

Discovery of Photoactive Anti-cancer Platinum(II) Complexes: Synthesis, Screening and Machine Learning Approaches

Molly Bartlett

A thesis presented for the degree of Doctor of Philosophy
Imperial College London
Department of Chemistry

March 2025

Abstract

A prevailing issue with current anti-cancer therapeutics is non-specific drug toxicity leading to a range of negative side effects. One way to avoid these off-target effects is through the use of pro-drugs: drugs which are non-toxic until activated by a trigger such as an enzyme, change in pH or light. However, the development of new drugs is a highly laborious process, which takes many years and costs millions of pounds. Recent advances in drug discovery have been accelerated by the use of data driven and machine learning approaches. This thesis presents a case study applying such approaches to the discovery of photoactive complexes as anti-cancer therapeutics.

Within this thesis 22 new platinum(II) salphen complexes were synthesised during the development of a high throughput (HT) synthesis and screening pipeline. Platinum(II) salphen complexes were selected for their unique photophysical properties and known binding affinity to alternate DNA secondary structures highly enriched in cancer cells. To determine the potential drug-like properties of the complexes, photoactivity, DNA binding, DNA damage, and cellular activity were screened. As a result, two lead candidates were identified which had good DNA binding affinity and selectivity, cell permeability, and localisation within the nucleus. These complexes were shown to be non-toxic within cancer cells until irradiated, at which point they exhibited IC_{50} values of 0.29 μ M and 2.57 μ M.

Alongside the experimental work, computational calculations were performed on a larger library of 263 platinum(II) salphen complexes for the application of machine learning (ML) prediction tasks. The objectives of these ML models were to better understand structure activity relationships (SAR) and predict photophysical properties of unseen platinum(II) salphen complexes.

Overall, this work has combined data driven experimental and computational approaches for the synthesis, screening and ML analysis of platinum(II) salphen complexes as potential novel photoactive anti-cancer therapeutics.

Contents

List of Figures	5
List of Schemes	10
List of Tables	11
List of Abbreviations	13
Acknowledgments	15
Declaration of Originality	16
Declaration of Copyright	16
List of Publications	17
1 Introduction	18
1.1 Drug discovery: past, present and future	18
1.2 Anti-cancer therapeutics	30
1.3 G-quadruplex DNA as an anti-cancer therapeutic target	35
1.4 Project overview and objectives	48
2 Synthesis of platinum(II) salphen complexes	51
2.1 Platinum(II) salphen complex library design	51
2.2 Platinum(II) salphen complex synthesis	54
2.3 High throughput synthesis	76
2.4 Synthesis conclusions	81
3 Photophysical and biological properties of platinum(II) salphen complexes	83
3.1 Photophysical properties	85
3.2 ROS generation	90
3.3 DNA binding	94
3.4 ROS induced DNA damage	109

3.5	Cellular behaviour	116
3.6	Correlation analysis of photophysical and biological properties	132
3.7	Screening conclusions	135
4	Computational modelling of platinum(II) salphen geometries and photophysical properties	139
4.1	Introduction to DFT and TDDFT	139
4.2	Relationships between DFT and photophysical properties	145
4.3	Correlation analysis of DFT energies and experimental properties . . .	149
4.4	DFT and TDDFT conclusions	152
5	Machine learning analysis on the photophysical properties of metal complexes	153
5.1	Computational generation of complexes	155
5.2	Molecular descriptors	157
5.3	Development of machine learning predictive models	171
5.4	Determination of features important for DFT and TDDFT energies . . .	197
5.5	Computational conclusions	201
6	Conclusions and future outlook	204
7	Experimental details	208
7.1	Material and methods	208
7.2	Synthetic methods	214
	References	237
A	Appendix	282
A.1	Synthetic spectra	282
A.2	Screening data	393
A.3	Computational data	446

List of Figures

1.1	Natural products and the first synthetic drug	19
1.2	Aspirin derived from a natural product	19
1.3	First drug developed by SBDD: Dorzolamide	20
1.4	Alternate DNA secondary structures	22
1.5	Drug discovery and development funnel	23
1.6	AI driven HT screening	28
1.7	First AI designed <i>de novo</i> drug: INS018-055	30
1.8	Anti-cancer alkylating agents	31
1.9	Anti-cancer photosensitiser agents	32
1.10	Photodynamic therapy schematic	33
1.11	First AI <i>de novo</i> designed cancer drug: ISM5939	34
1.12	G-quadruplex structure and topologies	35
1.13	G-quadruplex ligand binding modes	38
1.14	G-quadruplex ligands which entered clinical trials	39
1.15	Timeline of G-quadruplex ligand research publications	40
1.16	Note worthy organic G-quadruplex ligands	41
1.17	Metalloporphyrin derivatives of TMPyP4	42
1.18	Metal salphen geometries with different metal centres	44
1.19	Ruthenium polypyridyl G4 binders	45
1.20	Metallophthalocyanines G4 binders	46
1.21	Metal-terpyridine and NNC polypyridyl G4 binders	47
1.22	Metal-phenanthroline G4 binders	47
2.1	Metal salphen scaffold	51
2.2	CBA building blocks	52

2.3	DAB building blocks	53
2.4	Ether DAB hydrogenation reaction ¹ H NMR spectra	56
2.5	¹ H NMR spectra showing degradation of the ether linked DAB	57
2.6	¹ H NMR spectra of zinc templating and transmetalation	58
2.7	¹ H NMR spectrum of Pt_PEM_Ph	59
2.8	¹ H NMR spectra of Pt_PEP_Ph	60
2.9	¹ H NMR spectra of Pt_AEDEA_Ph	63
2.10	¹ H NMR spectrum of the mono-reacted AEDEA DAB with Ph CBA	64
2.11	¹ H NMR spectrum of Zn_AEDEA_Ph	65
2.12	¹ H NMR spectrum of Zn_AEDEA_Ph reaction mixture formed at a higher temperature	65
2.13	¹ H NMR spectrum of failed Zn_AEDEA_Ph transmetalation reaction . . .	66
2.14	¹ H NMR spectrum for isolated Zn_AEDEA_Ph	67
2.15	¹ H NMR spectra of Zn_AEDEA_Ph transmetalation to Pt_AEDEA_Ph. . .	67
2.16	Aromatic ¹ H NMR spectrum of the Zn_AEDEA_4OMe reaction mixture. .	68
2.17	¹ H NMR spectrum of isolated precipitate from Zn_AEDEA_4OMe transmetalation reaction	69
2.18	AEP and AEM DAB platinum(II) salphen complexes	69
2.19	¹ H NMR spectrum of Pt_AEP_4OMe	70
2.20	¹ H NMR spectrum for Pt_PhBr_Ph	73
2.21	Pt_PhBr_Ph to triazole reaction scheme	73
2.22	Pure isolated platinum(II) salphen complexes	75
2.23	HT synthesis DABs and additional CBAs	77
2.24	HT synthesis reaction plate layout	77
2.25	HT reaction OT2 schematic plan	78
2.26	HT synthesis ¹ H NMR spectra	79
2.27	HT synthesis ¹ H NMR spectra	79
2.28	HT reaction plates after each step	80
3.1	Schematic overview of chapter 3	83
3.2	Jablonski diagram	84
3.3	Platinum(II) salphen absorbance spectra	85
3.4	Platinum(II) salphen emission spectra	86

3.5	Platinum(II) salphen complex physical appearance	87
3.6	Platinum(II) salphen complex peak emission and absorbance wavelength strip plot	88
3.7	Peak emission wavelengths vs Hammett constants for the CBA functionalities	89
3.8	Platinum(II) salphen complex ROS generating abilities.	92
3.9	ROS generation against complex components strip plot.	93
3.10	Emission titration to determine binding affinity	95
3.11	Hand prepared HTG4 binding assay affinities	97
3.12	OT2 calibration curve	98
3.13	OT2 prepared HTG4 binding assay affinities	99
3.14	OT2 prepared DNA binding assay screen affinities	100
3.15	DNA binding affinity strip plots	101
3.16	Docking simulation of Pt_PPTMA_Ph with DNA targets of interest . . .	105
3.17	Docking simulation of complexes with different DAB with <i>c-myc</i> G4 DNA	106
3.18	Docking simulation of 5OMe and 5OCF ₃ complexes with HTG4	107
3.19	Docking simulation of Pt_PPTMA_Naph with LTR(III) G4 DNA	107
3.20	Topologies of plasmid DNA (pUC19)	110
3.21	Image of ROS DNA damage gel and the percentage of DNA cleavage achieved by each complex	111
3.22	Image of quenched ROS DNA damage gel and the percentage of DNA cleavage achieved by in the presence of each ROS quencher	113
3.23	CD spectra to track ROS induced G4 damage	114
3.24	Change in ellipticity for the CD spectra tracking ROS induced G4 damage	115
3.25	Change in ellipticity for the CD spectra of G4s incubated with complexes in the dark	116
3.26	Cytotoxicity dose response curves	118
3.27	Phototoxicity dose response curves	120
3.28	Phototoxicity IC ₅₀ values	120
3.29	Images of platinum(II) salphen complexes in HeLa cells from the HT imaging screen	123
3.30	Strip plot for the total cell fluorescence of complexes in HeLa cells . . .	124

3.31 Confocal images of complexes in HeLa cells	126
3.32 Emission spectra of complexes and co-stains used within confocal imaging experiments	127
3.33 Confocal co-stain experiments with Pt_AEP_4OMe	128
3.34 Confocal co-stain experiments with Pt_PEM_4OMe	129
3.35 Confocal co-stain experiments with Pt_PEM_Ph	129
3.36 Confocal co-stain experiments with Pt_PPTMA_4NEt ₂	130
3.37 Confocal images of platinum(II) salphen complexes at increasing incubation periods.	132
3.38 Pearson's correlation between experimental properties.	133
3.39 Lead complexes identified from experimental work	138
4.1 Simplified Jablonksi diagram	145
4.2 Scatter plots showing the peak emission wavelength verses T ₁ to S ₀ transition energy and HOMO LUMO gap for the platinum(II) salphen complexes	146
4.3 HOMO and LUMO orbital diagrams of platinum(II) salphen complexes .	148
4.4 Pearson's correlation coefficient between experimental properties and DFT and TDDFT values	150
5.1 Schematic overview of chapter 5	154
5.2 Platinum(II) salphen building blocks used within computational structure generation	155
5.3 Steps of computational structure generation of platinum(II) salphen complexes	156
5.4 Fingerprint descriptor generation schematic	158
5.5 t-SNE dimensionality reduction plots: fingerprint descriptors	161
5.6 t-SNE dimensionality reduction plots: Mordred descriptors	163
5.7 t-SNE dimensionality reduction plots: PCA reduced fingerprint descriptors	167
5.8 t-SNE dimensionality reduction plots: PCA reduced Mordred descriptors	169
5.9 Platinum(II) complexes used to validate prediction models	180
5.10 Visualisation of the best T ₁ to S ₀ transition energy prediction model. . .	182
5.11 Examples of zinc(II) salphen and platinum(II) NNC complexes	186

5.12 Prediction of zinc(II) salphen emission energies with the original platinum(II) salphen models	188
5.13 Visualisation of the best zinc(II) salphen emission energy TL prediction model	190
5.14 Visualisation of the best zinc(II) salphen emission energy prediction model	191
5.15 Visualisation of the best platinum(II) NNC emission energy TL prediction model	193
5.16 Visualisation of the best platinum(II) NNC emission energy prediction model	193
5.17 Visualisation of the best zinc(II) salphen emission energy DA prediction model	195
5.18 Visualisation of the best platinum(II) NNC emission energy DA prediction model	196
5.19 Visualisation of molecular features important for the HOMO and LUMO energies	198
5.20 Visualisation of molecular features important for the DFT and TDDFT transition energies	200

