different regions, identification of

adulterated products and the assessment of drug quality. The key advantage of this system lies in its ability to produce rapid results using only a spectrophotometer. However, further investigations are necessary before the sensor array can be effectively applied to testing drugs available on the market. This includes training the array to recognize varying concentrations of additive molecules, establishing a clear linear detection range, and making other essential optimizations to ensure its accuracy and reliability in real-world applications.

6.3. Conclusion

In conclusion we have developed a colorimetric sensor array using an extended library of TPA derivatives and their complexes with CB[7] and CB[8]. By harnessing the variations in the absorbance properties of the TPA derivatives upon host-guest complexation, followed by the addition of analytes, we have demonstrated the ability of the sensor array to discriminate a small group of few diverse molecules such as amino acids, proteins and small molecules. Furthermore, the array was effectively applied to distinguish commonly used anaesthetic drugs by extracting distinct output parameters from the absorbance spectra of the sensor array to create a colorimetric fingerprint. This result highlights the potential of this sensor array for counterfeit detection applications. Furthermore, the data treatment and statistical analysis of the colorimetric fingerprint was streamlined through the development of a combined LDA-PCA algorithm, which was designed to exclude non-contributing sensing elements, thereby enhancing the overall discrimination performance of the sensor array by LDA.

6.4. Materials and Methods

6.4.1. Protocol for discrimination of analytes by colorimetric sensor array

Stock solutions of commercial analytes were prepared in 1X PBS, while pharmaceutical molecules were dissolved in DMSO due to solubility requirements.

For the sensing experiments with the different analytes, 25 μl of 10 μM solution of TPA derivatives with 25 μl of 50 μM CB[7] or CB[8] solution in water were taken in UV-transparent 96-well plates purchased from Corning. To this solution, 2.5 μl of corresponding analyte solutions were added to obtain a final concentration of 100 μM spiked in 1X PBS. The UV-visible absorbance spectra were recorded on a TECAN SPARK CYTO400 spectrophotometer from 400 to 650 nm. Control spectra were recorded with the addition of 1X PBS in the absence of spiked analyte to the TPA-CB[n] complexes.

6.4.2. R codes for LDA-PCA routine

All R codes utilized for the discrimination studies will be provided in the Appendix 7-9 of this thesis.



Chapter 7

Macrocyclic fluorescence based sensor array:

***A methodological study for the diagnosis of the SARS-CoV-2
infection in human serum***

The work presented in this chapter is based on the following article:

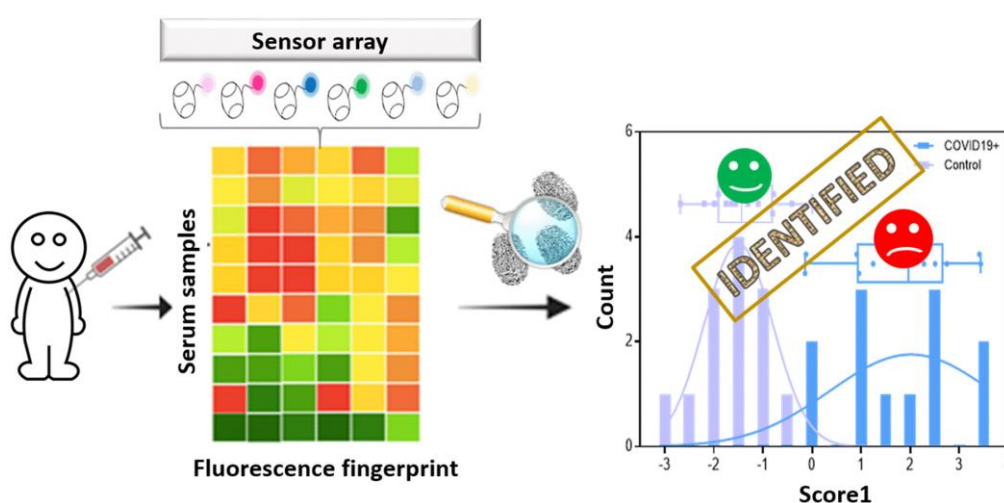
“A methodological study for the diagnostic of the SARS-CoV-2 infection in human serum with a macrocyclic sensor array”

Sensors and Diagnostics, 2024, 3, 863-871.

Bosco, M. S.; Topcu, Z.; Pradhan, S.; Sossah, A.; Tsatsaris, V.; Vauloup-Fellous, C.; Agasti, S.S.; Rozenholc, Y.; Gagey-Eilstein, N.

7.1. Presentation of the study

7.1.1 Abstract



The global outbreak of COVID-19 caused by SARS-CoV-2 has underscored the urgent need for rapid and sensitive diagnostic methods. While conventional nasopharyngeal swab-based tests face limitations in accuracy and turnaround time, alternative approaches leveraging blood samples hold promise. The alterations in blood composition induced by SARS-CoV-2 infection present an opportunity for novel diagnostic strategies, yet identifying specific biomarkers for diagnosis remains challenging. Therefore, a hypothesis-free diagnostic approach, based on the ‘Chemical Nose’ sensing method is proposed as a potential blood-based diagnostic strategy. This strategy focuses on detecting global changes in blood composition using an array-based sensing system combined with machine learning algorithms. In our study a host-guest-based direct sensing approach utilizing cucurbit[7]uril (CB[7]) as a receptor scaffold covalently conjugated to fluorophore moieties has been employed to develop the sensor array. This CB[7]-FL array selectively binds to various biomolecules in patient blood samples by a range of non-specific binding modes, generating unique fluorescence outputs that are characteristic of the interacting biomolecules. We have utilized this array to differentiate serum samples from healthy pregnant patients and SARS-CoV-2 infected pregnant patients with clinically relevant

accuracy. Our study not only serves as a proof-of concept for potential clinical diagnosis, but also provides a meticulous methodology to use this CB[7]-FL array-based sensor to monitor and analyse the fluorescence signals and fingerprints obtained from a small set of serum samples towards each sensor element of the CB[7]-FL array. This study therefore offers a streamlined methodology involving design of experiments, data extraction, treatment and analysis to run these types of chemical nose experiments, with key considerations to be kept in mind when using such chemical nose sensor arrays to diagnose more complex diseases and their associated severity and outcomes.

7.1.2. Covalently linked vs. Host-guest based sensing strategy

The strategy of host-guest based sensing, as discussed in previous chapters has been widely employed for developing IDA's. This approach allows ready access to libraries of fluorescent dyes with minimal synthetic efforts and less stringent selection criteria. The tunable optical properties and the ability to use several indicators with the same macrocyclic host based on variable binding affinities, enables generation of diversity in the recognition process with easy operation and controllable parameters. Despite these advantages, and the exhibition of high binding affinities and fast binding kinetics in deionized water, this approach has some drawbacks. For the creation of effective competitive binding conditions in IDA's, the binding affinity and working concentration of the indicator and analyte need to be finely tuned. Additionally, the presence of salts can significantly impact the non-covalent interactions that facilitate host-guest complexation, either through competitive or cooperative binding of cations to the carbonyl groups of the CB[n] portal, potentially causing the dissociation of weakly binding analytes. The non-covalent complexes formed between CB[n] and receptor dyes in IDA are further susceptible to dissociation upon dilution, similar to other biomolecular non-covalent complexes. Consequently, while IDA based sensing systems are highly effective for sensing in deionized water or minimal buffer conditions, they require adjustments to address possible complications that could arise in saline media or biofluids, which could complicate the sensing process⁹².

These limitations have led to the exploration of an alternative design strategy, which involves the covalent integration of the indicator dye and the macrocyclic receptor into a single, non-dissociable unimolecular chemosensor. For instance, emissive naphthalene units have been incorporated covalently into the cavity walls of cucurbit[n]uril derivatives, enabling them in detecting addictive over-the-counter drugs in human urine, which has variable salt concentrations²⁶⁹. However, using these systems requires the adoption of laborious synthesis, clever chemical design, and the selection of appropriate indicators that do not compete with CB[7] to create effective sensor libraries. This makes it challenging to introduce new recognition units with ease and enhance diversity in the recognition process of the sensor array. However, despite these drawbacks, the covalent tethering keeps the macrocycle and the dye in close proximity, ensuring a consistently high local concentration of the dye even upon dilution. This proximity in the case of intramolecular dye encapsulation systems (IIDA) maximises the

abundance of host-dye inclusion complex in the absence of the analyte, thus facilitating sensitive detection by displacement of the dye upon addition of the target analyte. Additionally, simple covalently bound dye molecules that produce unique optical outputs due to changes in their microenvironment upon analyte binding to the macrocycle, provide richer analyte information. This is because the generated sensing response is not solely related to the absolute binding affinity of the analyte but also depends on the different modes of interaction with the analyte. This allows for straightforward optimization of assay performance through concentration modifications across a wide range as compared to the host-guest systems. The overall design of these systems thus makes them resistant towards dilution and competitive salt binding, making them suitable for use in biofluids²⁷⁰

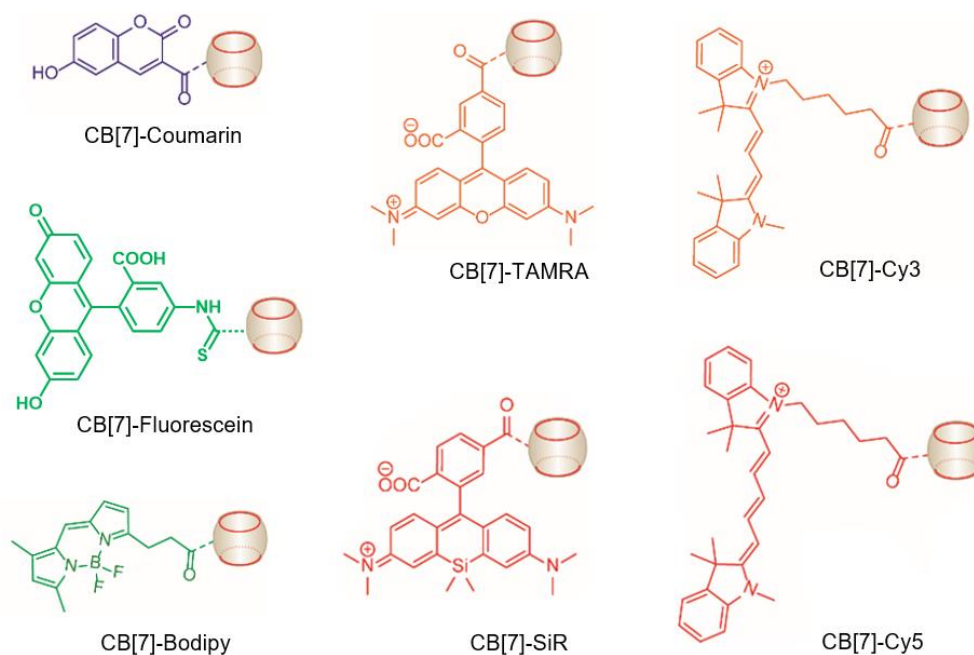


Figure 7.1: CB[7]-FL encoded library consisting of seven different fluorophores (Coumarin, Fluorescein, Bodipy, TAMRA, Cy3, Cy5 and SiR) covalently conjugated with CB[7] as the sensor elements of the macrocyclic sensor array (adapted from Agasti et al.²¹⁷).

7.1.3. Background of developed CB[7]-fluorophore sensor array

Agasti et al.,²¹⁷ have developed one such covalently-linked sensing array with CB[7] covalently tethered with a library of fluorescent reporters. The recognition of the chemical and topological features of biomolecules by CB[7] brings the tethered fluorophore into close proximity, thereby reporting the nature of its microenvironment in the presence of analytes through changes in the optical signature of the fluorophore. To achieve this design, the group synthesized a library of seven covalently conjugated CB[7]-FL reporter probes for fingerprinting biomolecular surfaces (Coumarin, Fluorescein, Bodipy, TAMRA, Cy3, Cy5 and SiR) (Figure 7.1). Amine-functionalized CB[7] was synthesized in a two-step

process from monohydroxylated CB[7] and then reacted with N-hydroxysuccinimide ester derivatives of the selected dyes to obtain a single point of conjugation of CB[7] with a yield of 25-54%.

The orthogonal mechanisms driving the photophysical changes of the different fluorophore conjugates were also probed in this study. Cy3 exhibits a restricted trans-to-cis isomerization mechanism, while SiR is influenced by a ground-state deactivation process, coupled with its nature to exist in equilibrium between a non-fluorescent spirolactone form and a fluorescent zwitterionic form. Cy5 was projected to undergo variations similar to Cy3, based on its trans-to-cis photoisomerization, alongside the additional possible formation of electronically coupled dye aggregates. TAMRA and Bodipy are expected to be influenced by the quenching effects of neighboring amino acids (Trp, Tyr, His) via static and collisional mechanisms with the dye molecule, while coumarin is susceptible to pH or static quenching mechanisms by surrounding amino acids. Consequently, the environmental sensitivity of the fluorophores in the vicinity of proteins provided this study a powerful strategy in interrogating their surfaces. The conjugated fluorophore system has been utilized to fingerprint diverse protein analytes as well as analyze the changes in the structural features for a given protein. Additionally, the energy landscape of protein folding and aggregation pathway, including native, aggregated, misfolded and fibrillary states have been fingerprinted. Importantly, the sensor has demonstrated the ability to discriminate between diverse amyloid structures of the A β 42 peptide associated with Alzheimer's disease from other protein sequences with 100% accuracy. The sensor was ultimately assessed for its capacity to discriminate clinically relevant information from patient blood samples, fingerprinting the serum of pregnant women vs. controls as a proof-of-principle study to demonstrate the clinical capability of the CB[7]-FL sensor array.

7.1.4. Diagnostic application of CB[7]-FL sensor array

Leveraging the recognition properties of the CB[7]-FL moieties, coupled with their compatibility with biological media facilitates their utilization for analyte detection in complex media like blood serum. Considering the earlier study and the documented alterations in the blood biochemistry among SARS-CoV-2 infected patients, we proposed that this CB[7]-FL sensor array could also be efficient in differentiating clinical serum samples from patients infected or not by SARS-CoV-2, with clinically relevant accuracy. Serum samples from pregnant women were obtained from the COVIPREG study conducted in nine maternities in the Paris area during the first two waves of the COVID-19 pandemic (04/28/2020 to 01/13/2021) before vaccines were available. The primary objective of this study was to evaluate SARS-CoV-2 excretion at maternal sites during delivery and postpartum, and to assess materno-fetal transmission following the guidelines of the World Health Organization (WHO). We selected serum samples from 12 SARS-CoV-2 infected women, confirmed by nasopharyngeal RT-PCR, and 14 healthy controls. To minimize inter-individual variability and emphasize blood composition changes due to SARS-CoV-2 infection, we selected women in their third trimester of pregnancy ($36 \pm$

4.3 weeks' gestation (WG) for the infected group and 37.4 ± 4.3 WG for the control group) without other known disease conditions. Among the infected group, 66.5% (8 of 12) were symptomatic with breathing difficulties, and 17% (2 of 12) required hospitalization.

The methodology and feasibility of the CB[7]-FL array-based sensing approach were thoroughly evaluated and documented through the analysis of fluorescence signals and fingerprints obtained from the small set of serum samples from pregnant women, both infected and not infected with SARS-CoV-2. This study not only provides an additional proof-of-concept for potential clinical diagnosis using the CB[7]-FL array but also offers a comprehensive, step-by-step methodology and a ready-to-use protocol for designing experiments, conducting measurements, and performing data extraction, processing, and analysis. The approach emphasizes a deeper understanding of the chemical nose strategy, encouraging a move beyond the use of general black-box statistical tools. This detailed methodology is particularly significant as it lays the groundwork for researchers to replicate and expand upon the study. Notably, the development of a pipeline of R-codes has streamlined the data processing, curing, formatting, and analysis directly from the raw fluorescence fingerprints, making the approach more accessible and user-friendly. Currently, the exploration of the chemical nose strategy for clinical diagnosis is hampered by the limited access to large sample cohorts and the challenges associated with big data management. These obstacles pose significant difficulties for researchers and clinicians in the field, restricting the broader application of this innovative approach. However, the extended applicability and utility of this analysis protocol could serve as a valuable tool, enabling more collaborative studies between researchers and clinicians. By facilitating the assessment of this hypothesis-free chemical diagnosis method with larger and more diverse cohorts, this protocol has the potential to advance the field significantly and contribute to the diagnosis of multifactorial disease.


 Cite this: *Sens. Diagn.*, 2024, 3, 863

A methodological study for the diagnosis of the SARS-CoV-2 infection in human serum with a macrocyclic sensor array†

 Monica Swetha Bosco,^{ab} Zeki Topçu,^d Soumen Pradhan,^{fg} Ariadne Sossah,^a
 Vassilis Tsatsaris,^{ac} Christelle Vauloup-Fellous,^e Sarit S. Agasti,^{id fg}
 Yves Rozenholc^{‡d} and Nathalie Gagey-Eilstein^{id ‡*ab}

This article reports the methodology and the proof of concept of a blood-based diagnostic strategy for the SARS-CoV-2 infection. The proposed method relies on the non-specific/selective array-based sensing strategy mimicking the human olfactory system using a cucurbit[7]uril macrocycle receptor conjugated with a library of environmentally sensitive fluorophores. The study cohort includes 26 samples, *i.e.* 12 cases and 14 controls. Statistical analysis methods such as linear discriminant and random forest were able to successfully classify and discriminate the two groups with almost 90% accuracy. This diagnostic result highlights the methodology and confirms the potential of this non-specific/selective sensing approach for non-invasive clinical diagnosis.

 Received 8th January 2024,
 Accepted 1st April 2024

DOI: 10.1039/d4sd00009a

rsc.li/sensors

1. Introduction

Since the end of 2019, the COVID-19 disease has been an ongoing threat spreading worldwide, with more than 600 million infections and more than 6 million deaths.¹ On March 11, 2020, it was classified as a global pandemic by the World Health Organization (WHO). Therefore, its diagnosis is considered indispensable to prevent and control the spreading of the disease. Currently, two types of diagnostic tests, based on nasopharyngeal or nasal swabs, are widely used: i) nucleic acid amplification tests based on polymerase chain reaction (PCR) technology that detect viral RNA; ii) antigen tests based on lateral flow immunoassays (LFAs) that detect viral proteins

(*i.e.* spike, envelope or nucleotide proteins). The accuracy, specificity, and sensitivity of a test as well as the time between the test and results are the main criteria to control the spreading of the disease. Antigen detection can provide results within 15 min with low accuracy and is subject to delivering false negatives, particularly when used in people with no signs or symptoms of the infection (up to 45% of false negatives).² This drawback results in high risk of virus dissemination. Consequently, PCR tests are often used for the confirmation of antigen tests and, therefore, might not suffer from false negatives. Even if the analytical performance of PCR tests approaches 100% by detecting 500–5000 RNA copies per mL, clinical performance approaches only 80% sensitivity due to biological and pre-analytical factors, particularly sample collection. Nasal and nasopharyngeal swabs are partly responsible for false negatives since the quality of specimen collection may be low and viral loads in the sample are neither homogeneous nor stable within the time of infection. Moreover, the delay between sampling and results can extend from 24 to 48 h due to time-consuming laboratory procedures that require certified laboratories, trained operators, and expensive equipment.³ While nasopharyngeal swabs are still the widely used specimen, many works explore diagnosis methods based on specimens from other types of samples such as the upper respiratory tract (throat and deep throat saliva), lower respiratory tract (sputum and bronchoalveolar lavage fluid), nasopharynx, feces, and blood.^{4–11} Herein, we propose a diagnostic strategy for the SARS-CoV-2 infection based on blood samples whose sampling method, stability, and homogeneity are highly controlled.

^a Université Paris Cité, INSERM UMR-S 1139, FHU PREMA, 4 Avenue de l'Observatoire, 75006 Paris, France. E-mail: nathalie.eilstein@u-paris.fr

^b Université Paris Cité, CNRS, INSERM, UTCBS, 4 Avenue de l'Observatoire, 75006 Paris, France

^c Department of Obstetric, Cochin Hospital, AP-HP, Université Paris Cité, FHU PREMA, 123 Bd Port-Royal, 75014 Paris, France

^d UR 7537, BioSTM, Université Paris Cité, 4 Avenue de l'Observatoire, 75006 Paris, France

^e Virology Laboratory, Hôpital Paul-Brousse, AP-HP, Université Paris-Saclay, INSERM U1193, Villejuif, France

^f New Chemistry Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, Karnataka 560064, India

^g Chemistry & Physics of Materials Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore, Karnataka 560064, India

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4sd00009a>

‡ These two authors contributed equally to the work.



Blood serologic tests to detect host-derived antibodies against viruses (IgM and IgG) are not efficient for diagnosing acute COVID-19 but rather previous infection and/or vaccination. For initial diagnoses based on viral RNA detection, the US-CDC does not recommend the use of blood samples.¹² However, SARS-CoV-2 infection is still responsible for pronounced changes in the blood composition. Many reviews investigate routine biochemical, immunological but also inflammatory or nucleic acid biomarkers as a promising avenue for early diagnosis and prediction of prognosis of COVID-19.^{13–16} A proteomic and metabolomic study found that 105 proteins and 373 metabolites were differentially expressed in the sera of COVID-19 patients compared to the control.¹⁷ However, no specific blood biomarker, even built as a combination of expressions, has been still clinically validated so far. Diagnostic strategy using proteomic or/and metabolomic signatures would have been too expensive to build for routine purpose. Routine blood parameters such as hematological (lymphocyte and neutrophil count), inflammatory (C-reactive protein), biochemical like (albumin, lactate dehydrogenase, alanine and aspartate aminotransferase and alkaline phosphatase) have been described as dysregulated in SARS-Cov-2 infected patients.¹⁸ Combining appropriate cutoffs for certain of these blood parameters could help in identifying COVID-19 positive patients but with a low accuracy.³ Also, when used in a machine learning model, they can help to differentiate the status of patients with 82–86% accuracy and 92–95% sensitivity.¹⁹ However, this strategy requires many blood parameters to be measured and therefore remains slow and expensive if transposed at the population level. Hence, the “Chemical Nose” hypothesis-free machine learning diagnostic strategy looking for global blood change composition rather than changes of some specific biomarkers or a combination of those will be suitable for a low-cost blood-based diagnostic strategy.

The “Chemical Nose” strategy mimics the human olfactory system using a set of non-specific sensors to sense the components of a simple or complex mixture, whose outputs are the inputs of one machine learning algorithm. Each sensor selectively binds to sample analytes based on electrostatic, hydrophobic, H-bonding or host-guest interactions. Recognition event between the sensor and the analyte is translated into an optical signal through the transduction element. It generates a pattern of outputs (fingerprint) for each analyte. Finally, the patterns are classified (unsupervised approach) and eventually identified (supervised approach) using one machine learning algorithm.^{20–22} This “Chemical Nose” strategy has been successfully applied in “fingerprinting” pure proteins, protein folding states or cancerous cell lines among others, thus allowing their classification and identification.^{23–27} It also has the potential to be developed for clinical diagnostics. Indeed, since the onset, progression and outcomes of a disease modify the blood composition,²⁸ sensing and fingerprinting blood samples with the set of non-specific

sensors would allow samples classification/identification with regard to their physiological or pathological status. Using a set of fluorophore conjugated polymers, Rotello *et al.* described the classification of healthy, mild or severe liver fibrosis patients, from a cohort of blood samples, with clinically relevant specificity and accuracy ([ROC-AUC] = 0.89).²⁹ A FRET-based polymer sensor array has also been used to discriminate cancerous or healthy mice by serum fingerprinting.³⁰ Recently, as a proof of concept, we demonstrated the use of a set of fluorophore-conjugated cucurbit[7]uril (CB[7]-FL) to discriminate the serum of pregnant women vs. the control.²⁵ Using a hydrophobic cavity with orthogonal H-bonding and electro-static/dipolar interaction utilizing two symmetry-equivalent uridyl carbonyl portals, CB[7] is an interesting receptor scaffold for examining biomolecules. Not only does it possess host-guest binding property toward a wide range of guest molecules but the H-bonding and electro-static/dipolar recognition elements provides an additional target probing mechanism.^{31–33} Moreover, these CB[7]-FL moieties offer a good compatibility with biological media due to good water solubility and strong fluorescence, easily detected in complex media.

Given the reported blood biochemistry changes in blood samples of SARS-CoV-2 infected patients, we hypothesized that this CB[7]-FL sensor array could be efficient in differentiating serum samples from patients infected or not infected by SARS-CoV-2, with clinically relevant accuracy. Herein, we propose, not only an additional proof-of concept for potential clinical diagnosis but, above all, a step-by step precise methodology to use this CB[7]-FL array-based sensor to monitor and analyse the fluorescence signals and fingerprints obtained from a small set of serum samples of pregnant women infected or not infected by SARS-CoV-2 towards each CB[7]-FL (Scheme 1). This work can be of high interest for researchers and clinicians who are interested in the assessment of hypothesis-free chemical diagnosis for clinical diagnosis with cohorts of large size.

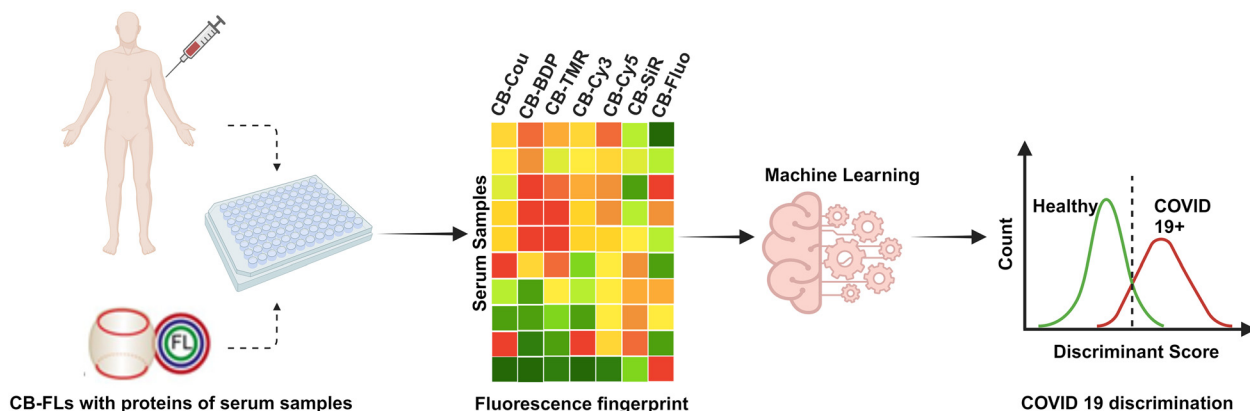
2. Material and methods

2.1. Participants and samples

Cohorts. All blood samples were collected in the frame of the COVIPREG study,³⁴ with approval from the national ethics comity, the CPP SUD MEDITERRANEE (No. 2020-A00924-35) on April 23rd, 2020. The trial is recorded in the clinical trial registry as NCT04355234. COVIPREG study is a prospective study conducted in France in 9 maternities of Paris area during the first two waves of COVID-19 pandemic (inclusion between 04/28/2020 and 01/13/2021) and before vaccine availability. All patients have signed a consent form to the use of their serums for research purposes. Serum samples were collected in pregnant women the same day they declared a positive SARS-CoV-2 nasopharyngeal RT-PCR.

Protocol. After centrifugation of the blood (1000 × *g* for 15 min at 4 °C), supernatant was transferred into a clean





Scheme 1 CB[7]-FL encoded library for fingerprinting serum samples of pregnant women infected or not infected by SARS-CoV-2. Analysis of the obtained serum-based fingerprints via machine learning algorithms to enable the differentiation of disease or healthy states.

polypropylene tube, aliquoted, frozen within 48 hours and stored at $-80\text{ }^{\circ}\text{C}$ until analysis. The control serum samples were used after one year of conservation in accordance with French law.

Subsampling. Our subsampling of the cohort is composed of 26 samples. The case group is composed of 12 samples whereas the control group is composed of 14 samples, collected from pregnant women in the pre-pandemic period (January 2019). Case vs. control were chosen to be aligned on gestational age. The baseline characteristics of the patients are provided in Table 1 and detailed characteristics for each patient are given in Table S1.†

2.2. CB[7]-FL synthesis

Conjugated cucurbit[7]urils were synthesized as described in previous publications.^{25,35}

2.3. Array methodology

100 nM solution of each CB[7]-FL was freshly prepared in 10 mM sodium phosphate buffer (pH 7.4). The concentration of the stock solution is controlled by Beer-Lambert law before further dilution. In the wells of a white, half-well, non-binding-surface plate (Corning®, Product Number: 3642), 25 μL of the desired CB[7]-FL solution was added. Fluorescence emission at the respective emission wavelength in each well ($I_{(S-)}$) was recorded on a microplate reader (PerkinELmer®, EnSpire) by exciting the fluorophores at their respective excitation wavelength. Emission and excitation wavelengths

for each CB[7]-FL are given in Table S2.† Next, 5 μL of pure serum was added to each fluorophore solutions, except in the control wells where 5 μL of PBS were added instead. Subsequently, after orbital stirring in the microplate reader and temperature control ($25\text{ }^{\circ}\text{C}$), fluorescence emission ($I_{(S+)}$ for sample wells and $I_{(PBS+)}$ for control wells) was recorded using the same excitation/emission wavelengths. Six consecutive measurements (each 5 min apart) were taken. Experiments were performed as replicates of four for each serum sample.

2.4. Data and feature extraction

R codes were written to 1/ read and arrange the raw data in a .csv file; 2/ build the “before and after serum addition” variation of fluorescence tables for each CB[7]-FL and 3/ combine these tables into one dataset made of 26 samples (12 cases and 14 controls) and 8 variables (the sample status and the 7 CB[7]-FL features with FL = Cy3, Cy5, TMR, SiR, Coum, Fluo, Bdp). These codes can be found in Annex 1 of ESI.† They are named respectively ‘lecture-raw-data.R’, ‘read-plate-data.R’ and ‘build-data.R’.

2.5. Data statistical analysis

The 7 variables for each sample in the two groups were compared and analyzed using different strategies of statistical analysis. Homemade R code (<https://cran.r-project.org>) was used and can be found in Annex 2 of ESI† (‘Covid_analysis.R’).

Table 1 Baseline clinical and pathological characteristics of the infected and control groups

	Confirmed infected group	Control group
<i>n</i>	12	14
Gestational age		
Mean (std)	36.5 (4.3)	37.4 (3.4)
Median (range)	38.5 (28–40)	38 (28–41)
Days of infection before sampling <i>N</i> (std)	6.08 (6.05)	Not applicable
Breathing symptoms	8 yes, 4 no	Not applicable
Hospitalization with oxygen and ICU	2 yes, 10 no	Not applicable



3. Results and discussions

The hypothesis-free approach using the CB[7]-FLs sensor array was exploratorily tested to determine whether it could “fingerprint” SARS-Cov2 infection in serum sample and later provide a potentially clinically PoC relevant assay. To this aim, we explored this sensor array-based diagnosis strategy by assessing a precise methodology, including the choice of clinical samples, careful design of experiments, observation and calculation of the fluorescence change, automatized data extraction and organization as well as statistical method comparison for serum samples' discrimination.

3.1. Clinical samples

Serum samples of pregnant women from the COVIPREG cohort³⁴ were used, among which we selected 12 infected by SARS-Cov-2 and 14 healthy controls. For the proof-of-concept of our strategy and since we do not have access to a large number of samples, we selected patients showing good clinical homogeneity in order to reduce inter-individual variability and therefore highlight variations in the blood composition, which could be expected from SARS-Cov2-infection. Therefore, we selected samples from women that were not known to suffer from other diseases and that were in their third trimester of pregnancy (36 ± 4.3 week of gestation (WG) for the infected group and 37.4 ± 4.3 of WG for the control group). Among the SARS-Cov2 infected group, 66.5% (8 over 12) were symptomatic with breathing symptoms and 17% (2 over 12) needed hospitalization. Clinical and biological data of the 26 patients of the two groups are presented in Table 1 (see also Table S1† for individual data).

3.2. Design of experiments

To run the sensing experiment, each of the seven element of the CB[7]-FL sensor array were mixed individually towards each serum sample in the wells of a microplate.

Initially, the complete emission spectra of each CB[7]-FL was recorded on the addition of serum and serum spiked with biologically relevant proteins (1.6 mg mL^{-1}). In some cases, significant fluorescence decrease was observed, with a more pronounced sensitivity of BDP, TMR, Cy5 and SiR. These modifications highlight the ability of CB[7]-FLs to detect subtle changes in complex matrix such as serum (Fig. S1†).

The choice of a half-well microplate has been made to minimize the needed volume of CB[7]-FL as well as serum. Indeed, with seven CB[7]-FLs to be tested, with measurements in quadruplicate, and with 26 sera to assess, the needed volume of each serum is $140 \mu\text{L}$ and the volume of 100 nM solution of CB-FL was more than 2.5 mL . The concentration of 100 nM has been fixed after preliminary studies and provide sufficient fluorescent signal, even after addition of the serum. For reproducibility concerns, this concentration has to be carefully controlled before

experiments.²⁵ Non-binding microplates were preferably used to avoid the adsorption of hydrophobic CB[7] on the well surface. To assess this phenomena, six consecutive measurements (each 5 min apart) were taken. The signal was stable overtime, as shown in Fig. S3.† Therefore, the mean of the six measurements overtime was used for the analysis. To test all the samples in quadruplicate, two microplates were necessary for each CB[7]-FL. To avoid an artificial discrimination due to samples position on the microplate, COVID+ and control samples were positioned on both plates, as shown in Fig. S4.†

3.3. Calculation of fluorescence change

Fluorescence signals of the fluorophore (FL) registered in quadruplicate before and after serum addition constitutes the output, and combining all the FL fluorescence changes later provides the fluorescence pattern for each sample. For each sample and each CB[7]-FL, the output, $\overline{\Delta I}$, is calculated as the mean over the quadruplicate of the changes in fluorescence at emission wavelength after and before serum addition in each well as given by the formula

$$\begin{aligned}\overline{\Delta I} &= \text{mean}(\Delta I) = \text{mean}(I_{(s+)} - \text{mean}(I_{(s-)})) \\ &= \text{mean}(I_{(s+)}) - \text{mean}(I_{(s-)})\end{aligned}$$

where $I_{(s+)}$ is the fluorescence intensity in each well after addition of serum, $I_{(s-)}$ is the fluorescence intensity before addition of serum in the corresponding well, and $\text{mean}(I_{(s-)})$ is the mean of the fluorescence intensities of the quadruplicate corresponding well before addition of serum. The generated data table is given in the ESI† (Table S3).

Obviously, it is impossible to discriminate two groups of samples (COVID+ vs. control) only by looking at the fingerprint on the heatmap plot (Fig. 1a). However, we observe that the mean of CB[7]-Cy5 for each of the two groups are significantly different ($p < 0.05$) (Fig. 1b). Individual values for each serum in each group are also plotted on Fig. 1c and S5.† These graphs offer a better visualization of the fluorescence variation and particularly differences between the two groups for CB[7]-Cy5. We can also observe the good reproducibility between the four replicates.

3.4. Statistical method for serum samples discrimination

Based on the data of fluorescence variations for each CB[7]-FL (Table S3†), statistical analysis methods have been evaluated towards the discrimination of the two groups of samples.

3.4.1. Random Forest analysis. Initially, Random Forest was used. Due to the low number of samples, splitting the data into train and test subset of data was not pertinent.

Using all features (7 CB[7]-FL), the classification rate for the 26 samples (12 COVID+ and 14 COVID-) evaluated using the out-of-the-bag strategy is 84.6%, with a predictive positive value of 74.5% and a negative predictive value of 92.9%.



Thanks to the use of Random Forest, we get an estimation of the feature importance in the classification: CB[7]-Cy5 brings

the highest contribution, followed by CB[7]-SiR and CB[7]-TMR (Fig. 2a).

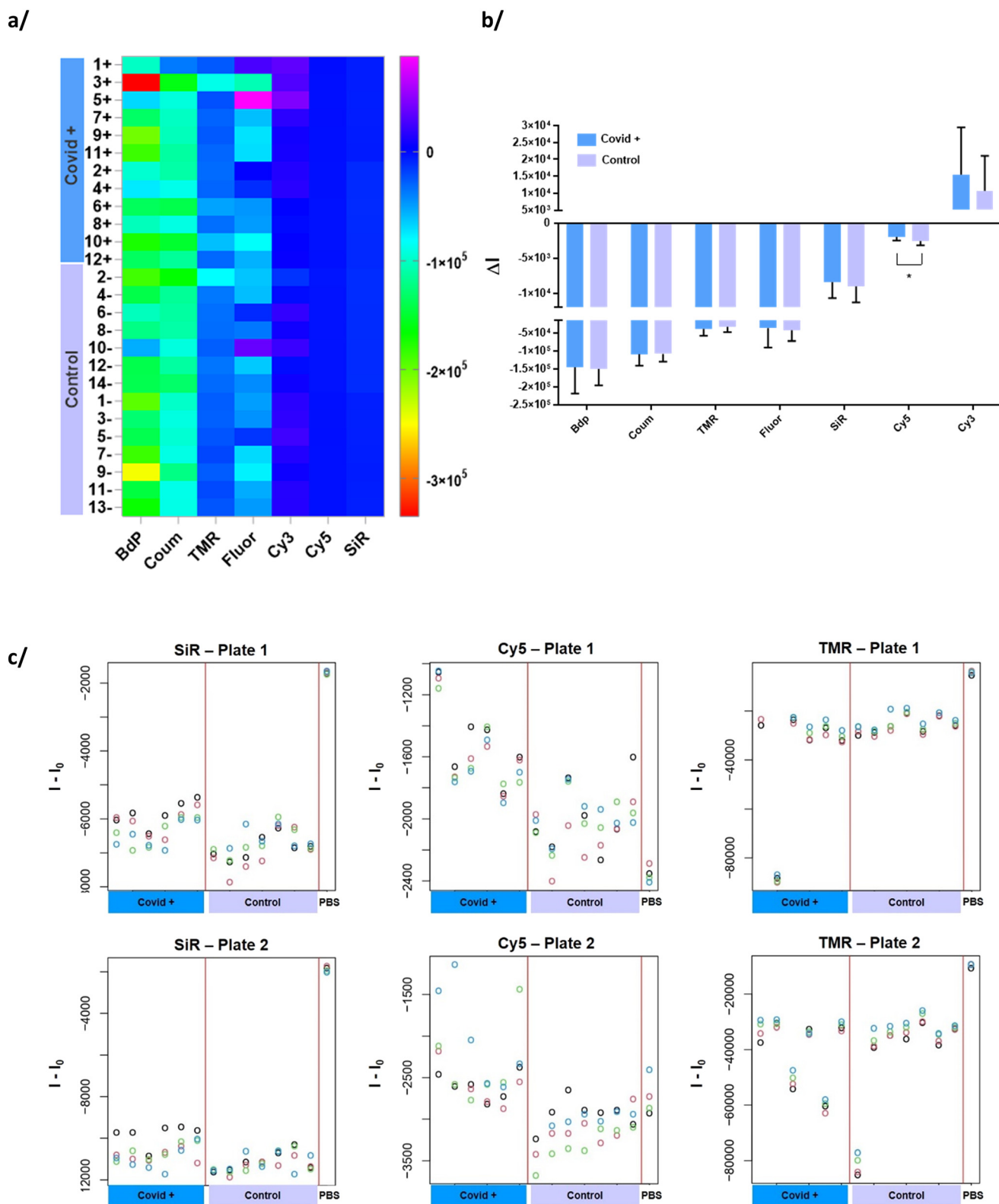


Fig. 1 a/ Heatmap plot fingerprinting each serum sample. b/ Means of fluorescence response of each CB[7]-FL against the two groups COVID+ (blue) and control (light purple). The symbol * indicates p value for the t -test used for comparison of the means smaller than 0.05. c/ Individual ΔI value (y axis) for each replicate (black, red, green and blue circles) of each serum and each CB[7]-FL for the two plates and dispatched such as 1–6: COVID samples, 7–13: control samples, 14: control well.



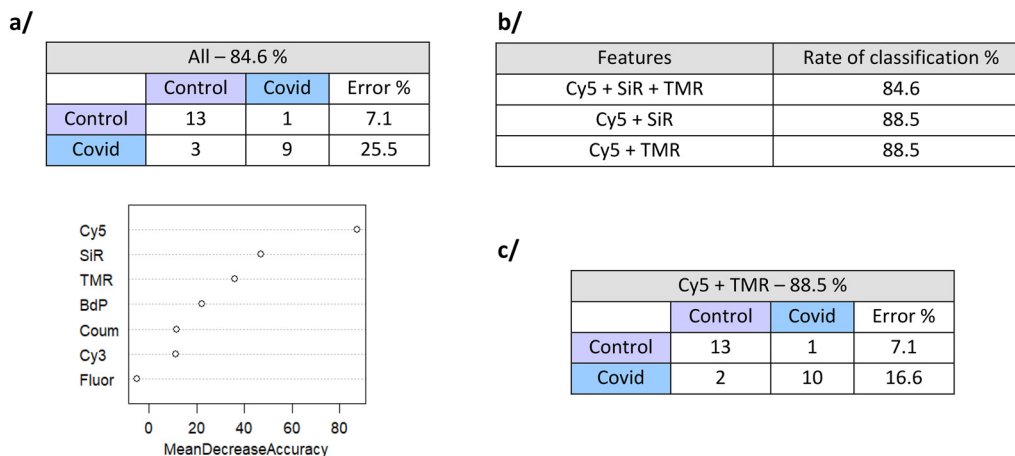


Fig. 2 a/ Confusion matrix for random forest on the whole variables and importance of each feature in the random forest algorithm of discrimination. b/ Error rate of discrimination for random forest run with the most important variables. c/ Confusion matrix for the random forest runs with Cy5 and SiR features.