List of Schemes

2.1	General platinum(II) salphen complexation reaction	53
2.2	Ether linked DAB synthesis	55
2.3	Ether linked DAB hydrogenation	55
2.4	Amide linked DAB first attempted synthetic route	61
2.5	Amide linked DAB first attempted synthetic route	62
2.6	Pt_AEDEA_Ph synthetic attempt 1.	63
2.7	Pt_PhBr_Ph synthetic route	72
3.1	ABDA ROS detection dye reaction with singlet oxygen	90
3.2	DHR-123 ROS detection dye reaction with radical oxygen species . . .	91
3.3	Reaction of MTS in the presence of metabolically active cells	117

List of Tables

2.1	Table showing the area of the UV peak and the masses detected for each isolated zinc(II) salphen intermediate complex by LCMS.	70
2.2	Table of LCMS data for the amide functionalised AEP and AEM platinum(II) salphen complexes.	71
3.1	Emission titration serial dilution map	96
3.2	Binding affinities and docking scores for complexes with HTG4 DNA . .	108
3.3	ROS generating abilities of complexes used within the DNA damage assays	110
3.4	Initial ROS induced cytotoxicity screen	119
3.5	Table summarising best and worst complexes IC_{50} , ROS generation and DNA binding	121
5.1	Fingerprint descriptor k-mean clusters	162
5.2	Mordred descriptor k-mean clusters	164
5.3	Fingerprint PCA k-mean clusters	168
5.4	SVM regresson kernels	175
5.5	T_1 to S_0 ML prediction model performance	181
5.6	T_1 to S_0 ML prediction model validation	183
5.7	Best ML model for each DFT and TDDFT energy	184
5.8	Platinum(II) salphen ML models for predicting zinc(II) salphen emission energy	187
5.9	Extended platinum(II) salphen ML models for predicting zinc(II) salphen emission energy	187
5.10	TL models of zinc(II) salphen emission energy prediction	190
5.11	TL models of platinum(II) NNC emission energy prediction	192

5.12 DA models of zinc(II) salphen emission energy prediction	195
5.13 DA models of platinum(II) NNC emission energy prediction	196
7.1 List of G4 sequences	209
7.2 Reaction conditions for the synthesis of the PPTMA solubilised platinum(II) salphen complexes.	220
7.3 Reaction conditions for the synthesis of the PEM solubilised platinum(II) salphen complexes.	220
7.4 Reaction conditions for the synthesis of the PhBr DAB platinum(II) salphen complexes	221
7.5 Reaction conditions for the synthesis of the AEP solubilised zinc(II) salphen complexes	221
7.6 Reaction conditions for the synthesis of the AEP solubilised platinum(II) salphen complexes	222

List of Abbreviations

- ABDA: Singlet oxygen sensor
- ADME: Absorbance, distribution, metabolism and excretion
- AI: Artificial intelligence
- CADD: Computer aided drug design
- CBA: Carboxaldehyde
- CD: Circular dichroism
- CTCF: Corrected total cell fluorescence
- CV: Monte Carlo cross validation
- DA: Domain adaptation
- DAB: Diamine backbone
- DFT: Density functional theory
- DHR-123: Radical oxygen species sensor
- ED/EW: Electron donating / Electron withdrawing
- FBDD: Fragment based drug design
- G4: G-quadruplex
- HT: High throughput
- HRMS: High resolution mass spectrometry
- KS: Kohn-Sham

- LCMS: Liquid chromatography mass spectrometry
- MAE: Mean absolute error
- ML: Machine learning
- MS: Mass spectrometry
- NMR: Nuclear magnetic resonance
- NN: Neural networks
- OLS: Ordinary least squares
- OT2: Opentron2
- PC: Pearson's correlation
- PCA: Principle component analysis
- PDB: Protein data bank
- PDT: Photodynamic therapy
- PET: Photoinduced electron transfer
- RF: Random forest
- ROS: Reactive oxygen species
- SAR: Structure activity relationships
- SBDD: Structure based drug design
- SMILES: Simplified molecular input line entry system
- SVM: Support vector machine regression
- TDDFT: Time dependent density functional theory
- TL: Transfer learning
- TPSA: Topological polar surface area
- UMAP: Uniformed Manifold Approximation and Projection
- XGB: XGBoost

Acknowledgments

During my time at Imperial, I have had the pleasure of meeting amazing people and having unforgettable experiences.

I am incredibly grateful for my primary supervisor Ramon Vilar, who leads his group with compassion. Thanks to his support I have been able to thrive throughout my PhD, continuously encouraged by his genuine enthusiasm. I would also like to thank my secondary supervisor, Mauricio Barahona, with whom I had many long and insightful conversations, leading me down fascinating avenues within the machine learning side of my project.

I am thankful to everyone who contributed to my skill development which made this project possible: Dr Tim Kench and Dr Aatikah Majid for endless support with synthesis and biological screening. Dr Jenna Robinson and Dr Petr Sherin with the cellular work. Dr Chris Roberts for his excellent training and support on the automation systems in ROAR. Dr Andrew Tarzia for all his support with using the STK toolkit. Wilson Wu and Joseph Burke for their work on the DFT automation pipeline. Having Joe as my MSci was incredibly rewarding, and I felt very proud to see him go on to do a PhD of his own.

I have treasured working within the Vilar group, with special shout outs to Aatikah, Libby, Wilson, and Amy, in addition to Sofia (honorary member), for all the laughs throughout the years. I was also lucky enough to be a part of the React CDT, which provided a strong sense of community, in addition to so many exciting opportunities and experiences. I've had the pleasure of travelling the world with the Cohort 2 Cult, and meeting some of my closest friends: Izzy, Sean, Tom and Andrew. I believe many of the friends I have made during my time at Imperial will be friends for life.

However, the PhD was not all roses, there were some very challenging times, during which I received endless support from my supervisors, friends, and family. I am also incredibly grateful for my partner, Jenna, who supported me throughout my PhD, encouraging me to the best version of myself, and to keep pushing through the challenges.

Declaration of Originality

I, Molly Bartlett, declare that the results presented in this thesis represent original work performed by myself. In cases where experiments were performed in collaboration with others, it has been declared within the text. All external sources have been referenced and acknowledged.

Declaration of Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence (CC BY NC-SA).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that you credit the author, do not use it for commercial purposes and share any derivative works under the same licence.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Where work published elsewhere has been used in this thesis, it has been used under a Creative Commons license.

List of Publications

- **M. Bartlett** J. Burke, P. Sherin, M. K. Kuimova, M. Barahona, R. Vilar*, Platinum(II)-Salphen Complexes as DNA Binders and Photosensitisers, *Chemistry - A European Journal*, 30, e202402465, 2024.
- V. Iannace, F. Sabaté, **M. Bartlett**, J. Berrones Reyes, A. Lázaro, A. Fantoni, R. Vilar*, L. Rodríguez, A. Dalla Cort., Syntheses, Characterization, and DNA Binding Studies of a Series of Copper(II), Nickel(II), Platinum(II) and Zinc(II) Complexes Derived from Schiff Base Ligands, *European Journal of Inorganic Chemistry*, 26, e202300144, 10, 2023.
- J. Robinson, S. G. Stenspil, K. Maleckaite, **M. Bartlett**, M. Di Antonio, R. Vilar and M. Kuimova, Cellular visualization of G-quadruplex RNA via fluorescence lifetime imaging microscopy, *Journal of the American Chemistry Society*, 146 (1), 1009-1018, 2023.

Chapter 1

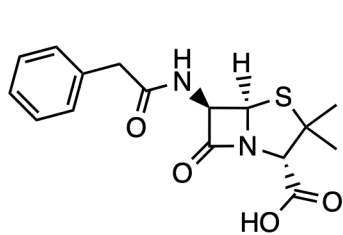
Introduction

1.1 Drug discovery: past, present and future

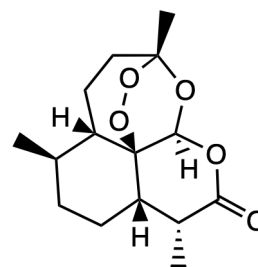
The use of medicines can be traced back to ancient civilizations, with one of the earliest recorded evidence being the Egyptian medical papyri [1], a collection of texts dating back to 1800 BCE, detailing the use of herbal remedies within medical procedures. However, it wasn't until Hippocrates of Kos in 400 BC that it was theorised that diseases were a result of natural causes as opposed to a result of superstition, or an act of God. Within his lifetime he developed and used over 300 medicines derived from plants, leading him to be known as the father of modern medicine [2].

The activity of many medicines originate from natural products, drug-like compounds which occur in nature. Some notable natural products include Penicillin [3], an antibiotic derived from *Penicillium* mould, Artemisinin [4, 5], an anti-malarial derived from the *Artemisia annua* plant, and Taxol [6], a cancer drug derived from the bark of the Pacific yew tree. It wasn't until 1869 that the first synthetic drug was developed: chloral hydrate [7], an anaesthetic agent, which is still used today as a sedative (Noctec®) (figure 1.1).

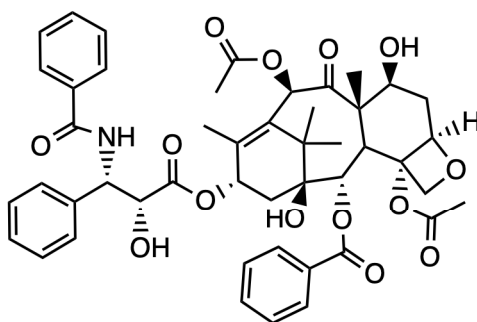
The discovery of new medicines can be approached by an array of methods. The first being serendipity, the discovery of drugs by chance, as was the case with Penicillin. Another method involves structural modification of known drugs or natural products to improve their therapeutic properties. An example of this is the anti-inflammatory medicine Aspirin: a structural derivative of salicylic acid [8], a natural product derived from willow bark, differing only by the addition of a single acyl group (figure 1.2).



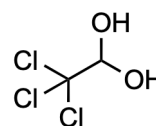
Penicillin



Artemisinin

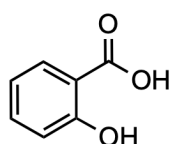


Taxol

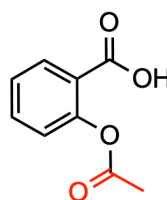


Chloral hydrate

Figure 1.1. Structures of natural products Penicillin, Artemisinin, and Taxol. In addition to the structure of Chloral hydrate, the first full synthetic medicine.



Salicylic acid

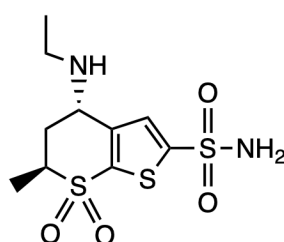


**Aspirin
(Acetylsalicylic acid)**

Figure 1.2. Structures of natural product salicylic acid, and its structural derivative: Aspirin.

The field of drug discovery evolved in 1894 following advancements in the understanding of drug target interactions [9]. Emil Fischer proposed the lock and key model where an enzyme/target is represented as a lock which is only activated by a substrate/key which fits the lock. Emphasising the importance of designing drugs with complementary shapes to the desired target. This became possible following the demystification of protein structures by X-ray crystallography in the 1950's, paving

the way for structure based drug design (SBDD) [10]. SBDD involves the use of computational tools which model the 3D shape of the target, allowing the design of drugs which participate in complementary target interactions. This was accelerated further in 1971 when the protein data bank (PDB) was established [11], providing a universal database of X-ray crystal protein structures. In 1995, the first drug discovered using SBDD was approved for human use: Dorzolamide [12], a carbonic anhydrase II inhibitor used to treat glaucoma (figure 1.3).