We next used Random Forest with combinations of these 3 most important features. It appears that the couple of features Cy5 and SiR or Cy5 and TMR offers estimated classification rate of 88.5% of error of discrimination and a predictive positive value of 83.4% and a negative predictive value of 92.9% (Fig. 2b and c). Stabilizations of the error rates with the number of trees are shown in Fig. S6.†

As an additional evidence toward the use of this type of non-specific sensor array for clinical sample classification, we also combined this dataset with the previous dataset discriminating pregnant and non-pregnant samples.²⁵ Pregnant samples from these two different data sets are discriminated with 97.1% of discrimination. Then, the samples of pregnant women infected by SARS-COV-2 were tested along with the pregnant and non-pregnant groups. When keeping the most important features (Cy5, Bdp, SiR, TMR, Cy3), only 6.9% of error of discrimination was obtained with a positive predictive value of 75% (Fig. S6 and Table S4†). These results provide additional evidence that the

sensor array strategy may be suitable for clinical diagnosis even if it is, of course, still mandatory to have access to big size clinical cohorts.

LDA analysis. We also used linear discriminant analysis (LDA) model combined with leave-one-out strategy for discrimination rate estimation. The histogram showing the LDA scores distribution within the two groups is given in Fig. S8.† When using all the features, the two groups can be discriminated with 77% accuracy (to compare with the 84.6% of the Random Forest) (Fig. 3a). Again, coefficients of linear discriminant analysis confirmed that Cy5, SiR and TMR have the highest weight in the discrimination. Using these three variables, the accuracy of discrimination is improved to 85% (88.5% for Random Forest) (see Fig. 3b). As for the Random Forest, the predictive positive value is better than the predictive negative value. We carefully looked to the misassigned COVID+ samples and tried to correlate the results to clinical data from Table S1.† However, no correlation was observed whatever the considered clinical data (Fig. 3c).

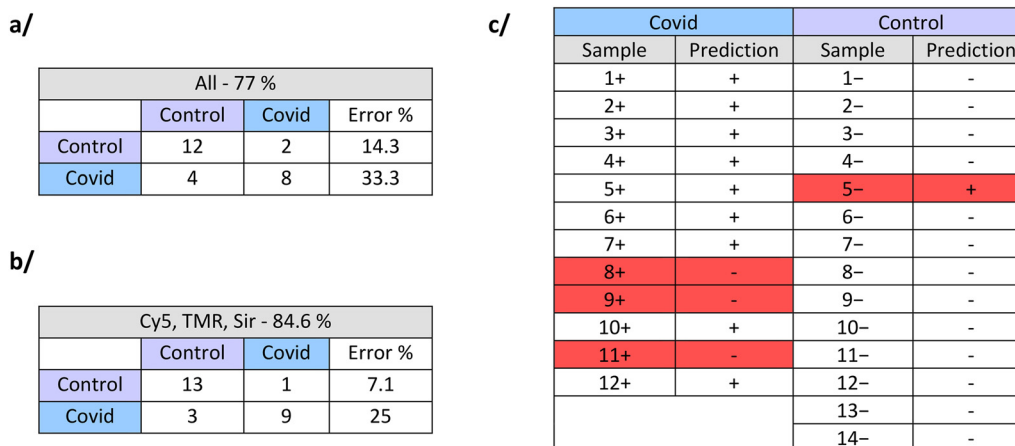


Fig. 3 a/ Confusion matrix for LDA runs with the whole variables and b/ runs with the most important variables. c/ Prediction table for individual samples. Misclassified samples are highlighted in red.



4. Conclusion

With a small case-control cohort of SARS-CoV-2 infected serum samples, this work highlights the strategy of training a non-specific sensing array to discriminate clinical samples. It offers the proper methodology, including the design of experiments, data extraction, treatment and analysis to run these types of chemical nose experiments and alert the readers to the particular points of attention to take care of when using chemical nose sensor array. From a statistical point-of-view, even if with the small size of our cohort, we demonstrate that a such strategy is efficient to discriminate SARS-CoV-2 infected samples from control samples. By implementing a step-by-step protocol, the methodology can be easily adapted to extract and take into account relevant experimental variables on large cohorts. Linear discriminant analysis clearly shows some lack of performance; however, more machine learning oriented strategy such as Random Forest seems to be well adapted to implement a strategy of training a non-specific sensing array, thanks to its flexibility, the out-of-the-bag error rate estimation, which does not require a train/test approach, and opportunity to retrieve feature (sensor) importance. Despite the higher expected variability in larger cohort, one can anticipate that these properties of the Random Forest will help to keep high discrimination performances. The available R-codes to extract and process from raw data is made available and can be easily tuned to new clinical samples. From a chemical point-of-view, our first results shows the promising efficacy of the macrocyclic-based non-specific CB[7]-FL sensor array to discriminate the two health states between SARS-CoV-2 infected serum and control and that such an approach may constitute a strategy of diagnosis based on blood-serum samples. Interestingly, starting from a fixed size sensor array, using the knowledge regarding importance of each feature could ultimately lead to the reduction of the sensor array dimension, providing a cheaper and easier point-of-care test. From a clinical point-of-view, we can point out that two (over the 3) false negative patients were asymptomatic (see Fig. 3c) and therefore wonder about the result of the SARS-Cov2 PCR test that may be subjected to some error. To conclude, we would like to point out that within a larger and clinically well-documented cohort, the proposed strategy could be used not only to diagnose the clinical binary status but also more complex disease outcomes like severity and eventual complications.

Ethical statement

All experiments were performed in accordance with ethical standards. All blood samples were collected with approval from the national ethics comity CPP SUD MEDITERRANEE (No. 2020-A00924-35) on April 23rd, 2020. The trial is recorded in the clinical trial registry as NCT04355234. Informed consents were obtained from human participants of this study.

Sample CRediT author statement

Nathalie Gagey-Eilstein: conceptualization, methodology, supervision, writing – original draft preparation, funding acquisition. Yves Rozenholc: data curation, formal analysis, software, supervision. Sarit S. Agasti: resources. Christelle Vauloup-Fellous, Vassilis Tsatsaris: resources, funding acquisition. Monica Swetha Bosco, Ariadne Sossah, Soumen Pradhan: investigation. Zeki Topcu: investigation, data curation, formal analysis.

Author contributions

N. G. E. and Y. R. contributed equally to the work.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research has received funding from the FHU Prema foundation to N. G. E., V. T. and C. V. F. and from the Agence Nationale pour la Recherche (ANR-20-CE17-0035) to N. G. E., Y. R. and V. T. SERB support (CVD/2020/000855) to S. S. A. is also acknowledged. COVIPREG study: sponsor was Assistance Publique – Hôpitaux de Paris (Délégation à la Recherche Clinique et à l'Innovation).

References

- 1 Coronavirus, <https://www.who.int/health-topics/coronavirus>, (accessed 2023-09-19).
- 2 R. W. Peeling, D. L. Heymann, Y.-Y. Teo and P. J. Garcia, Diagnostics for COVID-19: Moving from Pandemic Response to Control, *Lancet*, 2022, **399**(10326), 757–768, DOI: [10.1016/S0140-6736\(21\)02346-1](https://doi.org/10.1016/S0140-6736(21)02346-1).
- 3 D. Ferrari, A. Motta, M. Strollo, G. Banfi and M. Locatelli, Routine Blood Tests as a Potential Diagnostic Tool for COVID-19, *Clin. Chem. Lab. Med.*, 2020, **58**(7), 1095–1099, DOI: [10.1515/cclm-2020-0398](https://doi.org/10.1515/cclm-2020-0398).
- 4 R. F. Thudium, M. P. Stoico, E. Høgdall, J. Høgh, H. B. Krarup, M. A. H. Larsen, P. H. Madsen, S. D. Nielsen, S. R. Ostrowski, A. Palombini, D. B. Rasmussen and N. T. Foged, Early Laboratory Diagnosis of COVID-19 by Antigen Detection in Blood Samples of the SARS-CoV-2 Nucleocapsid Protein, *J. Clin. Microbiol.*, 2021, **59**(10), e01001-21, DOI: [10.1128/JCM.01001-21](https://doi.org/10.1128/JCM.01001-21).
- 5 A. Townsend, P. Rijal, J. Xiao, T. K. Tan, K.-Y. A. Huang, L. Schimanski, J. Huo, N. Gupta, R. Rahikainen, P. C. Matthews, D. Crook, S. Hoosdally, S. Dunachie, E. Barnes, T. Street, C. P. Conlon, J. Frater, C. V. Arancibia-Cárcamo, J. Rudkin, N. Stoesser, F. Karpe, M. Neville, R. Ploeg, M. Oliveira, D. J. Roberts, A. A. Lamikanra, H. P. Tsang, A. Bown, R. Vipond, A. J. Mentzer, J. C. Knight, A. J. Kwok, G. R. Screaton, J. Mongkolsapaya, W. Dejnirattisai, P. Supasa, P. Klenerman, C. Dold, J. K. Baillie, S. C. Moore,



- P. J. M. Openshaw, M. G. Semple, L. C. W. Turtle, M. Ainsworth, A. Allcock, S. Beer, S. Bibi, D. Skelly, L. Stafford, K. Jeffrey, D. O'Donnell, E. Clutterbuck, A. Espinosa, M. Mendoza, D. Georgiou, T. Lockett, J. Martinez, E. Perez, V. Gallardo Sanchez, G. Scozzafava, A. Sobrinodiaz, H. Thraves and E. Joly, A Haemagglutination Test for Rapid Detection of Antibodies to SARS-CoV-2, *Nat. Commun.*, 2021, **12**(1), 1951, DOI: [10.1038/s41467-021-22045-y](https://doi.org/10.1038/s41467-021-22045-y).
- 6 S. S. Adigal, N. V. Rayaroth, R. V. John, K. M. Pai, S. Bhandari, A. K. Mohapatra, J. Lukose, A. Patil, A. Bankapur and S. Chidangil, A Review on Human Body Fluids for the Diagnosis of Viral Infections: Scope for Rapid Detection of COVID-19, *Expert Rev. Mol. Diagn.*, 2021, **21**(1), 31–42, DOI: [10.1080/14737159.2021.1874355](https://doi.org/10.1080/14737159.2021.1874355).
- 7 W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu and W. Tan, Detection of SARS-CoV-2 in Different Types of Clinical Specimens, *JAMA, J. Am. Med. Assoc.*, 2020, **323**(18), 1843–1844, DOI: [10.1001/jama.2020.3786](https://doi.org/10.1001/jama.2020.3786).
- 8 R. Samson, G. R. Navale and M. S. Dharme, Biosensors: Frontiers in Rapid Detection of COVID-19, *3 Biotech*, 2020, **10**(9), 385, DOI: [10.1007/s13205-020-02369-0](https://doi.org/10.1007/s13205-020-02369-0).
- 9 P. Pokhrel, C. Hu and H. Mao, Detecting the Coronavirus (COVID-19), *ACS Sens.*, 2020, **5**(8), 2283–2296, DOI: [10.1021/acssensors.0c01153](https://doi.org/10.1021/acssensors.0c01153).
- 10 G. Giovannini, H. Haick and D. Garoli, Detecting COVID-19 from Breath: A Game Changer for a Big Challenge, *ACS Sens.*, 2021, **6**(4), 1408–1417, DOI: [10.1021/acssensors.1c00312](https://doi.org/10.1021/acssensors.1c00312).
- 11 B. Shan, Y. Y. Broza, W. Li, Y. Wang, S. Wu, Z. Liu, J. Wang, S. Gui, L. Wang, Z. Zhang, W. Liu, S. Zhou, W. Jin, Q. Zhang, D. Hu, L. Lin, Q. Zhang, W. Li, J. Wang, H. Liu, Y. Pan and H. Haick, Multiplexed Nanomaterial-Based Sensor Array for Detection of COVID-19 in Exhaled Breath, *ACS Nano*, 2020, **14**(9), 12125–12132, DOI: [10.1021/acsnano.0c05657](https://doi.org/10.1021/acsnano.0c05657).
- 12 T. Kilic, R. Weissleder and H. Lee, Molecular and Immunological Diagnostic Tests of COVID-19: Current Status and Challenges, *iScience*, 2020, **23**(8), 101406, DOI: [10.1016/j.isci.2020.101406](https://doi.org/10.1016/j.isci.2020.101406).
- 13 R. Mittal, N. Chourasia, V. K. Bharti, S. Singh, P. Sarkar, A. Agrawal, A. Ghosh, R. Pal, J. R. Kanwar and A. Kotnis, Blood-Based Biomarkers for Diagnosis, Prognosis, and Severity Prediction of COVID-19: Opportunities and Challenges, *J. Family Med. Prim. Care*, 2022, **11**(8), 4330–4341, DOI: [10.4103/jfmpc.jfmpc_2283_21](https://doi.org/10.4103/jfmpc.jfmpc_2283_21).
- 14 COVID-19 Multi-omics Blood Atlas (COMBAT) Consortium, A Blood Atlas of COVID-19 Defines Hallmarks of Disease Severity and Specificity, *Cell*, 2022, **185**(5), 916–938, DOI: [10.1016/j.cell.2022.01.012](https://doi.org/10.1016/j.cell.2022.01.012).
- 15 J. Park, H. Kim, S. Y. Kim, Y. Kim, J.-S. Lee, K. Dan, M.-W. Seong and D. Han, In-Depth Blood Proteome Profiling Analysis Revealed Distinct Functional Characteristics of Plasma Proteins between Severe and Non-Severe COVID-19 Patients, *Sci. Rep.*, 2020, **10**(1), 22418, DOI: [10.1038/s41598-020-80120-8](https://doi.org/10.1038/s41598-020-80120-8).
- 16 A. B. Palmos, V. Millischer, D. K. Menon, T. R. Nicholson, L. S. Taams, B. Michael, G. Sunderland, M. J. Griffiths, COVID Clinical Neuroscience Study Consortium, C. Hübel and G. Breen, Proteome-Wide Mendelian Randomization Identifies Causal Links between Blood Proteins and Severe COVID-19, *PLoS Genet.*, 2022, **18**(3), e1010042, DOI: [10.1371/journal.pgen.1010042](https://doi.org/10.1371/journal.pgen.1010042).
- 17 B. Shen, X. Yi, Y. Sun, X. Bi, J. Du, C. Zhang, S. Quan, F. Zhang, R. Sun, L. Qian, W. Ge, W. Liu, S. Liang, H. Chen, Y. Zhang, J. Li, J. Xu, Z. He, B. Chen, J. Wang, H. Yan, Y. Zheng, D. Wang, J. Zhu, Z. Kong, Z. Kang, X. Liang, X. Ding, G. Ruan, N. Xiang, X. Cai, H. Gao, L. Li, S. Li, Q. Xiao, T. Lu, Y. Zhu, H. Liu, H. Chen and T. Guo, Proteomic and Metabolomic Characterization of COVID-19 Patient Sera, *Cell*, 2020, **182**(1), 59–72.e15, DOI: [10.1016/j.cell.2020.05.032](https://doi.org/10.1016/j.cell.2020.05.032).
- 18 Y.-M. Chen, Y. Zheng, Y. Yu, Y. Wang, Q. Huang, F. Qian, L. Sun, Z.-G. Song, Z. Chen, J. Feng, Y. An, J. Yang, Z. Su, S. Sun, F. Dai, Q. Chen, Q. Lu, P. Li, Y. Ling, Z. Yang, H. Tang, L. Shi, L. Jin, E. C. Holmes, C. Ding, T.-Y. Zhu and Y.-Z. Zhang, Blood Molecular Markers Associated with COVID-19 Immunopathology and Multi-Organ Damage, *EMBO J.*, 2020, **39**(24), e105896, DOI: [10.15252/embj.2020105896](https://doi.org/10.15252/embj.2020105896).
- 19 M. Kukar, G. Gunčar, T. Vovko, S. Podnar, P. Černelč, M. Brvar, M. Zalaznik, M. Notar, S. Moškon and M. Notar, COVID-19 Diagnosis by Routine Blood Tests Using Machine Learning, *Sci. Rep.*, 2021, **11**(1), 10738, DOI: [10.1038/s41598-021-90265-9](https://doi.org/10.1038/s41598-021-90265-9).
- 20 J. J. Lavigne and E. V. Anslyn, Sensing A Paradigm Shift in the Field of Molecular Recognition: From Selective to Differential Receptors, *Angew. Chem., Int. Ed.*, 2001, **40**(17), 3118–3130, DOI: [10.1002/1521-3773\(20010903\)40:17<3118::AID-ANIE3118>3.0.CO;2-Y](https://doi.org/10.1002/1521-3773(20010903)40:17<3118::AID-ANIE3118>3.0.CO;2-Y).
- 21 Y. Geng, W. J. Peveler and V. Rotello, Array-based “Chemical Nose” Sensing in Diagnostics and Drug Discovery, *Angew. Chem., Int. Ed.*, 2019, **58**(16), 5190–5200, DOI: [10.1002/anie.201809607](https://doi.org/10.1002/anie.201809607).
- 22 K. J. Albert, N. S. Lewis, C. L. Schauer, G. A. Sotzing, S. E. Stitzel, T. P. Vaid and D. R. Walt, Cross-Reactive Chemical Sensor Arrays, *Chem. Rev.*, 2000, **100**, 2595–2626, DOI: [10.1021/cr980102w](https://doi.org/10.1021/cr980102w).
- 23 O. R. Miranda, C.-C. You, R. Phillips, I.-B. Kim, P. S. Ghosh, U. H. F. Bunz and V. M. Rotello, Array-Based Sensing of Proteins Using Conjugated Polymers, *J. Am. Chem. Soc.*, 2007, **129**(32), 9856–9857, DOI: [10.1021/ja0737927](https://doi.org/10.1021/ja0737927).
- 24 M. De, S. Rana, H. Akpınar, O. R. Miranda, R. R. Arvizo, U. H. F. Bunz and V. M. Rotello, Sensing of Proteins in Human Serum Using Conjugates of Nanoparticles and Green Fluorescent Protein, *Nat. Chem.*, 2009, **1**(6), 461–465, DOI: [10.1038/nchem.334](https://doi.org/10.1038/nchem.334).
- 25 N. Das Saha, S. Pradhan, R. Sasmal, A. Sarkar, C. M. Beraç, J. C. Kölsch, M. Pahwa, S. Show, Y. Rozenholc, Z. Topçu, V. Alessandrini, J. Guibourdenche, V. Tsatsaris, N. Gagey-Eilstein and S. S. Agasti, Cucurbit[7]Uril Macrocyclic Sensors for Optical Fingerprinting: Predicting Protein Structural Changes to Identifying Disease-Specific Amyloid Assemblies, *J. Am. Chem. Soc.*, 2022, **144**(31), 14363–14379, DOI: [10.1021/jacs.2c05969](https://doi.org/10.1021/jacs.2c05969).



- 26 A. Bajaj, O. R. Miranda, I.-B. Kim, R. L. Phillips, D. J. Jerry, U. H. F. Bunz and V. M. Rotello, Detection and Differentiation of Normal, Cancerous, and Metastatic Cells Using Nanoparticle-Polymer Sensor Arrays, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**(27), 10912–10916, DOI: [10.1073/pnas.0900975106](https://doi.org/10.1073/pnas.0900975106).
- 27 N. Das Saha, R. Sasmal, S. K. Meethal, S. Vats, P. V. Gopinathan, O. Jash, R. Manjithaya, N. Gagey-Eilstein and S. S. Agasti, Multichannel DNA Sensor Array Fingerprints Cell States and Identifies Pharmacological Effectors of Catabolic Processes, *ACS Sens.*, 2019, **4**(12), 3124–3132, DOI: [10.1021/acssensors.9b01009](https://doi.org/10.1021/acssensors.9b01009).
- 28 L. M. Killingsworth and J. C. Daniels, Plasma Protein Patterns in Health and Disease, *Crit. Rev. Clin. Lab. Sci.*, 1979, **11**(1), 1–30, DOI: [10.3109/10408367909105852](https://doi.org/10.3109/10408367909105852).
- 29 W. J. Peveler, R. F. Landis, M. Yazdani, J. W. Day, R. Modi, C. J. Carmalt, W. M. Rosenberg and V. M. Rotello, A Rapid and Robust Diagnostic for Liver Fibrosis Using a Multichannel Polymer Sensor Array, *Adv. Mater.*, 2018, **30**(28), 1800634, DOI: [10.1002/adma.201800634](https://doi.org/10.1002/adma.201800634).
- 30 N. D. B. Le, A. K. Singla, Y. Geng, J. Han, K. Seehafer, G. Prakash, D. F. Moyano, C. M. Downey, M. J. Monument, D. Itani, U. H. F. Bunz, F. R. Jirik and V. M. Rotello, Simple and Robust Polymer-Based Sensor for Rapid Cancer Detection Using Serum, *Chem. Commun.*, 2019, **55**(76), 11458–11461, DOI: [10.1039/C9CC04854E](https://doi.org/10.1039/C9CC04854E).
- 31 K. Kim, N. Selvapalam, Y. H. Ko, K. M. Park, D. Kim and J. Kim, Functionalized Cucurbiturils and Their Applications, *Chem. Soc. Rev.*, 2007, **36**(2), 267–279, DOI: [10.1039/B603088M](https://doi.org/10.1039/B603088M).
- 32 A. R. Urbach and V. Ramalingam, Molecular Recognition of Amino Acids, Peptides, and Proteins by Cucurbit[n]Urils Receptors, *Isr. J. Chem.*, 2011, **51**(5–6), 664–678, DOI: [10.1002/ijch.201100035](https://doi.org/10.1002/ijch.201100035).
- 33 J. W. Lee, M. H. Shin, W. Mobley, A. R. Urbach and H. I. Kim, Supramolecular Enhancement of Protein Analysis via the Recognition of Phenylalanine with Cucurbit[7]Uril, *J. Am. Chem. Soc.*, 2015, **137**(48), 15322–15329, DOI: [10.1021/jacs.5b10648](https://doi.org/10.1021/jacs.5b10648).
- 34 O. Picone, A. J. Vivanti, J. Sibiude, A.-G. Cordier, V. Alessandrini, G. Kayem, C. Borie, D. Luton, P. Manchon, C. Couffignal, C. Vauloup Fellous and COVIPREG study group, SARS-COV-2 Excretion and Maternal-Fetal Transmission: Virological Data of French Prospective Multi-Center Cohort Study COVIPREG during the First Wave, *J. Gynecol. Obstet. Hum. Reprod.*, 2023, **52**(4), 102547, DOI: [10.1016/j.jogoh.2023.102547](https://doi.org/10.1016/j.jogoh.2023.102547).
- 35 A. Som, M. Pahwa, S. Bawari, N. D. Saha, R. Sasmal, M. S. Bosco, J. Mondal and S. S. Agasti, Multiplexed Optical Barcoding of Cells via Photochemical Programming of Bioorthogonal Host–Guest Recognition, *Chem. Sci.*, 2021, **12**(15), 5484–5494, DOI: [10.1039/D0SC06860H](https://doi.org/10.1039/D0SC06860H).



A methodology study toward the diagnostic of a SARS-Cov-2 infection in human serum with a macrocyclic based sensor array

Monica Swetha Bosco^{1,2}, Zeki Topcu⁴, Soumen Pradhan^{6,7}, Ariadne Sossah¹, Vassilis Tsatsaris^{1,3}, Christelle
Vauloup-Fellous⁵, Sarit S. Agasti^{6,7}, Yves Rozenholc^{4*}, Nathalie Gagey-Eilstein^{1,2*}

¹Université Paris Cité, INSERM UMR-S 1139, FHU PREMA, 4 avenue de l'observatoire, 75006 PARIS,
France

²Université Paris Cité, CNRS, INSERM, UTCBS, 4 avenue de l'observatoire, 75006 PARIS, France

³Department of Obstetric, Cochin Hospital, AP-HP, Université Paris Cité, FHU PREMA, 123 Bd Port-Royal,
75014 Paris, France.

⁴UR 7537, BioSTM, Université Paris Cité, 4 avenue de l'Observatoire, 75006 Paris, France.

⁵ Université Paris-Saclay, INSERM U1193, Virology Laboratory, Hôpital Paul-Brousse, AP-HP, Villejuif,
France

⁶New Chemistry Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore,
Karnataka 560064, India.

⁷Chemistry & Physics of Materials Unit, Jawaharlal Nehru Centre for Advanced Scientific Research
(JNCASR), Bangalore, Karnataka 560064, India.

Correspondance e-mail: *nathalie.eilstein@u-paris.fr*

*These two authors contributed equally to the work.

TABLE OF CONTENTS

A. Clinical data

B. Sensing Covid in clinical specimen

A. Clinical data

Table S1: Detailed clinical informations for each samples

N° sample	Covid status	Age	WG*	Days between symptoms and test	Breath symptoms	Hospitalization	Oxygen need	ICU
1	I	40	40	0	Y	Y	Y	Y
2	I	31	40	4	Y	N	N	N
3	I	37	35	5	Y	N	N	N
4	I	28	39	0	Y	N	N	N
5	I	31	40	1	Y	N	N	N
6	I	36	37	16	Y	N	N	N
7	I	31	29	16	N	N	N	N
8	I	32	28	11	Y	N	N	N
9	I	32	40	11	N	N	N	N
10	I	31	33	1	Y	Y	Y	Y
11	I	25	38	1	N	N	N	N
12	I	28	39	7	N	N	N	N
1	NI	39	40					
2	NI	27	40					
3	NI	19	40					
4	NI	30	38					
5	NI	37	38					
6	NI	31	38			N.A		
7	NI	30	39					
8	NI	42	38					
9	NI	38	39					
10	NI	31	41					

11	NI	44	36
12	NI	26	36
13	NI	23	28
14	NI	43	33

*WK = Week of Gestation

B. Sensing Covid in clinical specimen

Table S2: Excitation and emission wavelengths used for each CB[7]-FL.

FL in CB[7]-FL	$\lambda_{\text{excitation}}$ (nm)	$\lambda_{\text{emission}}$ (nm)
Bodipy	500	515
Coumarin	418	433
Cyanine 3	515	560
Cyanine 5	630	650
Fluorescein	500	515
Silicorhodamine	645	660
TAMRA	554	574

Figure S1: Effect of complex matrix on fluorescence of CB[7]-FL. Variation in the fluorescence emission spectra of a/ CB[7]-BDP b/ CB[7]-TMR after addition of proteins in serum matrix (1.6 mg/mL) and control (serum only). c/ Histogram of the fluorescence emission at $\lambda_{em\ max}$ of CB-BDP, CB-TMR, CB-Cy3, CB-Cy5, CB-Cou and CB-SiR, in presence of serum before or after addition of proteins (BSA, pepsin and lysozyme) at 1.6 mg/mL. * indicate significant differences (p -value < 0.05)

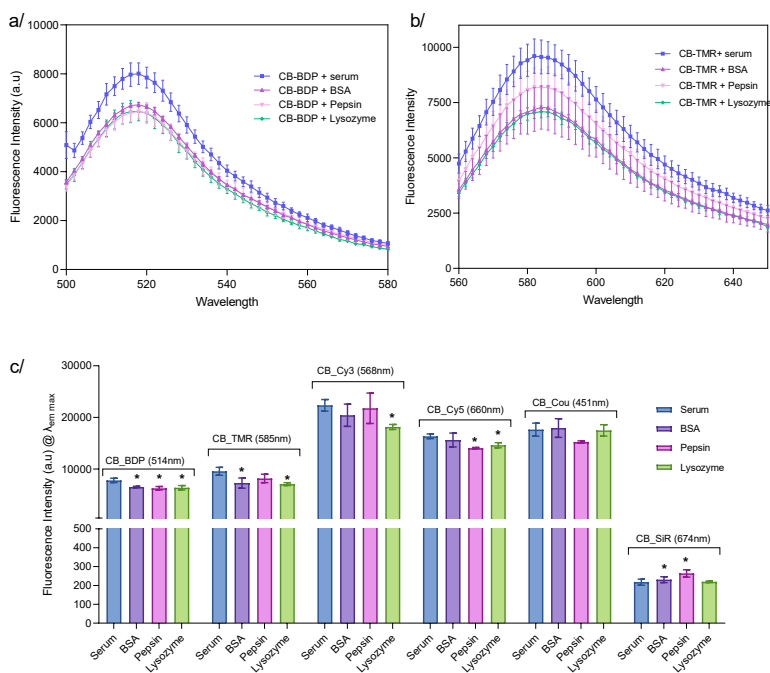


Figure S2: Stability of the fluorescence signals with different CB-[FL] during the 30 min of measurements

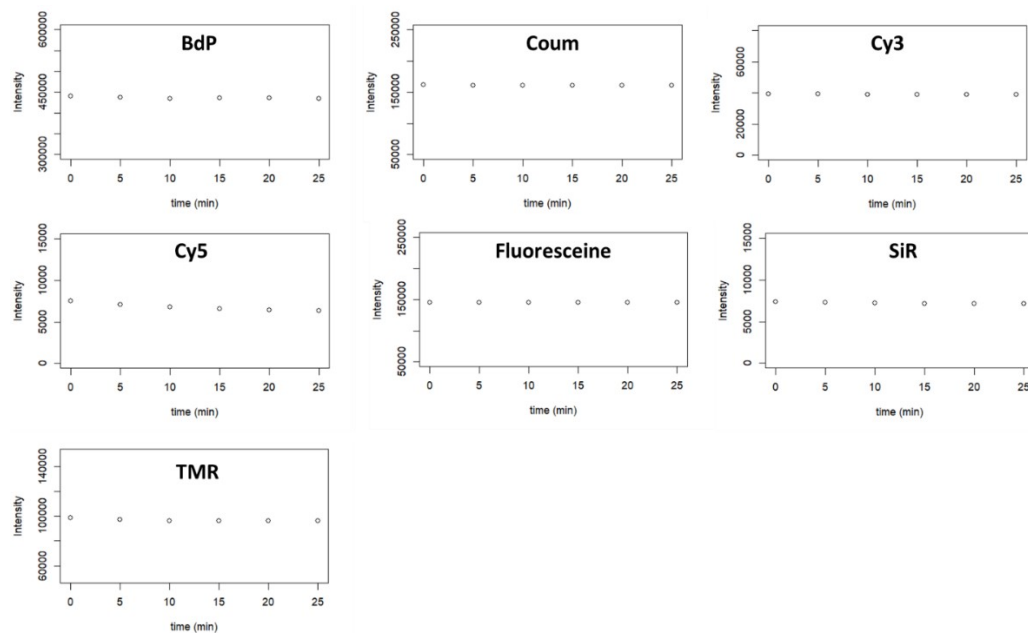


Figure S3: Microplates organization – 1+, 2+, 3+, 4+, 5+, 6+, 7+, 8+, 9+, 10+, 11+, 12+ are Covid+ samples and 1-, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 11-, 12-, 13-, 14- are control samples. Black positions in B11 to E11 are CB[7]-FL with PBS only instead of serum.

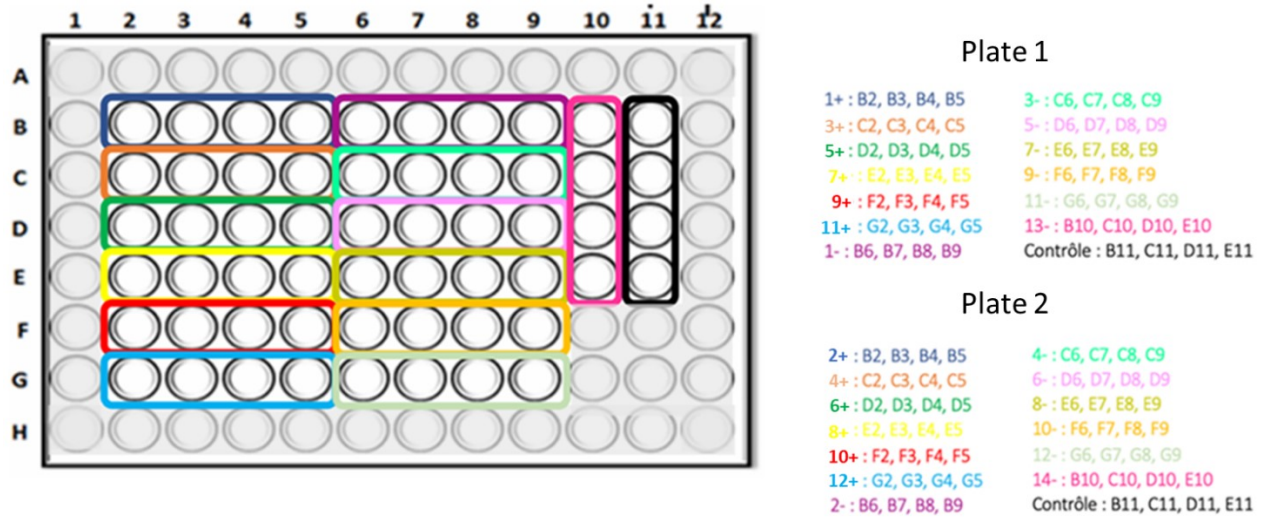


Table S3: Data of Fluorescence modulation, $\bar{\Delta}I$, of CB-FLs against serum samples (**1:** Sars-Cov2 infected serum samples, **2:** control serum samples)

Sample	Covid status	BdP	Coum	Cy3	Cy5	Fluor	SiR	TMR
1+	1	-100977,38	-37362,04	34368,25	-1088,58	27325,33	-6286,17	-26557,79
3+	1	-335107,08	-159768,08	29186,13	-1722,71	-106612,13	-6319,00	-88688,42
5+	1	-68893,33	-92973,00	45036,42	-1596,04	87689,71	-6640,63	-23434,75
7+	1	-133037,46	-101560,67	17072,88	-1464,50	-60315,75	-6412,75	-29726,63
9+	1	-203599,00	-104554,63	8662,88	-1840,08	-72260,67	-5843,04	-26606,33
11+	1	-185874,92	-113050,88	9921,58	-1672,29	-70588,13	-5735,38	-30772,83
1-	-1	-196079,15	-99345,79	16444,75	-2039,04	-49643,17	-7106,19	-27855,54
3-	-1	-128953,67	-92031,04	18264,88	-2251,75	-45384,92	-7304,50	-28776,29
5-	-1	-140275,71	-96240,96	23421,25	-1820,58	-14528,46	-6886,79	-24953,93
7-	-1	-184890,71	-90600,38	13964,49	-2043,92	-69207,21	-6803,79	-20248,42
9-	-1	-247232,75	-122156,88	7224,63	-2107,33	-77408,54	-6136,54	-27620,29
11-	-1	-145756,29	-89371,08	14840,13	-2010,71	-54077,79	-6547,71	-21355,75
13-	-1	-168824,88	-89789,21	14415,50	-1868,29	-48213,00	-6823,21	-25259,67
2+	1	-96936,25	-110928,13	14218,38	-2052,71	3529,75	-10645,21	-33003,96
4+	1	-75001,71	-88250,88	16230,75	-2226,08	-11588,71	-10648,71	-30618,96
6+	1	-135989,33	-140846,13	2282,37	-2508,29	-46848,88	-11085,13	-51040,96
8+	1	-104084,04	-90749,29	-148,21	-2690,08	-48809,88	-10669,42	-33807,46
10+	1	-173953,54	-150597,96	3774,75	-2693,25	-82865,58	-10146,21	-60229,29
12+	1	-134602,83	-115489,79	4471,17	-2173,54	-55977,00	-10245,50	-31619,71
2-	-1	-187206,58	-165789,67	-13957,38	-3575,56	-62294,42	-11550,54	-81514,96
4-	-1	-142263,46	-116576,04	-674,58	-3144,54	-59988,67	-11617,00	-36797,79
6-	-1	-103875,36	-112677,75	19119,21	-3051,79	-10778,67	-11148,29	-33746,67
8-	-1	-121953,79	-110793,17	7781,37	-3065,58	-36462,13	-11212,58	-33059,08
10-	-1	-54591,83	-92565,33	21856,67	-3088,04	36676,96	-10803,67	-28357,79
12-	-1	-140930,46	-112764,13	-678,58	-3033,33	-63730,25	-10807,67	-35957,21
14-	-1	-139832,83	-131591,54	7531,50	-2963,92	-43370,96	-11263,08	-32212,75

Figure S4: Fluorescence modulation after addition of serum in each well. Covid+ samples are on the left, Control sample on the middle and control well on the right.