Dorzolamide

Figure 1.3. Structure of the first drug developed by structure based drug design (SBDD): Dorzolamide

Despite its success, SBDD was limited by the availability and resolution of 3D protein structures with well-defined binding pockets. Additionally, the 3D structures were rigid, not accounting for the dynamic nature of proteins which can undergo conformational changes when interacting with substrates. This is known as the induced fit model [13, 14], which was proposed by Daniel Koshland in 1958. Furthermore, SBDD was limited by computational power as a result of the high complexity of protein-ligand interactions, consisting of a combination of forces including hydrogen bonding, van der Waals interactions, electrostatic forces, and hydrophobic effects. The scoring functions used within SBDD rely on simplified models and approximations which can lead to inaccurate binding affinities [15].

To overcome these inaccuracies, SBDD modelling techniques are often used in association with complementary approaches, such as fragment-based drug design (FBDD) [16]. FBDD is an experimental technique which uses small libraries (~1000) of compound fragments (< 300 Da) to screen for activity against target proteins. The outcomes of these screens are often hit compounds with low binding affinities, but high

atom-efficiency, providing a good starting point for the design of more potent drugs.

Despite the many strategies for drug development, there is one major limitation: chemical space. The number of drug-like molecules within organic chemical space is estimated to be approximately 10^{60} , and all the molecules synthesised to date have barely scratched the surface [17, 18]. Additionally, this estimation neglects transition metal complexes [19], which have shown promise as drug candidates across a range of therapeutic areas including; cancer [20], autoimmune diseases [21], diabetes, Alzheimer's disease [22], anti-bacterial agents [23], and photodynamic therapies (PDT) [24–26].

The presence of the metal facilitates intrinsic properties which can provide useful drug-like activity. Examples of these properties include catalytic activity, which can disturb biochemical pathways such as the disruption of NADH oxidation to NAD⁺ as an anti-cancer therapy [27]. This has been demonstrated across a range of metal complexes including ruthenium(II), rhodium(III), iridium(III) and osmium(II). Redox potentials, which enable the development of pro-drugs by switching complex geometries to give inactive and active forms of the drug [28, 29]. As well as unique photoactivity [30], such as fluorescence which can be used for fluorescent imaging microscopy, and the generation of reactive oxygen species (ROS) which can damage DNA and biomolecules [26].

In addition to the challenges associated with the exploration of chemical space, the substantial size of biological target space ($\sim 10^6$) [31, 32], makes for an inconceivable amount of potential drug-target interactions. So far, approximately only 3% of biological space has been targeted, and only 7% more is expected to be targetable using small molecules. However, this estimation neglects another set of potential therapeutic targets: nucleic acids.

Nucleic acids, such as DNA, offer therapeutic intervention at a genetic level, enabling the modulation of gene expression and the disruption of biological pathways which are otherwise inaccessible through drug-protein interactions [33]. The effects of drug-DNA interactions include gene regulation, alteration of DNA-protein interactions, and induced cytotoxic effects [34–36]. In addition to the canonical double helix structure of DNA

[37], DNA has been found to fold into many other unique secondary topologies such as hairpins [38–40], g-quadruplexs (G4) [41], and i-motifs [42] (figure 1.4). These alternate DNA folding topologies arise from specific genomic sequences which contain complementary nucleotides which organise into unique secondary structures. Hairpin structures form in regions of DNA which contain symmetry, allowing the topology to switch between inter-strand and intra-strand base pair interactions. G4 and i-motif topologies form in regions of guanine and cytosine rich sequences respectively.

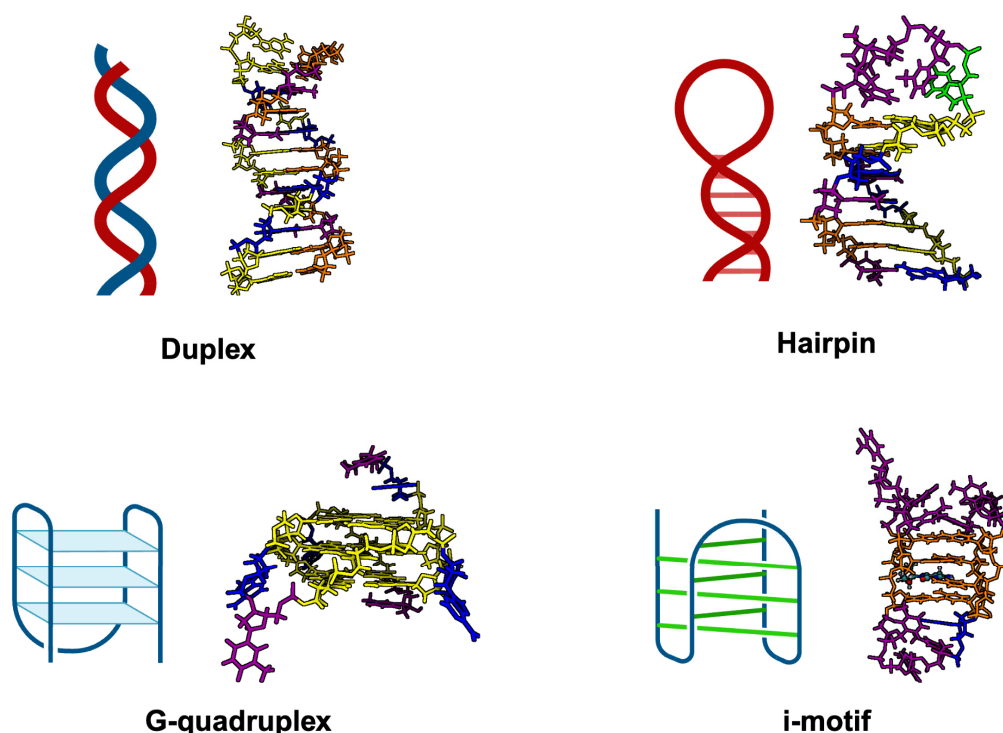


Figure 1.4. Alternate secondary DNA topologies, including schematic and X-ray crystal structures of duplex (2K0V), hairpin (1QE7), G-quadruplex (2LBY) and i-motif (1A83) DNA. Nucleic acid colours: guanine (yellow), cytosine (orange), adenine (blue), thymine (purple).

Unique DNA topologies have been gaining increasing interest as potential anti-cancer therapeutic targets. For example, small molecules have been developed which bind to the hairpin within genes related to replication [40]. Three G4 ligands have entered clinical trials as anti-cancer therapies [43–45], which is discussed further in section 1.3. i-motifs are also beginning to receive increasing interest as anti-cancer therapeutic targets [42].

To explore every possible combination of compound and target would be unachievable

and inefficient due to the high cost of experimental and computational labour, and limitations in access to chemical space. Which explains why drug development is such a costly and time intensive process, with the typical development timeline for a new drug being approximately 14 years, and costing over £2 billion [46]. Part of this problem is attrition: behind every successful drug is approximately tens of thousands of unsuccessful derivatives which did not progress through the drug discovery and development funnel (figure 1.5). Furthermore, as the most expensive part of this process is the clinical development stage, it is critical that the drugs which enter clinical trials are likely to succeed.

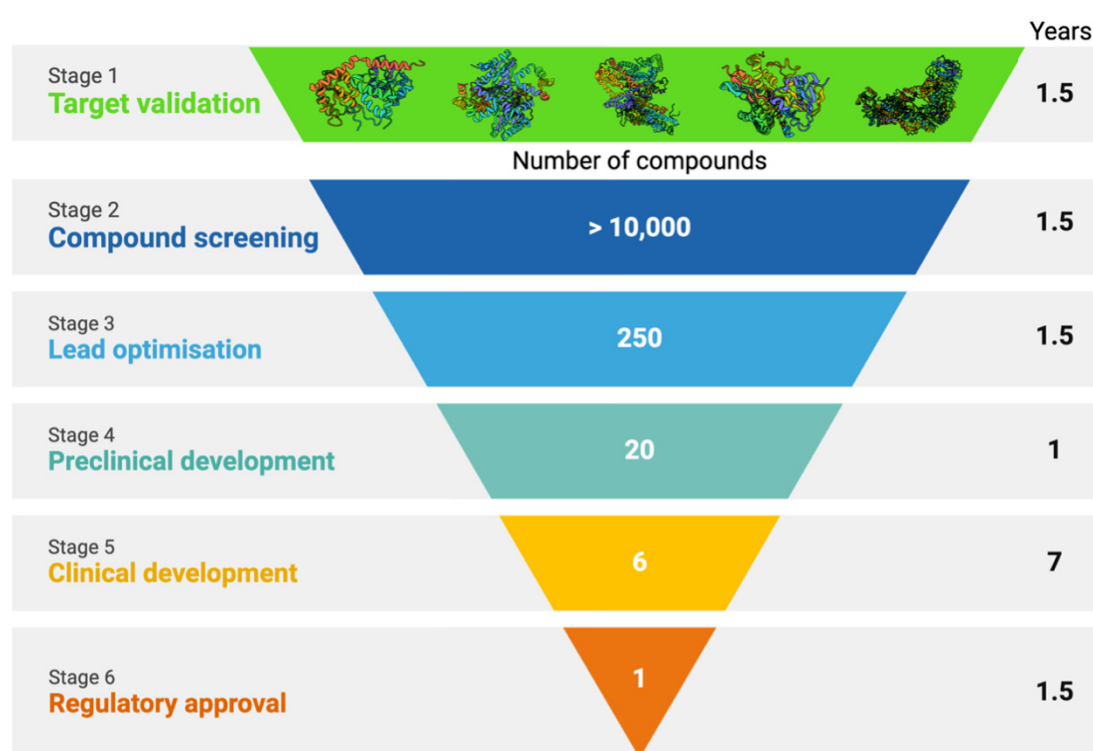


Figure 1.5. Drug discovery timeline showing the average number of years and compounds investigated at each stage of the drug development process (figure inspired by [46]).

This process first involves the selection of a druggable target (i.e protein or nucleic acid structure), followed by the screening of diverse sets of compounds to find a hit: a compound with some binding affinity for the target. This hit compound is then developed further to optimise the target interactions, improving the binding affinity and other drug-like properties to produce a set of lead compounds. The lead compounds are then taken forward into pre-clinical *in vivo* testing to determine which have the best

ADME (absorbance, distribution, metabolism and excretion) profiles. If the candidates show enough promise, they progress into clinical development within humans. If the drug was clinically successful, applications for regulatory approval are then submitted. Many drug candidates are lost to attrition due to their failure within the development stages. One approach to reduce attrition would be designing “better drugs” which are less likely to fail by considering efficacy, toxicity and ADME properties within early development stages. This has led to the raising implementation of automation and artificial intelligence (AI) within drug discovery.

AI refers to computational techniques which enable machines to perform tasks which would typically require human intelligence, such as pattern recognition and rule-based decision making. AI has the ability to detect trends within multidimensional space, making it well suited for drug discovery as it can handle the high dimensionality of molecular properties which afford desired drug activity alongside balanced ADME properties [47–49]. Machine learning (ML) is a subset of AI which involves the implementation of algorithms which learn patterns within data. These algorithms can be supervised or unsupervised: supervised algorithms learn using labelled data, mapping inputs to outputs, using the ground truth to learn patterns between them. Examples of supervised algorithms include random forests (RF), support vector machines (SVM) and neural networks (NN), which are discussed further in chapter 5. In contrast, unsupervised learning only receives unlabelled input data, and searches for patterns within it. Examples of unsupervised algorithms include clustering and principle component analysis (PCA). Within drug discovery, the inputs are molecular descriptors, machine-readable representation of compound structures and properties, and the outputs are the desired target property, such as drug-like activity.