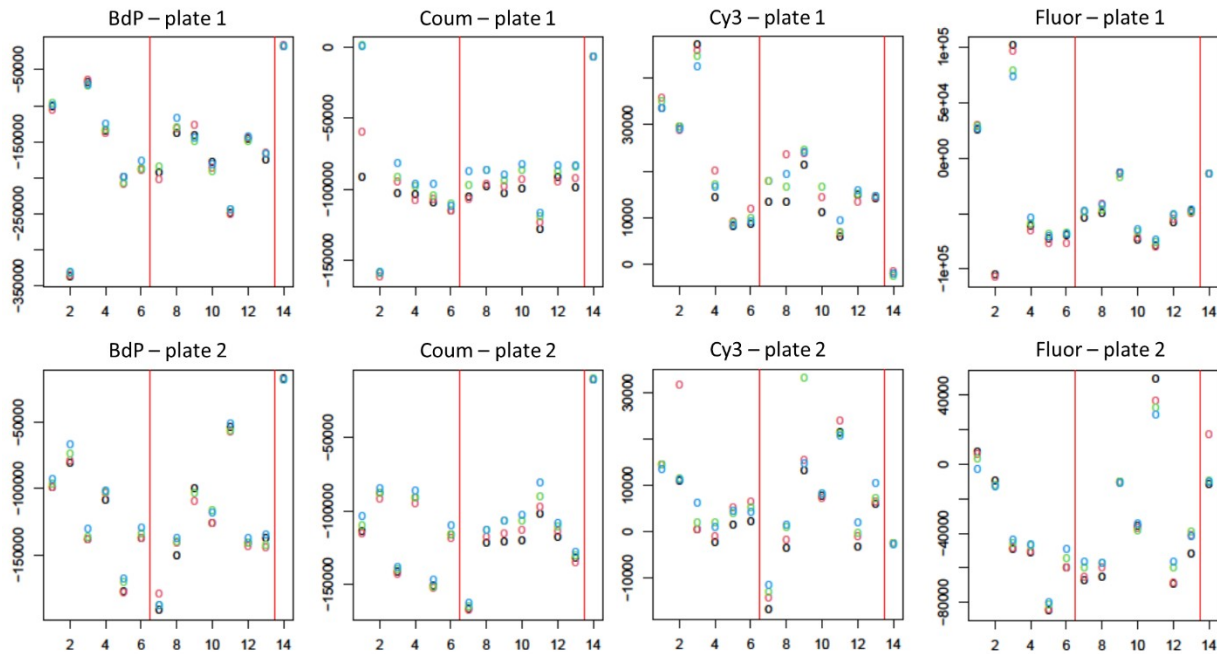


Figure S5: Stabilization of the error rates for the random forest on all the variables (a) and only on the important variables, Cy5 and SiR (b)

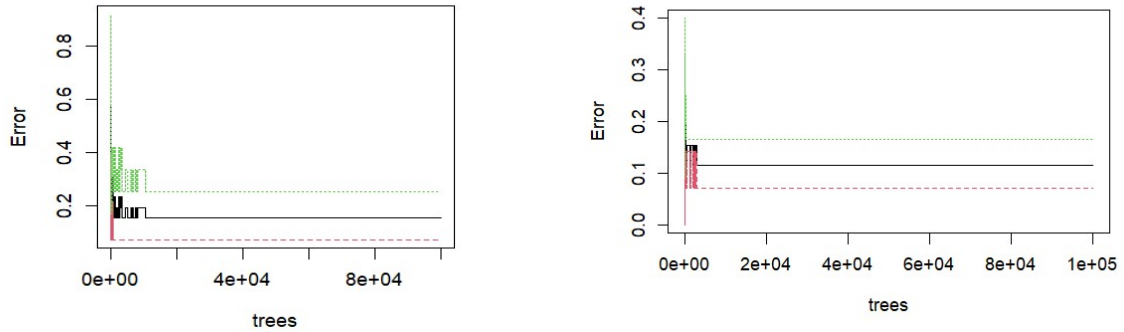


Table S4: Data of Fluorescence modulation, ΔI , of CB-FLs against serum samples (P: Pregnant (2 combined datasets), NP: Non pregnant, Covid: Sars-Cov2 infected serum samples)

	CB_Cy5	CB_Bdp	CB_SiR	CB_TMR	CB_Cy3	CB_Coum
P	-3446,389	-187206,58	-11550,542	-81514,96	-13957,375	-165789,67
P	-3144,542	-142263,46	-11617	-36797,79	-674,5833	-116576,04
P	-3051,792	-104629,06	-11148,292	-33746,67	19119,2083	-112677,75
P	-3065,583	-121953,79	-11212,583	-33059,08	7781,375	-110793,17
P	-3088,042	-54591,83	-10803,667	-28357,79	21856,6667	-92565,33
P	-3033,333	-140930,46	-10807,667	-35957,21	-678,5833	-112764,12
P	-2963,917	-139832,83	-11263,083	-32212,75	7531,5	-131591,54
P	-2039,042	-193913,33	-7026,778	-27855,54	16352,1111	-99345,79
P	-2251,75	-128953,67	-7304,5	-28776,29	18264,875	-92031,04
P	-1820,583	-140275,71	-6886,792	-24953,92	23421,25	-96240,96
P	-2043,917	-184890,71	-6803,792	-20248,42	14094,1667	-90600,38
P	-2107,333	-247232,75	-6136,542	-27620,29	7224,625	122156,88
P	-2010,708	-145756,29	-6547,708	-21355,75	14840,125	-89371,08
P	-1868,292	-168824,88	-6823,208	-25259,67	14415,5	-89789,21
P	1690,06	-88540,52	-5319,65	-8902,73	32807,21	-52670,48
P	1525,73	-85174,19	-5827,48	-10058,06	34345,21	-59142,98
P	1543,4	-77320,19	-6016,65	-8264,9	34334,38	-57700,98
P	1678,23	-73041,52	-5907,81	-5588,56	35764,04	-53545,81
P	1516,06	-74916,19	-6016,81	-8062,4	34382,38	-56137,65
P	1664,23	-90913,69	-5774,48	-7884,73	32524,71	-52595,65
P	1626,06	-67777,19	-5828,31	-9456,06	37104,38	-46513,98
P	1694,23	-77103,02	-6039,98	-9610,9	32856,21	-45789,65
P	1638,56	-68564,52	-6225,81	-10059,73	40317,88	-46861,98
P	1821,73	-63000,19	-6324,65	-9297,23	36709,38	-46239,48
P	1653,4	-68675,19	-6608,65	-9320,06	34823,38	-45587,65
P	1820,23	-72750,02	-6123,15	-9089,73	36033,04	-44187,31
P	1685,06	-95091,69	-6108,65	-7551,4	39311,04	-50310,81
P	1719,4	-92602,35	-6180,98	-8938,23	38436,38	-53123,31
P	1695,9	-91944,85	-6410,81	-7931,23	39311,38	-52402,98
P	1603,23	-94356,19	-6442,98	-7990,9	39081,54	-52295,81
P	1719,23	-96028,35	-6719,31	-7742,23	39747,21	-50034,31
P	1824,56	-88318,52	-6539,98	-7524,23	39579,38	-51266,81
P	1638,06	-93054,85	-6579,31	-8051,4	38059,21	-52160,81
P	1735,9	-98695,02	-6543,48	-8264,9	40313,88	-54326,31
P	1725,06	-96420,52	-6796,15	-7756,4	41745,54	-54301,65
P	1839,9	-92472,35	-6877,65	-7556,73	41340,38	-53055,15
P	1926,73	-95547,85	-6739,81	-7609,06	40231,54	-52496,81
P	1809,73	-102659,02	-6559,48	-7050,9	40872,88	-50680,15
P	1883,73	-102128,02	-6475,65	-8007,9	36731,38	-65553,48
P	1967,4	-105015,69	-6945,48	-8148,9	37092,21	-69410,81
P	1952,9	-98213,02	-6901,48	-8637,06	39244,21	-64575,15
P	1865,06	-98316,69	-6790,65	-10028,73	36665,38	-66733,98
P	1940,9	-104259,85	-6792,65	-8346,06	38603,21	-66893,31

P	1986,4	-96352,35	-6371,15	-8262,23	38085,04	-65476,15
P	2852,6	-86423,44	-4708,81	-13964,25	24509,75	-54130,4
P	2806,44	-86354,94	-4810,15	-13986,75	28210,75	-54654,4
P	2707,77	-86210,94	-4803,65	-13551,25	22747,92	-53321,4
P	2738,6	-86888,6	-4756,31	-13707,58	22496,08	-53053,56
P	2780,1	-86822,44	-4764,65	-13038,92	23144,08	-52723,23
P	2808,44	-84545,6	-4764,98	-13831,25	22237,25	-53559,23
P	2775,94	-75756,6	-5163,98	-5608,75	42138,75	-31130,9
P	3218,6	-75496,44	-5320,48	-6188,92	46812,42	-32529,23
P	3286,6	-77256,6	-5553,98	-5756,92	38543,92	-30847,23
P	4042,77	-78086,1	-5198,98	-5715,75	39119,25	-31317,4
P	4090,77	-79526,77	-5503,65	-6332,58	41220,75	-32222,56
P	3306,27	-78002,27	-5033,81	-5949,08	36943,92	-30601,73
NP	1863,06	-132815,69	-6410,15	-7433,56	38167,54	-59593,81
NP	1711,23	-125201,02	-6701,65	-8707,4	38356,54	-59085,31
NP	1886,06	-130831,02	-6763,98	-8596,73	38317,88	-59245,98
NP	1747,4	-125199,52	-6847,81	-7313,4	36398,71	-60088,48
NP	1989,73	-131704,69	-6839,81	-7156,23	38962,38	-57306,81
NP	1885,73	-131921,02	-6710,65	-6876,73	41057,88	-58338,98
NP	1964,4	-120134,85	-6626,65	-15123,4	24781,71	-76338,31
NP	1792,56	-120444,02	-6770,81	-15315,4	25153,21	-78879,65
NP	1949,4	-125060,52	-7287,65	-14868,06	24917,71	-80286,65
NP	2031,23	-119649,02	-6635,31	-15919,06	30390,54	-80713,65
NP	2047,4	-122046,19	-6998,48	-15043,23	28651,38	-77786,48
NP	1910,73	-118153,52	-7327,31	-14571,73	30095,21	-78445,48
NP	1931,4	-162399,19	-4454,65	-6509,06	40340,38	-73221,48
NP	1881,56	-178452,35	-4645,65	-7782,56	41036,21	-75121,15
NP	1827,73	-164926,52	-4803,15	-7622,73	41667,54	-76412,98
NP	1855,73	-173629,69	-4776,15	-7657,9	41996,88	-74895,15
NP	1889,4	-166619,19	-4669,15	-6716,73	54340,71	-72779,98
NP	1917,73	-149115,35	-4863,65	-6902,9	42758,21	-75225,31
NP	2244,4	-132602,02	-4441,48	-10666,9	54362,54	-65376,81
NP	2128,23	-124045,35	-4683,31	-12516,06	61515,54	-67507,31
NP	2312,56	-122827,02	-4781,98	-11967,9	48499,71	-66178,98
NP	2175,9	-123014,69	-4877,98	-11959,4	41833,04	-66063,98
NP	2090,73	-123506,69	-4784,65	-12788,73	40323,04	-64876,48
NP	2347,73	-120813,02	-4733,31	-11226,73	38655,88	-63284,31
NP	2128,73	-157387,85	-4472,31	-10704,23	45957,04	-67993,98
NP	2132,73	-155890,52	-4859,65	-10197,9	44088,38	-66203,15
NP	2069,23	-147190,19	-4761,98	-10349,9	40534,71	-64789,15
NP	2166,73	-145353,19	-4892,31	-11381,4	39208,71	-64387,81
NP	1998,56	-131797,69	-5149,15	-11420,06	38016,71	-65861,81
NP	2161,56	-120526,02	-4667,65	-11264,73	36441,21	-62606,81
NP	2498,44	-105088,77	-8875,98	-7632,92	46298,42	-46085,9
NP	3000,77	-102449,1	-8994,81	-7468,92	43131,25	-46287,23
NP	2133,77	-101896,94	-8792,15	-8336,92	36052,08	-48313,23
NP	2073,44	-105028,44	-8905,81	-7039,75	43603,25	-47071,4
NP	2884,94	-102947,6	-8882,31	-7446,42	38290,75	-44894,06
NP	3259,44	-103547,6	-8788,15	-6952,75	36692,75	-45425,73
NP	2947,27	-119070,27	-4643,65	-6228,25	51104,58	-62738,23
NP	3719,77	-121121,1	-4696,65	-6441,42	47888,08	-61390,56
NP	3003,27	-119983,77	-4811,48	-6591,08	45047,42	-58682,73
NP	2703,44	-119158,44	-4588,65	-6180,42	48608,25	-58685,9
NP	3009,27	-119665,94	-4431,48	-5429,75	39946,58	-56622,9
NP	3855,77	-117609,44	-4790,31	-6017,92	40124,08	-66290,9
NP	4967,1	-129962,1	-4687,31	-11029,08	43143,58	-51114,9
NP	5111,6	-127439,1	-4442,31	-10623,08	42798,75	-50427,4
NP	4487,44	-129477,6	-4635,15	-10932,92	42302,42	-51303,73
NP	4198,94	-129285,27	-4813,31	-11291,42	35822,42	-50653,9
NP	5128,6	-128218,44	-4603,15	-9688,92	39779,42	-51592,23
NP	4369,94	-125100,77	-4354,48	-9077,25	48471,58	-50071,9
Covid	-1088,583	-100977,38	-6286,167	-24584,42	34368,25	-37362,04
Covid	-1722,708	-335107,08	-6319	-88688,42	34368,25	-159768,08
Covid	-1596,042	-68893,33	-6640,625	-23434,75	45036,4167	-92973
Covid	-1464,5	-133037,46	-6412,75	-29726,62	17072,875	-101560,67
Covid	-1840,083	-203599	-5843,042	-26606,33	8662,875	-104554,62
Covid	-1672,292	-185874,92	-5735,375	-30772,83	9921,5833	-113050,88
Covid	-2052,708	-96936,25	-10645,208	-33003,96	14218,375	-110928,12
Covid	-2226,083	-75001,71	-10648,708	-30618,96	16230,75	-88250,88
Covid	-2508,292	-135989,33	-11085,125	-51040,96	2282,375	-140846,12
Covid	-2690,083	-104084,04	-10669,417	-33807,46	-148,2083	-90749,29

Covid	-2693,25	-173953,54	-10146,208	-60229,29	3774,75	-150597,96
Covid	-2173,542	-134602,83	-10245,5	-31619,71	4471,1667	-115489,79

Figure S6: Discrimination of Pregnant (P) Non-pregnant (NP) and Sars-COV-2 pregnant (Sars-COV-2 P) clinical samples from two sets of experiments. a/ Explanation of the different cohorts and combination b/ Confusion matrix for discrimination of P and NP with the random forest analysis using the whole variables. c/ Confusion matrix for discrimination of P, NP and Sars-COV-2 P with the random forest analysis using the whole variables and the most important variables.

a/

Cohort 1	Cohort 2	Combined cohort
48 NP		42 NP
42 P	14 P	56 P+P
	12 Sars-Cov-2 P	12 Sars-Cov-2 P

b/

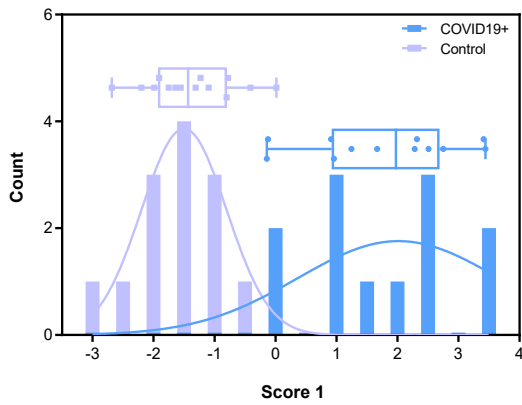
OOB estimate of error rate : 2.9 %		
	P	NP
P	54	2
NP	1	47

c/

	ALL - OOB estimate of error rate : 12.1 %			Cy5, Bdp, SiR, TMR, Cy3 OOB estimate of error rate : 6.9 %		
	P	NP	P Sars-COV-2	P	NP	P Sars-COV-2
P	51	2	3	52	2	2
NP	1	47	0	1	47	0
P Sars-COV-2	4	0	8	3	0	9

Figure S7: LDA models for the 2-group “control versus covid” model. The histogram is marked with normal distributions fitted to the full data consisting of average of replicates for each serum sample. The box-plot depicts the max/min of canonical scores obtained from the LDA.

a.

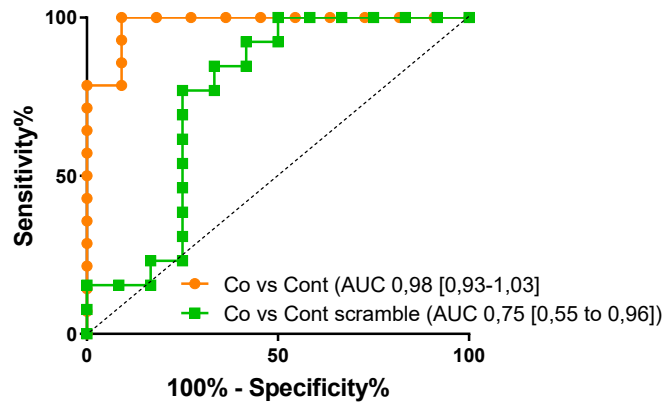
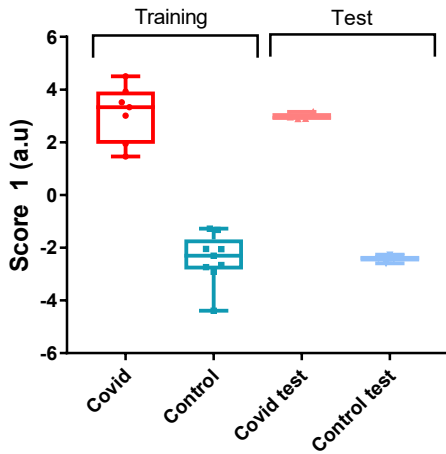


b.

	Control	Covid	%error
85%			
Control	13	1	7
Covid	2	10	16.6

<i>Random attribution</i>			
	Control	Covid	%correct
58%			
Control	5	9	64
Covid	5	7	42

Figure S8: Diagnosis of Covid-19 samples based on serum response toward the CB[7]-FL sensor array. A/ Box plot of the first canonical score of the training (70% of the samples) and test set (30% of the samples). The horizontal line in the boxes represents the median and the bottom and top of the boxes represent the 25th and 75th percentiles, respectively. B/ ROC curves for the array-based sensor in patients with Covid-19 (Co) compared with control (Cont). (Co) is for normal data and (Co) is for scrambled data.



Annex 1 : R-code for the data extraction

❖ Lecture-raw-data

```
rows.letters = LETTERS[2:7] # B,C,D,...,G
cols.info = 3:12
n.rows = length(rows.letters)
n.cols = 10
n.cycles = 6

raw.folder = './Raw Data Covid'
out.folder = './Table Data Covid'
ref.folder = './Ref Data Covid'

if (!dir.exists(out.folder)) dir.create(out.folder)
if (!dir.exists(ref.folder)) dir.create(ref.folder)

list.dir = dir(raw.folder,all.files=F,recursive=F,full.names=T)

for (name in list.dir) { # read directory content
  out.name = gsub(raw.folder,out.folder,name) # directory for output
  if (!dir.exists(out.name)) dir.create(out.name)

  list.file = list.files(path=name,pattern="*.csv",full.names=T)

  for (file in list.file) { # read each file
    print(paste('processing of',file))
    X = read.csv(file, header=F, blank.lines.skip=FALSE, stringsAsFactors=FALSE)
    if (dim(X)[2]==1) X = read.csv(file, header=F, blank.lines.skip=FALSE, stringsAsFactors=FALSE, sep=';')
    IntensityArray = array(0,dim=c(n.rows,n.cols,n.cycles),dimnames=list(rows.letters,1+(1:n.cols),1:n.cycles))
    i = tps = 1
    for (line in 1:dim(X)[1]) { # read line by line the file
      champs = X[line,] # get the contents of the line
      # print(champs[1])
      if (!(champs[1] %in% rows.letters)) next() # forget non intensity line
      # here we are on intensity line
      intensities = as.numeric(champs[cols.info])
      # print(intensities)
      IntensityArray[i,,tps] = intensities
      i = i+1

      if (champs[1]==rows.letters[n.rows]) {i=1; tps=tps+1}
      if (tps==n.cycles+1) break() # too many cycles read // stop
    }

    if (grepl('Coum',file)) print(IntensityArray)

    save(IntensityArray,file=gsub(raw.folder,out.folder,gsub('.csv','rda',file)))

    if (grepl('Cy3',file)) { # we deal with NA's on experiment 2569 Cy3 (1) I1
      print(file)
      ref.name = gsub(raw.folder,ref.folder,name) # directory for output
      if (!dir.exists(ref.name)) dir.create(ref.name)
      founded = FALSE
      while (!founded) {
        line = line+1
        old.champs = champs
        champs = X[line,]
        if (length(old.champs)<=9 | length(champs)<=9) next()
        founded = ((old.champs[9]=='Wavelength') & (as.numeric(champs[9])==560))
      }

      line = line+6
      Ref560Array = array(0,dim=c(n.rows,n.cols),dimnames=list(rows.letters,1+(1:n.cols)))
      i = 1
      for (l in line:(line+5)) {
        champs = X[l,]
        intensities = as.numeric(champs[cols.info])
        Ref560Array[i,] = intensities
        i = i+1
      }

      save(Ref560Array,file=gsub(raw.folder,ref.folder,gsub('.csv','ref560.rda',file)))
    } } }
```

❖ read-plate-data.R

```
the.method = 'Fatio'

plate2table = function(plate) {
  # print(plate)
```

```

table = rbind(plate[1:6,1:4], plate[1:6,4+(1:4)], plate[1:4,9], plate[1:4,10])
rownames(table) = c(LETTERS[1:13],0)
colnames(table) = 1:4
# print(table)
table
}

# directory des fichiers .rda
in.folder = '../Table Data Covid'

# directory des fichiers resultats
out.folder = sub('Table','Treated',in.folder)

if (!dir.exists(out.folder)) dir.create(out.folder)

list.dir = dir(in.folder,all.files=F,recursive=T,full.names=T)
list.I0 = grep('I0.rda',dir(in.folder,all.files=F,recursive=T,full.names=T), value=T)
list.I1 = grep('I1.rda',dir(in.folder,all.files=F,recursive=T,full.names=T), value=T)

for (name in list.I1)
{ # read directory content

  load(name)

  # get name of output sub-folder and create it if not existing
  in.sub.folder = substr(name,grepexpr('/',name)[[1]][2],grepexpr('/',name)[[1]][3])
  out.sub.folder = paste0(out.folder,in.sub.folder)
  fluo = gsub('/','',in.sub.folder); fluo = gsub(' ','',fluo)
  print(fluo)

  if (!dir.exists(out.sub.folder)) dir.create(out.sub.folder)

  MeanIntensityArray = apply(IntensityArray,1:2,mean,na.rm=TRUE,trim=0.1)

  IntensityTable = array(dim=c(14,4,dim(IntensityArray)[3]))
  for (i in 1:dim(IntensityArray)[3]) IntensityTable[,i] = plate2table(IntensityArray[,i])

  # Reorganize in a table instead of a plate // 4 values per line // control is the last line
  I = plate2table(MeanIntensityArray)

  bool = apply(I,1,function(l) all(is.nan(l))) # check for line made only of NaN
  if (any(bool)) {
    ref = sub('.rda',' ref560.rda',sub('Table','Ref',name))
    print(paste('use',ref))
    load(ref)
    I[bool,] = plate2table(Ref560Array)[bool,] # get line from ref
  }

  I.mean = rowMeans(I,na.rm=T) # mean over replicates

  # we deal with changes in the plate numbering from I1 to I0 in order to find I0 name
  last.barre = rev(gregexec('/',name)[[1]][1,])[1]
  end.name = substring(name,last.barre)
  next.white = gregexec(' ',end.name)[[1]][1,][1]
  end.name = substring(end.name,next.white+1)
  new.end = sub('I1','I0',end.name)

  # Do same for the associated control plate
  name.I0 = grep(new.end,list.I0,fixed=T,value=T)
  load(name.I0)
  MeanIntensityArray = apply(IntensityArray,1:2,mean,na.rm=TRUE,trim=0.1) # variance 1/6 - If only NA's create NaN
  IntensityTable0 = array(dim=c(14,4,dim(IntensityArray)[3]))
  for (i in 1:dim(IntensityArray)[3]) IntensityTable0[,i] = plate2table(IntensityArray[,i])
  I0 = plate2table(MeanIntensityArray)
  I0.mean = rowMeans(I0,na.rm=T) # mean over replicates

  out.name = paste0(out.sub.folder,substring(name,grepexpr('/',name)[[1]][3]+1))
  out.name = gsub('.rda',paste0('-','the.method','.rda'),out.name)
  if (the.method=='Eatio') {
    Eatio.all = I - I0.mean
    Eatio.mean = I.mean - I0.mean
    Eatio.sd = apply(Eatio.all,1,sd,na.rm=T)
  }
  use.all = eval(parse(text=paste(the.method,'all',sep='.')))
  use.mean = eval(parse(text=paste(the.method,'mean',sep='.')))
  use.sd = eval(parse(text=paste(the.method,'sd',sep='.')))
  covid.status = as.factor(c(rep(1,l=6),rep(-1,l=7))) # 6 first are covid+

  save(fluo, I, I0, use.all,use.mean,use.sd, covid.status, file=out.name)

}

```

❖ build-data.R

```

the.method = 'Eatio'

# directory des fichiers .rda
in.folder = '../Treated Data Covid'

# directory des fichiers resultats
out.folder = sub('Treated','Tree',in.folder)
if (!dir.exists(out.folder)) dir.create(out.folder)

list.dir = dir(in.folder,all.files=F,recursive=T,full.names=T,pattern = paste0('*-',the.method,'.rda'))

```

```

sample1 = sample2 = fluo1 = fluo2 = NULL
for (name in list.dir) { # read directory content
  print(name)
  tmp = load(name)
  if (grepl('(1)',name,fixed=T)) {
    sample1 = cbind(sample1,use.mean[names(use.mean)!='0'])
    fluo1 = c(fluo1,fluo)
    covid1 = covid.status
  } else {
    sample2 = cbind(sample2,use.mean[names(use.mean)!='0'])
    fluo2 = c(fluo2,fluo)
    covid2 = covid.status
  }
}
samples = rbind(sample1[,order(fluo1)],sample2[,order(fluo2)])
colnames(samples) = fluo1[order(fluo1)]
covid.status = c(covid1,covid2)
rownames(samples) = covid.status

save(samples,covid.status,file=paste0(out.folder,'/',the.method,'-covid-19-fluo-data.rda'))

```

Annex 2: R-code for the data analysis

```

library('randomForest')
library('FactoMineR')
library('plotly')
library('MASS')
library('plot3D')
library('ggplot2')
library('e1071')
library('caTools')
library('caret')
library('factoextra')

the.method = 'Eatio'

#Preparing the data
load(paste0('../Tree Data Covid/',the.method,'-covid-19-fluo-data.rda'))
n.data = dim(samples)[1]
the.data <- data.frame(covid.status, samples)

#### RANDOM FORESTS OOB ####
tree.full = randomForest(covid.status~., data=samples, ntree=50000, mtry=2, importance=T, proximity=T)
tree.full

importance(tree.full)
plot(tree.full)
legend('right',col=c(3,2,1),pch=16,legend = c('OOB','Covid -', 'Covid +'))
varImpPlot(tree.full)

tree.ok = randomForest(covid.status~Cy5+SiR, data=samples, ntree=50000, importance=T)
tree.ok
importance(tree.ok)
plot(tree.ok)
legend('right',col=c(3,2,1),pch=16,legend = c('OOB','Covid -', 'Covid +'))
varImpPlot(tree.ok)

#### LDA using LOCV ####
the.data <- data.frame(samples)
mLDA <- lda(covid.status~., data=the.data, CV=F)
mLDA
plot(mLDA)
mean(mLDA!=the.data$covid.status)

```



Chapter 8

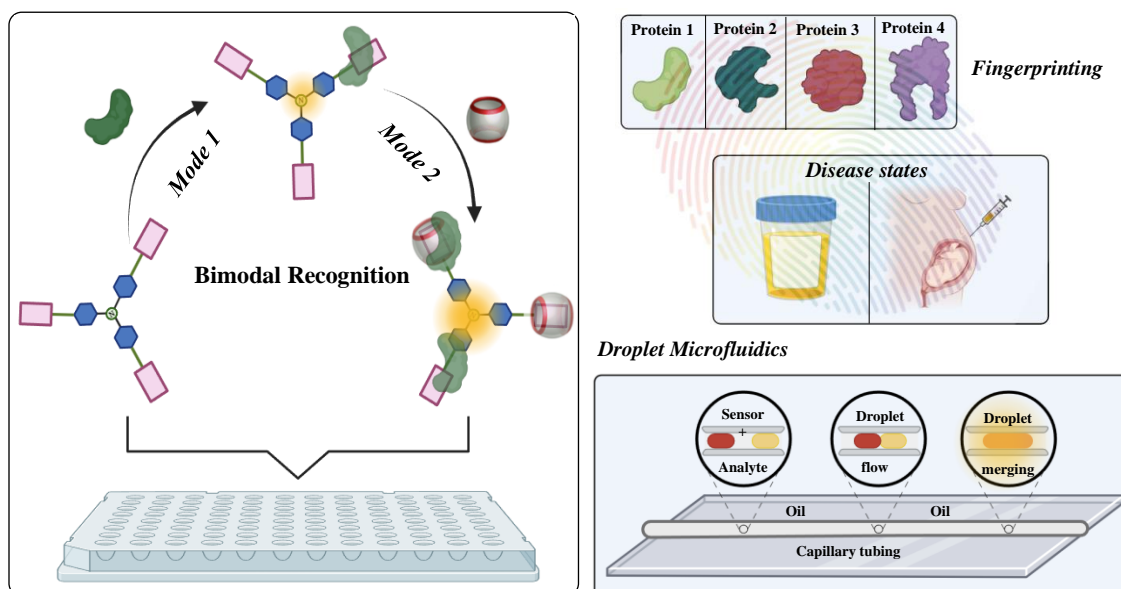
General Discussion: Future outlook and perspectives

8.1 Introduction

Array-based sensing or differential sensing inspired by nature provides an alternative approach to existing analyte detection technologies that employ a “lock-and-key” principle. This strategy brings together the molecular design of selective cross-reactive receptors, each with varying degrees of selectivity, and machine learning methodologies. The receptors generate a pattern of responses when exposed to different analytes. The collective data from these patterns are then analyzed using advanced chemometric methodologies, such as pattern recognition algorithms and multivariate statistical techniques, to accurately identify analytes or subtle differences between consistent mixtures of analytes. As a result, array-based sensing holds significant potential for a wide range of applications, from environmental monitoring to medical diagnostics, where detecting a diverse array of compounds with high accuracy is crucial.

8.2. Comprehensive overview of the project

My Ph.D. project has focused on the development of optical sensor arrays utilizing macrocyclic host-guest interactions facilitated by the family of cucurbit[n]uril (CB[n]) macrocycles and conjugated multibranching triphenylamine (TPA) derivatives.



The constructed sensor array consists of a set of sensor elements constituted of recognition elements and optical transducers. The CB[n]s were employed as scaffolds for analyte interactions, utilizing their hydrophobic cavities of different sizes for shape-based molecular recognition in combination with orthogonal H-bonding and electrostatic/dipolar interaction through uridyl carbonyl portals. The dynamic nature of supramolecular chemistry allows the use of the TPA derivatives as optical

transduction elements through competitive host-guest-driven indicator displacement assay with CB[n]s. The encapsulation of TPA inside the cavity of CB[n], stabilized by the interaction of its pyridinium acceptor groups with the C=O portals of the macrocycle, modulates its photophysical properties by effectively restricting the intermolecular rotation of the vinyl double bonds in the TPA branches. This restriction reduces radiationless decay channels and enhances fluorescence emission, thus optimizing the role of TPA as optical transducers in the sensor array. The interaction of analytes with these sensor elements leads to differential displacement of the TPA from the CB[n]s, resulting in further optical modulation of these derivatives. Additionally, the TPA derivatives offer additional orthogonal binding sites for interaction with hydrophobic regions of analytes, by docking onto their hydrophobic surfaces and pockets by a combination of weak interactions, thus pairing the two properties of transduction and recognition together to develop a bimodal recognition sensor array. These diverse interactions between analytes and the developed macrocyclic sensor array produce distinct, analyte-specific optical signatures, which could then be processed and analysed for identification and classification purposes.

The variation in the optical properties of the TPA-CB[n] system upon interaction with analytes was thoroughly studied, leading to the development of both macrocyclic fluorescence and colorimetric sensor arrays.

- The fluorescence sensor array has been constructed with four TPA-CB[7] complexes and has been validated by the successful generation, identification and classification of a diverse group of protein analytes spiked in simple physiological buffer and complex human serum. This strategy was further expanded to demonstrate the capability of the sensor array to predict pathological status from more complex body fluid samples. As a proof-of-concept, the sensor array effectively differentiated between healthy controls and samples from two separate disease models: a metabolic disorder phenylketonuria (PKU) and the pregnancy-related disorder preeclampsia (PE). This diagnostic fingerprinting by the sensor array was demonstrated in artificial urine and clinical patient serum, respectively. These findings highlight the potential of this sensor array as a non-invasive diagnostic tool, capable of detecting multifactorial diseases by profiling global proteomic changes rather than relying solely on disease-specific biomarkers.
- Additionally, a colorimetric array was developed using macrocyclic TPA-CB[7] and TPA-CB[8] complexes. This array successfully discriminated and classified common anaesthetic drug molecules that interact with CB[n]s, based on variations in absorbance facilitated by IDA. The ability of this array to provide quick and accurate identification of these drug molecules based on simple output parameters, highlights its potential for applications in counterfeit drug detection, where timely and precise substance identification is essential.
- To assess the capability of the developed sensor array in discriminating different analytes based on their unique optical fingerprints, it is crucial to develop effective data management and chemometric strategies. These strategies are necessary to transform the complex transduced fingerprints into

interpretable data. To tackle this challenge, a streamlined R program-based pipeline was created. This pipeline automates data management processes, including raw data extraction, formatting, and processing. Additionally, it handles statistical analysis for constructing dimensionality reduction and pattern recognition algorithms. Techniques such as LDA, PCA and random forest models were implemented to generate informative and non-redundant numerical representations of the processed fingerprint signals. These representations facilitate pattern recognition through suitable algorithms, enabling the identification, discrimination, and classification of different analyte classes.

- To further facilitate the ability of the sensor arrays to discriminate samples in large-scale cohorts particularly in applications like disease diagnosis, high-throughput and automated analytical platforms were explored. A droplet microfluidic platform was developed to enable controlled compartmentalization of nanoliter aqueous droplets of the sensor elements and analytes within an immiscible carrier fluid. By merging these individual droplets, the platform facilitates the detection of optical fingerprints generated by the sensor array, while significantly reducing the required sample volumes. This reduction in volume is especially advantageous when working with precious or limited analytes, such as clinical samples, as it allows for a larger number of replicates to be tested, thereby improving the reliability and robustness of the analysis. A pipetting robot platform was also integrated into the experimental workflow to provide automated, precise, and reliable handling of reduced sample volumes. This automation minimizes pipetting errors and ensures high consistency across experiments, enabling the processing of a larger number of analytes in significantly less time and enhancing high-throughput utility of the array-based sensors.

Additionally, by exploiting the covalent strategy of integration of recognition and transduction elements into a single, non-dissociable unimolecular chemosensor, we employed a developed library of seven covalently conjugated CB[7]-fluorophore (CB[7]-FL) reporter pairs to discriminate non-invasively clinically relevant information from patient blood samples. The CB[7] molecules recognize the chemical and structural features of biomolecules, bringing the covalently tethered fluorophore into close proximity, which then alters its optical signature in response to analytes, revealing the nature of the microenvironment. This strategy was adopted to propose a proof-of-concept for potential clinical diagnosis of SARS-CoV-2 from a small set of serum samples of pregnant women infected or not infected by SARS-CoV-2 towards each CB[7]-FL. Moreover, a comprehensive, step-by-step methodology and a ready-to-use protocol was proposed for designing experiments, conducting measurements, and performing data extraction, processing, and analysis using the chemical nose. This approach aimed to deepen the understanding of the chemical nose strategy, encouraging researchers to move beyond the use of general black-box statistical tools, thus facilitating the application of this hypothesis-free methodology for potential use with larger and more diverse cohorts, particularly for the diagnosis of multifactorial diseases.

8.3. Future perspectives of the project

The research presented in this thesis paves the way for several promising avenues of future exploration. While the development of the optical sensor array-based on CB[n]-TPA interactions has shown significant potential, there remain opportunities for further improvement and expansion.

Expansion of host-guest strategy for sensor design

While the macrocyclic host family of cucurbit[n]urils has been explored extensively, further optimization of the developed sensor array could benefit from exploring other macrocyclic host families such as cyclodextrins, crown ethers, calixarenes, and pillar[n]arenes. These alternative hosts offer distinct molecular encapsulation properties due to their unique shapes and charge preferences. Exploring these options could provide complementary or improved recognition capabilities compared to CB[n]s, enabling the detection of a broader spectrum of analytes with varying binding affinities and structural characteristics. This includes a wide range of substances such as charged amino acids, neurotransmitters, carbohydrates, lipids, and oligonucleotides. These analytes are crucial for maintaining healthy metabolic conditions, and their levels can indicate global changes associated with disease states. Thus, utilizing these macrocyclic hosts could extend the application of the sensor array to biofluids for the diagnosis of disease states associated with these analytes. Furthermore, introduction of other optical guest transducers in addition to the TPA derivatives could allow for the generation of improved IDA or ABA based transduction mechanisms. This addition could lead to the generation of more complex, multidimensional and informative output signals, providing richer data for the pattern recognition of analytes.

High-throughput measurements and automation of experimental workflow

The design of the droplet microfluidic platform has proven effective in enhancing the throughput of sensor array analysis. However, further advancements in the merging techniques are essential to ensure the reliability and high-throughput of this platform for analysis of large cohort of samples by the sensor array with minimal manual intervention. The integration of the microfluidic chip along the capillary could be the way to achieve this; however extensive testing and optimization are necessary to identify the most suitable materials for the fabrication of these chips. This is essential to ensure consistent merging and prevention of issues such as droplet splitting, air bubble formation, and oil leakage, while still being compatible with the sensor elements and analytes.

Application to clinical diagnosis

Although an initial proof-of-concept has been demonstrated for the disease models of phenylketonuria and preeclampsia using a small sample set, testing with larger cohorts is still necessary.

This will allow for a thorough evaluation of the reproducibility and robustness of the developed sensor array, and aid in identifying relevant classifiers for efficient pattern recognition. Such extensive testing is crucial for translating this sensing strategy into a viable clinical diagnostic tool. The clinical sample cohort used for the disease model of preeclampsia (APHERESE) consists of samples from 241 patients with 150 healthy and 91 preeclamptic patients, of whom 30% presented with very severe form of the disease. High-throughput analysis of this complete cohort on the developed analytical platforms will provide a robust approach for studying such a large and well-documented cohort of preeclamptic samples. With thorough supervised and unsupervised data analysis, there is a high likelihood of developing a reliable classifier not only for differentiation between preeclamptic and healthy patients but with access to clinical and biological patient data and disease outcomes, predictions between the severe and mild forms of the preeclamptic samples could also be evaluated. Thus, this powerful methodology, with appropriate adaptation, holds significant potential for addressing a range of clinical challenges associated with various multifactorial diseases.

8.4. Conclusions and future directions

Array-based sensing methods have shown significant promise in addressing various sensing challenges by using collections of semi-selective receptors combined with chemometric analysis. They are particularly effective in discriminating between structurally similar analytes and complex analyte mixtures without needing to fully characterize each component, which is especially advantageous for detecting mixtures that are difficult or impossible to analyse comprehensively, such as explosive mixtures or bacterial metabolites. This capability and predictive power makes array-based sensing a powerful alternative to the traditional, highly specific techniques. However, to enable broader applications beyond controlled environments, several challenges must be addressed moving forward.

- Unlike the mammalian olfactory system, which provides qualitative information about odours and is less dependent on concentration, many artificial arrays exhibit significant concentration dependence. This can lead to misidentification of analytes if the system is not trained across a range of concentrations. Moreover, if the response pattern of the array is not sufficiently unique for each analyte within the desired concentration range, or if there is even slight changes in the composition of the mixture—such as the addition of new components or unanticipated interferents—this can confuse the system and compromise the accuracy of the assay.
- To enhance the reliability and applicability of array-based sensing systems, several conditions must be fulfilled. The responses of analytes to the array must be reproducible, and the system must be able to match responses from unknown analytes to those in the training set accurately. This requires that the sensor array possesses sufficient semi-selectivity for the range of analytes of interest. If the

sensors are too non-selective, the data will have low dimensionality, limiting the system's ability to discriminate between different analytes effectively. Therefore, gaining a deeper understanding of sensor principles through multivariate analysis will offer valuable feedback for probe design, aiding in the development of more effective cross-reactive probes.

- Proper training of the array-based system is crucial for its success. The system can only identify analytes based on the information provided during training. Therefore by combining this library with high-throughput and automated measurement systems, the amount of information can be considerably scaled up, potentially reaching the so-called big data level. Advanced analysis of big data using sophisticated machine learning, deep learning, and artificial intelligence technologies will allow pattern-recognition-based sensors to accurately identify challenging analytes with high complexity or only subtle differences, to quantify and predict unknown analytes.

In summary, while array-based sensing has made significant strides in various sensing applications, further work is needed to address the challenges associated with the predictive power, concentration dependence, and the identification of complex mixtures by the array. Future research should aim to optimize training processes, improve the reproducibility and selectivity of sensor arrays, and develop strategies to handle interferences effectively. These advancements will be crucial for the practical deployment of array-based sensing technologies in real-world settings.





\

Appendix

Résumé

Introduction

La détection basée sur des réseaux ou la détection différentielle inspirée de la nature offre une approche alternative aux technologies existantes de détection d'analytes qui utilisent un principe spécifique « clé/serrure ». Cette stratégie associe la conception moléculaire de récepteurs sélectifs, chacun ayant un degré de sélectivité variable pour différents analytes, et des méthodologies d'apprentissage automatique. Les récepteurs génèrent un signal de réponses lorsqu'ils sont exposés à différents analytes. Les données collectées de ces signaux pour chaque récepteur sont ensuite analysées à l'aide de méthodologies chimométriques avancées, telles que des algorithmes de reconnaissance de signaux et des techniques statistiques multivariées, afin d'identifier avec précision les analytes ou les différences subtiles entre des mélanges cohérents d'analytes. Par conséquent, la détection basée sur les réseaux présente un potentiel important pour un large éventail d'applications, allant de la surveillance de l'environnement au diagnostic médical, où la détection d'un large éventail de composés avec une grande précision est cruciale.

Vue d'ensemble du projet

Mon projet de doctorat s'est concentré sur le développement de réseaux de capteurs optiques utilisant des interactions hôte-invité macrocycliques, utilisant la famille des macrocycles cucurbit[n]uril (CB[n]) et des dérivés conjugués multibranchés de la triphénylamine (TPA).

Le réseau de capteurs construit consiste en un ensemble d'éléments individuels chacun constitué d'un élément de reconnaissance et d'un transducteur optique. Les CB[n]s ont été utilisés comme élément de reconnaissance pour les interactions avec les analytes, en utilisant leurs cavités hydrophobes de différentes tailles pour la reconnaissance moléculaire basée sur la forme en combinaison avec la liaison H orthogonale et l'interaction électrostatique/dipolaire par l'intermédiaire des portails uridyl-carbonyle. La nature dynamique de la chimie supramoléculaire permet d'utiliser les dérivés de TPA fluorogéniques comme éléments de transduction optique par le biais d'un déplacement compétitif piloté par l'hôte et l'invité avec les CB[n]s. L'encapsulation du TPA dans la cavité du CB[n], stabilisée par l'interaction de ses groupes accepteurs pyridinium avec les portails C=O du macrocycle, module ses propriétés photophysiques en limitant efficacement la rotation intermoléculaire des doubles liaisons vinyle dans les branches du TPA. Cette restriction augmente l'émission de fluorescence, optimisant ainsi le rôle du TPA en tant que transducteur optique dans le réseau de capteurs. L'interaction des analytes avec ces éléments de capteur entraîne un déplacement différentiel du TPA par rapport aux CB[n]s, ce qui se traduit par une modulation optique supplémentaire de ces dérivés. En outre, les dérivés du TPA offrent

des sites de liaison orthogonaux supplémentaires pour l'interaction avec les régions hydrophobes des analytes, en s'arrimant à leurs surfaces et poches hydrophobes par une combinaison d'interactions faibles, associant ainsi les deux propriétés de transduction et de reconnaissance pour développer un réseau de capteurs à reconnaissance bimodale. Ces diverses interactions entre les analytes et le réseau de capteurs macrocycliques mis au point produisent des signatures optiques distinctes, spécifiques à l'analyte, qui peuvent ensuite être traitées et analysées à des fins d'identification et de classification.

La variation des propriétés optiques du système TPA-CB[n] lors de l'interaction avec les analytes a été étudiée de manière approfondie, ce qui a conduit au développement de réseaux de capteurs macrocycliques fluorescents et colorimétriques.

- Le réseau de capteurs de fluorescence a été construit avec quatre complexes TPA-CB[7] et a été validé par la génération, l'identification et la classification réussies d'un groupe diversifié d'analytes protéiques soit dans du tampon physiologique simple soit dans du sérum humain complexe. Cette stratégie a été élargie pour démontrer la capacité du réseau de capteurs à prédire l'état pathologique à partir d'échantillons de fluides corporels plus complexes. En guise de preuve de concept, le réseau de capteurs a permis de différencier efficacement les témoins sains des échantillons provenant de deux modèles pathologiques distincts : un trouble métabolique, la phénylcétonurie (PCU), et un trouble lié à la grossesse, la prééclampsie (PE). Cette empreinte diagnostique par le réseau de capteurs a été démontrée dans l'urine artificielle et le sérum de patients cliniques, respectivement. Ces résultats soulignent le potentiel de ce réseau de capteurs en tant qu'outil de diagnostic non invasif, capable de détecter des maladies multifactorielles en établissant le profil des changements protéomiques globaux plutôt qu'en s'appuyant uniquement sur des biomarqueurs spécifiques à la maladie.

- En outre, un réseau colorimétrique a été mis au point à l'aide de complexes macrocycliques TPA-CB[7] et TPA-CB[8]. Cette matrice a permis de discriminer et de classer les molécules de médicaments anesthésiques courants qui interagissent avec les CB[n]s, en se basant sur les variations d'absorbance facilitées par l'IDA. La capacité de ce réseau à fournir une identification rapide et précise de ces molécules médicamenteuses sur la base de paramètres de sortie simples souligne son potentiel pour des applications dans la détection de médicaments contrefaits, où l'identification rapide et précise de la substance est essentielle.

-Pour évaluer la capacité du réseau de capteurs développé à discriminer différents analytes sur la base de leurs empreintes optiques uniques, il est crucial de développer des stratégies efficaces de gestion des données et de chimiométrie. Ces stratégies sont nécessaires pour transformer les empreintes transduites complexes en données interprétables. Pour relever ce défi, un pipeline rationalisé basé sur un programme R a été créé. Ce pipeline automatise les processus de gestion des données, y compris l'extraction, le formatage et le traitement des données brutes. En outre, il prend en charge l'analyse statistique pour construire des algorithmes de réduction de la dimensionnalité et de reconnaissance des formes. Des techniques telles que les modèles LDA, PCA et de forêt aléatoire ont été mises en œuvre

pour générer des représentations numériques informatives et non redondantes des signaux d'empreintes digitales traités. Ces représentations facilitent la reconnaissance des formes grâce à des algorithmes appropriés, permettant l'identification, la discrimination et la classification de différentes classes d'analytes.

- Pour faciliter davantage la capacité des réseaux de capteurs à discriminer les échantillons dans des cohortes à grande échelle, en particulier dans des applications telles que le diagnostic de maladies, des plates-formes analytiques automatisées et à haut débit ont été explorées. Une plateforme microfluidique de gouttelettes a été développée pour permettre la compartimentation contrôlée de gouttelettes aqueuses de nanolitres d'éléments de capteurs et d'analytes dans un fluide porteur non miscible. En fusionnant ces gouttelettes individuelles, la plateforme facilite la détection des empreintes optiques générées par le réseau de capteurs, tout en réduisant considérablement les volumes d'échantillons requis. Cette réduction de volume est particulièrement avantageuse lorsque l'on travaille avec des analytes précieux ou limités, comme les échantillons cliniques, car elle permet de tester un plus grand nombre de répliqués, améliorant ainsi la fiabilité et la robustesse de l'analyse. Une plateforme robotisée de pipetage a également été intégrée au flux de travail expérimental afin de permettre une manipulation automatisée, précise et fiable de volumes d'échantillons réduits. Cette automatisation minimise les erreurs de pipetage et garantit une grande cohérence entre les expériences, ce qui permet de traiter un plus grand nombre d'analytes en beaucoup moins de temps et d'améliorer l'utilité à haut débit des capteurs en réseau.

En outre, en exploitant une stratégie covalente d'intégration des éléments de reconnaissance et de transduction en un seul chimiodétecteur unimoléculaire non dissociable, nous avons utilisé une bibliothèque développée de sept paires de rapporteurs CB[7]-fluorophore (CB[7]-FL) conjugués de manière covalente pour discriminer de manière non invasive des informations cliniquement pertinentes à partir d'échantillons de sang de patients. Les molécules CB[7] reconnaissent les caractéristiques chimiques et structurales des biomolécules, ce qui rapproche le fluorophore fixé de manière covalente, qui modifie alors sa signature optique en réponse aux analytes, révélant ainsi la nature du microenvironnement. Cette stratégie a été adoptée pour proposer une preuve de concept pour le diagnostic clinique potentiel du SARS-CoV-2 à partir d'un petit ensemble d'échantillons de sérum de femmes enceintes infectées ou non par le SARS-CoV-2 vers chaque CB[7]-FL. En outre, une méthodologie complète, étape par étape, et un protocole prêt à l'emploi ont été proposés pour concevoir des expériences, effectuer des mesures et procéder à l'extraction, au traitement et à l'analyse des données à l'aide du nez chimique. Cette approche visait à approfondir la compréhension de la stratégie du nez chimique, en encourageant les chercheurs à aller au-delà de l'utilisation d'outils statistiques généraux de boîte noire, facilitant ainsi l'application de cette méthodologie sans hypothèse pour une utilisation potentielle avec des cohortes plus importantes et plus diversifiées, en particulier pour le diagnostic de maladies multifactorielles.

Perspectives d'avenir du projet

Les recherches présentées dans cette thèse ouvrent la voie à plusieurs pistes prometteuses d'exploration future. Bien que le développement du réseau de capteurs optiques basé sur les interactions CB[n]-TPA ait montré un potentiel significatif, il reste des possibilités d'amélioration et d'expansion.

Élargissement de la stratégie hôte-invité pour la conception de capteurs

Bien que la famille d'hôtes macrocycliques des cucurbit[n]urils ait été largement explorée, l'optimisation du réseau de capteurs développé pourrait bénéficier de l'exploration d'autres familles d'hôtes macrocycliques telles que les cyclodextrines, les éthers-couronnes, les calixarènes et les pillar[n]arènes. Ces hôtes alternatifs offrent des propriétés d'encapsulation moléculaire distinctes en raison de leurs formes uniques et de leurs préférences en matière de charge. L'exploration de ces options pourrait fournir des capacités de reconnaissance complémentaires ou améliorées par rapport aux CB[n]s, permettant la détection d'un spectre plus large d'analytes ayant des affinités de liaison et des caractéristiques structurales variées. Cela inclut un large éventail de substances telles que les acides aminés chargés, les neurotransmetteurs, les hydrates de carbone, les lipides et les oligonucléotides. Ces analytes sont essentiels au maintien de conditions métaboliques saines et leurs niveaux peuvent indiquer des changements globaux associés à des états pathologiques. Ainsi, l'utilisation de ces hôtes macrocycliques pourrait étendre l'application du réseau de capteurs aux biofluides pour le diagnostic des états pathologiques associés à ces analytes. En outre, l'introduction d'autres transducteurs optiques invités en plus des dérivés du TPA pourrait permettre de générer des mécanismes de transduction améliorés basés sur l'IDA ou l'ABA. Cet ajout pourrait conduire à la génération de signaux de sortie plus complexes, multidimensionnels et informatifs, fournissant des données plus riches pour la reconnaissance de modèles d'analytes.

Mesures à haut débit et automatisation du flux de travail expérimental

La conception de la plateforme microfluidique à gouttelettes s'est avérée efficace pour améliorer le débit de l'analyse par réseau de capteurs. Cependant, des progrès supplémentaires dans les techniques de fusion sont essentiels pour garantir la fiabilité et le haut débit de cette plateforme pour l'analyse d'une grande cohorte d'échantillons par le réseau de capteurs avec une intervention manuelle minimale. L'intégration de la puce microfluidique le long du capillaire pourrait être le moyen d'y parvenir ; toutefois, des tests et une optimisation approfondis sont nécessaires pour identifier les matériaux les plus appropriés pour la fabrication de ces puces. Cela est essentiel pour assurer une fusion cohérente et prévenir des problèmes tels que la division des gouttelettes, la formation de bulles d'air et les fuites d'huile, tout en restant compatible avec les éléments du capteur et les analytes.

Application au diagnostic clinique

Bien qu'une première preuve de concept ait été démontrée pour les modèles pathologiques de la phénylcétonurie et de la prééclampsie à l'aide d'un petit ensemble d'échantillons, des essais sur des cohortes plus importantes sont encore nécessaires. Cela permettra une évaluation approfondie de la reproductibilité et de la robustesse du réseau de capteurs mis au point, et aidera à identifier les classificateurs pertinents pour une reconnaissance efficace des formes. Ces tests approfondis sont essentiels pour transformer cette stratégie de détection en un outil de diagnostic clinique viable. La cohorte d'échantillons cliniques utilisée pour le modèle de prééclampsie (APHERESE) se compose d'échantillons provenant de 241 patientes, dont 150 en bonne santé et 91 prééclamptiques, parmi lesquelles 30 % présentaient une forme très grave de la maladie. L'analyse à haut débit de cette cohorte complète sur les plateformes analytiques développées fournira une approche robuste pour l'étude d'une cohorte d'échantillons prééclamptiques aussi importante et bien documentée. Grâce à une analyse approfondie des données supervisées et non supervisées, il y a de fortes chances de développer un classificateur fiable non seulement pour différencier les patients prééclamptiques des patients sains, mais aussi pour évaluer les prédictions entre les formes sévères et légères des échantillons prééclamptiques grâce à l'accès aux données cliniques et biologiques des patients et à l'évolution de la maladie. Ainsi, cette méthodologie puissante, moyennant une adaptation appropriée, présente un potentiel important pour relever une série de défis cliniques associés à diverses maladies multifactorielles.

Conclusions et orientations futures

Les méthodes de détection basées sur des matrices se sont révélées très prometteuses pour relever divers défis de détection en utilisant des collections de récepteurs semi-sélectifs combinés à une analyse chimiométrique. Elles sont particulièrement efficaces pour faire la distinction entre des analytes structurellement similaires et des mélanges d'analytes complexes sans qu'il soit nécessaire de caractériser complètement chaque composant, ce qui est particulièrement avantageux pour détecter des mélanges difficiles ou impossibles à analyser de manière exhaustive, comme les mélanges d'explosifs ou les métabolites bactériens. Cette capacité et ce pouvoir prédictif font de la détection basée sur les réseaux une alternative puissante aux techniques traditionnelles hautement spécifiques. Cependant, pour permettre des applications plus larges au-delà des environnements contrôlés, plusieurs défis doivent être relevés pour aller de l'avant.

- Contrairement au système olfactif des mammifères, qui fournit des informations qualitatives sur les odeurs et dépend moins de la concentration, de nombreux réseaux artificiels présentent une dépendance importante vis-à-vis de la concentration. Cela peut conduire à une mauvaise identification des analytes

si le système n'est pas entraîné sur une gamme de concentrations. En outre, si le modèle de réponse du réseau n'est pas suffisamment unique pour chaque analyte dans la plage de concentration souhaitée, ou si la composition du mélange change, même légèrement, par exemple par l'ajout de nouveaux composants ou d'interférents imprévus, cela peut perturber le système et compromettre la précision de l'essai.

-Pour améliorer la fiabilité et l'applicabilité des systèmes de détection à base de réseaux, plusieurs conditions doivent être remplies. Les réponses des analytes au réseau doivent être reproductibles et le système doit être capable de faire correspondre avec précision les réponses d'analytes inconnus à celles de l'ensemble d'apprentissage. Pour ce faire, le réseau de capteurs doit posséder une semi-sélectivité suffisante pour la gamme d'analytes concernés. Si les capteurs sont trop peu sélectifs, les données auront une faible dimensionnalité, ce qui limitera la capacité du système à distinguer efficacement les différents analytes. Par conséquent, une meilleure compréhension des principes des capteurs grâce à l'analyse multivariée offrira un retour d'information précieux pour la conception des sondes, ce qui contribuera au développement de sondes à réaction croisée plus efficaces.

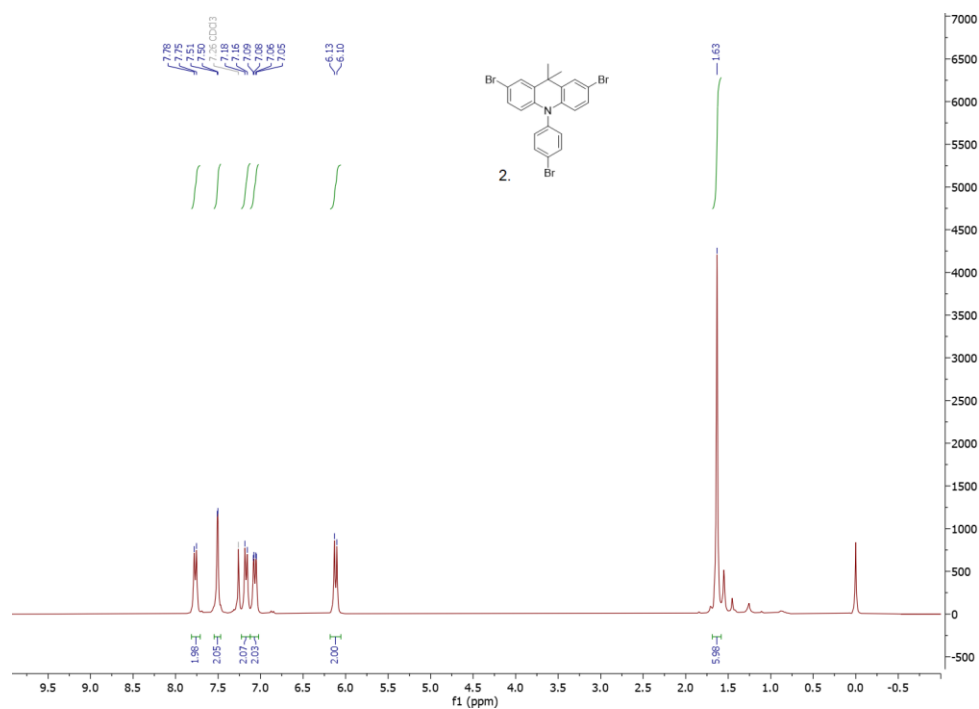
- Une formation adéquate du système basé sur les réseaux est cruciale pour son succès. Le système ne peut identifier les analytes que sur la base des informations fournies pendant la formation. Par conséquent, en combinant cette bibliothèque avec des systèmes de mesure automatisés et à haut débit, la quantité d'informations peut être considérablement augmentée, atteignant potentiellement ce que l'on appelle le niveau « big data ». L'analyse avancée des big data à l'aide de technologies sophistiquées d'apprentissage automatique, d'apprentissage profond et d'intelligence artificielle permettra aux capteurs basés sur la reconnaissance des formes d'identifier avec précision les analytes difficiles présentant une grande complexité ou seulement des différences subtiles, de quantifier et de prédire les analytes inconnus.

En résumé, bien que la détection basée sur les réseaux ait fait des progrès significatifs dans diverses applications de détection, des travaux supplémentaires sont nécessaires pour relever les défis associés au pouvoir prédictif, à la dépendance de la concentration et à l'identification de mélanges complexes par le réseau. Les recherches futures devraient viser à optimiser les processus de formation, à améliorer la reproductibilité et la sélectivité des réseaux de capteurs et à développer des stratégies pour traiter efficacement les interférences. Ces progrès seront cruciaux pour le déploiement pratique des technologies de détection basées sur les réseaux dans le monde réel.

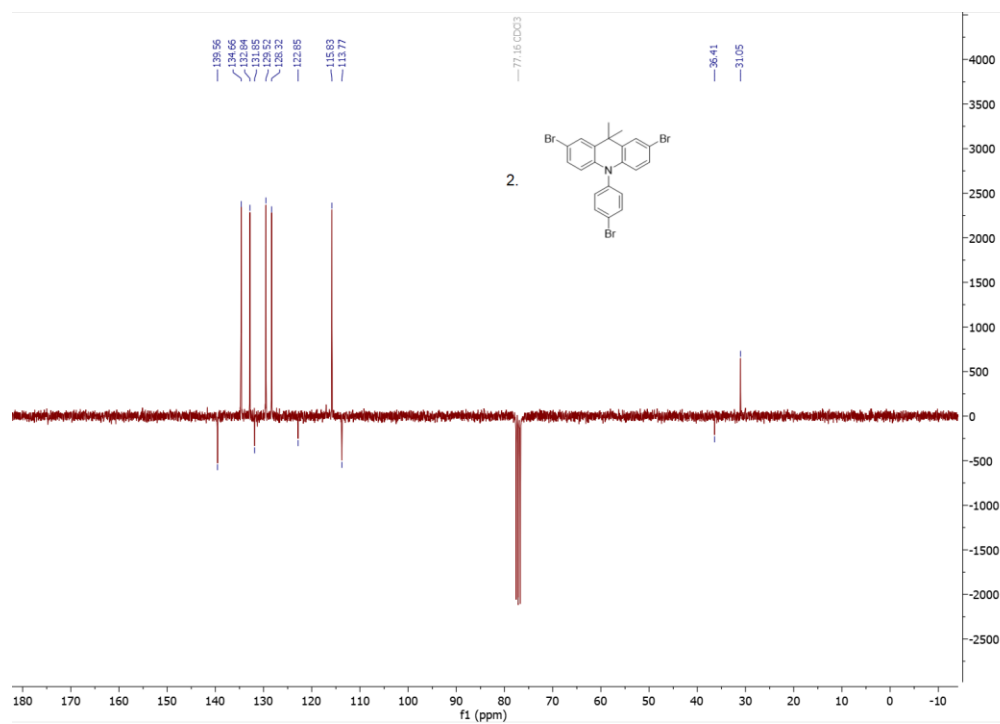
Chapter 2

Appendix 1: Characterization of TPA derivatives

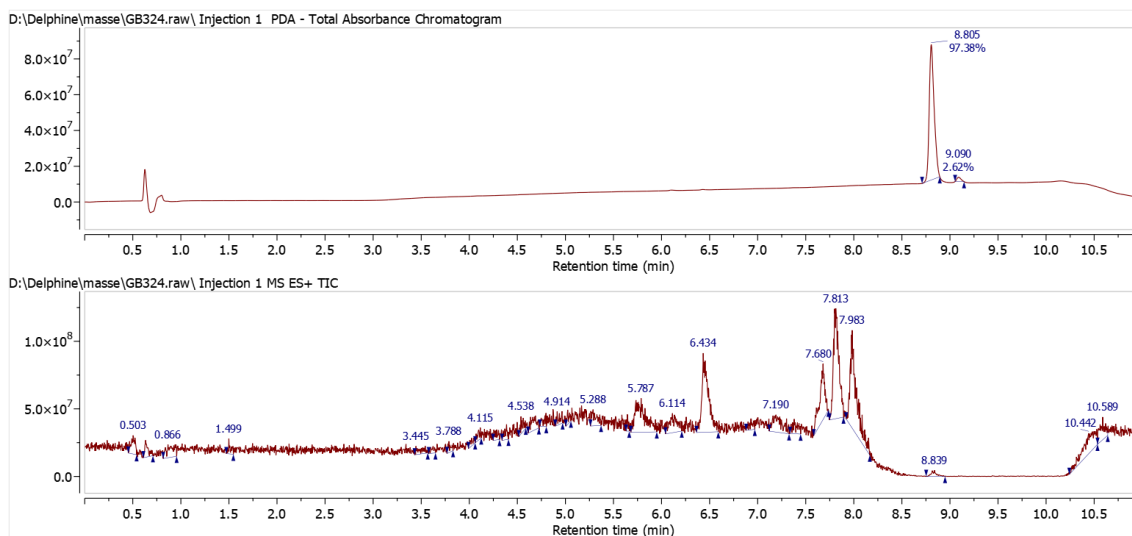
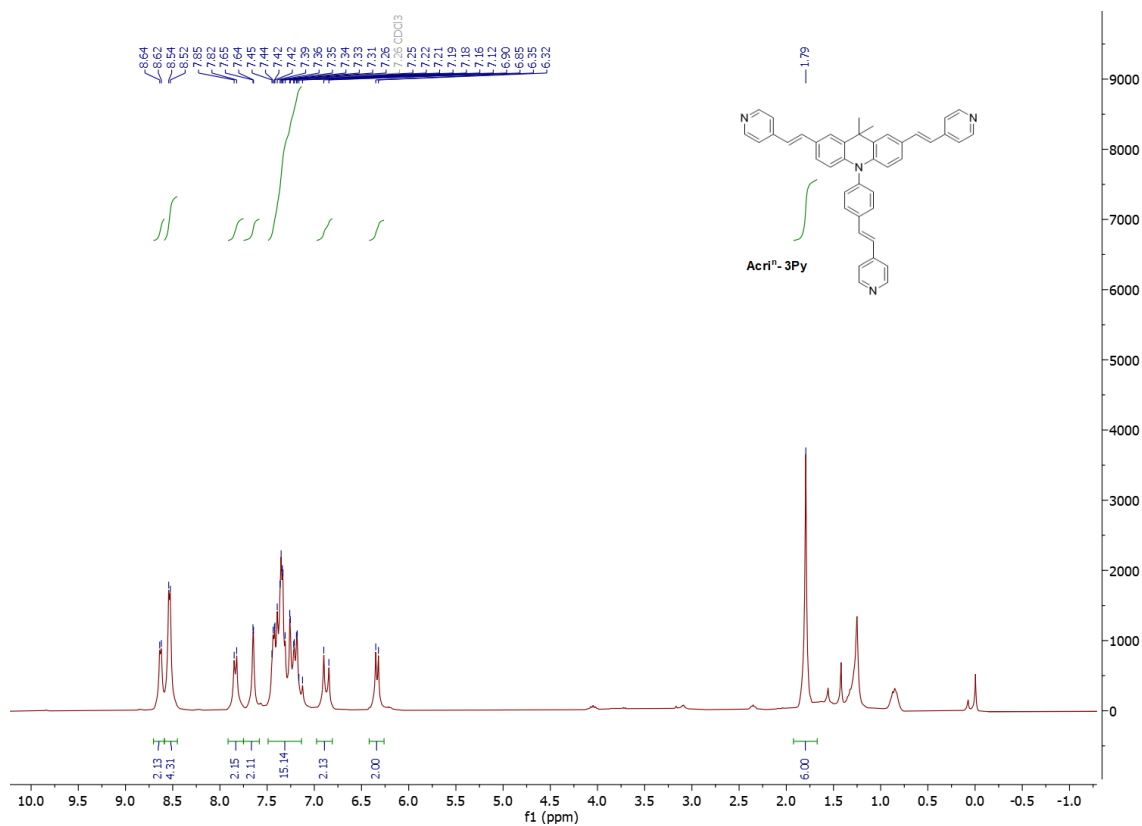
^1H NMR spectra of **2** in CDCl_3 (300 MHz):



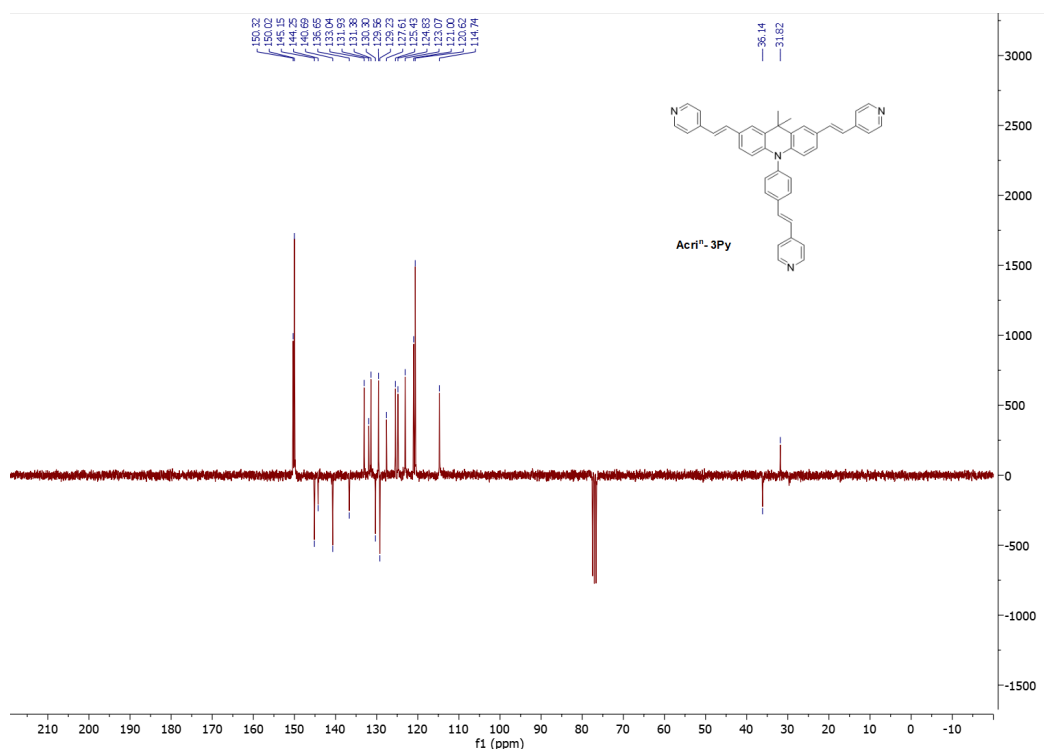
^{13}C APT NMR spectra of **2** in CDCl_3 (75 MHz):



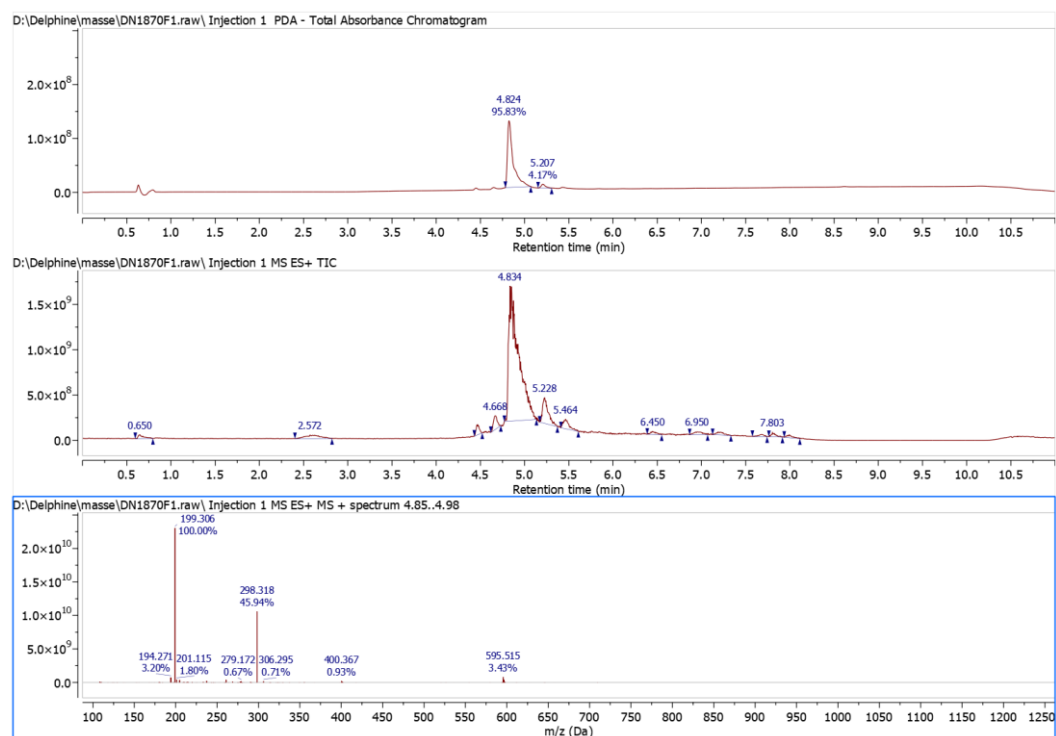
LC-MS of 2

 ^1H NMR spectra of Acrin_3Py in CDCl_3 (300 MHz):

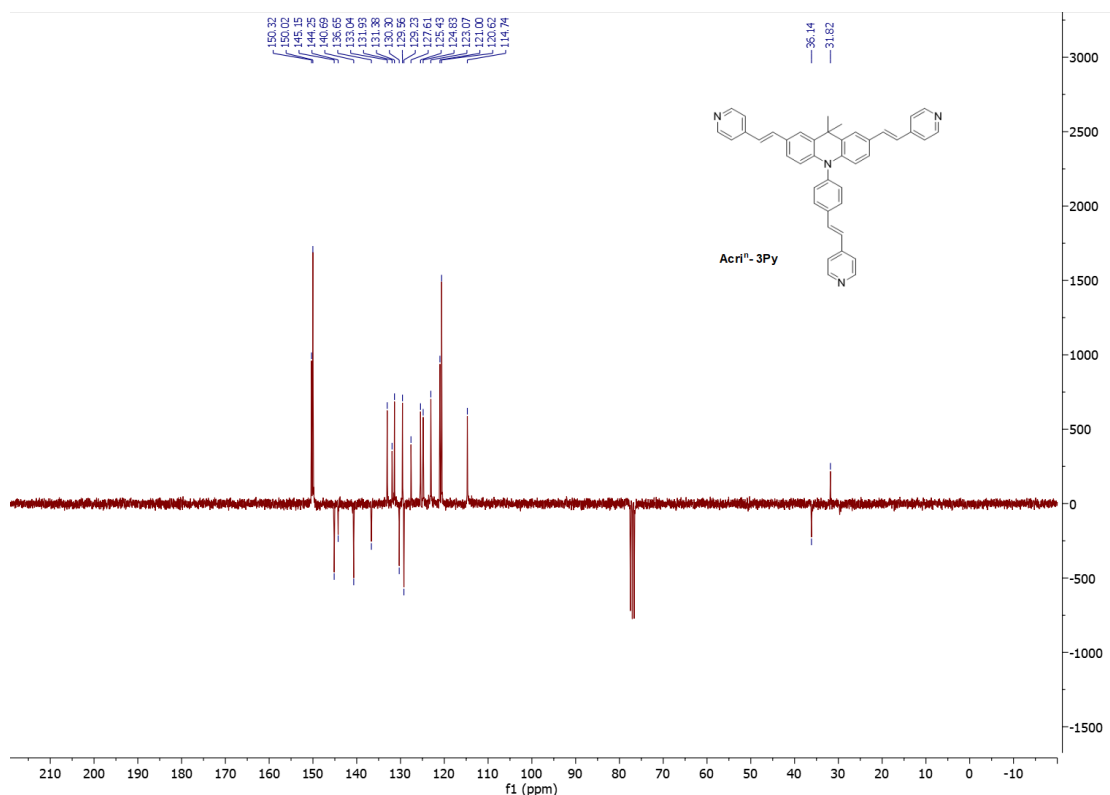
^{13}C APT NMR spectra of **Acrin-3Py** in CDCl_3 (75 MHz):



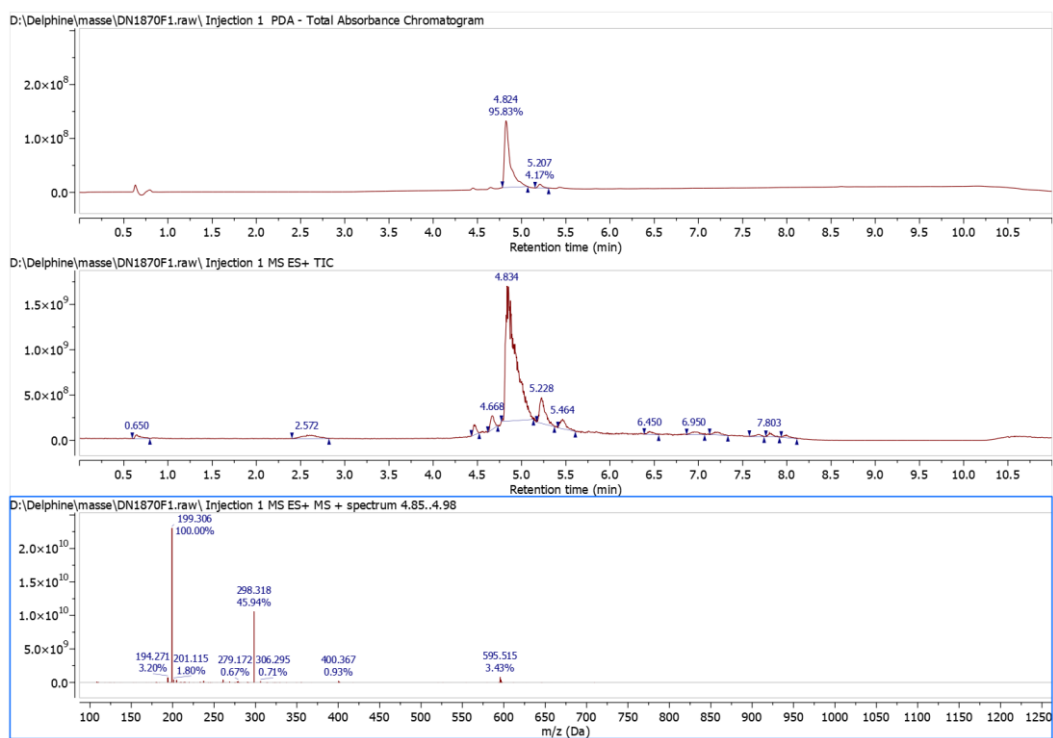
LC-MS of **Acrin-3Py**



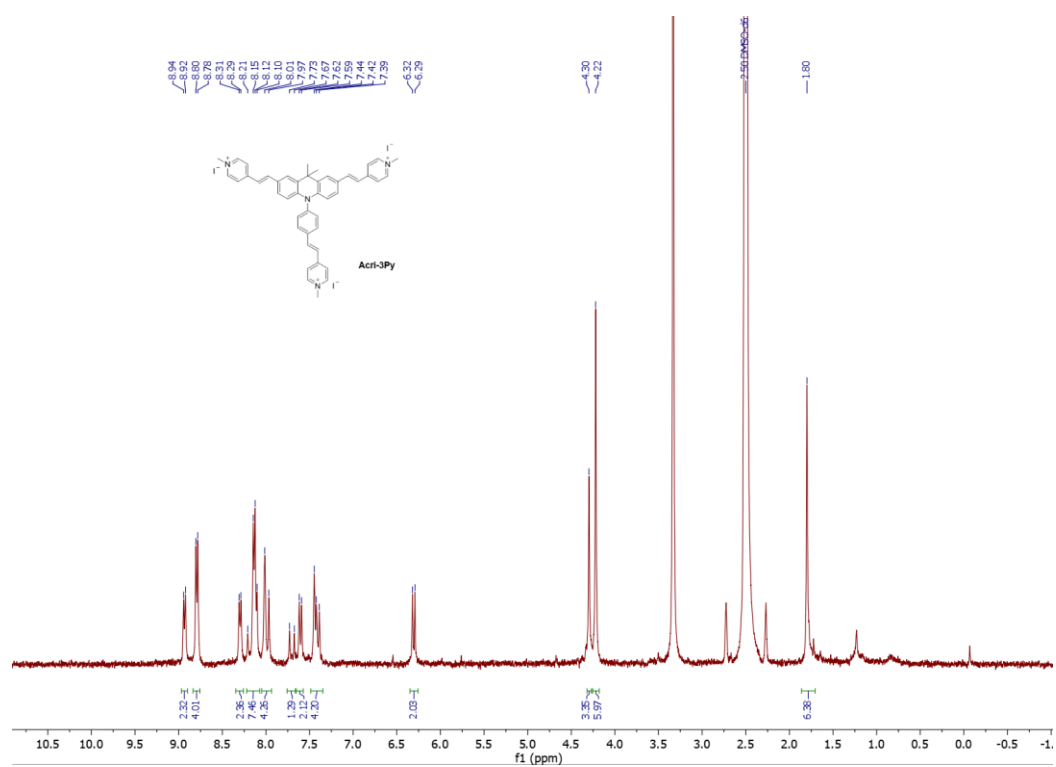
^{13}C APT NMR spectra of **Acrin-3Py** in CDCl_3 (75 MHz):



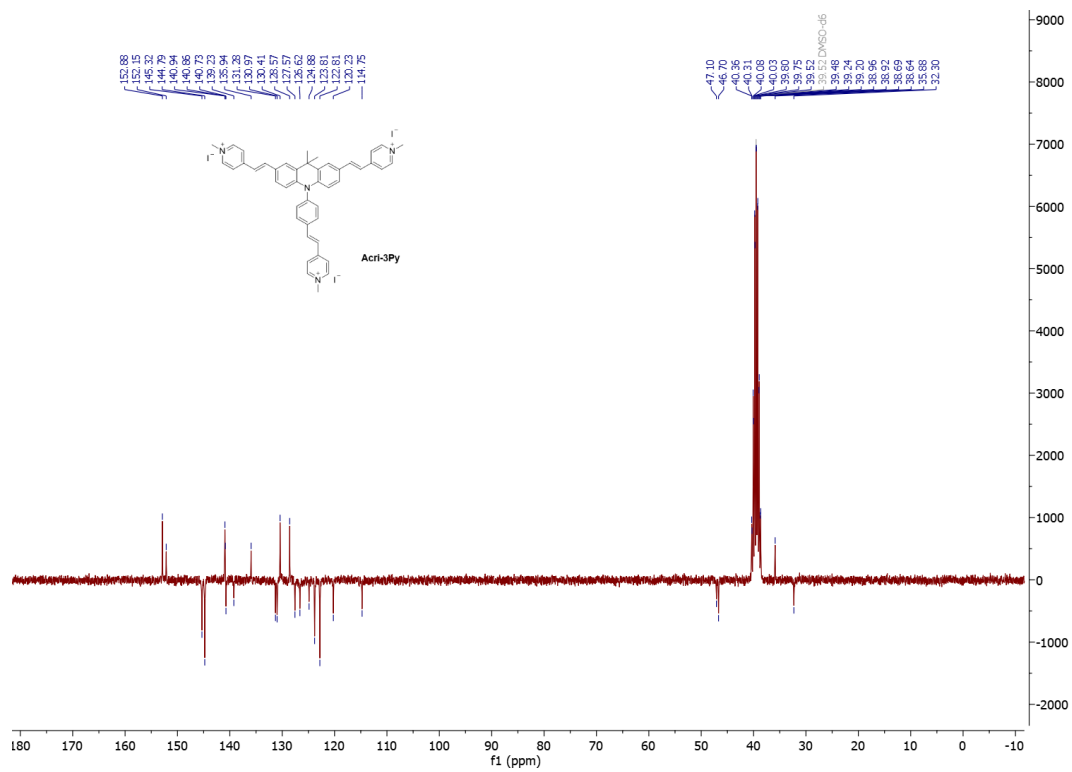
LC-MS of **Acrin-3Py**



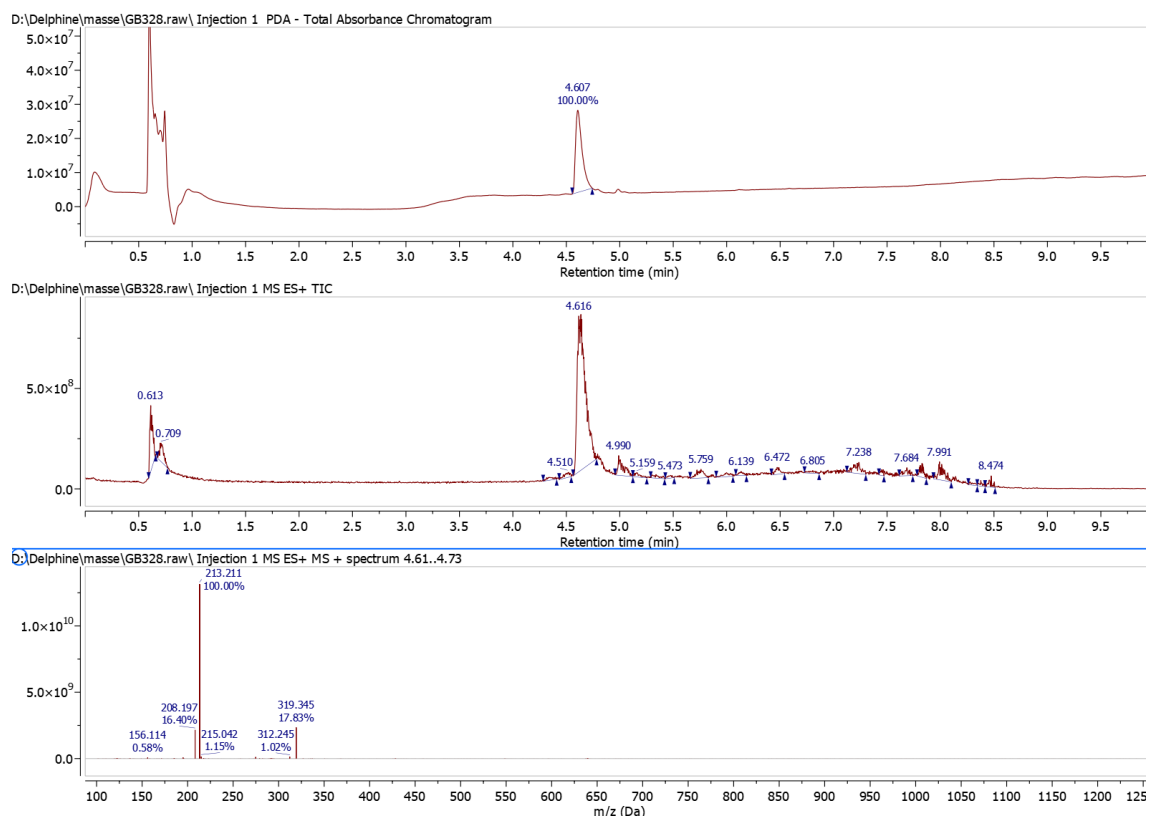
^1H NMR spectra of **Acrid-3Py** in DMSO-d₆ (300 MHz):