However, for AI to identify meaningful patterns, large high-quality datasets are required. The volume of data required is dependent on the selected algorithm and the complexity of the prediction task [50]. For less complex supervised models, such as RF and SVM, reasonable prediction accuracy can be achieved with datasets containing hundreds to thousands of data points, providing the input features are well correlated to the target property. Whereas, for deep learning algorithms, such as NNs, thousands to millions of data points may be required for the model to perform well. This is due to the complex

structure of NNs which have a large number of parameters which must be tuned during training, and many layers of interconnected nodes which adjust these parameters to minimise the error of the previous layer. If trained on too few data points, models will often overfit to the training data, meaning they will memorise the training examples as opposed to learning meaningful patterns from the data.

One method for generating large training datasets is using automated high throughput (HT) synthesis and screening techniques. Automation involves the use of robotic platforms to synthesise and test thousands of compounds in HT [51, 52]. To generate molecules for HT screening, combinatorial chemistry techniques are implemented in HT to generate libraries of thousands of diverse compounds. Combinatorial chemistry involves performing parallel reactions between diverse sets of molecular building blocks to produce structurally diverse final compounds from a single process. To be applicable for HT synthesis, reactions must be feasible at miniature scales to be performed in 96, 384, or 1536 well plates, in volumes as low as micro and nano litre. Reactions must be simple, reliable, and high yielding to avoid labour intensive purification. Key combinatorial chemistry techniques include click chemistry [53], cross-coupling reactions [54], and solid phase peptide synthesis [55–57]. These techniques share the key characteristics of being regio- and stereo-specific, producing minimal by-products which can be easily purified. Additionally, these reactions tend to be air and water stable, with simple reaction conditions.

Click chemistry is a synthetic approach which utilises simple and selective chemical reactions to join together molecular building blocks, often compared to "snapping together Lego pieces". Click chemistry was recognised for its power in accelerating discovery with the 2022 Nobel Prize in Chemistry [53]. Examples of click chemistry reactions include copper-catalysed azide alkyne cycloadditions to form 1,2,3-triazole rings between azides and terminal alkynes [58], thiol-ene reactions for polymer synthesis and functionalisation [59], and Diels-Alder cycloadditions for ring formation or expansions [60]. Examples of cross-coupling reactions include Suzuki-Miyaura cross-couplings which build carbon-carbon bonds [61], joining diverse functionalised aromatic systems. In addition to Buchwald-Hartwig aminations [62], which can introduce a variety of functionalised amines to aromatic systems. Methods are also

being developed for the HT synthesis of transition metal complexes [63, 64], which are of increasing interest as potential drug candidates. Examples of HT synthesis for metal complexes include combinatorial libraries of cis-platin like platinum complexes for anti-cancer therapeutics [65], iridium complexes as photosensitisers for PDT [64], and ruthenium complexes as anti-cancer and anti-microbial agents [23].

The resultant compounds synthesised within HT synthetic approaches can then be taken forward for HT screening to rapidly assess the activity of these large compound libraries. HT screening assays can be used to measure a range of biological responses including biochemical and cell-based assays [66]. The most common readouts of the biochemical HT screening assays are photophysical endpoint measurements (fluorescence or absorbance). Measurements include, total fluorescent intensity, which can be categorised as: fluorogenic assays, a switch on of fluorescence, or fluorescence quench assays, a switch off of fluorescence. Examples of fluorogenic assays include detecting the presence of fluorescent compounds after incubation with a target, and detection of compounds which become fluorescent upon binding to a target [67–70]. An example of a fluorescent quench assay is fluorescent indicator displacement [71–73], which involves competitive binding between the study drug and a fluorescent substrate, measuring the switch off of fluorescence upon substrate displacement by the study drug.

Other fluorescent HT screening assays include fluorescent resonance energy transfer (FRET) assays [39, 74–77], which utilise fluorescent donors that are quenched by fluorescent acceptors when in close proximity (2 – 6 nm). These assays can probe binding interactions [74], protein-protein interactions [75], and the stability of DNA topologies [38, 39, 77]. Additional non-fluorescence based assays include the use of mass spectrometry to detect bound compounds from a mixed pool of compounds incubated with the target protein [78–80], and nuclear magnetic resonance (NMR) spectroscopy to detect changes to the target protein's NMR signals to assess binding [81].

Cell-based assays provide insights into a drugs behaviour in a multi-component system, enabling assessment of target specificity and action (agonist or inhibitor), cellular permeability, and overall cytotoxicity [82]. Common assays include cell viability assays,

secondary messengers and reporter genes, and microscopy-based assays. Cell viability assays use dyes which change colour, or fluorescence, in the presence of metabolically active cells to determine a drug's toxicity [83–86]. Secondary messenger and reporter gene assays rely on the cell producing a signal from a biochemical pathway which activates an added dye [87, 88], or the expression of a gene which codes for a fluorescent readout respectively [89–91]. HT microscopy is used to screen cells with distinctive phenotypes, such as the presence of parasites and visible damage [92, 93], or in the presence of a fluorescent stain [94]. Common fluorescent microscopy assays include assessing cellular localisation using organelle specific costains, measuring immune response with immunofluorescent stains, and measuring the presence of cytokines using cytokine sensitive dyes [95, 96].

Together, HT synthesis and screening have proven their ability to accelerate the progress of drug discovery by drastically increasing the number of experiments it is possible to conduct within a set time frame. Additionally, the use of robotic platforms reduces the potential for human error, improving the consistency of assay preparation, ensuring reliable and reproducible results. However, HT synthesis and screening techniques also have some disadvantages: firstly, researchers require access to extensive libraries of compounds, and even with HT synthesis, access to compounds remains limited, and within the restricted chemical space [97]. HT techniques additionally have very high start-up costs, with expensive robotic platforms, and the high cost of running experiments in HT [66, 98].

The key benefit of these HT techniques is their ability to generate large experimental datasets, which, when used in combination with AI, can drastically accelerate discovery and reduce overall costs. An example of how AI and ML can be used to minimise the number of compounds which are required within a HT screen is "AI-driven" iterative HT screening: HT screens are performed in batches and the results are analysed by ML active/inactive classification models between iterations to select the next batch of compounds to screen, also known as reinforcement learning [99]. G. Dreiman, et al., demonstrated the power of this approach using a large set of PubChem HT screen datasets which explored activity across 12 unique biological targets. Using a random, but diverse selection of compounds from each dataset (10%) as a starting point, a

selection of ML algorithms were trained to predict the hit probability for the remaining compounds. Following each prediction round, algorithms selected an additional 5% of the dataset predicted to be active, and retrained the model. Figure 1.6 shows the percentage of active compounds recovered from each dataset with each iteration by each ML algorithm compared to random selection. This study demonstrated that using ML, 78% of the active compounds could be identified from screening only 35% of the entire dataset, hugely reducing the screening burden.

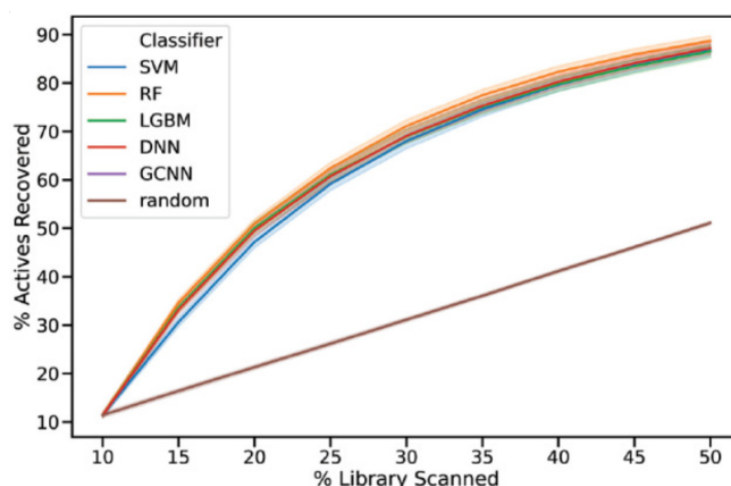


Figure 1.6. Mean recovery of active compounds against the percentage of the PubChem library using classification ML methods: support vector machines (SVM), random forests (RF), light gradient boosting machine (LGBM), deep neural network (DNN), and graph convolutional neural network (GCNN). Iterations started with 10% of the dataset, followed by steps of 5%. Shading shows a 68% confidence interval. Figure taken from “Changing the HTS Paradigm: AI-Driven Iterative Screening for Hit Finding” [99].

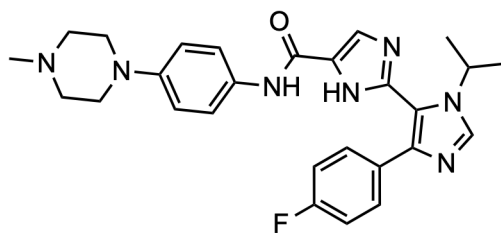
AI tools have also revolutionised computer aided drug design (CADD), especially following the development of AlphaFold, a protein structure prediction AI, which has increased the structural coverage of the human proteome from 46% to 76% [100]. AlphaFold also has the ability to model protein-ligand interactions, providing an 50% improvement in accuracy of binding interactions when compared to traditional physics-based SBDD modelling methods, in just a fraction of the time [101].

In addition to AlphaFold, in which docking is a secondary function, there are many dedicated AI docking models which leverage extensive protein-ligand interaction datasets to predict ligand pose conformation and binding energy. These models are

typically trained on descriptors of protein-ligand binding structures and associated binding data to learn the statistical patterns which lead to binding interactions, opposed to any mechanistic understanding of the system [102]. An example of these AI docking models is the Deep Docking platform which uses previously established docking data as a training set to predict virtual hits for untested compounds. These predicted hits are then iteratively docked in a reinforcement learning process, resulting in a 100-fold acceleration of virtual library screening [103].

Whilst these AI models have advanced the field of CADD, the results still fall short of true experimental measurements [104]. This is due to challenges associated with performance fluctuations, the model's ability to predict protein flexibility and induced-fit effects, in addition to challenges with hallucinations where the AI predicts binding poses which are not sterically allowed, or clash with the binding site [105]. Furthermore, these models provide no insights into the *in vivo* behaviour of drug candidates, therefore are best used as tools to guide the synthesis of hit compounds for more vigorous experimental testing, as opposed to a replacement to all experimental research.

AI can also be used to design novel drugs through a similar approach to SBDD. This is known as AI *de novo* drug design, and leverages the 3D structure of the target and known substrates, using ML to generate and optimise novel structures which fit the binding pockets [106]. The most successful AI *de novo* drug to date is INS018-055 (figure 1.7), a first in class anti-fibrotic inhibitor, now in phase II clinical trials for the treatment of idiopathic pulmonary fibrosis [107]. Both the biological target, and chemical structure of the drug were identified using AI. This included the use of PandaOmics, an AI and bioinformatics toolkit [108], to explore gene pathways and score their disease association based on experimental and clinical data. Upon selection of the new target, Chemistry42, an AI toolkit for molecular design and optimisation [109], was then used to generate novel molecules and assess their ability to interact with the new target. The most impressive part of the discovery of this drug was the very short timeline leading to the pre-clinical development. In as few as 18 months after the project began, the pre-clinical candidate was selected, and nine months later, the drug entered phase I clinical trials, a process which would typically take 5 to 7 years [110].



INS018-055

Figure 1.7. Structure of the first AI *de novo* designed drug: INS018-055.

In addition to the examples above, AI has been applied to a plethora of other drug discovery tasks, including target identification and validation [111], and the prediction of drug-like properties such as efficacy [112, 113], ADME [114], toxicity [115], drug-drug interactions [116], and drug repurposing [117, 118]. It is therefore clear that AI is an extremely powerful tool which can be applied to a range of diverse drug discovery tasks. Furthermore, these techniques are still in their infancy and will continue to grow in capacity, speed, and accuracy, in time helping society to discover more life changing medicines [97].

1.2 Anti-cancer therapeutics

Each year, there are approximately 20 million new cancer cases and 10 million cancer related deaths worldwide. Despite cancer research being one of the most heavily funded research areas [119], there are currently only 321 regulatory approved cancer drugs [120]. This highlights a huge unmet need for new cancer medicines. Cancer can affect a wide array of cell types within the body ranging from epithelial tissue, bone, muscle, blood, and specific organ cells. Due to this, and each cancer possessing unique genetic mutations, there is no single unified therapeutic approach [121]. Conventional cancer treatments involve surgical removal of the cancer cells, followed by radiotherapy or chemotherapy [122]. These second lines of defence use ionising radiation X-rays, or drugs which target rapidly dividing cells respectively. There are additionally more advanced cancer treatments such as immunotherapy [123], antibody-drug conjugates [124], hormone therapy [125], gene therapy [126], and peptide therapy [127], however, this thesis will focus on small molecule drug interventions.

The cytotoxicity of many chemotherapy agents occurs by inducing DNA damage by directly forming covalent bonds with DNA and proteins, or by disrupting DNA synthesis and repair pathways [122]. In some cases, such as with oxaliplatin, this DNA damage can trigger downstream immunogenic responses that contribute to apoptosis [128]. The class of cancer drugs which damage DNA directly are known as alkylating agents, and include both organic molecules (e.g. chlorambucil [129], doxorubicin [130]), and platinum based drugs (e.g. cisplatin [131], oxaliplatin [132], carboplatin [133]) (figure 1.8).

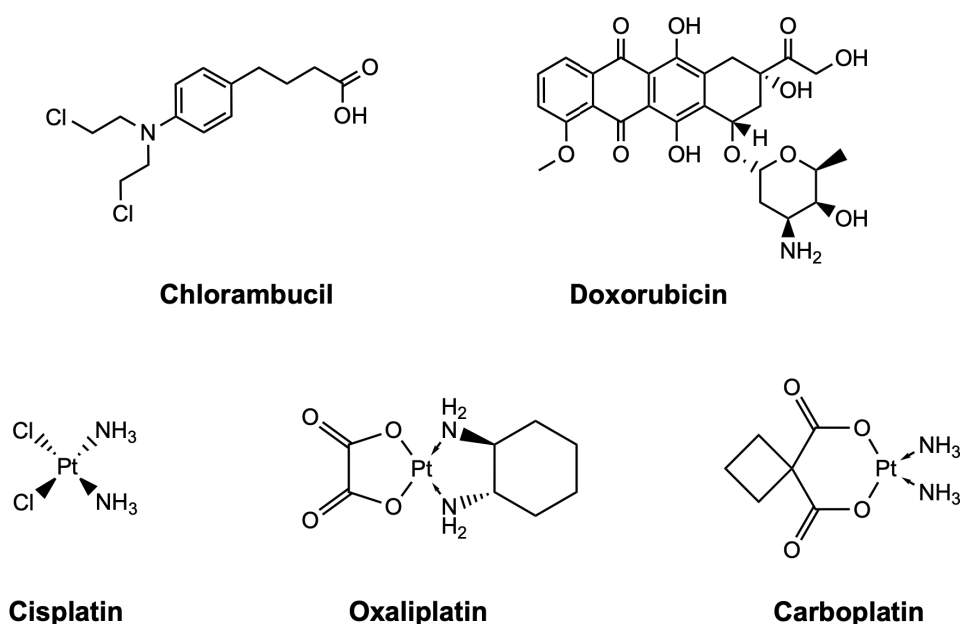


Figure 1.8. Molecular structures of a selection of alkylating agents used in chemotherapy

However, these therapies are known to result in many unpleasant side effects as a result of their off-target activity affecting healthy proliferating cells. This has driven the development of therapies which target specific proteins involved in cancer cell growth, proliferation, and survival [134, 135]. These include targeting the endocrine, hormone, or metabolic channels important for cancer cells, such as oestrogen receptor antagonists, androgen receptor antagonists, and PI3K inhibitors.

Whilst the majority of these treatments are mediated by organic small molecules, inorganic complexes have emerged as a promising class of anti-cancer therapeutics, following on from the early success of cisplatin and its derivatives (figure 1.8). As of 2022 there were 12 clinically approved metal containing compounds used within

cancer therapy, and a further 26 within clinical trials [136]. 9 out of 12 of the clinically approved complexes are derivatives of cisplatin, therefore all exhibit their cytotoxicity via the same DNA damage mechanism. The three other clinically approved metal complexes are macrocycles which are photosensitisers: molecules which absorb light and transfer the absorbed energy to oxygen, generating ROS which react with nearby biomolecules inducing cell death [137–139]. Figure 1.9 shows two of the clinically approved metal-based photosensitisers, and an additional metal-based photosensitiser currently within clinical trials; padeliporfin, Photosens, and TLD-1433 [140].

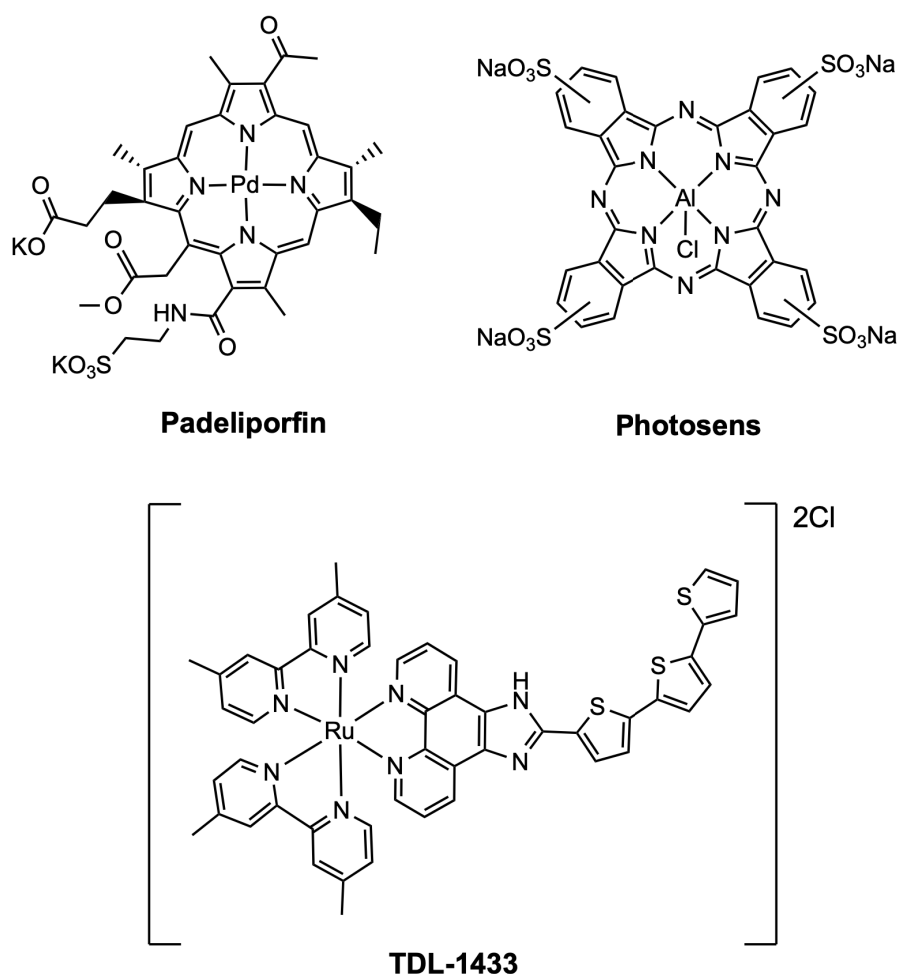


Figure 1.9. Molecular structures of a selection of metal-based photosensitisers used in cancer therapies: padeliporfin, Photosens. In addition to one complex within clinical trials: TLD-1433.

Photosensitisers are used within PDT: a two-step cancer treatment which involves the administration of a photosensitiser which distributes non-specifically throughout the body. Then upon activation from a source of light energy, the photosensitiser undergoes photophysical processes to generate ROS which damage cells within the

activated area, leading to localised cell death [141] (figure 1.10). The photophysics of this process are described in more detail within chapter 3.

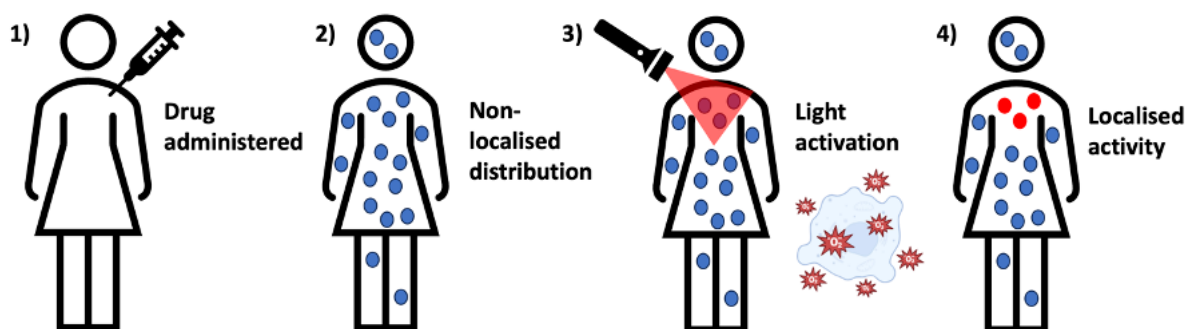


Figure 1.10. Schematic showing the steps in PDT. 1) Photosensitiser drug is administered. 2) Drug is distributed around the body. 3) The area of intended activity is irradiated using a light source. 4) Activity of the photosensitisers is localised to the point of irradiation.

The benefits of PDT is that therapeutic activity is specific to the area being irradiated, localising the cytotoxic effects of the photosensitiser, thereby reducing off target side effects [141]. Although PDT is already used in the clinic and promises further developments, there are still some limitations: PDT can only be applied to cancer which can be irradiated externally or via endoscopic procedures (e.g. skin, oesophagus, lung, bladder). Additionally, the therapy is dependent on tissue oxidation and light accessibility to generate the ROS, making PDT unsuitable for deep tumours [142, 143].

Despite all these potential treatment options, many cancers still have poor overall survival rates, and most therapies are accompanied with many harmful side effects. On top of this, there is a major issue of cancers developing drug resistance to these small molecule therapies over time. The mechanisms of resistance can include mutations which enable DNA repair, drug efflux, modifications to the drug-target active site. As well as direct deactivation of the drug through an enzyme catalysed chemical modification, such as phosphorylation. One approach to counter the emergence of drug resistance is through the use of combination therapy which target multiple treatment pathways, with the hypothesis that if a population of cancer cells mutate to form resistance to intervention of one pathway, they would remain susceptible to in the other pathway [77, 134]. Unfortunately, even with this approach resistance can still occur as a

result of the complex epigenetic nature of cancer cells which allows their rapid evolution of resistance mechanisms [144]. Therefore, it is of interest to explore therapies which work via unique modes of action and the discovery of new therapeutic targets. Despite this need, between 2009 and 2020 only 16% of new cancer drugs were based on new mechanisms of action [145].