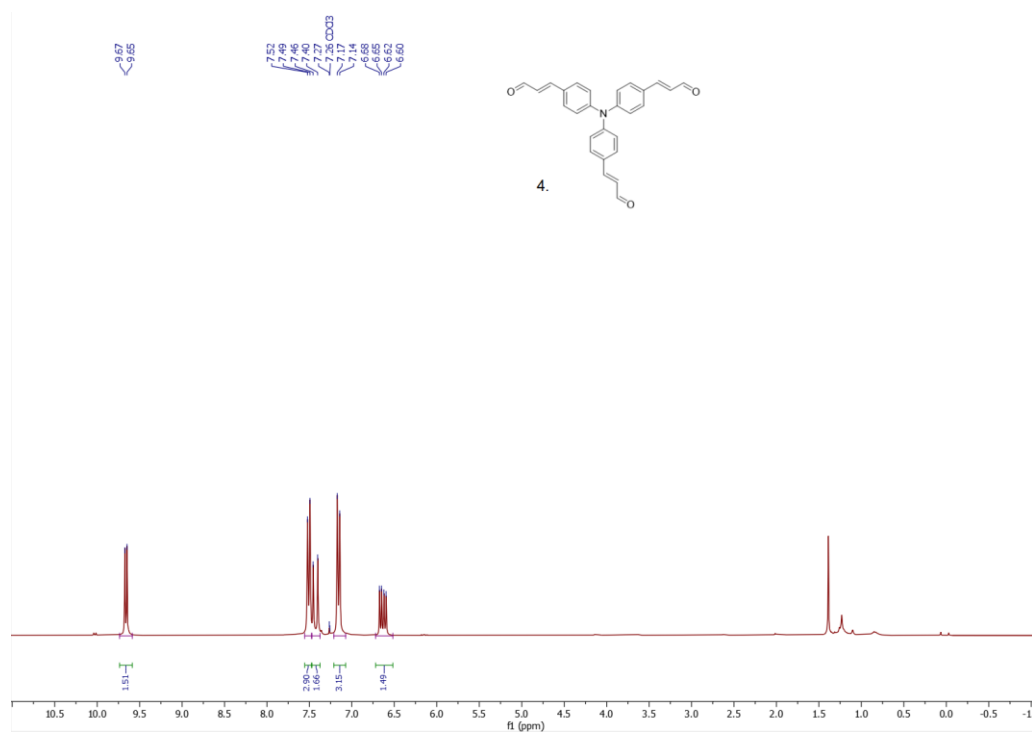
^{13}C APT NMR spectra of **Acrid-3Py** in DMSO-d₆ (75 MHz):



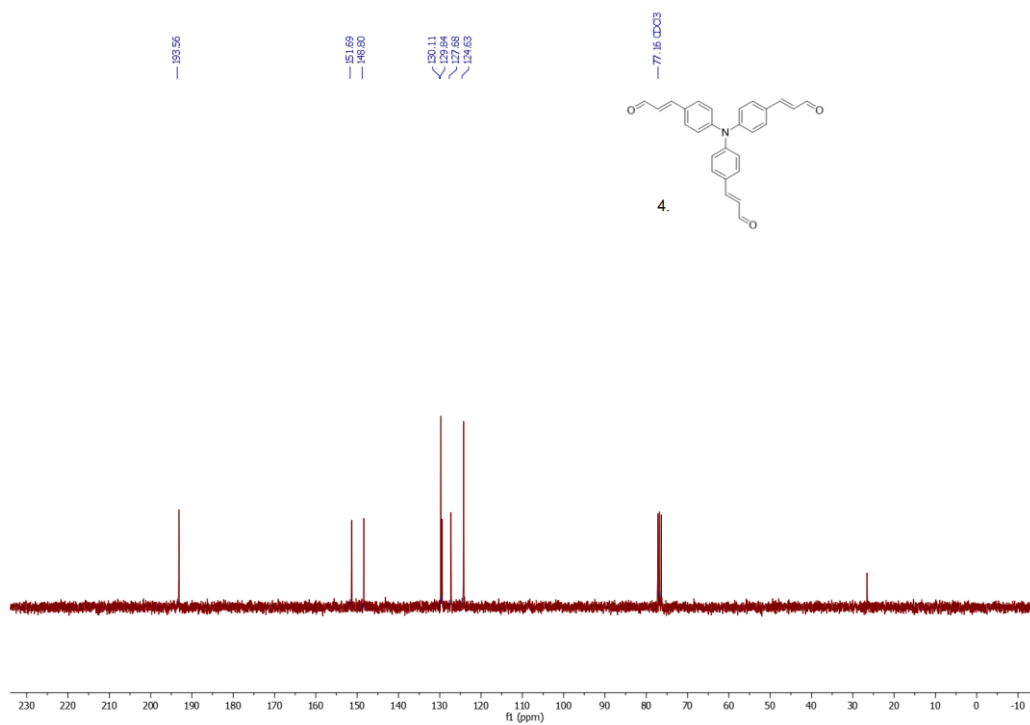
LC-MS of Acri-3Py



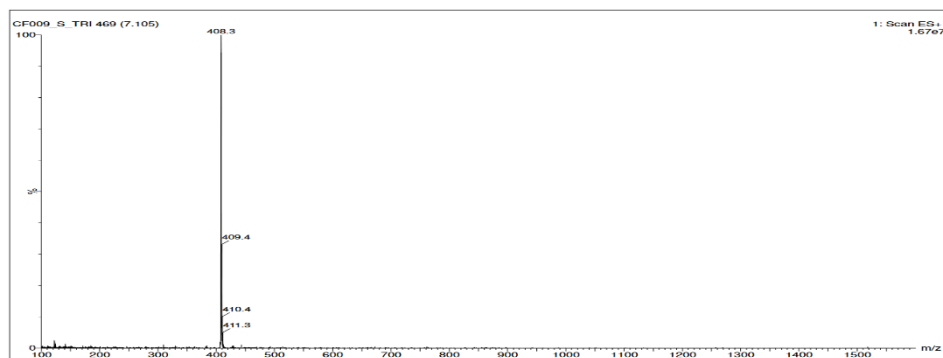
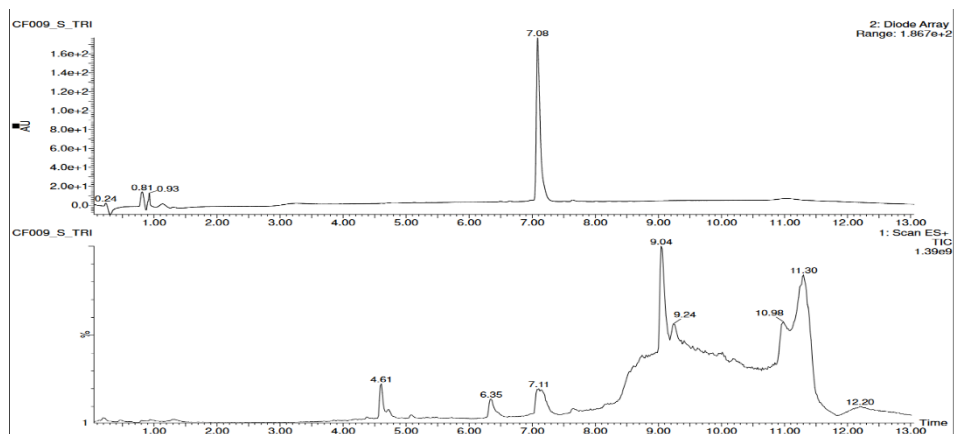
^1H NMR spectra of **4** in CDCl_3 (300 MHz):



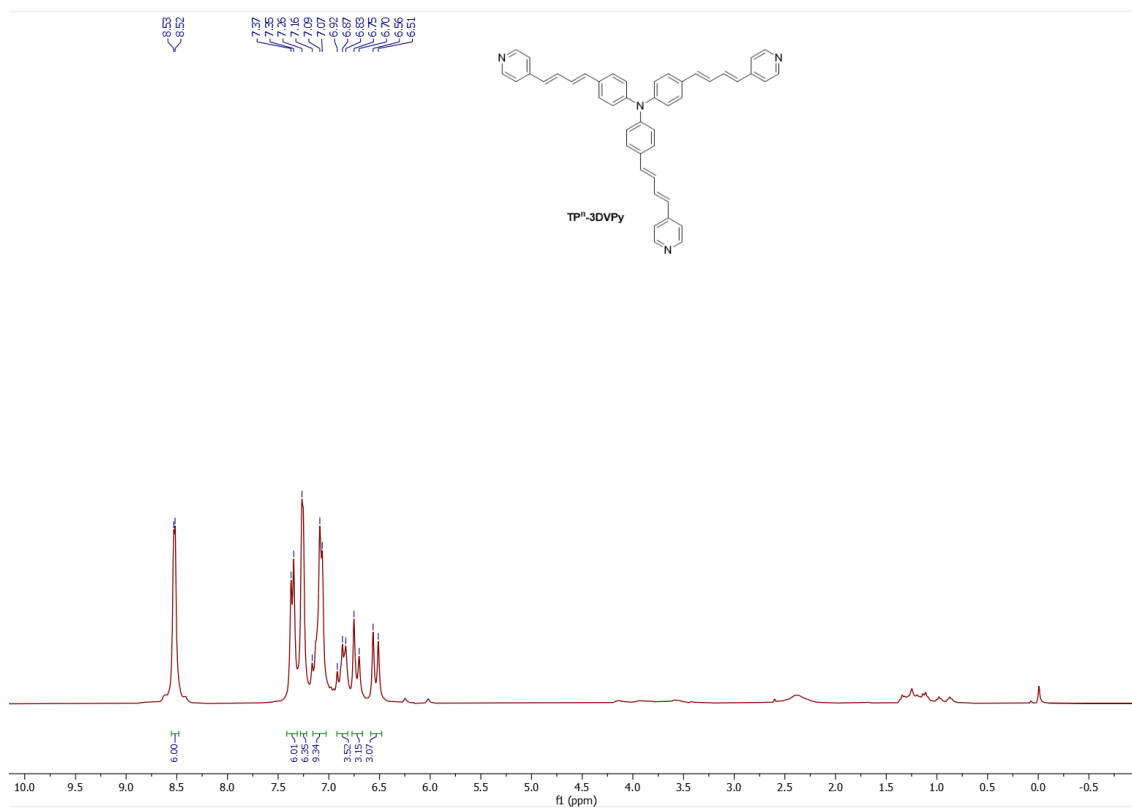
^{13}C NMR spectra of **4** in CDCl_3 (75 MHz):



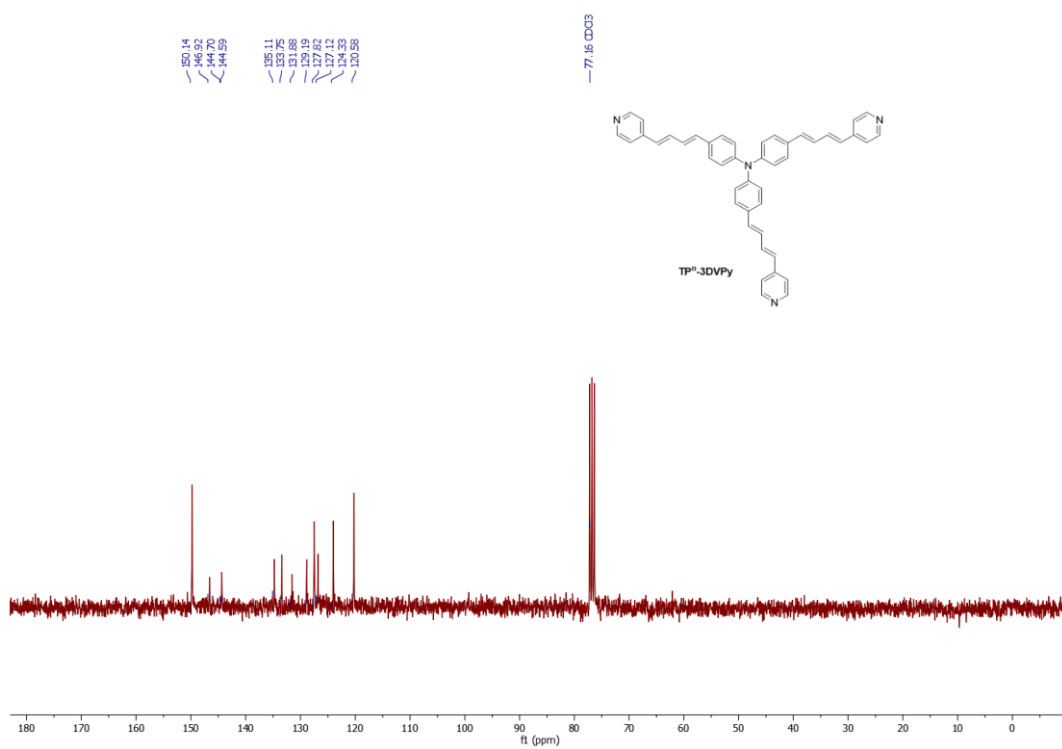
LC-MS of **4**



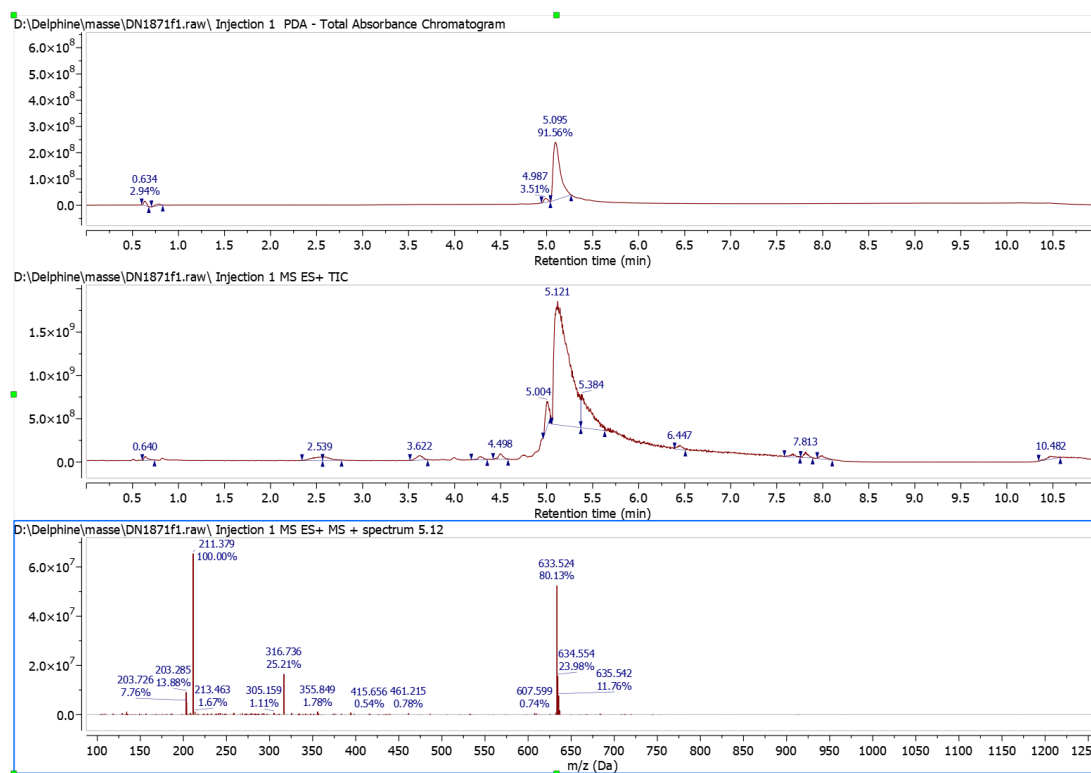
^1H NMR spectra of **TPn-3DVPy** in CDCl_3 (300 MHz):



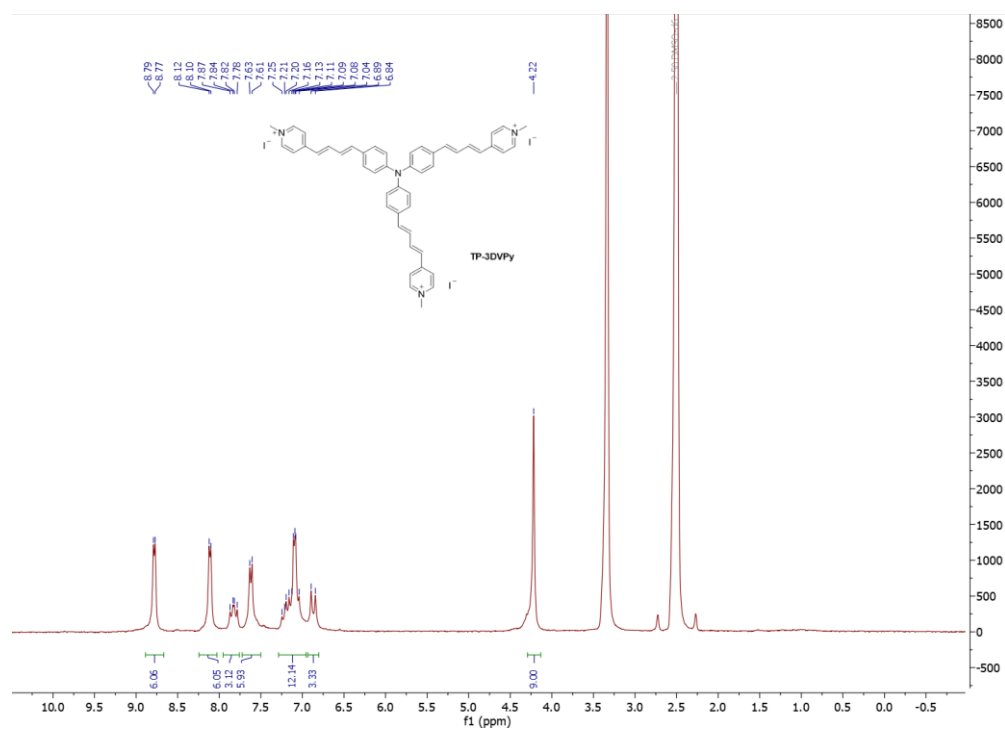
^{13}C NMR spectra of **TPn-3DVPy** in CDCl_3 (75 MHz):



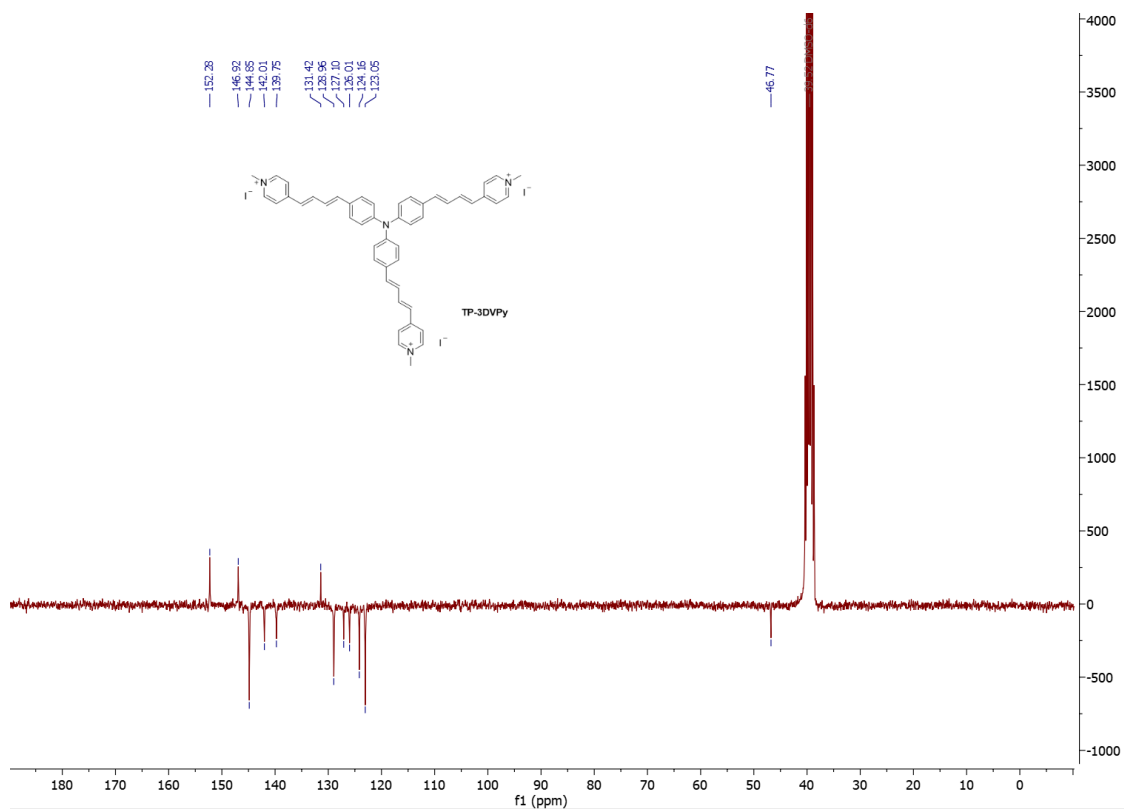
LC-MS of TPn-3DVPy



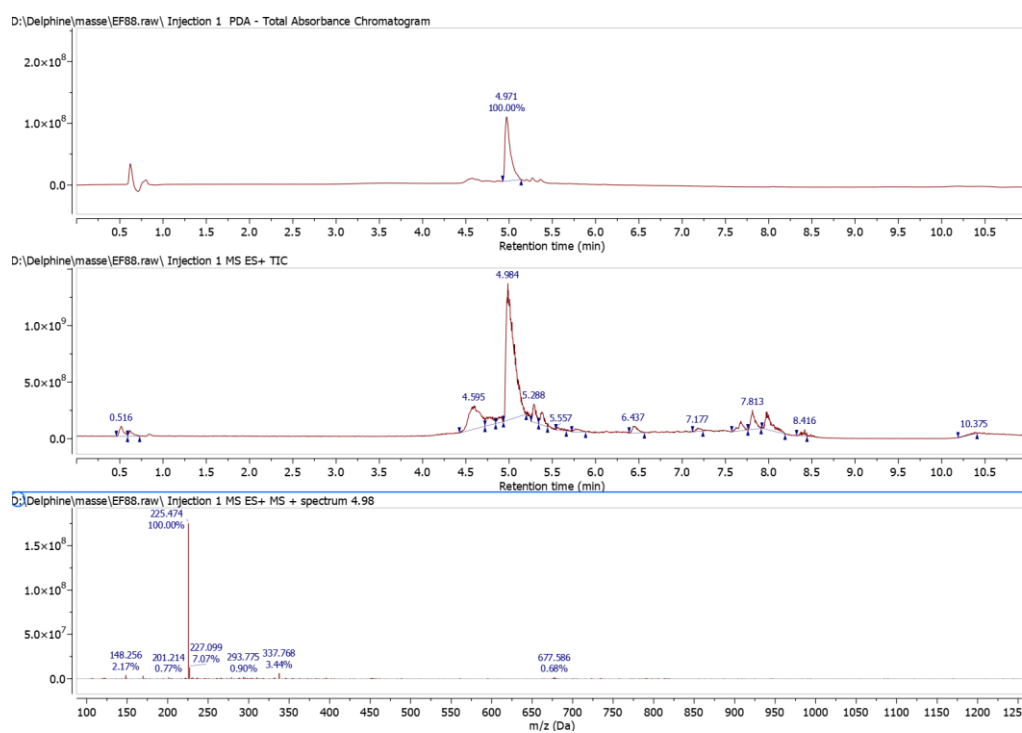
^1H NMR spectra of TP-3DVPy in DMSO-d₆ (300 MHz):



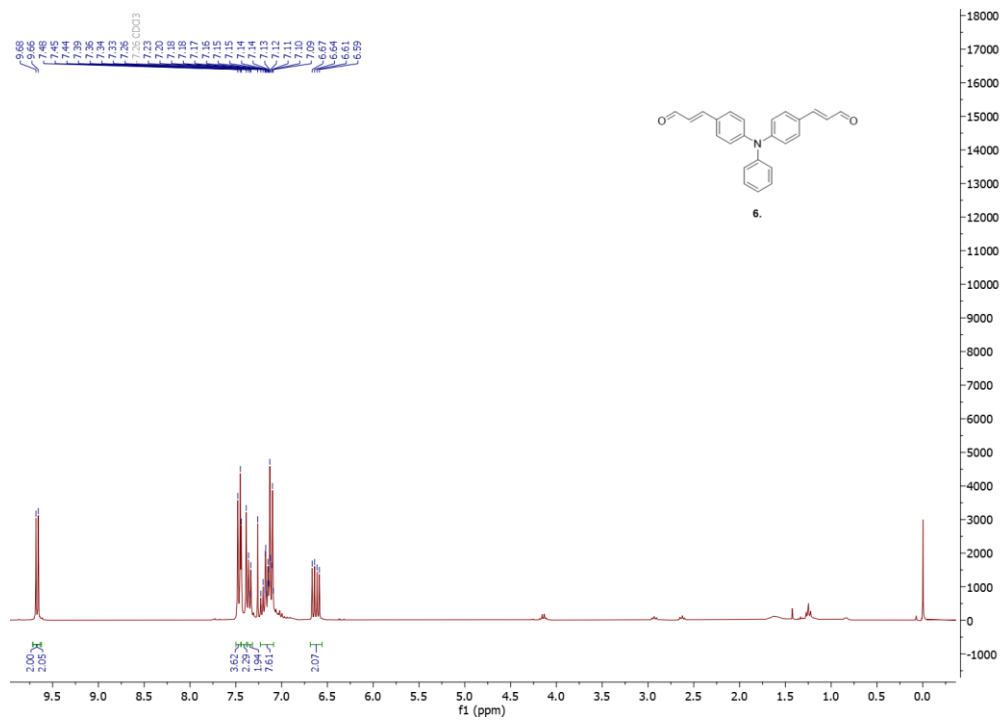
^{13}C APT NMR spectra of **TP-3DVPy** in DMSO-d₆ (75 MHz):



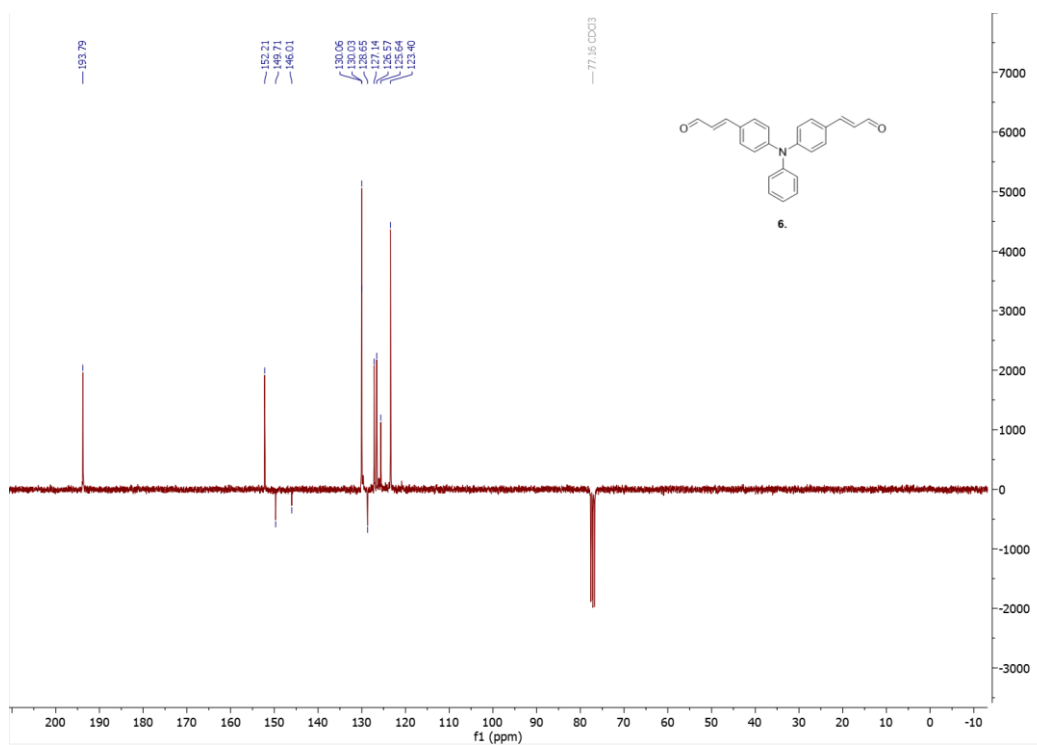
LC-MS of **TP-3DVPy**



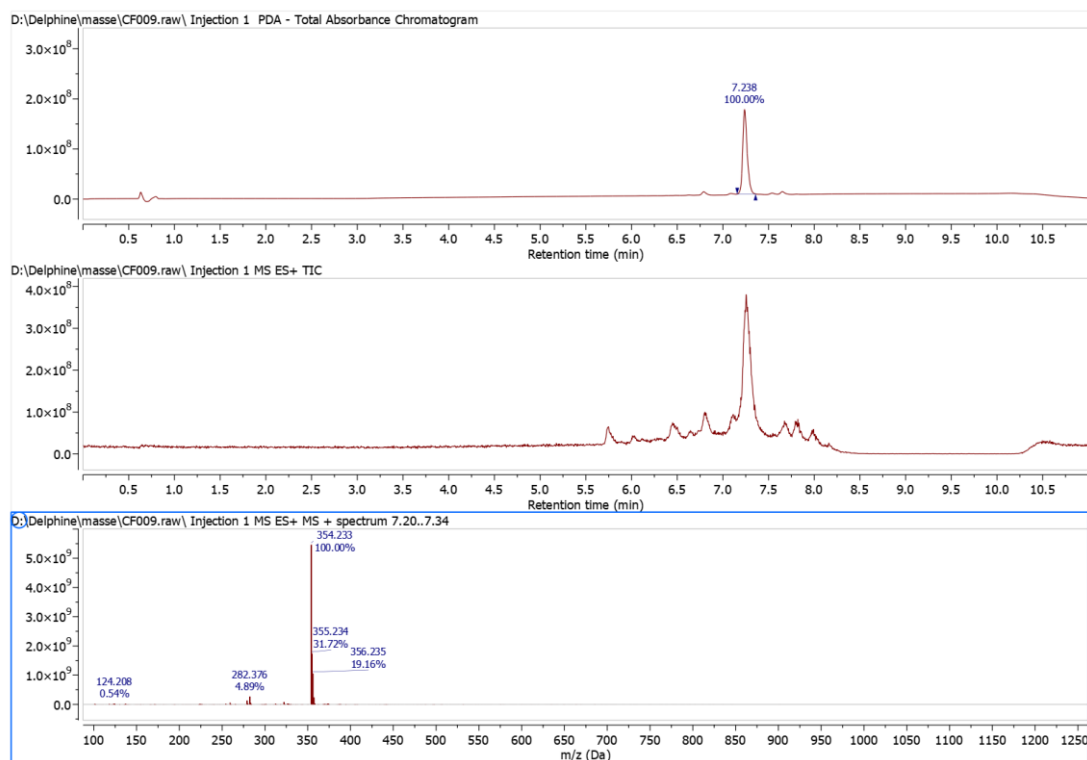
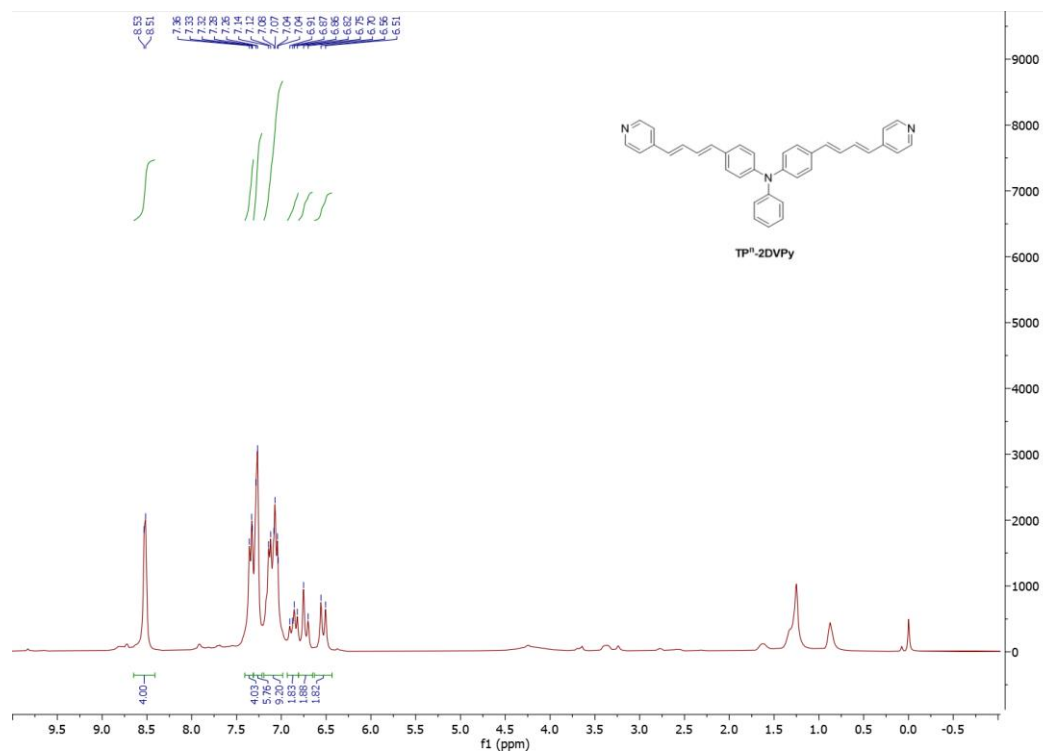
^1H NMR spectra of **6** in CDCl_3 (300 MHz):



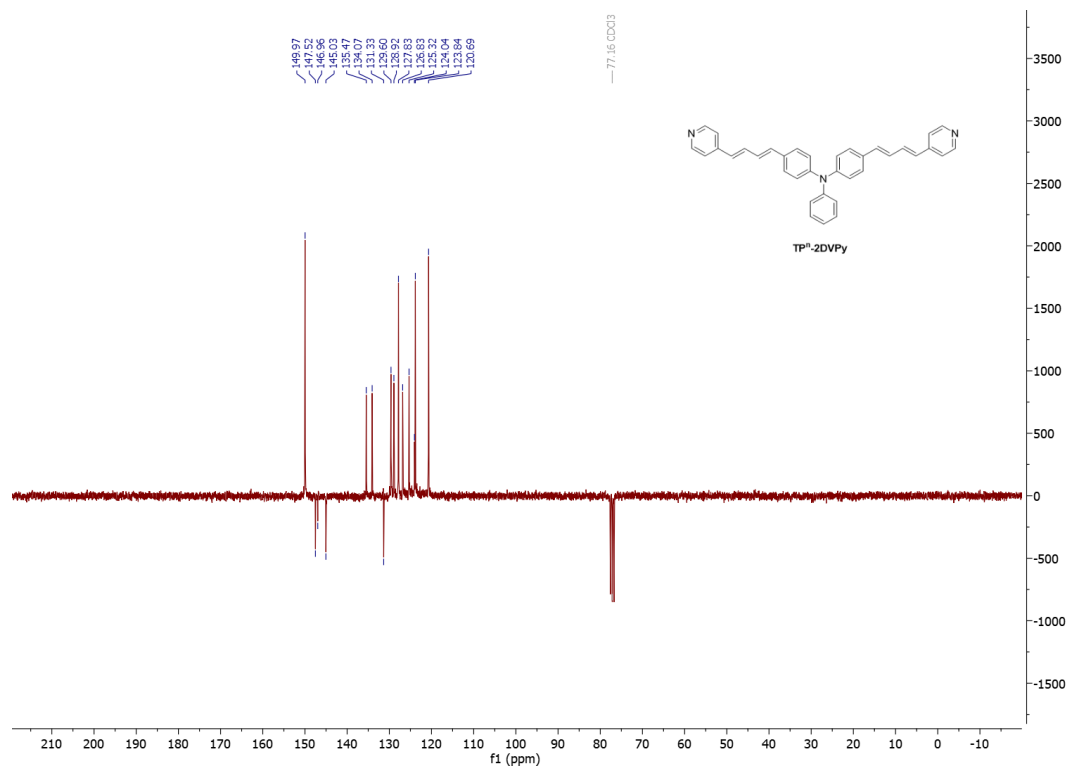
^{13}C APT NMR spectra of **6** in CDCl_3 (75 MHz):



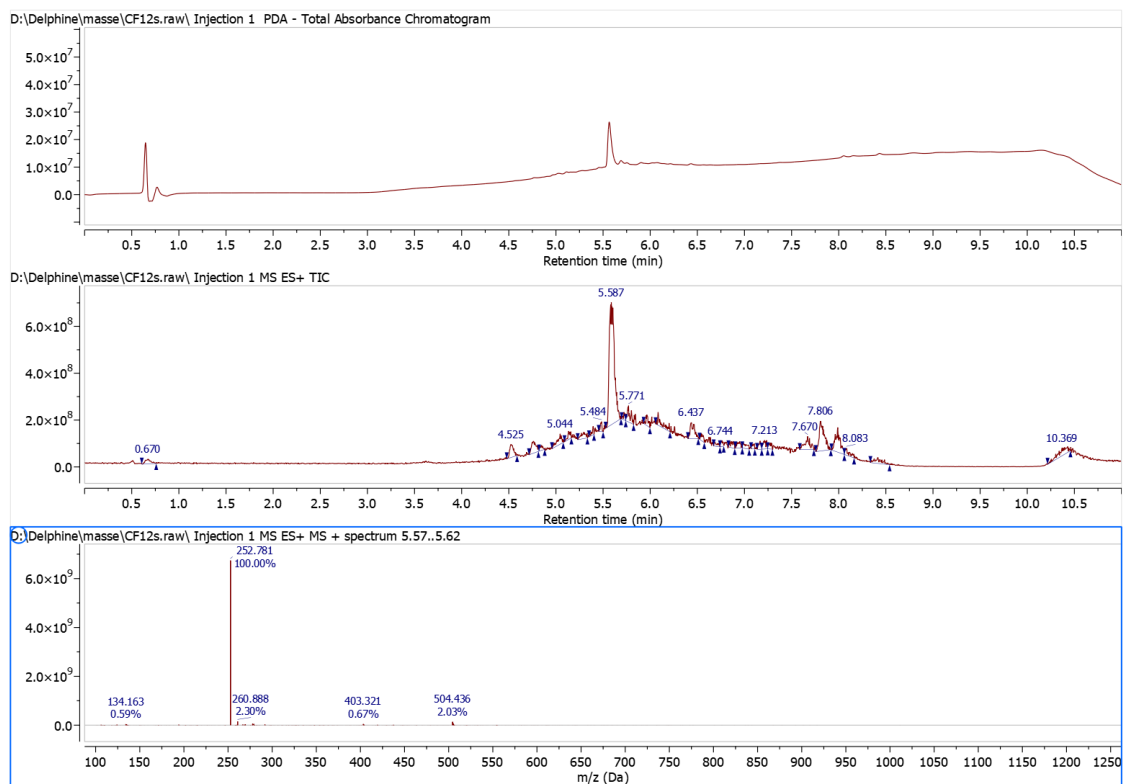
LC-MS of 6

 ^1H NMR spectra of TPn-3DVPy in CDCl_3 (300 MHz)

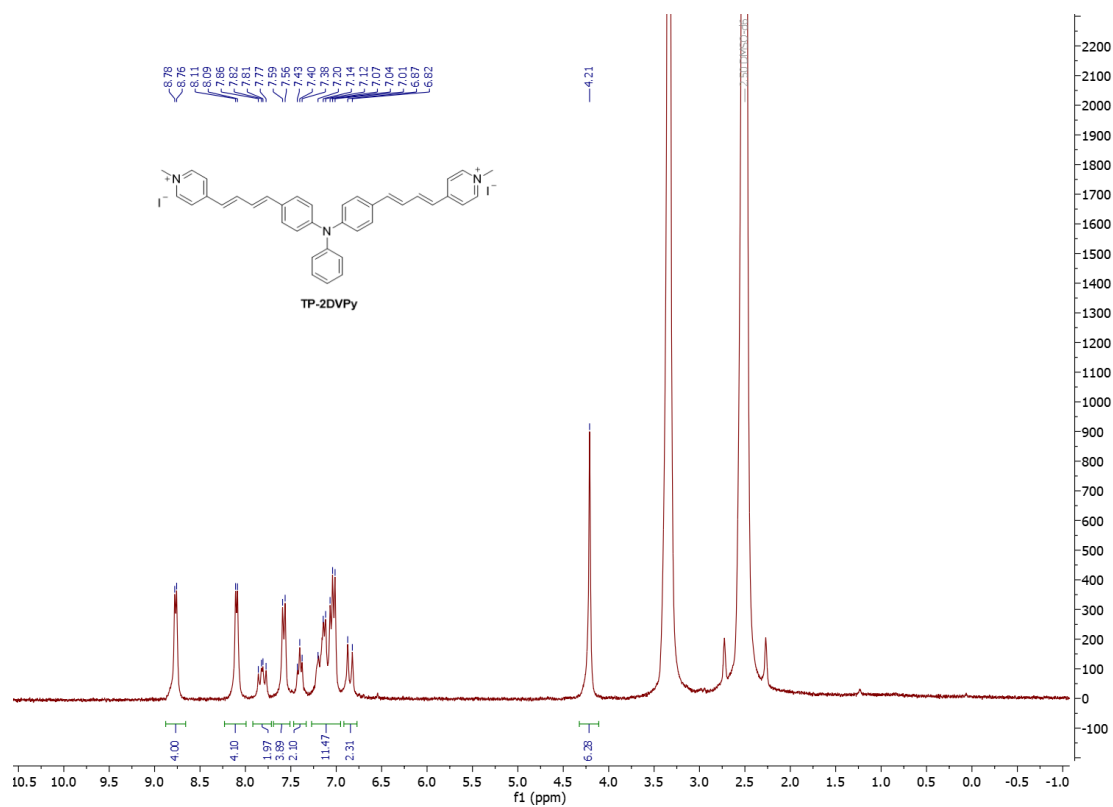
^{13}C APT NMR spectra of **TPn-3DVPy** in CDCl_3 (75 MHz):