AI has been implemented to address this challenge by identifying new cancer targets and determining their druggability with the aim of finding new mechanisms of action [146–151]. Examples of AI include graph convolution network based AI model which uses neural net (NN) algorithms to learn patterns between molecular features and drug target interactions which are represented as graphs, predicting new drug target interactions based on the learnt relationships [146]. AI has also been implemented to virtually screen previously established cancer drugs to new targets for drug repurposing [117, 118, 152–154]. Drug repurposing is a very desirable approach for drug development as the drugs have previously passed the clinical development and regulatory approval stages, the most time consuming part of the development of a new drug [155–161]. Finally, AI, namely Chemistry42, was also utilised for the *de novo* design of a new cancer drug. As of November 2024, the first small molecule AI *de novo* designed cancer drug, ISM5939, was approved for clinical trials [162] (figure 1.11).

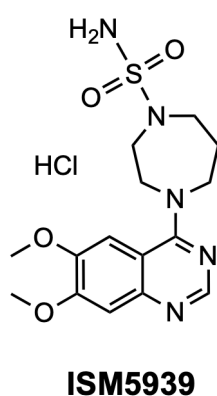


Figure 1.11. First AI *de novo* designed cancer drug: ISM5939

Most small molecule cancer therapeutics have been designed to target proteins. However, DNA targeting therapies are gaining increasing interest as potential cancer therapies with novel modes of action [163, 164]. G4 DNA has been investigated as a potential target for a wide variety of anti-cancer therapeutics [165–171], and will be the

DNA target of focus for this thesis.

1.3 G-quadruplex DNA as an anti-cancer therapeutic target

G4s assemble through Hoogsteen hydrogen bonding between four co-planar guanine residues to form G-quartets. The G-quartets stack upon one another to form the G4 structure [172]. G4s are able to fold and unfold dynamically, and are typically stabilised by the presence cations such as K^+ or Na^+ [173]. G4s are polymorphic, presenting as a range of topologies which depend on the sequences of the oligonucleotide flanking bases, and the nature of cations which template their formation [174]. Three key G4 topologies include parallel, anti-parallel, and hybrid [175, 176]. Parallel G4s consist of oligonucleotide strands which all run in the same direction, with propeller-type loops positioned at the side of the G4. Anti-parallel G4s contain sets of two oligonucleotide strands which run in opposite directions, and hybrid G4s contains mixtures of strands running in different directions (figure 1.12).

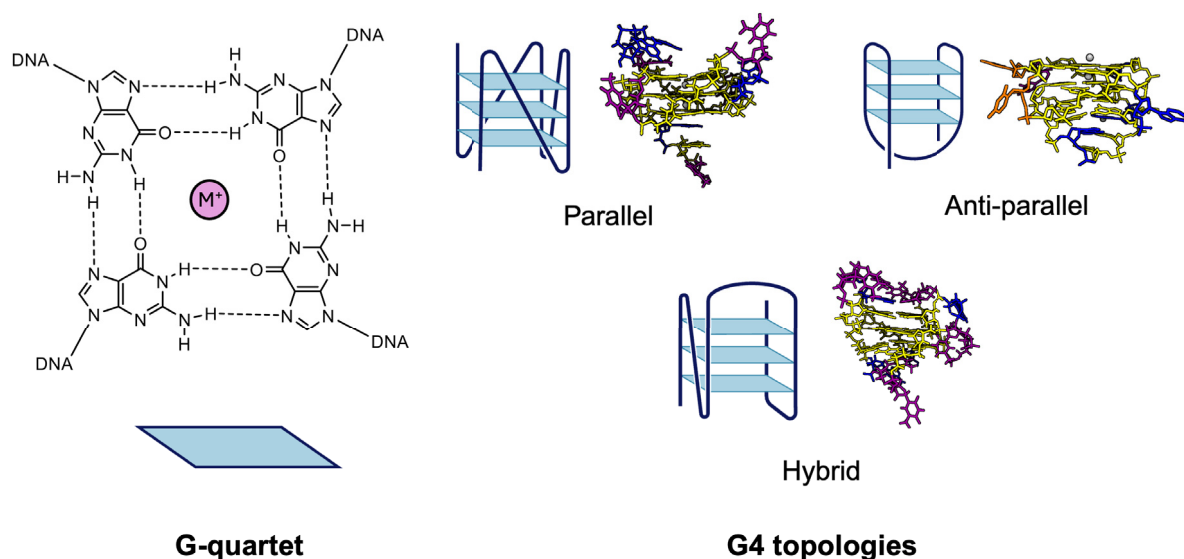


Figure 1.12. Illustration of G4s built from stacks of G-quartets which form by 4 co-planar guanine residues hydrogen bonding, and examples of G4 topologies: parallel (c-myc: 217V), anti-parallel(c-kit1: 4WO2) and hybrid (HTG4: 2GKU). Nucleic acid colours: guanine (yellow), cytosine (orange), adenine (blue), thymine (purple).

The first evidence of the biological relevance of G4s was obtained in 2002: an X-ray crystal structure of a G4 within human telomeric DNA [177]. However, it wasn't until

2013, following the development of a G4 selective antibody, BG4, that it was possible to map the presence of G4s within human cells [178]. In 2015 HT sequencing of G4s (G4-seq) was performed on the human genome identifying over 700,000 regions which had the potential to form G4 structures [179]. This technique identified more than 250,000 additional G4 sequences than had been previously predicted to form by computational methods. This was likely due to G4s forming in regions of the genome which did not fit the typical G4 motif that the computational models searched for: $G_{3+N_{1-7}}$ where G is a guanine base which form the G-quartets, and N is any base within the connecting loops [180]. Examples of these computational programmes include Quadparser and QGRS Mapper [41, 181]. More sophisticated programs were later developed which expanded the motif scope, such as G4Hunter [182].

Many of the guanine rich regions of DNA which form G4s are associated with the regulation of oncogene transcription and translation, and in chromosomal telomere sequences [171, 183–186]. As such, their presence has been linked to telomere maintenance and the regulation of transcription. Additionally, the presence of G4 structures were found to be upregulated within cancer tissues when compared to healthy tissue [187], leading to the hypothesis that cancer cells may be more susceptible to G4 targeting therapies than healthy cells, which could reduce off target effects.

G4s in telomeres are of interest as therapeutic targets due to the key role telomeres play within cell replication and uncontrolled proliferation of cancer cells [188, 189]. Telomeres shorten by 100-200 base pairs each time a cell completes its replication cycle, until they become critically short. At this point the cell will undergo apoptosis, or enter a state of senescence where it is no longer active. The extension of telomeres is mediated by a telomerase enzyme accompanied by reverse transcription proteins, which are overexpressed in 85% of cancer cells. As a result, the cell can continue to divide without telomeres becoming critically short, resulting in uncontrolled cell proliferation [189, 190].

The role of G4s in telomeres is complicated. Whilst many studies have found that stabilisation of G4s inhibits telomerase activity, more recent studies have found that G4s can also have positive effects on telomere maintenance [35, 191]. Telomerase has

been identified as being able to extend parallel intermolecular G4s. This G4 unfolding could play a role in the binding of the enzyme to the telomeric DNA [192]. Therefore it has been postulated that the dynamic nature of G4s play a role in regulating telomerase activity, and that G4 ligand binding hyper-stabilises these G4 structures, preventing the essential unfolding required for telomerase activity. G4s have also been found to be extended by telomere capping proteins TRF2 and TET1, which can be displaced by G4 binding ligands, providing an alternate pathway for telomere dysregulation. The telomeric G4 sequences of interest within this thesis include HTG4 and HTelo which dynamically coexist sharing the same telomeric G4 repeat sequence (TTAGGG)_n, differing by the included flanking bases and loop arrangements [193–195].

G4s within promoter regions of oncogenes are also potential targets for anti-cancer therapeutics, with evidence showing that the stabilisation of promoter G4s in *c-myc*, *VEGF*, *BCL2*, *KRAS*, and *c-kit* oncogenes resulted in a reduction in gene expression [196–200]. The oncogene promoters of interest for this thesis include *c-myc* and *c-kit*. *c-myc* is a transcription factor in cell growth, division, and apoptosis. Across a range of cancerous tumour cells, the *c-myc* oncogene is over-expressed. Stabilisation of the *c-myc* G4 leads to a down-regulation of the gene and as a result a reduced production of hTERT, a catalytic subunit of telomerase and other genes relating to the proliferation of cells [201]. These observations have been attributed with decreased growth rates of tumorous cells. *c-kit* expression has been shown to play a key role in signal transduction cascades controlling cell growth and differentiation of a variety of cancers [202–205]. Studies have shown that upon stabilisation of the *c-kit1* G4, there was a downregulation in expression [206, 207].

1.3.1 G4 ligands as potential anti-cancer therapeutics

Compounds can bind to G4s in a range of different modes: the key binding interaction involves π - π stacking between the aromatic surfaces of the terminal G-quartets and the aromatic surface of the ligand. Other binding modes include groove binding through electrostatic interactions with the charged phosphate backbone, and intercalation where the ligand inserts between G-quartets [208, 209] (figure 1.13).

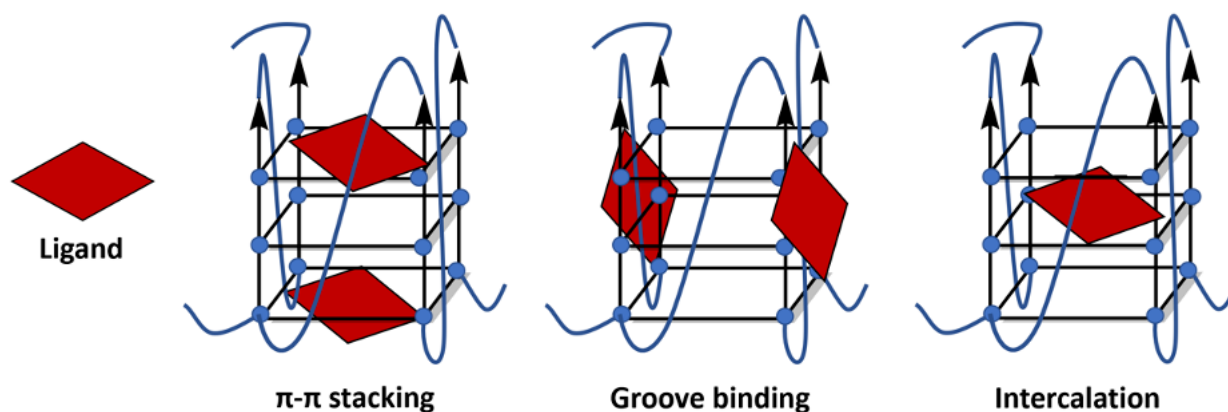


Figure 1.13. Illustration of the different possible binding modes between ligands and G4s.

It has been observed that the presence of a positive charge improves binding through electrostatic interactions with the positively occupied ion channel, and the negatively charged phosphate backbone of the oligonucleotide. Based on these binding motifs ligands tend to be planar polyaromatic compounds with charged arms capable of π - π stacking interactions with G-quartets. The unique environment of each G4 forming sequence results in distinct G-quadruplex topologies which provides a medium for selective targeting of genes within the genome. Research into G4 stabilisation as an anti-cancer therapeutic approach first began in 1997. However, since then only three G4 binding ligands have made it into clinical trials: CX-5461, CX-3543, and QN-302 [44, 45, 210] (figure 1.14).

CX-5461 is currently within phase I and phase II trials for haematologic cancers, and has been shown to be well tolerated whilst demonstrating antitumor activity [44]. CX-5461 was a first-in-class selective inhibitor of ribosomal DNA transcription. CX-3543, also known as Quarfloxin, reached phase II trials for colon cancers and neuroendocrine carcinomas [210]. CX-3543 was initially thought to selectively target *c-myc* G4 for its therapeutic activity. However, CX-3543 was also found to interact with ribosomal RNA which led to off target effects [211]. Additionally, CX-3543 was found to have poor bioavailability, limiting its clinical effectiveness, and as a result was withdrawn from clinical development [77]. QN-302 is the newest of the G4 binding ligands to enter clinical trials. It is currently in phase I trials for its use in pancreatic cancer [45]. QN-302 is considered a pan-quadruplex targeting agent as it binds to G4 sequences within multiple gene promoter regions and multiple gene pathways. This wide G4 targeting

scope results in QN-302 down-regulating the expression of multiple cancer-related genes.

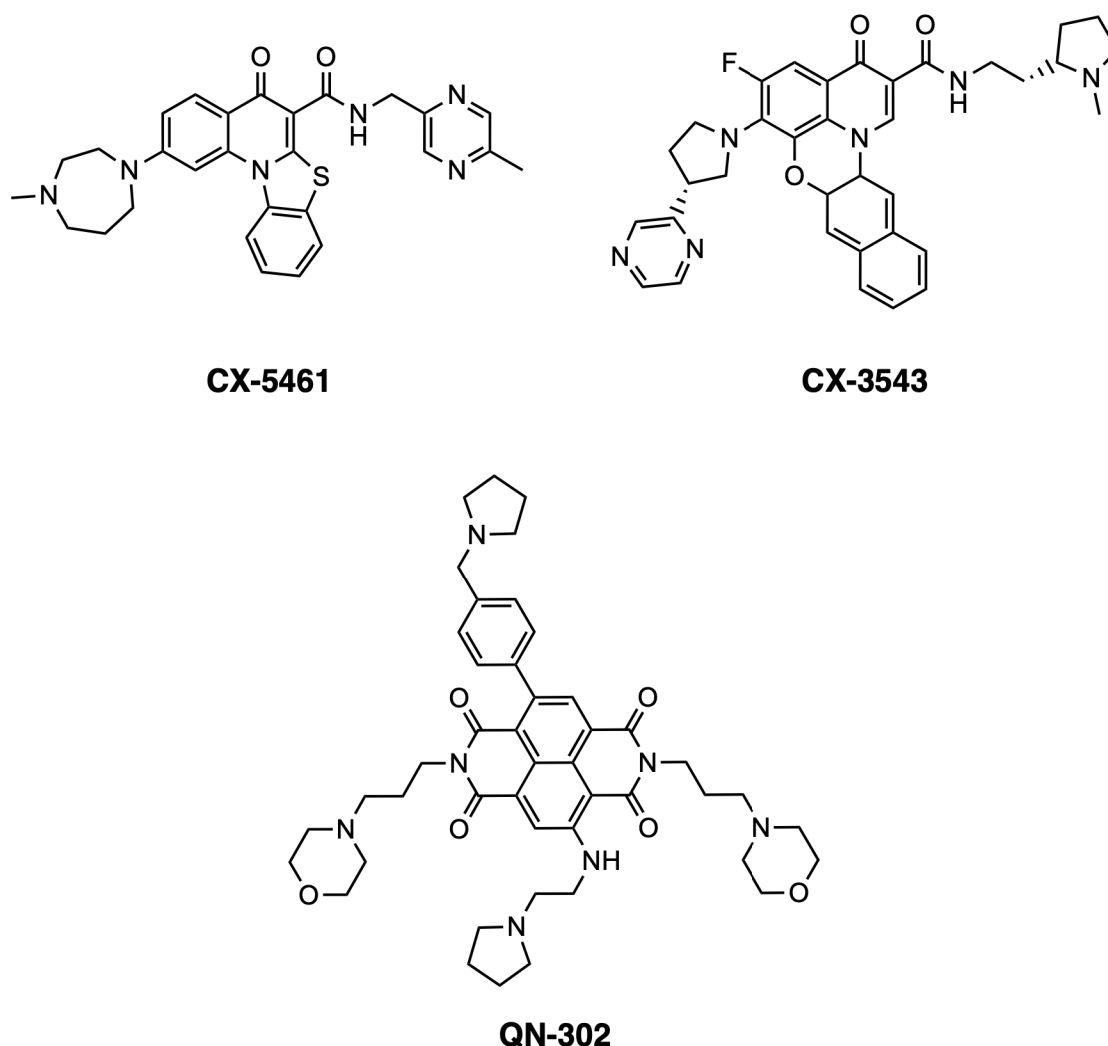


Figure 1.14. Structures of the three G4 ligands which have entered clinical trials: CX-5451, CX-3543, and QN-302. [44, 45, 210]

With the promising preliminary clinical performances of G4 ligands, their research popularity has increased every year, with over 4500 G4 ligands identified to date, many of which are implicated as potential anti-cancer therapeutics [212]. A recent perspective on the use of G4 ligands as cancer therapeutics presented an informative timeline showing the trend of increasing G4 ligand publications from 1996 to April 2024 [213] (figure 1.15).

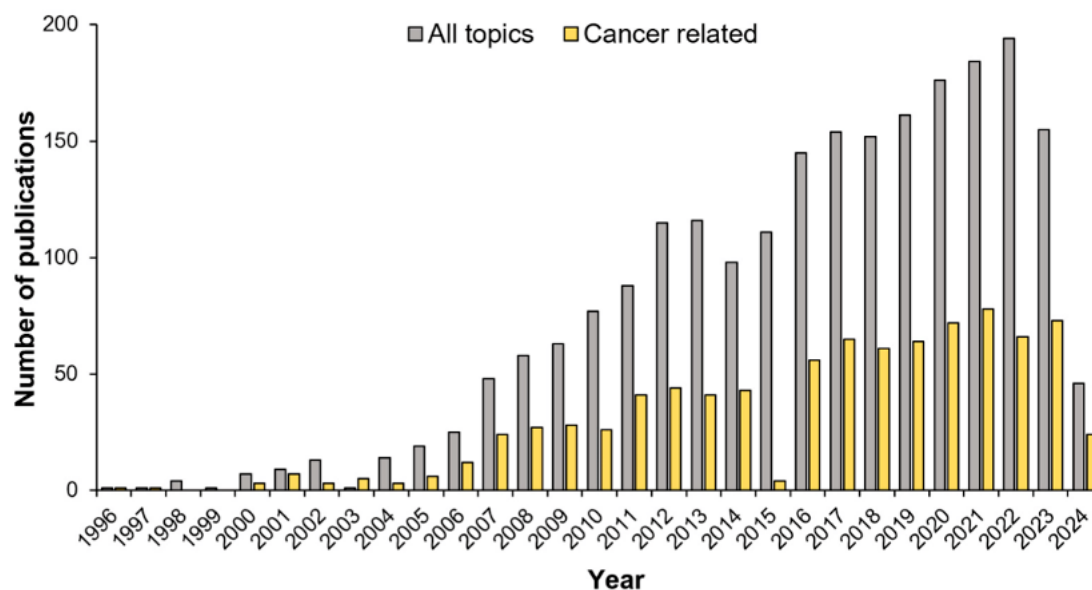


Figure 1.15. Timeline showing the number of G4 ligand publications per year, showing all publications in grey, and cancer specific publications in yellow. Figure taken directly from “Therapeutic Use of G4-Ligands in Cancer: State-of-the-Art and Future Perspectives” [213]. Data was collected from the PubMed database until April 2024.

Additional organic G4 binding ligands which have been identified as having activity relevant to cancer therapies include BRACO-19, TMPyP4, Pyridostatin (PDS), RHPS4, PhenDC3 and cAQ-mBen [214–219] (figure 1.16). These compounds have been reported to inhibit telomerase activity (BRACO-19, RHPS4, PhenDC3, and cAQ-mBen), downregulate *c-myc* protein levels (TMPyP4), induce cell cycle arrest (PDS), and re-sensitise drug resistant cancer cells (PDS and PhenDC3) [220].

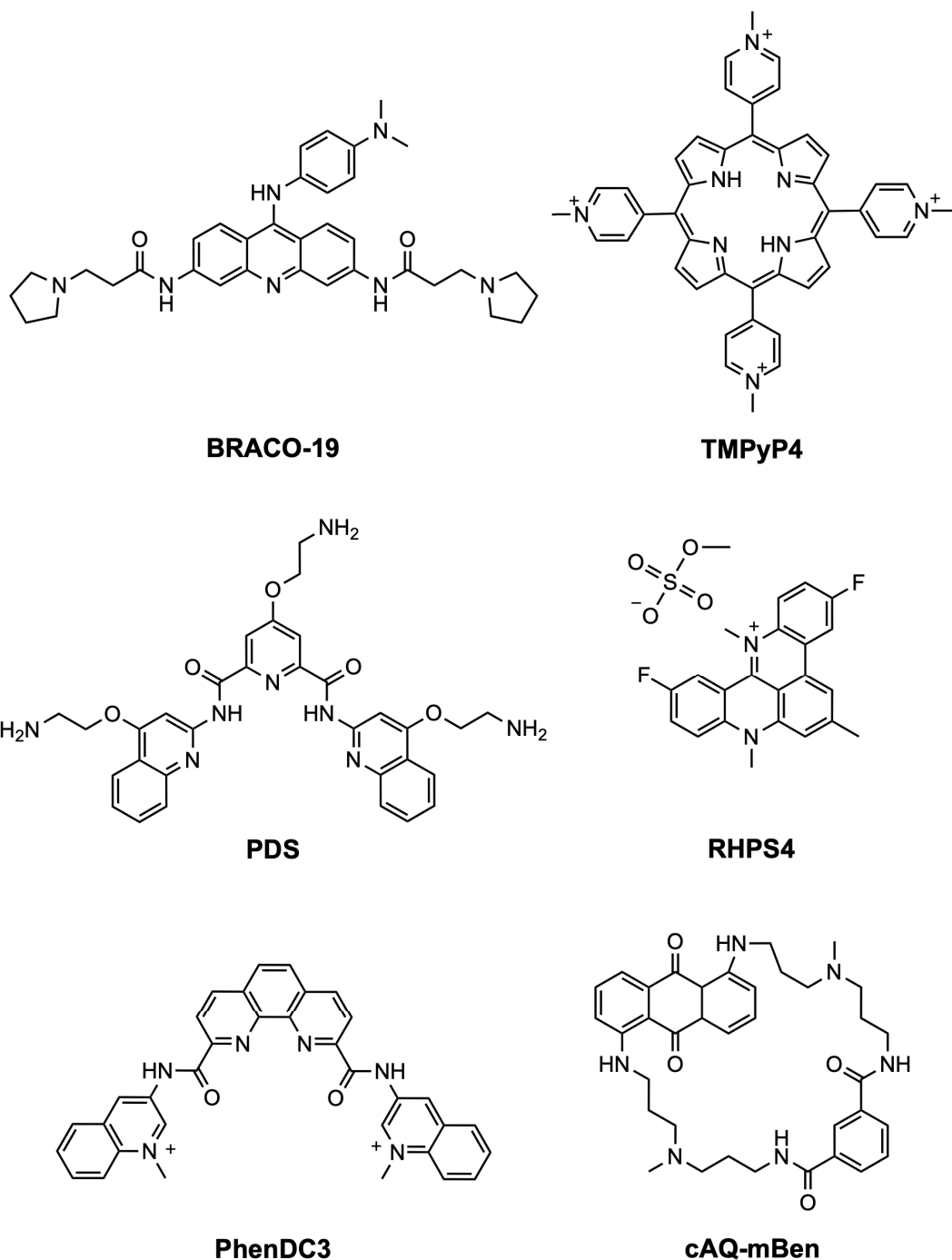


Figure 1.16. Structures of a selection of G4 binding ligands identified as having anti-cancer activity: BRACO-19, TMPyP4, PDS, RHPS4, PhenDC3 and cAQ-mBen.