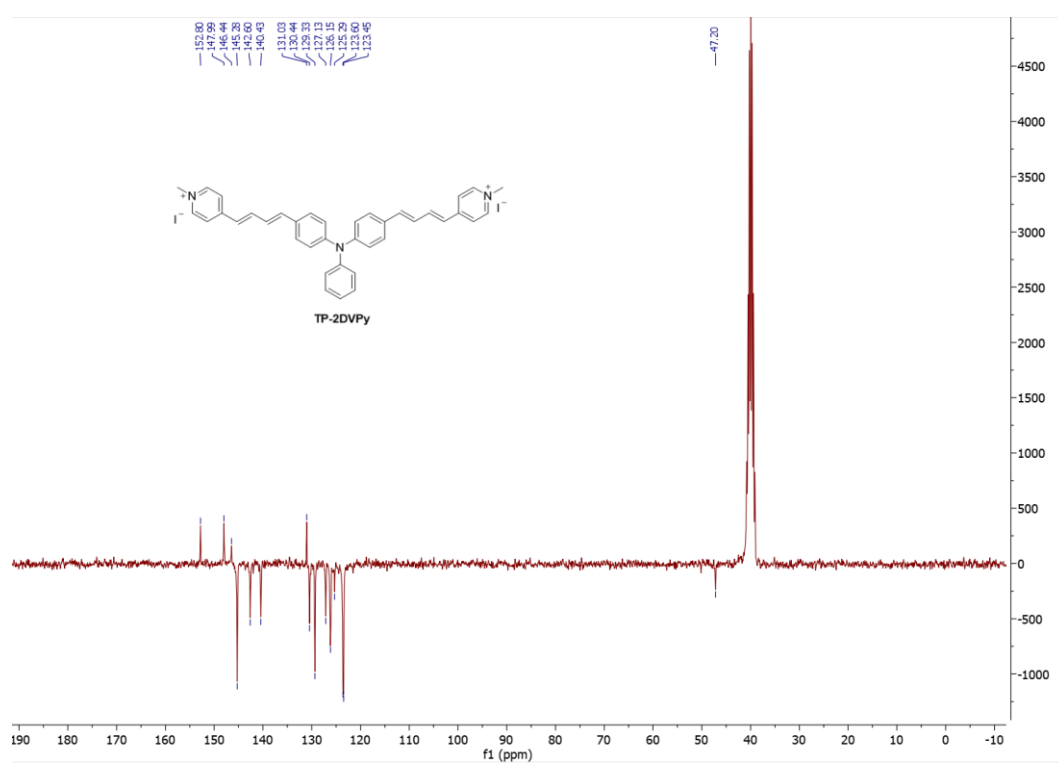
LC-MS of **TPn-3DVPy**:



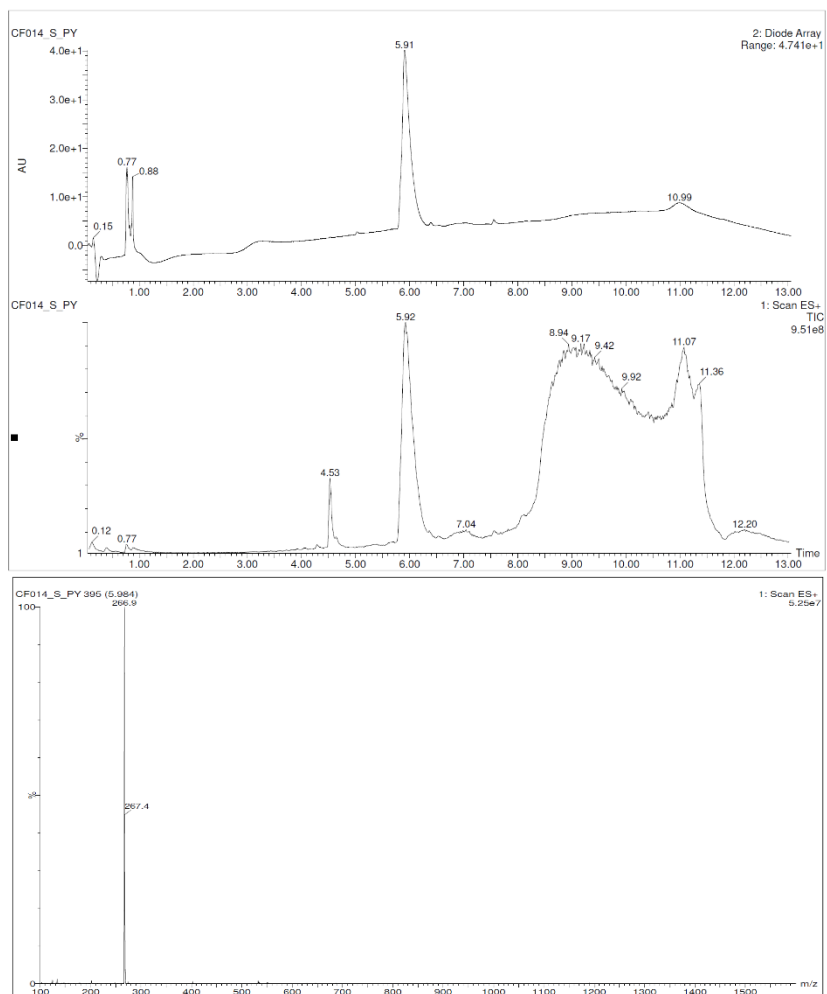
^1H NMR spectra of **TP-2DVPy** in DMSO-d₆ (300 MHz):



^{13}C APT NMR spectra of **TP-2DVPy** in DMSO-d₆ (75 MHz):



LC-MS of TP-2DVPy:



Chapter 4

Appendix 2: R-code for analysis of 14 proteins in PBS, serum and treated serum

```

# Repl.by.Proteins is a list with 4 components namely â..6â, â..8â, â..14â, â..18â™
#
# Each component, for example Repl.by.Proteins[[â..6â..]], is a table n x 13 x 17
# It has for colnames : Â« MeasA.WavelengthEms Â», Â« B Â», Â« C Â», Â« D Â», Â« E Â», Â« F Â», Â« G Â», Â«
B-CB7 Â», Â« C-CB7 Â», Â« D-CB7 Â», Â« E-CB7 Â», Â« F-CB7 Â», Â« G-CB7 Â»
#
# MeasA.WavelengthEms is the wavelength and over columns measures the signal intensity at a given wavelength
#
# It provides 6 replicates ('A','B','C','D','E','F','G') under two conditions (with or without CB7) for the
17 proteins of interest
#
# Finally, Repl.by.Proteins provides 6 replicates with or without CB7 for 4 components (â..6â,â..8â,
â..4â,â..18â) that is 6 replicates for 2 x 4 = 8 conditions

library(abind)
library(stringr)

input.dir = '../Fluorescence Data_TPcomps_27_01_2022/'
ncol.1 = 11
ncol.2 = 10
sub.dirs = paste0(input.dir,c('TP 6','TP 9','TP 14','TP 18'))

proteins_ctrl = c('Peroxidase','Pepsin','Diastase','Amylase','Lipase','Lysozyme','g-Glogulin','Myoglobin',
'Phosphatase','Ctrl.1','Pancreatin','Catalase','Chymotrypsin','BSA','Trypsin','Kidrolase',
'Fetuin','Dispase','Ctrl.2')

use.serum = FALSE
if (use.serum) sub.dirs = paste0(sub.dirs,'_Serum')

read_exp = function(name,use.CB7=F) {
  if (!use.CB7) name = str_remove(name,'_CB7')
  tmp = read.csv(name,skip=8,sep=',',header=T)
  stop = which(grepl('Basic',tmp$Well))
  tmp = tmp[-(stop:dim(tmp)[1]),]
  smp = tapply(tmp$Well,tmp$Well,function(w)
subset(tmp,tmp$Well==w)[,c("MeasA.WavelengthEms","MeasA.Result")])

  all_wells = str_sub(names(smp),1,1)
  wells = unique(all_wells)
  all.proteins = NULL
  for (well in wells) {
    tmp.protein = NULL
    for(i in which(all_wells==well))
      tmp.protein = abind(tmp.protein,smp[[i]],along=3)
    if (is.null(all.proteins)) {
      all.proteins = tmp.protein
    } else all.proteins = abind(all.proteins,tmp.protein[,-1,,drop=F],along=2)
  }
  dimnames(all.proteins)[[2]] = c('Wavelength','B','C','D','E','F','G')
  # print(head(all.proteins))
  all.proteins = apply(all.proteins,1:3,as.numeric)

  all.proteins
}

All = NULL
for (sub.dir in sub.dirs) {
  files = list.files(sub.dir,pattern='*_CB7.csv',recursive=T,full.names=T)
  name = str_split(files,'/')[[1]][3]
  print(name)
  # read the good lines in name
  All[[name]] = abind(read_exp(files[1]),read_exp(files[2]),along=3)
  All[[paste0(name,'_CB7')]] = abind(read_exp(files[1],use.CB7=T),read_exp(files[2],use.CB7=T),along=3)
  if (dim(All[[paste0(name,'_CB7')]])[3]==17) # the controls in CB7 are missing, thanks to Monika
    All[[paste0(name,'_CB7')]] = abind(All[[paste0(name,'_CB7')]][,1:9],All[[name]][,10,drop=F],
  All[[paste0(name,'_CB7')]][,10:17],All[[name]][,19,drop=F],along=3)
}

for (i in 1:length(All)) dimnames(All[[i]])[[3]] = proteins_ctrl

normalized.with.controls = function(TP) {
  TP[,2:7,1:9] = TP[,2:7,1:9] - apply(TP[,2:7,10,drop=F],1,mean)
  TP[,2:7,11:18] = TP[,2:7,11:18] - apply(TP[,2:7,19,drop=F],2,mean)
  TP = TP[,-c(10,19)]
}

```

```

All.norm = lapply(All,normalized.with.controls)

prepare.table =function(TP.name) {
  TP = All.norm[[TP.name]]
  out = NULL
  for (i in 1:dim(TP)[3]) {
    new = t(TP[,-1,i])
    rownames(new) = rep(dimnames(TP)[[3]][i],1=dim(new)[1])
    out = rbind(out, new)
  }
  colnames(out) = paste(str_remove(TP.name, ' '),TP[,1,1])
  out
}

All.in.table = NULL
for (name in names(All.norm)) All.in.table = cbind(All.in.table,prepare.table(name))

save(All,All.norm,All.in.table,file='Repl-by-Proteins.rda')

df=data.frame(id=c(1,1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4,4,5,5,5,5,5,6,6,6,6,6,7,7,7,7,7,8,8,
8,8,8,9,9,9,9,10,10,10,10,10,11,11,11,11,11,12,12,12,12,12,13,13,13,13,13,14,14,14,14,14,
14,15,15,15,15,15,16,16,16,16,16,17,17,17,17,17,17),

protname=c("peroxidase","peroxidase","peroxidase","peroxidase","peroxidase","peroxidase","pepsin","pepsin",
"pepsin","pepsin","pepsin","pepsin","diastase","diastase","diastase","diastase","diastase","diastase","amyl
ase","amylase","amylase","amylase","amylase","amylase","lipase","lipase","lipase","lipase","lipase","lipase
","lysozyme","lysozyme","lysozyme","lysozyme","lysozyme","lysozyme","globulin","globulin","globulin","globu
lin","globulin","globulin","myoglobin","myoglobin","myoglobin","myoglobin","myoglobin","myoglobin","phospha
tase","phosphatase","phosphatase","phosphatase","phosphatase","phosphatase","pancreatin","pancreatin","panc
reatin","pancreatin","pancreatin","pancreatin","catalase","catalase","catalase","catalase","catalase","cata
lase","chymotrypsin","chymotrypsin","chymotrypsin","chymotrypsin","chymotrypsin","chymotrypsin","BSA","BSA"
,"BSA","BSA","BSA","BSA","trypsin","trypsin","trypsin","trypsin","trypsin","trypsin","trypsin","trypsin","kidrolase","kidrolase
","kidrolase","kidrolase","kidrolase","kidrolase","fetuin","fetuin","fetuin","fetuin","fetuin","fetuin","di
spase","dispase","dispase","dispase","dispase","dispase"))

A <- cbind(df, All.in.table)
A <- A[,-1]

save(A, All,All.norm,All.in.table,file='Repl-by-Proteins.rda')

#####
#### PBS, serum, pretreated serum and 14 proteins ####

load("C:/Users/natha/Desktop/ANR PEfingerprint/shared datas/TPA paper/R-files/Repl-by-Proteins.rda")

A_16 = A[-(37:42),]
A_14 = A_16[-(73:84),]

save(A_14,file='Repl-by-14Proteins.rda')

library("readxl")
library('ggplot2')
library('randomForest')
library('MASS')
library('caTools')
library('caret')
library('factoextra')
library('tidyverse')

ind.var = grep('TP',names(A_14))

#### RANDOM FORESTS OOB whole spectra ####

prot.forest = randomForest(A_14[ind.var],
as.factor(A_14$protname),ntree=20000,replace=FALSE,importance=TRUE,proximity=TRUE, do.trace = TRUE)
prot.forest
importance(prot.forest)
plot(prot.forest)
varImpPlot(prot.forest)

#####

#### create sub folder with selected wavelengths

####~10(13) & 5(26) wavelengths for each TP/ TP + CB combination

protname = A_14[,1]
subset_col = seq(2, ncol(A_14), by=26)
sub_wl = cbind(protname, A_14[,subset_col])

ind.var = grep('TP',names(sub_wl))

####LDA####

```

```

whole.data=as.data.frame(sub_wl)

n.whole.data=dim(whole.data)[1]
n.train=round(0.7*n.whole.data)

ind=sample(n.whole.data,n.train,replace = FALSE)
n.s8=dim(sub_wl)[1]

library(MASS)
model.lda = lda(sub_wl[2:41], as.factor(whole.data$protname),CV=TRUE) # using leave-one-out cross-validation
mean(model.lda$class!=whole.data$protname)
model.lda

#define data to plot
lda_plot <- cbind(whole.data[2:41], predict(model.lda)$x)

#create plot
ggplot(lda_plot, aes(LD1, LD2,color=whole.data$protname)) + geom_point(size = 3) +
  stat_ellipse(level = 0.95, geom = "polygon", alpha = 0.2) +
  labs(x = "LDA Score 1", y = "LDA Score 2", color = "Group") +
  theme_bw()

#### train and test####
set.seed(31)

error.rates.lda=replicate(1000,{

  Ind.train=sample(n.whole.data,n.train,replace=F)
  model.lda.train=lda(whole.data[,2:41],as.factor(whole.data$protname),subset=Ind.train)
  mean(predict(model.lda.train,whole.data[-Ind.train,2:41])$class!=whole.data$protname[-Ind.train])

})
mean(error.rates.lda)

#### RANDOM FORESTS OOB for 10/5 wavelengths ####

library("readxl")
library('ggplot2')
library('randomForest')
library('MASS')
library('caTools')
library('caret')
library('factoextra')
library('tidyverse')

prot.forest = randomForest(sub_wl[ind.var],
as.factor(sub_wl$protname),ntree=20000,replace=FALSE,importance=TRUE,proximity=TRUE, do.trace = TRUE)
prot.forest
importance(prot.forest)
plot(prot.forest)
varImpPlot(prot.forest)

the.good.var = names(sort(importance(prot.forest)[,'MeanDecreaseAccuracy'],decreasing = T)[1:16])
mini.foret = randomForest(sub_wl[,the.good.var],as.factor(sub_wl$protname),ntree=10000,importance=TRUE)
mini.foret
varImpPlot(mini.foret)

####write output CSV for Systat with 5 WL####

write.csv(sub_wl, "5_wl_PBS.csv", row.names = FALSE) library('readxl')

```

Appendix 3: R-code for analysis of different concentrations for Phenylketonuria model

```

library('writexl')
library('ggplot2')
library('e1071')
library('caTools')
library('caret')
library('factoextra')
library('tidyverse')
library('randomForest')
library('FactoMineR')
library('plotly')
library('MASS')

```

```

library('plot3D')
library('reshape2')
library('dplyr')
library('tidyr')

TP_Phen = read_excel("Phen_discrimination_I_Io.xlsx")
colnames(TP_Phen)[1] <- "conc"
TP_Phen = TP_Phen[-c(55:108),]

ind.var = grep('TP',names(TP_Phen))

#### RANDOM FORESTS OOB ####

prot.forest = randomForest(TP_Phen[ind.var],
as.factor(TP_Phen$conc),ntree=20000,replace=FALSE,importance=TRUE,proximity=TRUE, do.trace = TRUE)
prot.forest
importance(prot.forest)
plot(prot.forest)
varImpPlot(prot.forest)

#####
#### create sub folder with selected wavelengths

####~10(13) & 5(26) wavelengths for each TP/ TP + CB combination

conc = TP_Phen[,1]
subset_col = seq(2, ncol(TP_Phen), by=26)
sub_wl = cbind(conc, TP_Phen[,subset_col])

ind.var = grep('TP',names(sub_wl))

####LDA#####
whole.data=as.data.frame(sub_wl)

n.whole.data=dim(whole.data)[1]
n.train=round(0.7*n.whole.data)

ind=sample(n.whole.data,n.train,replace = FALSE)
n.s8=dim(sub_wl)[1]

library(MASS)
model.lda = lda(sub_wl[2:21], as.factor(whole.data$conc),CV=TRUE) # using leave-one-out cross-validation
mean(model.lda$class!=whole.data$conc)
model.lda

#### train and test ####

set.seed(31)
error.rates.lda=replicate(1000,{

  Ind.train=sample(n.whole.data,n.train,replace=F)
  model.lda.train=lda(whole.data[,2:21],as.factor(whole.data$conc),subset=Ind.train)
  mean(predict(model.lda.train,whole.data[-Ind.train,2:21])$class!=whole.data$conc[-Ind.train])

})
mean(error.rates.lda)

#### RANDOM FORESTS OOB for 10/5 wavelengths ####

library("readxl")
library('ggplot2')
library('randomForest')
library('MASS')
library('caTools')
library('caret')
library('factoextra')
library('tidyverse')

prot.forest = randomForest(sub_wl[ind.var],
as.factor(sub_wl$conc),ntree=20000,replace=FALSE,importance=TRUE,proximity=TRUE, do.trace = TRUE)
prot.forest
importance(prot.forest)
plot(prot.forest)
varImpPlot(prot.forest)

the.good.var = names(sort(importance(prot.forest)[,'MeanDecreaseAccuracy'],decreasing = T)[1:16])
mini.foret = randomForest(sub_wl[,the.good.var],as.factor(sub_wl$conc),ntree=10000,importance=TRUE)
mini.foret
varImpPlot(mini.foret)

####write output CSV for Systat with 5 WL####

write.csv(sub_wl, "5_wl_Phen_5wl_I_Io.csv", row.names = FALSE)

```

```
#####
#### LDA to identify predicted vs identified class ####

####train set

Phen_Train = read_excel("Phen_Train_I_Io.xlsx")
colnames(Phen_Train)[1] <- "conc"
ind.var = grep('TP',names(Phen_Train))

conc = Phen_Train[,1]
subset_col = seq(2, ncol(Phen_Train), by=26)
sub_wl = cbind(conc, Phen_Train[,subset_col])
ind.var = grep('TP',names(sub_wl))

whole.data=as.data.frame(sub_wl)
n.whole.data=dim(whole.data)[1]
n.train=round(0.7*n.whole.data)
ind=sample(n.whole.data,n.train,replace = FALSE)
n.s8=dim(sub_wl)[1]

####Test set

Phen_Test = read_excel("Phen_Test_I_Io.xlsx")
colnames(Phen_Test)[1] <- "conc"
ind.var = grep('TP',names(Phen_Test))

conc = Phen_Test[,1]
subset_col = seq(2, ncol(Phen_Test), by=26)
sub_wlt = cbind(conc, Phen_Test[,subset_col])
ind.var = grep('TP',names(sub_wlt))

####train and test aet prediction

set.seed(51)
error.rates.lda <- replicate(1000, {
  ind.train <- sample(nrow(whole.data), size = nrow(whole.data), replace = FALSE)
  ind.test <- sample(nrow(sub_wlt), size = nrow(sub_wlt), replace = FALSE)

  model.lda.train <- lda(whole.data[ind.train, 2:21], as.factor(whole.data$conc[ind.train]))
  test_predictions <- predict(model.lda.train, newdata = sub_wlt[ind.test, 2:21])

  mean(test_predictions$class != sub_wlt$conc[ind.test])
})

mean_test_error <- mean(error.rates.lda)
cat("Mean Test Error:", mean_test_error, "\n")
conf_matrix <- table(Actual = sub_wlt$conc[ind.test], Predicted = test_predictions$class)
print(conf_matrix)

####prediction confusion matrix

ind.train <- sample(nrow(whole.data), size = nrow(whole.data), replace = FALSE)
ind.test <- sample(nrow(sub_wlt), size = nrow(sub_wlt), replace = FALSE)

model.lda.train <- lda(whole.data[ind.train, 2:21], as.factor(whole.data$conc[ind.train]))
test_predictions <- predict(model.lda.train, newdata = sub_wlt[ind.test, 2:21])

mean(test_predictions$class != sub_wlt$conc[ind.test])

conf_matrix <- table(Actual = sub_wlt$conc[ind.test], Predicted = test_predictions$class)
print(conf_matrix)
```

Appendix 4: R-code for analysis of different concentrations of Preeclampsia samples

```
library('readxl')
library('writexl')
library('ggplot2')
library('e1071')
library('caTools')
library('caret')
library('factoextra')
library('tidyverse')
library('randomForest')
library('FactoMineR')
library('plotly')
library('MASS')
library('plot3D')
library('reshape2')
library('dplyr')
```

```

library('tidyr')

TP_PE = read_excel("PE_nonPE_data_TP 6,9,14,18_wl selection.xlsx", sheet=1)
ind.var = grep('w',names(TP_PE))

#### RANDOM FORESTS OOB ####

prot.forest = randomForest(TP_PE[ind.var],
as.factor(TP_PE$ID),ntree=20000,replace=FALSE,importance=TRUE,proximity=TRUE, do.trace = TRUE)
prot.forest
importance(prot.forest)
plot(prot.forest)
varImpPlot(prot.forest)

#####
#### create sub folder with selected wavelengths #####

####~10(13) & 5(26) wavelengths for each TP/ TP + CB combination

ID = TP_PE[,1]
subset_col = seq(2, ncol(TP_PE), by=26)
sub_wl = cbind(ID, TP_PE[,subset_col])

ind.var = grep('w',names(sub_wl))

####LDA####

whole.data=as.data.frame(sub_wl)

n.whole.data=dim(whole.data)[1]
n.train=round(0.7*n.whole.data)

ind=sample(n.whole.data,n.train,replace = FALSE)
n.s8=dim(sub_wl)[1]

library(MASS)
model.lda = lda(sub_wl[2:33], as.factor(whole.data$ID),CV=TRUE) # using leave-one-out cross-validation
mean(model.lda$class!=whole.data$ID)
model.lda

#### train and test####

set.seed(31)
error.rates.lda=replicate(1000,{

  Ind.train=sample(n.whole.data,n.train,replace=F)
  model.lda.train=lda(whole.data[,2:33],as.factor(whole.data$ID),subset=Ind.train)
  mean(predict(model.lda.train,whole.data[-Ind.train,2:33])$class!=whole.data$ID[-Ind.train])

})
mean(error.rates.lda)

#### RANDOM FORESTS OOB for 10/5 wavelengths ####

library("readxl")
library('ggplot2')
library('randomForest')
library('MASS')
library('caTools')
library('caret')
library('factoextra')
library('tidyverse')

prot.forest = randomForest(sub_wl[ind.var],
as.factor(sub_wl$ID),ntree=20000,replace=FALSE,importance=TRUE,proximity=TRUE, do.trace = TRUE)
prot.forest
importance(prot.forest)
plot(prot.forest)
varImpPlot(prot.forest)

the.good.var = names(sort(importance(prot.forest)[,'MeanDecreaseAccuracy'],decreasing = T)[1:16])
mini.foret = randomForest(sub_wl[,the.good.var],as.factor(sub_wl$ID),ntree=10000,importance=TRUE)
mini.foret
varImpPlot(mini.foret)

####write output CSV for Systat with 5 WL####

write.csv(sub_wl, "TP_PE_nonPE_16wl_corrected.csv", row.names = FALSE)

```

Chapter 5

Appendix 5: R-code for analysis of proteins on droplet microfluidics platform

```

library(ggplot2)
library(randomForest)
library(e1071)
library(caTools)
library(caret)

train = read_xlsx("../Data/Train set_TP_CB_prot_1_2.xlsx")
test=read_xlsx("../Data/Test set_TP_CB_prot_1_2.xlsx")
test =as.data.frame(test)
train =as.data.frame(train)

n=dim(train)[1]
ind=sample(n,n,replace = FALSE)

colnames(train)=c("id","tpsix","tpsixcb","tpnine","tpninecb","tpfourteen","tpfourteencb", "tpeighteen",
"tpeighteencb")
colnames(test)=c("id","tpsix","tpsixcb","tpnine","tpninecb","tpfourteen","tpfourteencb", "tpeighteen",
"tpeighteencb")

train[2:9] <- lapply(train[2:9], as.numeric)
test[2:9] <- lapply(test[2:9], as.numeric)
train[1] <- lapply(train[1], as.factor)
test[1] <- lapply(test[1], as.factor)

#### RANDOM FORESTS ####

library(randomForest)

L=randomForest(as.factor(train$id)~.,data=train[-
1],ntree=1500,mtry=2,replace=FALSE,importance=TRUE,proximity=TRUE, do.trace = F)
L
predict(L,newdata=test[-1])
mean(as.factor(predict(L,newdata=test[-1])!=test$id))
mean((Id.test.predict)!=as.factor(test$id))

#### LDA####

library(MASS)

model.lda = lda(train[2:9],as.factor(train$id),CV=TRUE) # using leave-one-out cross-validation
mean(model.lda$class!=train$id)

set.seed(31)
error.rates.lda=replicate(1000,{

  Ind.train=sample(n,n,replace=F)
  model.lda.train=lda(train[,2:9],as.factor(train$id),subset=Ind.train)
  mean(predict(model.lda.train,test[,2:9])$class!=test$id)

})

mean(error.rates.lda)

model.lda.train=lda(train[,2:9],as.factor(train$id))

model.lda.test=predict(model.lda.train,test[-1])

```

Appendix 6: Protocol for sensing of protein on 384-well plate using the pipetting protocol

Initial test condition:

2 TPA TP derivatives

1 Cucurbit[7]uril CB[7] derivative

12 analytes A (11 proteins + 1 PBS control)

Final conditions in the well of a 384-well fluorescent plate:

2nd step: V= 32.5 μ L, TP 20 μ L + 10 μ L Water or 10 μ L CB[7] + 2.5 μ L analyte A

Read the final fluorescence

Consumables used

1 micronic tube plate (V=1.4 ml per tube)

1 96-well conical-bottom plate (GREINER 651201)

1 384-well plate for fluorescence

Boxes of 20 μ L tips

Boxes of 200 μ L tips for 384-well plates

1) Plate contents Reagents and Analytes

REAGENT plate (micronic 96-well plate)

	1	2	3	4	5	6	7	8	9	10	11	12
A	TP1	TP2	Water	CB[7]								
B	TP1	TP2	Water	CB[7]								
C	TP1	TP2	Water	CB[7]								
D	TP1	TP2	Water	CB[7]								
E	TP1	TP2	Water	CB[7]								
F	TP1	TP2	Water	CB[7]								
G	TP1	TP2	Water	CB[7]								
H	TP1	TP2	Water	CB[7]								

TP well filled with 650 μ L of TPA derivative solution = total volume 650*8 = 5.2 mL

Water well filled with 500 μ L of water = total volume 500*8= 4 mL

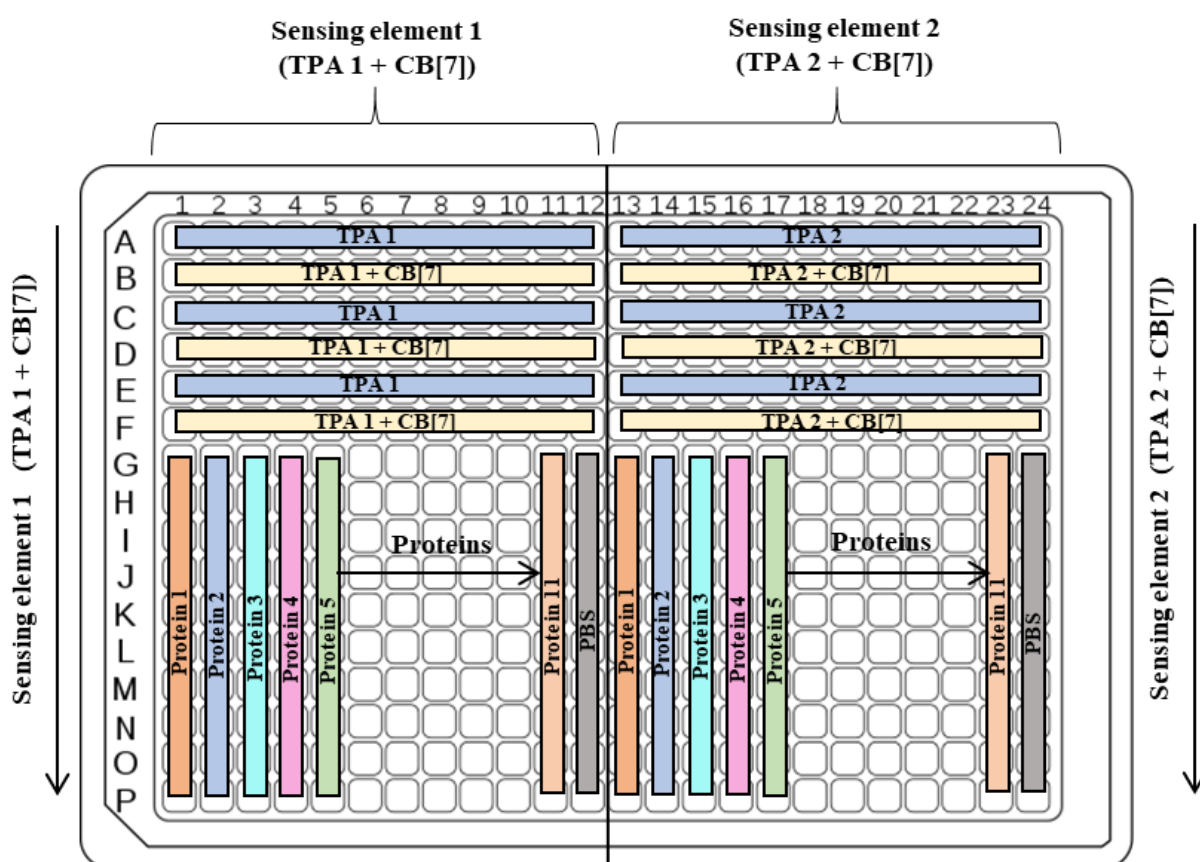
CB[7] well filled with 500 μ L of CB[7] = a total volume of 500*8= 4 mL

Note: Eliminate the vacuum-dispense mixing step, as this leaves a few drops on the tips when the plates are lifted. The plates will now need to be shaken on an orbital shaker before fluorescence reading.

Run time: approx. 38 min

4) Final plate layout for fluorescence measurements

Schematic illustration of 384-well plate layout prepared using the pipetting robot for the discrimination of 11 proteins spiked in 1X PBS.



Chapter 6

Appendix 7: R-code for PCA analysis of 8 analytes with 11-channel sensor array

```

library (Facto Mine R)
library (factoextra)
library (caret)
library (glmnet)
library (dplyr)
library ("corrplot")

data = read.csv2("PCA_5 -0806.csv ", header = TRUE, strings As Factors = TRUE)
data = data [c(1:24),c(2:12)] # to extract variables
group = data [c(1:24),1]
group

# list of functions
fviz _eig(res.pca , addlabels = TRUE , ylim = c(0,100))# contribution to dimension
fviz _cos2(res.pca , choice = "var", axes = 1:2) # quality of representation

var = get_pca_var (res.pca)
head(var$contrib , 11) # Contribution in number
corrplot (var$cos2, is.corr = FALSE)# quality representation for each dimension
corrplot (var$contrib, is.corr = FALSE) # contribution for each dimension

```

Appendix 8: R-code for LDA analysis of 8 analytes with 11-channel sensor array

```

library ( ' e1071 ' )
library ( ' caret ' )
library ( ' factoextra ' )
library ( ' tidyverse ' )
library ( ' random Forest' )
library ( ' Facto Mine R' )
library ( ' plotly ' )
library ( ' MASS ' )
library ( ' plot 3 D ' )
library ( ' car ' )

df <- read.csv2("PCA_05 -0806.csv", header = TRUE, strings As Factors = TRUE)

# Extract group labels and variable data
group_labels <- rep (1:8, each = 3) # Assume 8 groups, each represented by 3 lines
variables <- df [1:24, c(2:12)]

# Initialize vectors to store predicted classes and error rates "leave one out"
classes_predites <- vector("character", length(group_labels))
rate_errors <- numeric (length(group_labels))
rate_errors_groups <- numeric (8) # Initialize a vector to store the error rate for each group

# Perform cross validation leave one out
for (i in 1:length (group_labels)) {
  # Omit i th sample for test

  data_training <- variables [-i,]
  labels_training <- group_labels [-i]
  sample_test <- variables [i,]

  # Train the LDA model
  model_lda <- lda (data_training, tags_training)

  # Predict test sample class
  class_predict <- predict (modele_lda , echantillon_test)$class

  # Store the predicted class and calculate the error rate
  classes_predicted [i] <- class_predict
  rate_errors [i] <- class_predict != group_labels [i]

  # Calculate the error rate for each group
  if (classe_predite != group_labels [i]) {
    rate_errors_groups [group_labels [i]] <- rate_errors_groups[group_labels [i]] + 1
  }
}

# Calculate the overall leave - one - out error rate
rate_error_leave_one_out <- mean (rate_error)

# Calculate the error rate for each group

```

```

rate_errors_groups <- rate_errors_groups / 3

# Display predicted classes and error rate leave one out
cat ("Predicted classes :n")
print (classes_predicted)
cat ("Leave-One-Out Error Rate:", rate_error_leave_one_out , "n")

# Display the error rate for each group
cat ("Error rate per group :n")
for (i in 1:8) {
cat ("Group", i, ": ", rate_errors_groups [i], "\n")
}

# ##### GRAPH #####
# Perform LDA on the entire dataset
modele_lda_tout <- lda (variables, group_labels)

# Get LDA scores for all samples
scores_lda <- predict (modele_lda_tout)$x

# Create a data frame with LDA scores and group labels
data <- data.frame (LDA1 = scores_lda [,1] , LDA2 = scores_lda [,2] , LDA3 = scores_lda [,3] , Group = factor
(group_labels))

# Get group names from group labels
group <- df[c (1 : 24) ,1]
names_groups <- levels (factor (group))
names_groups

palette_personnalisee <- c(" # 0000 FF", "# FF 0000 ", " #00 FF00 ", "# FFA 500 ", "# FF 00 FF ", " #00 FFFF
", "# FFFF 00 ", " #800080 ")

## Create a graph
ggplot(data, aes(x = LDA1,y = LDA2,color = Group)) +
  geom_point (size=3) +
  stat_ellipse (level = 0.95, geom = " polygon ", alpha = 0.2) +
  labs (x = " Score LDA 1 ", y = " Score LDA 2 ", color = " Group ") +
  scale_color_manual (values = palette_personnalisee , labels = names_groups)
+ theme_bw ()

# ##### GRAPH 3 D #####
# Create an interactive 3D graph from plotly
plot_ly(data, x =~LDA1, y =~LDA2, z =~LDA3, color = ~ Group, colors = palette_personnalisee) %>%
  add_markers (size = 4, opacity = 0.8 , text = ~ Group) % >%
  layout(scene = list (xaxis = list (title = "Score LDA 1"),
                        yaxis = list (title = "Score LDA 2"),
                        zaxis = list (title = "Score LDA 3")),
          margin = list (l = 0 , r = 0 , b = 0 , t = 0) ,
          legend = list (x = 1 , y = 0.8)) % >%
config (display Mode Bar = T)

```

Appendix 9: Optimization routine using PCA and LDA for biomolecule discrimination with sensors.

Example of discrimination of 8 biomolecules with 12 sensors

```

rm(list =ls ())

# Load necessary libraries
library (MASS) # For LDA
library (Facto Mine R) # For PCA
library (factoextra) # To visualize PCA results

# Read data
df <- read.csv2 (" Shift for RT.csv ", header = TRUE , stringsAsFactors = TRUE)

#Extract group labels and variable data
group_labels <- rep (1:8 , each = 4) # Assuming 8 groups, each represented by 4 lines
variables <- df[c(1:32) , c(2:12)] # Assuming that the variables are in columns 2 to 12

# Define a function for LDA analysis with leave - one - out validation
lda_analysis <- function (data, labels) {
  predicted_classes <- vector ("character", length(labels)) # Initialize vector to store predicted classes
}

# Perform cross validation leave - one - out
for (i in 1:length (labels)) {

  # Omit the i-th sample for tests
  training_data <- data [-i,]
}

```

```

training_labels <- labels [-i]
test_sample <- data [i,, drop = FALSE]

# Train the LDA model
lda_model <- lda (training_data, training_labels)
# Predict the class of the test sample
predicted_class <- predict (lda_model, test_sample$class)

# Store predicted class
predicted_classes [i] <- predicted_class
}

# Calculate the error rate
error_rate <- mean(predicted_classes != labels)

return(list(predicted_classes = predicted_classes, error_rate = error_rate))
}

# Define a function for PCA analysis
pca_analysis <- function ( data , labels ) {
  res.pca <- PCA (data, scale.unit = TRUE, graph = FALSE) # Perform PCA
  coords <- as.data.frame ( res.pca$ind$coord)
  pca_data <- cbind (coords, group = labels) # Combining PCA results with group labels
  return(list(pca = res.pca, pca_data = pca_data))
}

# Perform initial LDA analysis
lda_result <- lda_analysis (variables, group_labels)
best_error_rate <- lda_result$error_rate
best_sensor_array <- colnames (variables)

# Perform initial PCA analysis
pca_result <- pca_analysis (variables, group_labels)
var <- get_pca_var ( pca_result$pca ) # Get information on PCA variables
fviz_eig (pca_result $pca, addlabels = TRUE, ylim = c(0,100)) # Plot eigenvalues
fviz_cos2 (pca_result$pca , choice = "var", axes = 1:2)

# Iterate on different numbers of sensors to remove
for (num_sensors_to_remove in 1:4) {

  #Perform LDA analysis with the current number of sensors to be deleted
  variables_subset <- data.frame (variables, check.names = FALSE)
  best_error_rate_subset <- Inf
  best_sensor_array_subset <- colnames (variables_subset)

  while (ncol (variables_subset) > num_sensors_to_remove) {
    # Obtain sensors with the lowest contribution to dimensions
    var_contrib <- var$contrib [, 1]
    min_contrib_indices <- order (var_contrib) [1: num_sensors_to_remove]
    sensors_to_remove <- colnames (variables_subset)[min_contrib_indices]

    #Remove sensors from variables
    variables_subset <- subset (variables_subset , select = - c(min_contrib_indices))

    # Perform LDA analysis with leave - one - out validation and new sensor set
    training_data <- as.matrix (variables_subset)
    lda_result <- lda_analysis (training_data , group_labels)

    # Check whether the new sensor array improves the error rate
    if (lda_result$error_rate < best_error_rate_subset) {
      best_error_rate_subset <- lda_result$error_rate
      best_sensor_array_subset <- colnames (variables_subset)
    } else {
      # No improvement, stop deleting sensors
      break
    }
  }

  # Check whether the current number of sensors to be removed has resulted in a better error rate
  if (best_error_rate_subset < best_error_rate) {
    best_error_rate <- best_error_rate_subset
    best_sensor_array <- best_sensor_array_subset
  }
}

# Display the best sensor array and the corresponding error rate
cat ("Best sensorarray:", best_sensor_array , "\n")

cat ("Best Error Rate:", best_error_rate , "\n")

```





References

- (1) Tripathy, A.; Nine, M. J.; Losic, D.; Silva, F. S. Nature Inspired Emerging Sensing Technology: Recent Progress and Perspectives. *Mater. Sci. Eng. R Rep.* **2021**, *146*, 100647. <https://doi.org/10.1016/j.mser.2021.100647>.
- (2) Jones, G.; Holderied, M. W. Bat Echolocation Calls: Adaptation and Convergent Evolution. *Proc. R. Soc. B Biol. Sci.* **2007**, *274* (1612), 905–912.
- (3) Tuthill, J. C.; Azim, E. Proprioception. *Curr. Biol.* **2018**, *28* (5), R194–R203.
- (4) Bleckmann, H. Nature as a Model for Biomimetic Sensors; Lakhtakia, A., Ed.; San Diego, California, 2012; p 833902. <https://doi.org/10.1117/12.915729>.
- (5) Bleckmann, H.; Schmitz, H. Nature as a Model for Technical Sensors.
- (6) Gayomali, C. Study: 53% of Youngsters Would Choose Technology over Sense of Smell. May 27, 2011.
- (7) Glusman, G.; Yanai, I.; Rubin, I.; Lancet, D. The Complete Human Olfactory Subgenome. *Genome Res.* **2001**, *11* (5), 685–702. <https://doi.org/10.1101/gr.171001>.
- (8) Rinaldi, A. The Scent of Life: The Exquisite Complexity of the Sense of Smell in Animals and Humans. *EMBO Rep.* **2007**, *8* (7), 629–633.
- (9) Sankaran, S.; Khot, L. R.; Panigrahi, S. Biology and Applications of Olfactory Sensing System: A Review. *Sens. Actuators B Chem.* **2012**, *171–172*, 1–17. <https://doi.org/10.1016/j.snb.2012.03.029>.
- (10) Persaud, K.; Dodd, G. Analysis of Discrimination Mechanisms in the Mammalian Olfactory System Using a Model Nose. *Nature* **1982**, *299* (5881), 352–355. <https://doi.org/10.1038/299352a0>.
- (11) Geng, Y.; Peveler, W. J.; Rotello, V. M. Array-based “Chemical Nose” Sensing in Diagnostics and Drug Discovery. *Angew. Chem. Int. Ed.* **2019**, *58* (16), 5190–5200. <https://doi.org/10.1002/anie.201809607>.
- (12) Behera, P.; De, M. Surface-Engineered Nanomaterials for Optical Array Based Sensing. *ChemPlusChem* **2023**, e202300610.
- (13) Sugai, H.; Tomita, S.; Kurita, R. Pattern-Recognition-Based Sensor Arrays for Cell Characterization: From Materials and Data Analyses to Biomedical Applications. *Anal. Sci.* **2020**, *36* (8), 923–934. <https://doi.org/10.2116/analsci.20R002>.
- (14) Wright, A. T.; Anslyn, E. V. Differential Receptor Arrays and Assays for Solution-Based Molecular Recognition. *Chem Soc Rev* **2006**, *35* (1), 14–28. <https://doi.org/10.1039/B505518K>.
- (15) Askim, J. R.; Mahmoudi, M.; Suslick, K. S. Optical Sensor Arrays for Chemical Sensing: The Optoelectronic Nose. *Chem. Soc. Rev.* **2013**, *42* (22), 8649. <https://doi.org/10.1039/c3cs60179j>.
- (16) Li, Z.; Askim, J. R.; Suslick, K. S. The Optoelectronic Nose: Colorimetric and Fluorometric Sensor Arrays. *Chem. Rev.* **2019**, *119* (1), 231–292. <https://doi.org/10.1021/acs.chemrev.8b00226>.
- (17) Lim, S.; Kuang, Y.; Ardoña, H. A. M. Evolution of Supramolecular Systems towards Next-Generation Biosensors. *Front. Chem.* **2021**, *9*, 723111.
- (18) Bigdeli, A.; Ghasemi, F.; Golmohammadi, H.; Abbasi-Moayed, S.; Nejad, M. A. F.; Fahimi-Kashani, N.; Jafarinejad, S.; Shahrajabian, M.; Hormozi-Nezhad, M. R. Nanoparticle-Based Optical Sensor Arrays. *Nanoscale* **2017**, *9* (43), 16546–16563. <https://doi.org/10.1039/C7NR03311G>.
- (19) Saha, K.; Agasti, S. S.; Kim, C.; Li, X.; Rotello, V. M. Gold Nanoparticles in Chemical and Biological Sensing. *Chem. Rev.* **2012**, *112* (5), 2739–2779. <https://doi.org/10.1021/cr2001178>.

- (20) Chen, L.-Y.; Wang, C.-W.; Yuan, Z.; Chang, H.-T. Fluorescent Gold Nanoclusters: Recent Advances in Sensing and Imaging. *Anal. Chem.* **2015**, *87* (1), 216–229.
- (21) Peveler, W. J.; Roldan, A.; Hollingsworth, N.; Porter, M. J.; Parkin, I. P. Multichannel Detection and Differentiation of Explosives with a Quantum Dot Array. *ACS Nano* **2016**, *10* (1), 1139–1146. <https://doi.org/10.1021/acsnano.5b06433>.
- (22) Wang, K.; Dong, Y.; Li, B.; Li, D.; Zhang, S.; Wu, Y. Differentiation of Proteins and Cancer Cells Using Metal Oxide and Metal Nanoparticles–Quantum Dots Sensor Array. *Sens. Actuators B Chem.* **2017**, *250*, 69–75. <https://doi.org/10.1016/j.snb.2017.04.152>.
- (23) Yang, X.; Li, J.; Pei, H.; Li, D.; Zhao, Y.; Gao, J.; Lu, J.; Shi, J.; Fan, C.; Huang, Q. Pattern Recognition Analysis of Proteins Using DNA-Decorated Catalytic Gold Nanoparticles. *Small* **2013**, *9* (17), 2844–2849. <https://doi.org/10.1002/smll.201202772>.
- (24) Yang, X.; Li, J.; Pei, H.; Zhao, Y.; Zuo, X.; Fan, C.; Huang, Q. DNA–Gold Nanoparticle Conjugates-Based Nanoplasmonic Probe for Specific Differentiation of Cell Types. *Anal. Chem.* **2014**, *86* (6), 3227–3231. <https://doi.org/10.1021/ac500381e>.
- (25) You, C.-C.; Miranda, O. R.; Gider, B.; Ghosh, P. S.; Kim, I.-B.; Erdogan, B.; Krovi, S. A.; Bunz, U. H. F.; Rotello, V. M. Detection and Identification of Proteins Using Nanoparticle–Fluorescent Polymer ‘Chemical Nose’ Sensors. *Nat. Nanotechnol.* **2007**, *2* (5), 318–323. <https://doi.org/10.1038/nnano.2007.99>.
- (26) Miranda, O. R.; Chen, H.-T.; You, C.-C.; Mortenson, D. E.; Yang, X.-C.; Bunz, U. H. F.; Rotello, V. M. Enzyme-Amplified Array Sensing of Proteins in Solution and in Biofluids. *J. Am. Chem. Soc.* **2010**, *132* (14), 5285–5289. <https://doi.org/10.1021/ja1006756>.
- (27) You, L.; Zha, D.; Anslyn, E. V. Recent Advances in Supramolecular Analytical Chemistry Using Optical Sensing. *Chem. Rev.* **2015**, *115* (15), 7840–7892. <https://doi.org/10.1021/cr5005524>.
- (28) Mitchell, L.; New, E. J.; Mahon, C. S. Macromolecular Optical Sensor Arrays. *ACS Appl. Polym. Mater.* **2021**, *3* (2), 506–530. <https://doi.org/10.1021/acsapm.0c01003>.
- (29) Bunz, U. H. Poly (Aryleneethynylene) s. *Macromol. Rapid Commun.* **2009**, *30* (9-10), 772–805.
- (30) Bunz, U. H. Poly (Aryleneethynylene) s: Syntheses, Properties, Structures, and Applications. *Chem. Rev.* **2000**, *100* (4), 1605–1644.
- (31) Leclerc, M. Polyfluorenes: Twenty Years of Progress. *J. Polym. Sci. Part Polym. Chem.* **2001**, *39* (17), 2867–2873.
- (32) Scherf, U.; List, E. J. Semiconducting Polyfluorenes—towards Reliable Structure–Property Relationships. *Adv. Mater.* **2002**, *14* (7), 477–487.
- (33) Perepichka, I. F.; Perepichka, D. F.; Meng, H.; Wudl, F. Light-emitting Polythiophenes. *Adv. Mater.* **2005**, *17* (19), 2281–2305.
- (34) Roncali, J. Conjugated Poly (Thiophenes): Synthesis, Functionalization, and Applications. *Chem. Rev.* **1992**, *92* (4), 711–738.
- (35) Freudenberg, J.; Hinkel, F.; Jansch, D.; Bunz, U. H. F. Chemical Tongues and Noses Based upon Conjugated Polymers. *Top. Curr. Chem.* **2017**, *375* (4), 67. <https://doi.org/10.1007/s41061-017-0155-2>.
- (36) Liu, K.; Kang, Y.; Wang, Z.; Zhang, X. 25th Anniversary Article: Reversible and Adaptive Functional Supramolecular Materials: “Noncovalent Interaction” Matters. *Adv. Mater.* **2013**, *25* (39), 5530–5548. <https://doi.org/10.1002/adma.201302015>.

- (37) Diehl, K. L.; Anslyn, E. V. Array Sensing Using Optical Methods for Detection of Chemical and Biological Hazards. *Chem. Soc. Rev.* **2013**, *42* (22), 8596. <https://doi.org/10.1039/c3cs60136f>.
- (38) Kneipp, K.; Kneipp, H.; Itzkan, I.; Dasari, R. R.; Feld, M. S. Ultrasensitive Chemical Analysis by Raman Spectroscopy. *Chem. Rev.* **1999**, *99* (10), 2957–2976.
- (39) Banholzer, M. J.; Millstone, J. E.; Qin, L.; Mirkin, C. A. Rationally Designed Nanostructures for Surface-Enhanced Raman Spectroscopy. *Chem. Soc. Rev.* **2008**, *37* (5), 885–897.
- (40) Nie, S.; Emory, S. R. Probing Single Molecules and Single Nanoparticles by Surface-Enhanced Raman Scattering. *science* **1997**, *275* (5303), 1102–1106.
- (41) Stockman, M. I. Electromagnetic Theory of SERS. In *Surface-enhanced Raman scattering: physics and applications*; Springer, 2006; pp 47–65.
- (42) Anzenbacher, P.; Lubal, P.; Bucek, P.; Palacios, M. A.; Kozelkova, M. E. A Practical Approach to Optical Cross-Reactive Sensor Arrays. *Chem. Soc. Rev.* **2010**, *39* (10), 3954–3979. <https://doi.org/10.1039/b926220m>.
- (43) Wang, R. Comparison of Decision Tree, Random Forest and Linear Discriminant Analysis Models in Breast Cancer Prediction; IOP Publishing, 2022; Vol. 2386, p 012043.
- (44) Lehn, J. Supramolecular Chemistry—Scope and Perspectives Molecules, Supermolecules, and Molecular Devices (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **1988**, *27* (1), 89–112. <https://doi.org/10.1002/anie.198800891>.
- (45) Tian, J.-H.; Xu, H.; Hu, X.-Y.; Guo, D.-S. Supramolecular Fluorescence Biosensing Based on Macrocycles. *Supramol. Mater.* **2024**, *3*, 100063. <https://doi.org/10.1016/j.supmat.2024.100063>.
- (46) Pinalli, R.; Pedrini, A.; Dalcanale, E. Biochemical Sensing with Macrocyclic Receptors. *Chem. Soc. Rev.* **2018**, *47* (18), 7006–7026. <https://doi.org/10.1039/C8CS00271A>.
- (47) Jiang, C.; Song, Z.; Yu, L.; Ye, S.; He, H. Fluorescent Probes Based on Macrocyclic Hosts: Construction, Mechanism and Analytical Applications. *TrAC Trends Anal. Chem.* **2020**, *133*, 116086. <https://doi.org/10.1016/j.trac.2020.116086>.
- (48) Ma, X.; Zhao, Y. Biomedical Applications of Supramolecular Systems Based on Host–Guest Interactions. *Chem. Rev.* **2015**, *115* (15), 7794–7839. <https://doi.org/10.1021/cr500392w>.
- (49) Wang, L.; Li, L.; Fan, Y.; Wang, H. Host–Guest Supramolecular Nanosystems for Cancer Diagnostics and Therapeutics. *Adv. Mater.* **2013**, *25* (28), 3888–3898. <https://doi.org/10.1002/adma.201301202>.
- (50) Yang, X.; Yuan, D.; Hou, J.; Sedgwick, A. C.; Xu, S.; James, T. D.; Wang, L. Organic/Inorganic Supramolecular Nano-Systems Based on Host/Guest Interactions. *Coord. Chem. Rev.* **2021**, *428*, 213609. <https://doi.org/10.1016/j.ccr.2020.213609>.
- (51) Challa, R.; Ahuja, A.; Ali, J.; Khar, R. K. Cyclodextrins in Drug Delivery: An Updated Review. *AAPS PharmSciTech* **2005**, *6* (2), E329–E357. <https://doi.org/10.1208/pt060243>.
- (52) Uzunova, V. D.; Cullinane, C.; Brix, K.; Nau, W. M.; Day, A. I. Toxicity of Cucurbit[7]Uril and Cucurbit[8]Uril: An Exploratory in Vitro and in Vivo Study. *Org. Biomol. Chem.* **2010**, *8* (9), 2037. <https://doi.org/10.1039/b925555a>.
- (53) Hettiarachchi, G.; Nguyen, D.; Wu, J.; Lucas, D.; Ma, D.; Isaacs, L.; Briken, V. Toxicology and Drug Delivery by Cucurbit[n]Uril Type Molecular Containers. *PLoS ONE* **2010**, *5* (5), e10514. <https://doi.org/10.1371/journal.pone.0010514>.
- (54) Lou, X.-Y.; Song, N.; Yang, Y.-W. Fluorescence Resonance Energy Transfer Systems in Supramolecular Macrocyclic Chemistry. *Molecules* **2017**, *22* (10), 1640. <https://doi.org/10.3390/molecules22101640>.

- (55) Crini, G. Review: A History of Cyclodextrins. *Chem. Rev.* **2014**, *114* (21), 10940–10975. <https://doi.org/10.1021/cr500081p>.
- (56) Loftsson, T.; Brewster, M. E.; M??sson, M. Role of Cyclodextrins in Improving Oral Drug Delivery: *Am. J. Drug Deliv.* **2004**, *2* (4), 261–275. <https://doi.org/10.2165/00137696-200402040-00006>.
- (57) Duchêne, D.; Bochot, A. Thirty Years with Cyclodextrins. *Int. J. Pharm.* **2016**, *514* (1), 58–72. <https://doi.org/10.1016/j.ijpharm.2016.07.030>.
- (58) Perret, F.; Lazar, A. N.; Coleman, A. W. Biochemistry of the Para-Sulfonato-Calix[n]Arenes. *Chem. Commun.* **2006**, No. 23, 2425. <https://doi.org/10.1039/b600720c>.
- (59) Arduini, A.; Pochini, A.; Raverberi, S.; Ungaro, R. P-t-Butyl-Calix[4]Arene Tetracarboxylic Acid. A Water Soluble Calixarene in a Cone Structure. *J. Chem. Soc. Chem. Commun.* **1984**, No. 15, 981. <https://doi.org/10.1039/c39840000981>.
- (60) Shahgaldian, P.; Coleman, A. W.; Kalchenko, V. I. Synthesis and Properties of Novel Amphiphilic Calix-[4]-Arene Derivatives. *Tetrahedron Lett.* **2001**, *42* (4), 577–579. [https://doi.org/10.1016/S0040-4039\(00\)02003-7](https://doi.org/10.1016/S0040-4039(00)02003-7).
- (61) Ukhatskaya, E. V.; Kurkov, S. V.; Matthews, S. E.; Loftsson, T. Encapsulation of Drug Molecules into Calix[n]Arene Nanobaskets. Role of Aminocalix[n]Arenes in Biopharmaceutical Field. *J. Pharm. Sci.* **2013**, *102* (10), 3485–3512. <https://doi.org/10.1002/jps.23681>.
- (62) Mutihac, L.; Lee, J. H.; Kim, J. S.; Vicens, J. Recognition of Amino Acids by Functionalized Calixarenes. *Chem. Soc. Rev.* **2011**, *40* (5), 2777. <https://doi.org/10.1039/c0cs00005a>.
- (63) Nimse, S. B.; Kim, T. Biological Applications of Functionalized Calixarenes. *Chem Soc Rev* **2013**, *42* (1), 366–386. <https://doi.org/10.1039/C2CS35233H>.
- (64) Kumar, R.; Sharma, A.; Singh, H.; Suating, P.; Kim, H. S.; Sunwoo, K.; Shim, I.; Gibb, B. C.; Kim, J. S. Revisiting Fluorescent Calixarenes: From Molecular Sensors to Smart Materials. *Chem. Rev.* **2019**, *119* (16), 9657–9721. <https://doi.org/10.1021/acs.chemrev.8b00605>.
- (65) Pan, Y.; Hu, X.; Guo, D. Biomedical Applications of Calixarenes: State of the Art and Perspectives. *Angew. Chem. Int. Ed.* **2021**, *60* (6), 2768–2794. <https://doi.org/10.1002/anie.201916380>.
- (66) Cao, L.; Šekutor, M.; Zavalij, P. Y.; Mlinarić-Majerski, K.; Glaser, R.; Isaacs, L. Cucurbit[7]uril·Guest Pair with an Attomolar Dissociation Constant. *Angew. Chem.* **2014**, *126* (4), 1006–1011. <https://doi.org/10.1002/ange.201309635>.
- (67) Moghaddam, S.; Yang, C.; Rekharsky, M.; Ko, Y. H.; Kim, K.; Inoue, Y.; Gilson, M. K. New Ultrahigh Affinity Host–Guest Complexes of Cucurbit[7]Urill with Bicyclo[2.2.2]Octane and Adamantane Guests: Thermodynamic Analysis and Evaluation of M2 Affinity Calculations. *J. Am. Chem. Soc.* **2011**, *133* (10), 3570–3581. <https://doi.org/10.1021/ja109904u>.
- (68) Biedermann, F.; Uzunova, V. D.; Scherman, O. A.; Nau, W. M.; De Simone, A. Release of High-Energy Water as an Essential Driving Force for the High-Affinity Binding of Cucurbit[n]Urills. *J. Am. Chem. Soc.* **2012**, *134* (37), 15318–15323. <https://doi.org/10.1021/ja303309e>.
- (69) Macartney, D. H. Encapsulation of Drug Molecules by Cucurbiturils: Effects on Their Chemical Properties in Aqueous Solution. *Isr. J. Chem.* **2011**, *51* (5–6), 600–615. <https://doi.org/10.1002/ijch.201100040>.
- (70) Feng, W.; Jin, M.; Yang, K.; Pei, Y.; Pei, Z. Supramolecular Delivery Systems Based on Pillararenes. *Chem. Commun.* **2018**, *54* (97), 13626–13640. <https://doi.org/10.1039/C8CC08252A>.

- (71) Sathiyajith, C.; Shaikh, R. R.; Han, Q.; Zhang, Y.; Meguellati, K.; Yang, Y.-W. Biological and Related Applications of Pillar[n]Arenes. *Chem. Commun.* **2017**, 53 (4), 677–696. <https://doi.org/10.1039/C6CC08967D>.
- (72) Xue, M.; Yang, Y.; Chi, X.; Zhang, Z.; Huang, F. Pillararenes, A New Class of Macrocycles for Supramolecular Chemistry. *Acc. Chem. Res.* **2012**, 45 (8), 1294–1308. <https://doi.org/10.1021/ar2003418>.
- (73) Ogoshi, T.; Kanai, S.; Fujinami, S.; Yamagishi, T.; Nakamoto, Y. Para -Bridged Symmetrical Pillar[5]Arenes: Their Lewis Acid Catalyzed Synthesis and Host–Guest Property. *J. Am. Chem. Soc.* **2008**, 130 (15), 5022–5023. <https://doi.org/10.1021/ja711260m>.
- (74) Xu, X.; Jerca, V. V.; Hoogenboom, R. Structural Diversification of Pillar[n]Arene Macrocycles. *Angew. Chem. Int. Ed.* **2020**, 59 (16), 6314–6316. <https://doi.org/10.1002/anie.202002467>.
- (75) Hu, C.; Jochmann, T.; Chakraborty, P.; Neumaier, M.; Levkin, P. A.; Kappes, M. M.; Biedermann, F. Further Dimensions for Sensing in Biofluids: Distinguishing Bioorganic Analytes by the Salt-Induced Adaptation of a Cucurbit[7]Uril-Based Chemosensor. *J. Am. Chem. Soc.* **2022**, 144 (29), 13084–13095. <https://doi.org/10.1021/jacs.2c01520>.
- (76) Krämer, J.; Kang, R.; Grimm, L. M.; De Cola, L.; Picchetti, P.; Biedermann, F. Molecular Probes, Chemosensors, and Nanosensors for Optical Detection of Biorelevant Molecules and Ions in Aqueous Media and Biofluids. *Chem. Rev.* **2022**, 122 (3), 3459–3636. <https://doi.org/10.1021/acs.chemrev.1c00746>.
- (77) Liu, Y.; Minami, T.; Nishiyabu, R.; Wang, Z.; Anzenbacher, P. Sensing of Carboxylate Drugs in Urine by a Supramolecular Sensor Array. *J. Am. Chem. Soc.* **2013**, 135 (20), 7705–7712. <https://doi.org/10.1021/ja4015748>.
- (78) Das Saha, N.; Pradhan, S.; Sasmal, R.; Sarkar, A.; Berač, C. M.; Kölsch, J. C.; Pahwa, M.; Show, S.; Rozenholc, Y.; Topçu, Z.; Alessandrini, V.; Guibourdenche, J.; Tsatsaris, V.; Gagey-Eilstein, N.; Agasti, S. S. Cucurbit[7]Uril Macrocyclic Sensors for Optical Fingerprinting: Predicting Protein Structural Changes to Identifying Disease-Specific Amyloid Assemblies. *J. Am. Chem. Soc.* **2022**, 144 (31), 14363–14379. <https://doi.org/10.1021/jacs.2c05969>.
- (79) Mei, Y.; Zhang, Q.-W.; Gu, Q.; Liu, Z.; He, X.; Tian, Y. Pillar [5] Arene-Based Fluorescent Sensor Array for Biosensing of Intracellular Multi-Neurotransmitters through Host–Guest Recognitions. *J. Am. Chem. Soc.* **2022**, 144 (5), 2351–2359.
- (80) Ikeda, H.; Murayama, T.; Ueno, A. Skeleton-Selective Fluorescent Chemosensor Based on Cyclodextrin Bearing a 4-Amino-7-Nitrobenz-2-Oxa-1, 3-Diazole Moiety. *Org. Biomol. Chem.* **2005**, 3 (23), 4262–4267.
- (81) Liu, Y.; Shi, J.; Guo, D.-S. Novel Permethylated β -Cyclodextrin Derivatives Appended with Chromophores as Efficient Fluorescent Sensors for the Molecular Recognition of Bile Salts. *J. Org. Chem.* **2007**, 72 (22), 8227–8234.
- (82) Minaker, S. A.; Daze, K. D.; Ma, M. C.; Hof, F. Antibody-Free Reading of the Histone Code Using a Simple Chemical Sensor Array. *J. Am. Chem. Soc.* **2012**, 134 (28), 11674–11680.
- (83) Liu, Y.; Perez, L.; Mettry, M.; Easley, C. J.; Hooley, R. J.; Zhong, W. Self-Aggregating Deep Cavitand Acts as a Fluorescence Displacement Sensor for Lysine Methylation. *J. Am. Chem. Soc.* **2016**, 138 (34), 10746–10749.
- (84) Zhong, W.; Hooley, R. J. Combining Excellent Selectivity with Broad Target Scope: Biosensing with Arrayed Deep Cavitand Hosts. *Acc. Chem. Res.* **2022**, 55 (7), 1035–1046.
- (85) Chen, J.; Hickey, B. L.; Wang, L.; Lee, J.; Gill, A. D.; Favero, A.; Pinalli, R.; Dalcanale, E.; Hooley, R. J.; Zhong, W. Selective Discrimination and Classification of G-Quadruplex Structures with a Host–Guest Sensing Array. *Nat. Chem.* **2021**, 13 (5), 488–495. <https://doi.org/10.1038/s41557-021-00647-9>.

- (86) Tian, J.-H.; Hu, X.-Y.; Hu, Z.-Y.; Tian, H.-W.; Li, J.-J.; Pan, Y.-C.; Li, H.-B.; Guo, D.-S. A Facile Way to Construct Sensor Array Library via Supramolecular Chemistry for Discriminating Complex Systems. *Nat. Commun.* **2022**, *13* (1), 4293. <https://doi.org/10.1038/s41467-022-31986-x>.
- (87) Ogoshi, T.; Harada, A. Chemical Sensors Based on Cyclodextrin Derivatives. *Sensors* **2008**, *8* (8), 4961–4982. <https://doi.org/10.3390/s8084961>.
- (88) Pagliari, S.; Corradini, R.; Galaverna, G.; Sforza, S.; Dossena, A.; Montalti, M.; Prodi, L.; Zaccheroni, N.; Marchelli, R. Enantioselective Fluorescence Sensing of Amino Acids by Modified Cyclodextrins: Role of the Cavity and Sensing Mechanism. *Chem. Eur. J.* **2004**, *10* (11), 2749–2758.
- (89) Ueno, A.; Minato, S.; Osa, T. Host-Guest Sensors of 6A, 6B-, 6A, 6C-, 6A, 6D-, and 6A, 6E-Bis (2-Naphthylsulfenyl)-. Gamma.-Cyclodextrins for Detecting Organic Compounds by Fluorescence Enhancements. *Anal. Chem.* **1992**, *64* (10), 1154–1157.
- (90) Suzuki, I.; Ohkubo, M.; Ueno, A.; Osa, T. Detection of Organic Compounds by Dual Fluorescence of Bis (1-Pyrenecarbonyl)-. GAMMA.-Cyclodextrins. *Chem. Lett.* **1992**, No. 2, 269–272.
- (91) Ueno, A.; Moriwaki, F.; Iwama, Y.; Suzuki, I.; Osa, T.; Ohta, T.; Nozoe, S. . . Gamma.-Cyclodextrin Template Method for Controlling Stereochemistry of Bimolecular Interactions and Reactions. *J. Am. Chem. Soc.* **1991**, *113* (18), 7034–7036.
- (92) Hu, C.; Grimm, L.; Prabodh, A.; Bakshi, A.; Siennicka, A.; Levkin, P. A.; Kappes, M. M.; Biedermann, F. Covalent Cucurbit[7]Uril–Dye Conjugates for Sensing in Aqueous Saline Media and Biofluids. *Chem. Sci.* **2020**, *11* (41), 11142–11153. <https://doi.org/10.1039/D0SC03079A>.
- (93) Zhang, S.; Assaf, K. I.; Huang, C.; Hennig, A.; Nau, W. M. Ratiometric DNA Sensing with a Host–Guest FRET Pair. *Chem. Commun.* **2019**, *55* (5), 671–674. <https://doi.org/10.1039/C8CC09126A>.
- (94) Haab, B. B. Applications of Antibody Array Platforms. *Curr. Opin. Biotechnol.* **2006**, *17* (4), 415–421. <https://doi.org/10.1016/j.copbio.2006.06.013>.
- (95) Peng, P.; Liu, C.; Li, Z.; Xue, Z.; Mao, P.; Hu, J.; Xu, F.; Yao, C.; You, M. Emerging ELISA Derived Technologies for in Vitro Diagnostics. *TrAC Trends Anal. Chem.* **2022**, *152*, 116605. <https://doi.org/10.1016/j.trac.2022.116605>.
- (96) Yu, Q.; Wang, Q.; Li, B.; Lin, Q.; Duan, Y. Technological Development of Antibody Immobilization for Optical Immunoassays: Progress and Prospects. *Crit. Rev. Anal. Chem.* **2015**, *45* (1), 62–75. <https://doi.org/10.1080/10408347.2014.881249>.
- (97) Stewart, S.; Syrett, A.; Pothukuchy, A.; Bhadra, S.; Ellington, A.; Anslyn, E. Identifying Protein Variants with Cross-Reactive Aptamer Arrays. *ChemBioChem* **2011**, *12* (13), 2021–2024. <https://doi.org/10.1002/cbic.201100046>.
- (98) Chen, J.; Hooley, R. J.; Zhong, W. Applications of Synthetic Receptors in Bioanalysis and Drug Transport. *Bioconj. Chem.* **2022**, *acs.bioconjchem.2c00096*. <https://doi.org/10.1021/acs.bioconjchem.2c00096>.
- (99) Chen, J.; Hickey, B. L.; Wang, L.; Lee, J.; Gill, A. D.; Favero, A.; Pinalli, R.; Dalcanale, E.; Hooley, R. J.; Zhong, W. Selective Discrimination and Classification of G-Quadruplex Structures with a Host–Guest Sensing Array. *Nat. Chem.* **2021**, *13* (5), 488–495. <https://doi.org/10.1038/s41557-021-00647-9>.
- (100) Liu, Y.; Perez, L.; Gill, A. D.; Mettry, M.; Li, L.; Wang, Y.; Hooley, R. J.; Zhong, W. Site-Selective Sensing of Histone Methylation Enzyme Activity via an Arrayed Supramolecular Tandem Assay. *J. Am. Chem. Soc.* **2017**, *139* (32), 10964–10967. <https://doi.org/10.1021/jacs.7b05002>.
- (101) Zeng, L.; Kaoud, T. S.; Zamora-Olivares, D.; Bohanon, A. L.; Li, Y.; Pridgen, J. R.; Ekpo, Y. E.; Zhuang, D. L.; Nye, J. R.; Telles, M.; Winkler, M.; Rivera, S.; Marini, F.; Dalby, K. N.; Anslyn, E. V. Multiplexing the Quantitation of MAP Kinase Activities Using Differential Sensing. *J. Am. Chem. Soc.* **2022**, *144* (9), 4017–4025. <https://doi.org/10.1021/jacs.1c12757>.

- (102) Hipp, M. S.; Kasturi, P.; Hartl, F. U. The Proteostasis Network and Its Decline in Ageing. *Nat. Rev. Mol. Cell Biol.* **2019**, *20* (7), 421–435. <https://doi.org/10.1038/s41580-019-0101-y>.
- (103) Killingsworth, L. M.; Daniels, J. C. Plasma Protein Patterns in Health and Disease. *CRC Crit. Rev. Clin. Lab. Sci.* **1979**, *11* (1), 1–30. <https://doi.org/10.3109/10408367909105852>.
- (104) Fan, J.; Qi, L.; Han, H.; Ding, L. Array-Based Discriminative Optical Biosensors for Identifying Multiple Proteins in Aqueous Solution and Biofluids. *Front. Chem.* **2020**, *8*, 572234. <https://doi.org/10.3389/fchem.2020.572234>.
- (105) Dai, X.; Cheng, H.; Bai, Z.; Li, J. Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. *J. Cancer* **2017**, *8* (16), 3131–3141. <https://doi.org/10.7150/jca.18457>.
- (106) Holliday, D. L.; Speirs, V. Choosing the Right Cell Line for Breast Cancer Research. *Breast Cancer Res.* **2011**, *13* (4), 215. <https://doi.org/10.1186/bcr2889>.
- (107) Martin, O. A.; Anderson, R. L.; Narayan, K.; MacManus, M. P. Does the Mobilization of Circulating Tumour Cells during Cancer Therapy Cause Metastasis? *Nat. Rev. Clin. Oncol.* **2017**, *14* (1), 32–44. <https://doi.org/10.1038/nrclinonc.2016.128>.
- (108) Wu, L.; Ji, H.; Guan, Y.; Ran, X.; Ren, J.; Qu, X. A Graphene-Based Chemical Nose/Tongue Approach for the Identification of Normal, Cancerous and Circulating Tumor Cells. *NPG Asia Mater.* **2017**, *9* (3), e356–e356. <https://doi.org/10.1038/am.2017.11>.
- (109) Kapur, S. K.; Katz, A. J. Review of the Adipose Derived Stem Cell Secretome. *Biochimie* **2013**, *95* (12), 2222–2228. <https://doi.org/10.1016/j.biochi.2013.06.001>.
- (110) Brown, K. J.; Formolo, C. A.; Seol, H.; Marathi, R. L.; Duguez, S.; An, E.; Pillai, D.; Nazarian, J.; Rood, B. R.; Hathout, Y. Advances in the Proteomic Investigation of the Cell Secretome. *Expert Rev. Proteomics* **2012**, *9* (3), 337–345. <https://doi.org/10.1586/epr.12.21>.
- (111) Tabar, V.; Studer, L. Pluripotent Stem Cells in Regenerative Medicine: Challenges and Recent Progress. *Nat. Rev. Genet.* **2014**, *15* (2), 82–92. <https://doi.org/10.1038/nrg3563>.
- (112) Jaroch, K.; Jaroch, A.; Bojko, B. Cell Cultures in Drug Discovery and Development: The Need of Reliable in Vitro-in Vivo Extrapolation for Pharmacodynamics and Pharmacokinetics Assessment. *J. Pharm. Biomed. Anal.* **2018**, *147*, 297–312. <https://doi.org/10.1016/j.jpba.2017.07.023>.
- (113) Meissner, R.; Eker, B.; Kasi, H.; Bertsch, A.; Renaud, P. Distinguishing Drug-Induced Minor Morphological Changes from Major Cellular Damage via Label-Free Impedimetric Toxicity Screening. *Lab. Chip* **2011**, *11* (14), 2352. <https://doi.org/10.1039/c1lc20212j>.
- (114) Latchney, S. E.; Majewska, A. K. Persistent Organic Pollutants at the Synapse: Shared Phenotypes and Converging Mechanisms of Developmental Neurotoxicity. *Dev. Neurobiol.* **2021**, *81* (5), 623–652. <https://doi.org/10.1002/dneu.22825>.
- (115) Zheng, W.; Thorne, N.; McKew, J. C. Phenotypic Screens as a Renewed Approach for Drug Discovery. *Drug Discov. Today* **2013**, *18* (21–22), 1067–1073. <https://doi.org/10.1016/j.drudis.2013.07.001>.
- (116) Ahmed, M. U.; Saaem, I.; Wu, P. C.; Brown, A. S. Personalized Diagnostics and Biosensors: A Review of the Biology and Technology Needed for Personalized Medicine. *Crit Rev Biotechnol.*
- (117) Petryayeva, E.; Algar, W. R. Toward Point-of-Care Diagnostics with Consumer Electronic Devices: The Expanding Role of Nanoparticles. *RSC Adv.* **2015**, *5* (28), 22256–22282. <https://doi.org/10.1039/C4RA15036H>.
- (118) De, M.; Rana, S.; Akpınar, H.; Miranda, O. R.; Arvizo, R. R.; Bunz, U. H. F.; Rotello, V. M. Sensing of Proteins in Human Serum Using Conjugates of Nanoparticles and Green Fluorescent Protein. *Nat. Chem.* **2009**, *1* (6), 461–465. <https://doi.org/10.1038/nchem.334>.

- (119) Xu, S.; Wu, Y.; Sun, X.; Wang, Z.; Luo, X. A Multicoloured Au NCs Based Cross-Reactive Sensor Array for Discrimination of Multiple Proteins. *J. Mater. Chem. B* **2017**, *5* (22), 4207–4213. <https://doi.org/10.1039/C7TB00367F>.
- (120) Bajaj, A.; Miranda, O. R.; Kim, I.-B.; Phillips, R. L.; Jerry, D. J.; Bunz, U. H. F.; Rotello, V. M. Detection and Differentiation of Normal, Cancerous, and Metastatic Cells Using Nanoparticle-Polymer Sensor Arrays. *Proc. Natl. Acad. Sci.* **2009**, *106* (27), 10912–10916. <https://doi.org/10.1073/pnas.0900975106>.
- (121) Bajaj, A.; Rana, S.; Miranda, O. R.; Yawe, J. C.; Jerry, D. J.; Bunz, U. H. F.; Rotello, V. M. Cell Surface-Based Differentiation of Cell Types and Cancer States Using a Gold Nanoparticle-GFP Based Sensing Array. *Chem. Sci.* **2010**, *1* (1), 134. <https://doi.org/10.1039/c0sc00165a>.
- (122) Bajaj, A.; Miranda, O. R.; Phillips, R. L.; Kim, I.-B.; Jerry, D.; Bunz, U.; Rotello, V. M. Array-Based Sensing of Normal, Cancerous, and Metastatic Cells Using Conjugated Fluorescent Polymers. *J. Am. Chem. Soc.* **2010**, *132* (3), 1018–1022. <https://doi.org/10.1021/ja9061272>.
- (123) Rana, S.; Le, N. D. B.; Mout, R.; Saha, K.; Tonga, G. Y.; Bain, R. E. S.; Miranda, O. R.; Rotello, C. M.; Rotello, V. M. A Multichannel Nanosensor for Instantaneous Readout of Cancer Drug Mechanisms. **2017**.
- (124) Rana, S.; Le, N. D. B.; Mout, R.; Saha, K.; Tonga, G. Y.; Bain, R. E. S.; Miranda, O. R.; Rotello, C. M.; Rotello, V. M. A Multichannel Nanosensor for Instantaneous Readout of Cancer Drug Mechanisms. *Nat. Nanotechnol.* **2015**, *10* (1), 65–69. <https://doi.org/10.1038/nnano.2014.285>.
- (125) Peveler, W. J.; Landis, R. F.; Yazdani, M.; Day, J. W.; Modi, R.; Carmalt, C. J.; Rosenberg, W. M.; Rotello, V. M. A Rapid and Robust Diagnostic for Liver Fibrosis Using a Multichannel Polymer Sensor Array. *Adv. Mater.* **2018**, *30* (28), 1800634. <https://doi.org/10.1002/adma.201800634>.
- (126) Das Saha, N.; Sasmal, R.; Meethal, S. K.; Vats, S.; Gopinathan, P. V.; Jash, O.; Manjithaya, R.; Gagey-Eilstein, N.; Agasti, S. S. Multichannel DNA Sensor Array Fingerprints Cell States and Identifies Pharmacological Effectors of Catabolic Processes. *ACS Sens.* **2019**, *4* (12), 3124–3132. <https://doi.org/10.1021/acssensors.9b01009>.
- (127) Motiei, L.; Margulies, D. Molecules That Generate Fingerprints: A New Class of Fluorescent Sensors for Chemical Biology, Medical Diagnosis, and Cryptography. *Acc. Chem. Res.* **2023**, *56* (13), 1803–1814. <https://doi.org/10.1021/acs.accounts.3c00162>.
- (128) Hatai, J.; Motiei, L.; Margulies, D. Analyzing Amyloid Beta Aggregates with a Combinatorial Fluorescent Molecular Sensor. *J. Am. Chem. Soc.* **2017**, *139* (6), 2136–2139. <https://doi.org/10.1021/jacs.6b10809>.
- (129) Pode, Z.; Peri-Naor, R.; Georgeson, J. M.; Ilani, T.; Kiss, V.; Unger, T.; Markus, B.; Barr, H. M.; Motiei, L.; Margulies, D. Protein Recognition by a Pattern-Generating Fluorescent Molecular Probe. *Nat. Nanotechnol.* **2017**, *12* (12), 1161–1168. <https://doi.org/10.1038/nnano.2017.175>.
- (130) Shen, Y.; Huang, Y.; Zhang, P.; Guo, B.; Jiang, H.; Tan, C.; Jiang, Y. Fluorescence Sensor Array for Discrimination of Urine Proteins and Differentiation Diagnosis of Urinary System Diseases. *ACS Appl. Bio Mater.* **2020**, *3* (9), 5639–5643. <https://doi.org/10.1021/acsabm.0c00845>.
- (131) Liu, Z.; Zhu, X.; Lu, Q.; Liu, M.; Li, H.; Zhang, Y.; Liu, Y.; Yao, S. Recognition Engineering-Mediated Multichannel Sensor Array for Gut Microbiota Sensing. *Anal. Chem.* **2023**, *95* (14), 5911–5919. <https://doi.org/10.1021/acs.analchem.2c04997>.
- (132) Meyer, E. Über Die Condensation Des Harnstoffs Mit Glyoxal Und Des Glykolorils Mit Formaldehyd, 1904.
- (133) Behrend, R.; Meyer, E.; Rusche, F. I. Ueber Condensationsproducte aus Glycoluril und Formaldehyd. *Justus Liebigs Ann. Chem.* **1905**, *339* (1), 1–37. <https://doi.org/10.1002/jlac.19053390102>.
- (134) Freeman, W.; Mock, W.; Shih, N. Cucurbituril. *J. Am. Chem. Soc.* **1981**, *103* (24), 7367–7368.

- (135) Kim, J.; Jung, I.-S.; Kim, S.-Y.; Lee, E.; Kang, J.-K.; Sakamoto, S.; Yamaguchi, K.; Kim, K. New Cucurbituril Homologues: Syntheses, Isolation, Characterization, and X-Ray Crystal Structures of Cucurbit[*n*]Urils (*n* = 5, 7, and 8). *J. Am. Chem. Soc.* **2000**, *122* (3), 540–541. <https://doi.org/10.1021/ja993376p>.
- (136) Day, A.; Arnold, A. P.; Blanch, R. J.; Snushall, B. Controlling Factors in the Synthesis of Cucurbituril and Its Homologues. *J. Org. Chem.* **2001**, *66* (24), 8094–8100. <https://doi.org/10.1021/jo015897c>.
- (137) Day, A.; Arnold, A. P.; Blanch, R. J.; Snushall, B. Controlling Factors in the Synthesis of Cucurbituril and Its Homologues. *J. Org. Chem.* **2001**, *66* (24), 8094–8100. <https://doi.org/10.1021/jo015897c>.
- (138) Barrow, S. J.; Kasera, S.; Rowland, M. J.; Del Barrio, J.; Scherman, O. A. Cucurbituril-Based Molecular Recognition. *Chem. Rev.* **2015**, *115* (22), 12320–12406. <https://doi.org/10.1021/acs.chemrev.5b00341>.
- (139) Márquez, C.; Hudgins, R. R.; Nau, W. M. Mechanism of Host–Guest Complexation by Cucurbituril. *J. Am. Chem. Soc.* **2004**, *126* (18), 5806–5816. <https://doi.org/10.1021/ja0319846>.
- (140) Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. The Cucurbit[*n*]Urils Family. *Angew. Chem. Int. Ed.* **2005**, *44* (31), 4844–4870. <https://doi.org/10.1002/anie.200460675>.
- (141) Assaf, K. I.; Nau, W. M. Cucurbiturils: From Synthesis to High-Affinity Binding and Catalysis. *Chem. Soc. Rev.* **2015**, *44* (2), 394–418. <https://doi.org/10.1039/C4CS00273C>.
- (142) Alešković, M.; Šekutor, M. Overcoming Barriers with Non-Covalent Interactions: Supramolecular Recognition of Adamantyl Cucurbit[*n*]Urils Assemblies for Medical Applications. *RSC Med. Chem.* **2024**, *15* (2), 433–471.
- (143) Mecozzi, S.; Rebek, Jr., J. The 55 % Solution: A Formula for Molecular Recognition in the Liquid State. *Chem. - Eur. J.* **1998**, *4* (6), 1016–1022. [https://doi.org/10.1002/\(SICI\)1521-3765\(19980615\)4:6<1016::AID-CHEM1016>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1521-3765(19980615)4:6<1016::AID-CHEM1016>3.0.CO;2-B).
- (144) Nau, W. M.; Florea, M.; Assaf, K. I. Deep Inside Cucurbiturils: Physical Properties and Volumes of Their Inner Cavity Determine the Hydrophobic Driving Force for Host–Guest Complexation. *Isr. J. Chem.* **2011**, *51* (5–6), 559–577. <https://doi.org/10.1002/ijch.201100044>.
- (145) Barrow, S. J.; Kasera, S.; Rowland, M. J.; Del Barrio, J.; Scherman, O. A. Cucurbituril-Based Molecular Recognition. *Chem. Rev.* **2015**, *115* (22), 12320–12406. <https://doi.org/10.1021/acs.chemrev.5b00341>.
- (146) Mock, W. L.; Shih, N. Y. Structure and Selectivity in Host-Guest Complexes of Cucurbituril. *J. Org. Chem.* **1986**, *51* (23), 4440–4446. <https://doi.org/10.1021/jo00373a018>.
- (147) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. Binding Affinities of Host–Guest, Protein–Ligand, and Protein–Transition-State Complexes. *Angew. Chem. Int. Ed.* **2003**, *42* (40), 4872–4897. <https://doi.org/10.1002/anie.200200565>.
- (148) Green, N. M. [5] Avidin and Streptavidin. In *Methods in enzymology*; Elsevier, 1990; Vol. 184, pp 51–67.
- (149) Pazy, Y.; Kulik, T.; Bayer, E. A.; Wilchek, M.; Livnah, O. Ligand Exchange between Proteins: Exchange of Biotin and Biotin Derivatives between Avidin and Streptavidin. *J. Biol. Chem.* **2002**, *277* (34), 30892–30900.
- (150) Jeon, W. S.; Moon, K.; Park, S. H.; Chun, H.; Ko, Y. H.; Lee, J. Y.; Lee, E. S.; Samal, S.; Selvapalam, N.; Rekharsky, M. V. Complexation of Ferrocene Derivatives by the Cucurbit [7] Uril Host: A Comparative Study of the Cucurbituril and Cyclodextrin Host Families. *J. Am. Chem. Soc.* **2005**, *127* (37), 12984–12989.

- (151) Rekharsky, M. V.; Mori, T.; Yang, C.; Ko, Y. H.; Selvapalam, N.; Kim, H.; Sobransingh, D.; Kaifer, A. E.; Liu, S.; Isaacs, L. A Synthetic Host-Guest System Achieves Avidin-Biotin Affinity by Overcoming Enthalpy–Entropy Compensation. *Proc. Natl. Acad. Sci.* **2007**, *104* (52), 20737–20742.
- (152) Moghaddam, S.; Yang, C.; Rekharsky, M.; Ko, Y. H.; Kim, K.; Inoue, Y.; Gilson, M. K. New Ultrahigh Affinity Host– Guest Complexes of Cucurbit [7] Uril with Bicyclo [2.2. 2] Octane and Adamantane Guests: Thermodynamic Analysis and Evaluation of M2 Affinity Calculations. *J. Am. Chem. Soc.* **2011**, *133* (10), 3570–3581.
- (153) Masson, E.; Ling, X.; Joseph, R.; Kyeremeh-Mensah, L.; Lu, X. Cucurbituril Chemistry: A Tale of Supramolecular Success. *RSC Adv* **2012**, *2* (4), 1213–1247. <https://doi.org/10.1039/C1RA00768H>.
- (154) Moghaddam, S.; Yang, C.; Rekharsky, M.; Ko, Y. H.; Kim, K.; Inoue, Y.; Gilson, M. K. New Ultrahigh Affinity Host– Guest Complexes of Cucurbit [7] Uril with Bicyclo [2.2. 2] Octane and Adamantane Guests: Thermodynamic Analysis and Evaluation of M2 Affinity Calculations. *J. Am. Chem. Soc.* **2011**, *133* (10), 3570–3581.
- (155) Sinn, S.; Biedermann, F. Chemical Sensors Based on Cucurbit[*n*]Uril Macrocycles. *Isr. J. Chem.* **2018**, *58* (3–4), 357–412. <https://doi.org/10.1002/ijch.201700118>.
- (156) Kim, K.; Selvapalam, N.; Ko, Y. H.; Park, K. M.; Kim, D.; Kim, J. Functionalized Cucurbiturils and Their Applications. *Chem Soc Rev* **2007**, *36* (2), 267–279. <https://doi.org/10.1039/B603088M>.
- (157) Yin, H.; Cheng, Q.; Bardelang, D.; Wang, R. Challenges and Opportunities of Functionalized Cucurbiturils for Biomedical Applications.
- (158) Karak, A.; Manna, S. K.; Mahapatra, A. K. Triphenylamine-Based Small-Molecule Fluorescent Probes. *Anal. Methods* **2022**, *14* (10), 972–1005. <https://doi.org/10.1039/D2AY00134A>.
- (159) Dumat, B.; Bordeau, G.; Aranda, A. I.; Mahuteau-Betzer, F.; Harfouch, Y. E.; Metgé, G.; Charra, F.; Fiorini-Debuisschert, C.; Teulade-Fichou, M.-P. Vinyl-Triphenylamine Dyes, a New Family of Switchable Fluorescent Probes for Targeted Two-Photon Cellular Imaging: From DNA to Protein Labeling. *Org. Biomol. Chem.* **2012**, *10* (30), 6054. <https://doi.org/10.1039/c2ob25515d>.
- (160) Mantooth, S. M.; Munoz-Robles, B. G.; Webber, M. J. Dynamic Hydrogels from Host–Guest Supramolecular Interactions. *Macromol. Biosci.* **2019**, *19* (1), 1800281.
- (161) Allain, C.; Schmidt, F.; Lartia, R.; Bordeau, G.; Fiorini-Debuisschert, C.; Charra, F.; Tauc, P.; Teulade-Fichou, M.-P. Vinyl-Pyridinium Triphenylamines: Novel Far-Red Emitters with High Photostability and Two-Photon Absorption Properties for Staining DNA. *ChemBioChem* **2007**, *8* (4), 424–433. <https://doi.org/10.1002/cbic.200600483>.
- (162) Xu, L.; Lin, W.; Huang, B.; Zhang, J.; Long, X.; Zhang, W.; Zhang, Q. The Design Strategies and Applications for Organic Multi-Branched Two-Photon Absorption Chromophores with Novel Cores and Branches: A Recent Review. *J. Mater. Chem. C* **2021**, *9* (5), 1520–1536. <https://doi.org/10.1039/D0TC05910B>.
- (163) Gong, Y.; Bi, X.; Chen, N.; Forconi, M.; Kuthirummal, N.; Teklu, A.; Gao, B.; Koenemann, J.; Harris, N.; Brennan, C.; Thomas, M.; Barnes, T.; Hu, M. Significant Enhancement of Two-Photon Excited Fluorescence in Water-Soluble Triphenylamine-Based All-Organic Compounds. *J. Phys. Chem. B* **2022**, *126* (29), 5513–5522. <https://doi.org/10.1021/acs.jpcc.2c03514>.
- (164) Koner, A. L.; Nau, W. M. Cucurbituril Encapsulation of Fluorescent Dyes. *Supramol. Chem.* **2007**, *19* (1–2), 55–66. <https://doi.org/10.1080/10610270600910749>.
- (165) Mandal, A. K.; Suresh, M.; Das, P.; Das, A. Restricted Conformational Flexibility of a Triphenylamine Derivative on the Formation of Host–Guest Complexes with Various Macrocyclic Hosts. *Chem. Eur. J.* **2012**, *18* (13), 3906–3917.

- (166) Gangopadhyay, M.; Mandal, A. K.; Maity, A.; Ravindranathan, S.; Rajamohanan, P. R.; Das, A. Tuning Emission Responses of a Triphenylamine Derivative in Host–Guest Complexes and an Unusual Dynamic Inclusion Phenomenon. *J. Org. Chem.* **2016**, *81* (2), 512–521. <https://doi.org/10.1021/acs.joc.5b02353>.
- (167) Mahuteau-Betzer, F.; Auvray, M.; Bolze, F.; Naud-Martin, D.; Poulain, M.; Bossuat, M.; Clavier, G. On the Road for More Efficient Biocompatible Two-photon Excitable Fluorophores. *Chem. – Eur. J.* **2022**, chem.202104378. <https://doi.org/10.1002/chem.202104378>.
- (168) Messina, M. A.; Meli, C.; Conoci, S.; Petralia, S. A Facile Method for Urinary Phenylalanine Measurement on Paper-Based Lab-on-Chip for PKU Therapy Monitoring. *The Analyst* **2017**, *142* (24), 4629–4632. <https://doi.org/10.1039/C7AN01115F>.
- (169) Paudics, A.; Hessz, D.; Bojtár, M.; Gyarmati, B.; Szilágyi, A.; Kállay, M.; Bitter, I.; Kubinyi, M. Binding Modes of a Phenylpyridinium Styryl Fluorescent Dye with Cucurbiturils. *Molecules* **2020**, *25* (21), 5111.
- (170) Kolman, V.; Babinský, M.; Kulhánek, P.; Marek, R.; Sindelar, V. Redistribution of Electron Density in Pyridinium and Pyrazinium Guests Induced by Complexation with Cucurbit [6] Uril. *New J. Chem.* **2011**, *35* (12), 2854–2859.
- (171) He, S.; Zhou, C.; Zhang, H.; Zhou, X. Binding Modes of Cucurbit [6] Uril and Cucurbit [7] Uril with a Series of Bis-Pyridinium Compounds. *J. Incl. Phenom. Macrocycl. Chem.* **2013**, *76*, 333–344.
- (172) Vedernikov, A. I.; Lobova, N. A.; Kuz'mina, L. G.; Fomina, M. V.; Strelenko, Y. A.; Howard, J. A.; Gromov, S. P. Self-Assembly of Cucurbiturils and Cyclodextrins to Supramolecular Millstones with Naphthalene Derivatives Capable of Translocations in the Host Cavities. *New J. Chem.* **2019**, *43* (9), 3673–3689.
- (173) Moon, K.; Kaifer, A. E. Modes of Binding Interaction between Viologen Guests and the Cucurbit [7] Uril Host. *Org. Lett.* **2004**, *6* (2), 185–188.
- (174) Assaf, K. I.; Alnajjar, M. A.; Nau, W. M. Supramolecular Assemblies through Host–Guest Complexation between Cucurbiturils and an Amphiphilic Guest Molecule. *Chem. Commun.* **2018**, *54* (14), 1734–1737.
- (175) Huang, Z.; Qin, K.; Deng, G.; Wu, G.; Bai, Y.; Xu, J.-F.; Wang, Z.; Yu, Z.; Scherman, O. A.; Zhang, X. Supramolecular Chemistry of Cucurbiturils: Tuning Cooperativity with Multiple Noncovalent Interactions from Positive to Negative. *Langmuir* **2016**, *32* (47), 12352–12360. <https://doi.org/10.1021/acs.langmuir.6b01709>.
- (176) Zhang, Y.; Zhou, T.; Zhang, K.; Dai, J.; Zhu, Y.; Zhao, X. Encapsulation Enhanced Dimerization of a Series of 4-Aryl-N-Methylpyridinium Derivatives in Water: New Building Blocks for Self-Assembly in Aqueous Media. *Chem. Asian J.* **2014**, *9* (6), 1530–1534.
- (177) Xu, W.; Zhu, X.; Bian, B.; Xiao, X.; Tao, Z.; Redshaw, C. A Study of the Interaction between Cucurbit [7] Uril and Alkyl Substituted 4-Pyrrolidinopyridinium Salts. *Chemistry* **2020**, *2* (2), 262–273.
- (178) Xu, W.; Kan, J.; Yang, B.; Prior, T. J.; Bian, B.; Xiao, X.; Tao, Z.; Redshaw, C. A Study of the Interaction Between Cucurbit [8] Uril and Alkyl-Substituted 4-Pyrrolidinopyridinium Salts. *Chem. Asian J.* **2019**, *14* (1), 235–242.
- (179) Delgado-Pinar, E.; Pont, I.; García-España, E.; Seixas de Melo, J. S. Cucurbituril Hosts as Promoters of Aggregation Induced Emission of Triphenylamine Derivatives. *Phys. Chem. Chem. Phys.* **2022**, *24* (4), 2403–2411. <https://doi.org/10.1039/D1CP04821J>.
- (180) Yu, J.; Wang, H.; Dai, X.; Chen, Y.; Liu, Y. Multivalent Supramolecular Assembly Based on a Triphenylamine Derivative for Near-Infrared Lysosome Targeted Imaging. *ACS Appl. Mater. Interfaces* **2022**, *14* (3), 4417–4422. <https://doi.org/10.1021/acsami.1c19698>.

- (181) Liu, H.; Lin, M.; Cui, Y.; Gan, W.; Sun, J.; Li, B.; Zhao, Y. Single-Crystal Structures of Cucurbituril-Based Supramolecular Host–Guest Complexes for Bioimaging. *Chem. Commun.* **2021**, 57 (79), 10190–10193. <https://doi.org/10.1039/D1CC04823F>.
- (182) Thordarson, P. Determining Association Constants from Titration Experiments in Supramolecular Chemistry. *Chem. Soc. Rev.* **2011**, 40 (3), 1305–1323.
- (183) Long, B. M.; Pfeffer, F. M. On the Use of ‘Shortcuts’ in the Method of Continuous Variation (Job’s Method). *Supramol. Chem.* **2015**, 27 (1–2), 136–140.
- (184) Schneider, H.-J.; Yatsimirsky, A. K. Principles and Methods in Supramolecular Chemistry. *No Title* **2000**.
- (185) Ghai, R.; Falconer, R. J.; Collins, B. M. Applications of Isothermal Titration Calorimetry in Pure and Applied Research—Survey of the Literature from 2010. *J. Mol. Recognit.* **2012**, 25 (1), 32–52.
- (186) Burnouf, D.; Ennifar, E.; Guedich, S.; Puffer, B.; Hoffmann, G.; Bec, G.; Disdier, F.; Baltzinger, M.; Dumas, P. kinITC: A New Method for Obtaining Joint Thermodynamic and Kinetic Data by Isothermal Titration Calorimetry. *J. Am. Chem. Soc.* **2012**, 134 (1), 559–565.
- (187) Ladbury, J. E.; Klebe, G.; Freire, E. Adding Calorimetric Data to Decision Making in Lead Discovery: A Hot Tip. *Nat. Rev. Drug Discov.* **2010**, 9 (1), 23–27.
- (188) Song, C.; Zhang, S.; Huang, H. Choosing a Suitable Method for the Identification of Replication Origins in Microbial Genomes. *Front. Microbiol.* **2015**, 6, 1049.
- (189) Callies, O.; Daranas, A. H. Application of Isothermal Titration Calorimetry as a Tool to Study Natural Product Interactions. *Nat. Prod. Rep.* **2016**, 33 (7), 881–904.
- (190) Rajarathnam, K.; Rösgen, J. Isothermal Titration Calorimetry of Membrane Proteins—Progress and Challenges. *Biochim. Biophys. Acta BBA-Biomembr.* **2014**, 1838 (1), 69–77.
- (191) Brown, A. Analysis of Cooperativity by Isothermal Titration Calorimetry. *Int. J. Mol. Sci.* **2009**, 10 (8), 3457–3477.
- (192) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L.-N. Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter. *Anal. Biochem.* **1989**, 179 (1), 131–137.
- (193) Chodera, J. D.; Mobley, D. L. Entropy–Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annu. Rev. Biophys.* **2013**, 42 (1), 121–142.
- (194) Chennoufi, R.; Trinh, N.-D.; Simon, F.; Bordeau, G.; Naud-Martin, D.; Moussaron, A.; Cinquin, B.; Bougherara, H.; Rambaud, B.; Tauc, P.; Frochot, C.; Teulade-Fichou, M.-P.; Mahuteau-Betzer, F.; Deprez, E. Interplay between Cellular Uptake, Intracellular Localization and the Cell Death Mechanism in Triphenylamine-Mediated Photoinduced Cell Death. *Sci. Rep.* **2020**, 10 (1), 6881. <https://doi.org/10.1038/s41598-020-63991-9>.
- (195) Alnajjar, M. A.; Nau, W. M.; Hennig, A. A Reference Scale of Cucurbit[7]Uril Binding Affinities. *Org. Biomol. Chem.* **2021**, 19 (39), 8521–8529. <https://doi.org/10.1039/D1OB01304A>.
- (196) Urbach, A. R.; Ramalingam, V. Molecular Recognition of Amino Acids, Peptides, and Proteins by Cucurbit[n]Uril Receptors. *Isr. J. Chem.* **2011**, 51 (5–6), 664–678. <https://doi.org/10.1002/ijch.201100035>.
- (197) Chinai, J. M.; Taylor, A. B.; Ryno, L. M.; Hargreaves, N. D.; Morris, C. A.; Hart, P. J.; Urbach, A. R. Molecular Recognition of Insulin by a Synthetic Receptor. *J. Am. Chem. Soc.* **2011**, 133 (23), 8810–8813. <https://doi.org/10.1021/ja201581x>.
- (198) Houtman, J. C. D.; Brown, P. H.; Bowden, B.; Yamaguchi, H.; Appella, E.; Samelson, L. E.; Schuck, P. Studying Multisite Binary and Ternary Protein Interactions by Global Analysis of Isothermal Titration

- Calorimetry Data in SEDPHAT: Application to Adaptor Protein Complexes in Cell Signaling. *Protein Sci.* **2007**, *16* (1), 30–42. <https://doi.org/10.1110/ps.062558507>.
- (199) Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. High-Precision Isothermal Titration Calorimetry with Automated Peak-Shape Analysis. *Anal. Chem.* **2012**, *84* (11), 5066–5073. <https://doi.org/10.1021/ac3007522>.
- (200) Scheuermann, T. H.; Brautigam, C. A. High-Precision, Automated Integration of Multiple Isothermal Titration Calorimetric Thermograms: New Features of NITPIC. *Methods* **2015**, *76*, 87–98. <https://doi.org/10.1016/j.ymeth.2014.11.024>.
- (201) Brautigam, C. A. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. In *Methods in Enzymology*; Elsevier, 2015; Vol. 562, pp 109–133. <https://doi.org/10.1016/bs.mie.2015.05.001>.
- (202) Héberger, K. Chemoinformatics—Multivariate Mathematical–Statistical Methods for Data Evaluation. In *Medical applications of mass spectrometry*; Elsevier, 2008; pp 141–169.
- (203) Molinaro, A. M.; Simon, R.; Pfeiffer, R. M. Prediction Error Estimation: A Comparison of Resampling Methods. *Bioinformatics* **2005**, *21* (15), 3301–3307.
- (204) Yang, J.; Lu, S.; Chen, B.; Hu, F.; Li, C.; Guo, C. Machine Learning-Assisted Optical Nano-Sensor Arrays in Microorganism Analysis. *TrAC Trends Anal. Chem.* **2023**, *159*, 116945.
- (205) Bhalla, Deepanshu. Random Forest in R: A Step-by-Step Guide. <https://www.listendata.com/2014/11/random-forest-with-r.html>.
- (206) Pardo, M.; Sberveglieri, G. Random Forests and Nearest Shrunken Centroids for the Classification of Sensor Array Data. *Sens. Actuators B Chem.* **2008**, *131* (1), 93–99.
- (207) Pang, Z.; Chong, J.; Zhou, G.; de Lima Morais, D. A.; Chang, L.; Barrette, M.; Gauthier, C.; Jacques, P.-É.; Li, S.; Xia, J. MetaboAnalyst 5.0: Narrowing the Gap between Raw Spectra and Functional Insights. *Nucleic Acids Res.* **2021**, *49* (W1), W388–W396. <https://doi.org/10.1093/nar/gkab382>.
- (208) Rana, S.; Singla, A. K.; Bajaj, A.; Elci, S. G.; Miranda, O. R.; Mout, R.; Yan, B.; Jirik, F. R.; Rotello, V. M. Array-Based Sensing of Metastatic Cells and Tissues Using Nanoparticle-Fluorescent Protein Conjugates. *ACS Nano* **2012**, *6* (9), 8233–8240. <https://doi.org/10.1021/nn302917e>.
- (209) Stewart, S.; Ivy, M. A.; Anslyn, E. V. The Use of Principal Component Analysis and Discriminant Analysis in Differential Sensing Routines. *Chem Soc Rev* **2014**, *43* (1), 70–84. <https://doi.org/10.1039/C3CS60183H>.
- (210) Leech, N.; Barrett, K.; Morgan, G. A. *SPSS for Intermediate Statistics: Use and Interpretation*; Routledge, 2013.
- (211) Zou, K. H.; O'Malley, A. J.; Mauri, L. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation* **2007**, *115* (5), 654–657.
- (212) Hosmer Jr, D. W.; Lemeshow, S.; Sturdivant, R. X. *Applied Logistic Regression*; John Wiley & Sons, 2013.
- (213) Nahm, F. S. Receiver Operating Characteristic Curve: Overview and Practical Use for Clinicians. *Korean J. Anesthesiol.* **2022**, *75* (1), 25–36.
- (214) Kubota, R.; Hamachi, I. Protein Recognition Using Synthetic Small-Molecular Binders toward Optical Protein Sensing in Vitro and in Live Cells. *Chem. Soc. Rev.* **2015**, *44* (13), 4454–4471. <https://doi.org/10.1039/C4CS00381K>.

- (215) Krainer, G.; Saar, K. L.; Arter, W. E.; Welsh, T. J.; Czekalska, M. A.; Jacquat, R. P.; Peter, Q.; Traberg, W. C.; Pujari, A.; Jayaram, A. K. Direct Digital Sensing of Protein Biomarkers in Solution. *Nat. Commun.* **2023**, *14* (1), 653.
- (216) Kim, N.; Thomas, M. R.; Bergholt, M. S.; Pence, I. J.; Seong, H.; Charchar, P.; Todorova, N.; Nagelkerke, A.; Belessiotis-Richards, A.; Payne, D. J.; Gelmi, A.; Yarovsky, I.; Stevens, M. M. Surface Enhanced Raman Scattering Artificial Nose for High Dimensionality Fingerprinting. *Nat. Commun.* **2020**, *11* (1), 207. <https://doi.org/10.1038/s41467-019-13615-2>.
- (217) Das Saha, N.; Pradhan, S.; Sasmal, R.; Sarkar, A.; Berač, C. M.; Kölsch, J. C.; Pahwa, M.; Show, S.; Rozenholc, Y.; Topçu, Z.; Alessandrini, V.; Guibourdenche, J.; Tsatsaris, V.; Gagey-Eilstein, N.; Agasti, S. S. Cucurbit[7]Uril Macrocyclic Sensors for Optical Fingerprinting: Predicting Protein Structural Changes to Identifying Disease-Specific Amyloid Assemblies. *J. Am. Chem. Soc.* **2022**, *144* (31), 14363–14379. <https://doi.org/10.1021/jacs.2c05969>.
- (218) Lee, J. W.; Lee, H. H. L.; Ko, Y. H.; Kim, K.; Kim, H. I. Deciphering the Specific High-Affinity Binding of Cucurbit[7]Uril to Amino Acids in Water. *J. Phys. Chem. B* **2015**, *119* (13), 4628–4636. <https://doi.org/10.1021/acs.jpcc.5b00743>.
- (219) Martins, J. N.; Lima, J. C.; Basílio, N. Selective Recognition of Amino Acids and Peptides by Small Supramolecular Receptors. *Molecules* **2020**, *26* (1), 106. <https://doi.org/10.3390/molecules26010106>.
- (220) Leeman, M.; Choi, J.; Hansson, S.; Storm, M. U.; Nilsson, L. Proteins and Antibodies in Serum, Plasma, and Whole Blood—Size Characterization Using Asymmetrical Flow Field-Flow Fractionation (AF4). *Anal. Bioanal. Chem.* **2018**, *410*, 4867–4873.
- (221) Messina, M. A.; Maugeri, L.; Spoto, G.; Puccio, R.; Ruggieri, M.; Petralia, S. Fully Integrated Point-of-Care Platform for the Self-Monitoring of Phenylalanine in Finger-Prick Blood. *ACS Sens.* **2023**, *8* (11), 4152–4160. <https://doi.org/10.1021/acssensors.3c01384>.
- (222) Van Spronsen, F. J.; Blau, N.; Harding, C.; Burlina, A.; Longo, N.; Bosch, A. M. Phenylketonuria. *Nat. Rev. Dis. Primer* **2021**, *7* (1), 36. <https://doi.org/10.1038/s41572-021-00267-0>.
- (223) Himmelreich, N.; Blau, N.; Thöny, B. Molecular and Metabolic Bases of Tetrahydrobiopterin (BH4) Deficiencies. *Mol. Genet. Metab.* **2021**, *133* (2), 123–136. <https://doi.org/10.1016/j.ymgme.2021.04.003>.
- (224) Walkowiak, D.; Bukowska-Posadzy, A.; Kałużny, Ł.; Ołtarzewski, M.; Staszewski, R.; Musielak, M.; Walkowiak, J. Therapy Compliance in Children with Phenylketonuria Younger than 5 Years: A Cohort Study. *Adv. Clin. Exp. Med.* **2019**, *28* (10), 1385–1391.
- (225) Cazzorla, C.; Bensi, G.; Biasucci, G.; Leuzzi, V.; Manti, F.; Musumeci, A.; Papadia, F.; Stoppioni, V.; Tummolo, A.; Vendemiale, M. Living with Phenylketonuria in Adulthood: The PKU ATTITUDE Study. *Mol. Genet. Metab. Rep.* **2018**, *16*, 39–45.
- (226) Alptekin, I. M.; Koc, N.; Gunduz, M.; Cakiroglu, F. P. The Impact of Phenylketonuria on PKU Patients' Quality of Life: Using of the Phenylketonuria-Quality of Life (PKU-QOL) Questionnaires. *Clin. Nutr. ESPEN* **2018**, *27*, 79–85.
- (227) Jafari, P.; Beigi, S. M.; Yousefi, F.; Aghabalazadeh, S.; Mousavizadegan, M.; Hosseini, M.; Hosseinkhani, S.; Ganjali, M. R. Colorimetric Biosensor for Phenylalanine Detection Based on a Paper Using Gold Nanoparticles for Phenylketonuria Diagnosis. *Microchem. J.* **2021**, *163*, 105909. <https://doi.org/10.1016/j.microc.2020.105909>.
- (228) Langenbeck, U.; Baum, F.; Mench-Hoinowski, A.; Luthe, H.; Behbehani, A. W. Predicting the Phenylalanine Blood Concentration from Urine Analyses. An Approach to Noninvasive Monitoring of Patients with Phenylketonuria. *J. Inherit. Metab. Dis.* **2005**, *28* (6), 855–861. <https://doi.org/10.1007/s10545-005-0160-4>.

- (229) Messina, M. A.; Maugeri, L.; Forte, G.; Ruggieri, M.; Petralia, S. A Highly Sensitive Colorimetric Approach Based on Tris (Bipyridine) Ruthenium (II/III) Mediator for the Enzymatic Detection of Phenylalanine. *Front. Chem.* **2023**, *11*, 1164014. <https://doi.org/10.3389/fchem.2023.1164014>.
- (230) Sánchez, S. 1 A Preclinical Biosensor for Detecting Phenylalanine Photometrically 2 in Plasma and Whole Blood Samples.
- (231) Weiss, D. J.; Dorris, M.; Loh, A.; Peterson, L. Dehydrogenase Based Reagentless Biosensor for Monitoring Phenylketonuria. *Biosens. Bioelectron.* **2007**, *22* (11), 2436–2441. <https://doi.org/10.1016/j.bios.2006.09.001>.
- (232) Sun, B.; Wang, Z.; Wang, X.; Qiu, M.; Zhang, Z.; Wang, Z.; Cui, J.; Jia, S. Paper-Based Biosensor Based on Phenylalanine Ammonia Lyase Hybrid Nanoflowers for Urinary Phenylalanine Measurement. *Int. J. Biol. Macromol.* **2021**, *166*, 601–610. <https://doi.org/10.1016/j.ijbiomac.2020.10.218>.
- (233) Deon, M.; Sitta, A.; Faverzani, J. L.; Guerreiro, G. B.; Donida, B.; Marchetti, D. P.; Mescka, C. P.; Ribas, G. S.; Coitinho, A. S.; Wajner, M.; Vargas, C. R. Urinary Biomarkers of Oxidative Stress and Plasmatic Inflammatory Profile in Phenylketonuric Treated Patients. *Int. J. Dev. Neurosci.* **2015**, *47* (Part_B), 259–265. <https://doi.org/10.1016/j.ijdevneu.2015.10.001>.
- (234) Sibai, B.; Dekker, G.; Kupferminc, M. Pre-Eclampsia. *The Lancet* **2005**, *365* (9461), 785–799.
- (235) Steegers, E. A.; Von Dadelszen, P.; Duvekot, J. J.; Pijnenborg, R. Pre-Eclampsia. *The lancet* **2010**, *376* (9741), 631–644.
- (236) Maynard, S.; Epstein, F. H.; Karumanchi, S. A. Preeclampsia and Angiogenic Imbalance. *Annu. Rev. Med.* **2008**, *59* (1), 61–78. <https://doi.org/10.1146/annurev.med.59.110106.214058>.
- (237) Baumwell, S.; Karumanchi, S. A. Pre-Eclampsia: Clinical Manifestations and Molecular Mechanisms. *Nephron Clin. Pract.* **2007**, *106* (2), c72–c81.
- (238) Trapiella-Alfonso, L.; Broussy, S.; Liu, W.-Q.; Vidal, M.; Lecarpentier, E.; Tsatsaris, V.; Gagey-Eilstein, N. Colorimetric Immunoassays for the Screening and Specificity Evaluation of Molecules Disturbing VEGFs/VEGFRs Interactions. *Anal. Biochem.* **2018**, *544*, 114–120.
- (239) Trapiella-Alfonso, L.; Alexandre, L.; Fraichard, C.; Pons, K.; Dumas, S.; Huart, L.; Gaucher, J.-F.; Hebert-Schuster, M.; Guibourdenche, J.; Fournier, T. VEGF (Vascular Endothelial Growth Factor) Functionalized Magnetic Beads in a Microfluidic Device to Improve the Angiogenic Balance in Preeclampsia. *Hypertension* **2019**, *74* (1), 145–153.
- (240) Tsatsaris, V.; Goffin, F.; Munaut, C.; Brichant, J.-F.; Pignon, M.-R.; Noel, A.; Schaaps, J.-P.; Cabrol, D.; Frankenne, F.; Foidart, J.-M. Overexpression of the Soluble Vascular Endothelial Growth Factor Receptor in Preeclamptic Patients: Pathophysiological Consequences. *J. Clin. Endocrinol. Metab.* **2003**, *88* (11), 5555–5563.
- (241) Mol, B. W.; Roberts, C. T.; Thangaratinam, S.; Magee, L. A.; De Groot, C. J.; Hofmeyr, G. J. Pre-Eclampsia. *The Lancet* **2016**, *387* (10022), 999–1011.
- (242) Magee, L. A.; Pels, A.; Helewa, M.; Rey, E.; von Dadelszen, P.; Audibert, F.; Bujold, E.; Côté, A.-M.; Douglas, M. J.; Eastabrook, G. Diagnosis, Evaluation, and Management of the Hypertensive Disorders of Pregnancy: Executive Summary. *J. Obstet. Gynaecol. Can.* **2014**, *36* (5), 416–438.
- (243) ACOG Committee on Obstetric Practice. ACOG Practice Bulletin. Diagnosis and Management of Preeclampsia and Eclampsia. Number 33, January 2002. American College of Obstetricians and Gynecologists. *Int. J. Gynaecol. Obstet. Off. Organ Int. Fed. Gynaecol. Obstet.* **2002**, *77* (1), 67–75.
- (244) Menzies, J.; Magee, L.; Macnab, Y.; Ansermino, J.; Li, J.; Douglas, M.; Gruslin, A.; Kyle, P.; Lee, S.; Moore, M. Current CHS and NHBPEP Criteria for Severe Preeclampsia Do Not Uniformly Predict Adverse Maternal or Perinatal Outcomes. *Hypertens. Pregnancy* **2007**, *26* (4), 447–462.

- (245) Gillon, T. E.; Pels, A.; von Dadelszen, P.; MacDonell, K.; Magee, L. A. Hypertensive Disorders of Pregnancy: A Systematic Review of International Clinical Practice Guidelines. *PLoS One* **2014**, *9* (12), e113715.
- (246) Pennington, K. A.; Schlitt, J. M.; Jackson, D. L.; Schulz, L. C.; Schust, D. J. Preeclampsia: Multiple Approaches for a Multifactorial Disease. *Dis. Model. Mech.* **2012**, *5* (1), 9–18.
- (247) Zeisler, H.; Llorba, E.; Chantraine, F.; Vatish, M.; Staff, A. C.; Sennström, M.; Olovsson, M.; Brennecke, S. P.; Stepan, H.; Allegranza, D. Predictive Value of the sFlt-1: PlGF Ratio in Women with Suspected Preeclampsia. *N. Engl. J. Med.* **2016**, *374* (1), 13–22.
- (248) Teh, S.-Y.; Lin, R.; Hung, L.-H.; Lee, A. P. Droplet Microfluidics. *Lab. Chip* **2008**, *8* (2), 198–220.
- (249) Liu, W.; Zhu, Y. “Development and Application of Analytical Detection Techniques for Droplet-Based Microfluidics”-A Review. *Anal. Chim. Acta* **2020**, *1113*, 66–84.
- (250) Ferraro, D.; Serra, M.; Filippi, D.; Zago, L.; Guglielmin, E.; Pierno, M.; Descroix, S.; Viovy, J.-L.; Mistura, G. Controlling the Distance of Highly Confined Droplets in a Capillary by Interfacial Tension for Merging On-Demand. *Lab. Chip* **2019**, *19* (1), 136–146.
- (251) Germain, M. E.; Knapp, M. J. Optical Explosives Detection: From Color Changes to Fluorescence Turn-On. *Chem. Soc. Rev.* **2009**, *38* (9), 2543–2555.
- (252) Zhang, Z.; Kim, D. S.; Lin, C.-Y.; Zhang, H.; Lammer, A. D.; Lynch, V. M.; Popov, I.; Miljanić, O. Š.; Anslyn, E. V.; Sessler, J. L. Expanded Porphyrin-Anion Supramolecular Assemblies: Environmentally Responsive Sensors for Organic Solvents and Anions. *J. Am. Chem. Soc.* **2015**, *137* (24), 7769–7774.
- (253) Zhong, X.; Huo, D.; Fa, H.; Luo, X.; Wang, Y.; Zhao, Y.; Hou, C. Rapid and Ultrasensitive Detection of Biogenic Amines with Colorimetric Sensor Array. *Sens. Actuators B Chem.* **2018**, *274*, 464–471.
- (254) Li, F.; Stewart, C.; Yang, S.; Shi, F.; Cui, W.; Zhang, S.; Wang, H.; Huang, H.; Chen, M.; Han, J. Optical Sensor Array for the Early Diagnosis of Alzheimer’s Disease. *Front. Chem.* **2022**, *10*, 874864. <https://doi.org/10.3389/fchem.2022.874864>.
- (255) Jin, Y.; Du, N.; Huang, Y.; Shen, W.; Tan, Y.; Chen, Y. Z.; Dou, W.-T.; He, X.-P.; Yang, Z.; Xu, N.; Tan, C. Fluorescence Analysis of Circulating Exosomes for Breast Cancer Diagnosis Using a Sensor Array and Deep Learning. *ACS Sens.* **2022**, *7* (5), 1524–1532. <https://doi.org/10.1021/acssensors.2c00259>.
- (256) Akuse, R. M.; Eke, F. U.; Ademola, A. D.; Fajolu, I. B.; Gbelee, H. O.; Ihejiahi, U.; Bugaje, M. A.; Anochie, I. C.; Asinobi, A. O.; Okafor, H. U.; Adeleke, S. I.; Audu, L. I.; Otuneye, A.; Disu, E.; Idris, H.; Aikhonbare, H.; Yakubu, A.; Ogala, W.; Ogunrinde, O.; Wammanda, R.; Orogade, A.; Anyiam, J.; Eseigbe, E.; Umar, L.; Musa, H.; Onalo, R.; West, B.; Paul, N.; Lesi, F.; Ladapo, T.; Boyede, O.; Okeowo, R.; Mustapha, A.; Akinola, I.; Chima-Oduko, O.; Awobusuyi, O. Diagnosing Renal Failure Due to Diethylene Glycol in Children in a Resource-Constrained Setting. *Pediatr. Nephrol.* **2012**, *27* (6), 1021–1028. <https://doi.org/10.1007/s00467-011-2082-8>.
- (257) Stanton, C.; Koski, A.; Cofie, P.; Mirzabagi, E.; Grady, B. L.; Brooke, S. Uterotonic Drug Quality: An Assessment of the Potency of Injectable Uterotonic Drugs Purchased by Simulated Clients in Three Districts in Ghana. *BMJ Open* **2012**, *2* (3), e000431. <https://doi.org/10.1136/bmjopen-2011-000431>.
- (258) Committee on Understanding the Global Public Health Implications of Substandard, Falsified, and Counterfeit Medical Products; Board on Global Health; Institute of Medicine. *Countering the Problem of Falsified and Substandard Drugs*; Gostin, L. O., Buckley, G. J., Eds.; National Academies Press: Washington, D.C., 2013; p 18272. <https://doi.org/10.17226/18272>.
- (259) Nickerson, J. W.; Chikumba, E. Access to Medicines for Improving Access to Safe Anesthetic Care. *Anesth. Analg.* **2018**, *126* (4), 1405–1408. <https://doi.org/10.1213/ANE.0000000000002746>.

- (260) WHO Model List of Essential Medicines - 23rd list, 2023. <https://www.who.int/publications-detail-redirect/WHO-MHP-HPS-EML-2023.02> (accessed 2024-03-05).
- (261) Ofori-Parku, S. S. Fighting the Global Counterfeit Medicines Challenge: A Consumer-Facing Communication Strategy in the US Is an Imperative. *J. Glob. Health* **2022**, *12*, 03018. <https://doi.org/10.7189/jogh.12.03018>.
- (262) Imran, E.; Moeen, F.; Abbas, B.; Yaqoob, B.; Wajahat, M.; Khan, Q.; Khurshid, Z. Comparative Analysis of Different Local Anesthetic Solutions Available in Market: An In Vitro and Clinical Study. *Eur. J. Dent.* **2021**, *15* (04), 660–668. <https://doi.org/10.1055/s-0041-1727553>.
- (263) Kovacs, S.; Hawes, S. E.; Maley, S. N.; Mosites, E.; Wong, L.; Stergachis, A. Technologies for Detecting Falsified and Substandard Drugs in Low and Middle-Income Countries. *PLoS ONE* **2014**, *9* (3), e90601. <https://doi.org/10.1371/journal.pone.0090601>.
- (264) Selinger, A. J.; Krämer, J.; Poarch, E.; Hore, D.; Biedermann, F.; Hof, F. Mixed Host Co-Assembled Systems for Broad-Scope Analyte Sensing. *Chem. Sci.* **2024**.
- (265) Han, J.; Wang, B.; Bender, M.; Kushida, S.; Seehafer, K.; Bunz, U. H. F. Poly(Aryleneethynylene) Tongue That Identifies Nonsteroidal Anti-Inflammatory Drugs in Water: A Test Case for Combating Counterfeit Drugs. *ACS Appl. Mater. Interfaces* **2017**, *9* (1), 790–797. <https://doi.org/10.1021/acsami.6b11690>.
- (266) Liu, Y.; Minami, T.; Nishiyabu, R.; Wang, Z.; Anzenbacher, P. Sensing of Carboxylate Drugs in Urine by a Supramolecular Sensor Array. *J. Am. Chem. Soc.* **2013**, *135* (20), 7705–7712. <https://doi.org/10.1021/ja4015748>.
- (267) Frosch, T.; Wyrwich, E.; Yan, D.; Domes, C.; Domes, R.; Popp, J.; Frosch, T. Counterfeit and Substandard Test of the Antimalarial Tablet Riamet® by Means of Raman Hyperspectral Multicomponent Analysis. *Molecules* **2019**, *24* (18), 3229. <https://doi.org/10.3390/molecules24183229>.
- (268) Wyman, I. W.; Macartney, D. H. Host–Guest Complexations of Local Anaesthetics by Cucurbit [7] Uril in Aqueous Solution. *Org. Biomol. Chem.* **2010**, *8* (1), 247–252.
- (269) Minami, T.; Esipenko, N. A.; Akdeniz, A.; Zhang, B.; Isaacs, L.; Anzenbacher, P. Multianalyte Sensing of Addictive Over-the-Counter (OTC) Drugs. *J. Am. Chem. Soc.* **2013**, *135* (40), 15238–15243. <https://doi.org/10.1021/ja407722a>.
- (270) Dsouza, R. N.; Pischel, U.; Nau, W. M. Fluorescent Dyes and Their Supramolecular Host/Guest Complexes with Macrocycles in Aqueous Solution. *Chem. Rev.* **2011**, *111* (12), 7941–7980. <https://doi.org/10.1021/cr200213s>.