Metal complexes are postulated to yield good G4 binding ligands due to the large potential for ligand modification whilst maintaining the required binding geometry, which is enforced by the metal centre [221–223]. Diverse geometries grant access to a greater chemical space, enabling exploration in three dimensions. Drugs which contain more 3D complexity, such as in the form of increased sp^3 hybridisation, tend to result in more

successful drug candidates [224]. The metal centre also has an electron withdrawing effect which causes the complexes aromatic system to be electron poor, which is more favourable for π - π stacking interactions with the electron rich guanines within the G-quartet. Additionally, the electropositive metal can position itself in the centre of the G-quartet, increasing the electrostatic stabilisation of the G4 otherwise provided by the charged ion channel. The earliest examples metal complex G4 ligands were the metalloporphyrins, established in 1999 with copper(II) and indium(III) derivatives of TMPyP4, which bound to telomeric G4 DNA [225]. Since then, broader studies have explored a range of additional metalloporphyrin derivatives (figure 1.17).

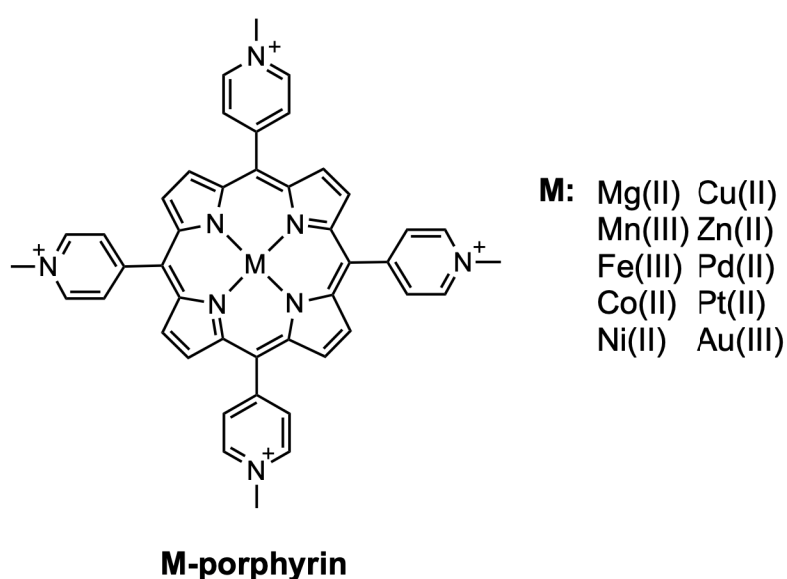


Figure 1.17. Structure of G4 binding TMPyP4 metal derivatives.

However, it was found that presence of the metal centre in the metalloporphyrin complexes did not significantly increase telomerase inhibition when compared to the metal free TMPyP4 compound [226]. The best performing metal complexes were row 1 metals which resulted in square planar geometries (zinc(II) or copper(II)), as they did not disrupt the planar binding surface of the TMPyP4 ligand. Square planar metal complexes with metals in rows 2 and 3 (palladium(II) and platinum(II)) appeared to reduce the activity, potentially due to their increased atomic radii and the limited space within the porphyrin ring, which may have led to slight distortion of the ligand planarity. In the cases where the metal centre introduced axial ligands in an octahedral geometry (iron(II), manganese(III) and magnesium(II)), activity was significantly reduced due to

the aromatic surfaces of the complex being blocked by the axial ligands, preventing π - π end stacking with the G4.

The next class of metal complexes investigated as G4 binders were metal salphen and metal salen complexes. These complexes were first explored as duplex DNA binders in 1997 [227]. However, in 2006, nickel(II) salphen complexes were found to selectively bind to G4s over duplex DNA [228]. The metal salen complexes, which have an ethyl group between the nitrogen atoms instead of an aromatic ring, generally were unable to bind to G4s selectively over duplex DNA [229]. Metal salphen complexes have a higher binding affinity and selectivity to G4s due to their extended aromatic system, and increased planarity compared to the salen complexes [222]. The metal salphen complexes were the first examples of G4 binders where the ligands without the metal centre were unable to bind to G4s. Since then, salphen complexes have been reported with a many metal centres which exhibit a range of geometries. These include nickel(II), copper(II) and platinum(II) with square planar geometries, zinc(II) with a distorted square planar geometry, vanadium(II) oxide with a square based pyramidal geometry, and platinum(IV) and cobalt(III) with octahedral geometries (figure 1.18).

The square planar metal salphen complexes demonstrate the best G4 binding of all metal salphen geometries [230]. When comparing HTelo G4 binding between complexes with the same ligand system and alternate square planar metal centres, the binding affinity was determined as nickel(II) > platinum(II) > copper(II). Conversely when binding to *c-kit2* G4 DNA the binding affinity was platinum(II) > nickel(II) > copper(II), showing that different G4 topologies have a preference for different metal centres. Nickel(II) salphen complexes were also found to inhibit telomerase activity [231]. The distorted square planar zinc(II) salphen complexes displayed poor binding affinity across both G4 topologies.

Unlike other metals, the platinum(II) metal centre not only provides the optimal square planar geometry, but also a range of useful intrinsic properties. As a result of the square planar geometry, heavy metal effect, and d^8 electron configuration, platinum(II) salphen complexes are phosphorescent following the absorbance of energy from a photon [232]. This emissive property allows the platinum(II) salphen complexes to be imaged in cells using fluorescent imaging microscopy [233, 234]. Additionally, from the energy state

which gives rise to the phosphorescence, the complexes can relax via the generation of ROS, opening the potential for the development of platinum(II) salphen G4 binding photosensitisers [235], a growing research area in the development of cancer therapy [236–242].

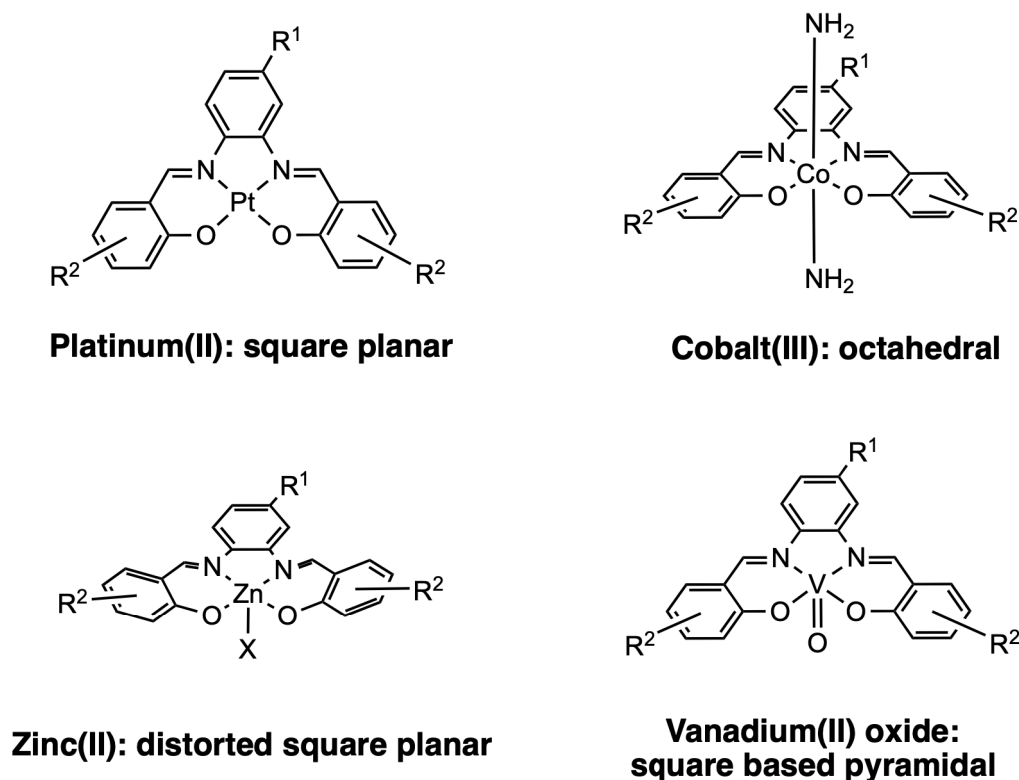
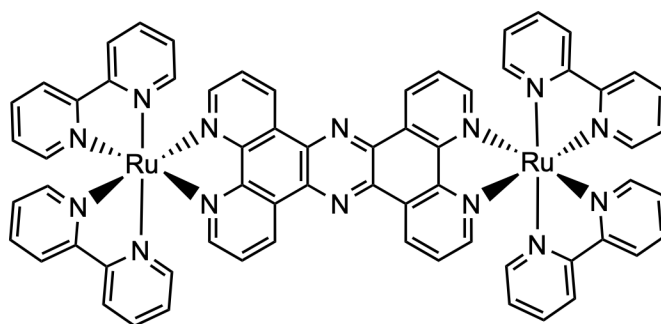


Figure 1.18. Examples of the geometries obtained with platinum(II), cobalt(III), zinc(II), and vanadium(II) oxide centres with a salphen ligand. X = coordinated solvent.

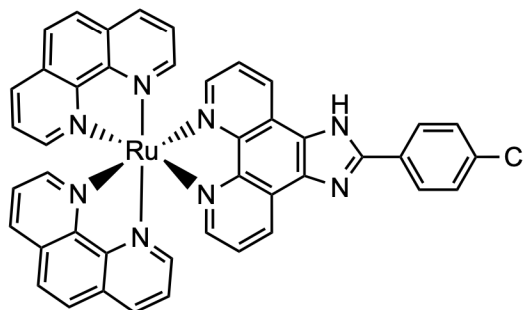
It would be expected for the octahedral complexes to also have very poor binding affinity due to the previously established requirement of an accessible planar aromatic surface for G4 binding. This was indeed the case for the platinum(IV) salphen derivatives with axial chloride ligands, which showed no binding to HTelo or *c-myc* G4 DNA [29]. However, this complex could then be reduced back to the active platinum(II) salphen complex upon addition of bio-reductants, making platinum(IV) salphen complexes possible candidates for G4 binding pro-drugs. Conversely the cobalt(III) salphen complexes with NH_3 axial ligands were able to bind to G4, with binding observed with HTelo, *c-myc* and *c-kit2* sequences, with significant selectivity over duplex DNA [243].

Ruthenium(II) polypyridyl complexes are another class of octahedral G4 binding metal

complexes. Instead of π - π stacking with G-quartets, the ruthenium(II) complexes bind to G4s through interactions with the G4 loops and grooves. The first example of these complexes was a di-ruthenium complex which was reported to have an increase in emission intensity upon binding to G4 and duplex DNA [244]. More recent studies have been investigating mono-ruthenium complexes which explore a range of polypyridyl ligands including phenanthroline and 1,4,5,8-tetraazaphenanthrene (TAP) ligands with extended aromatic systems [245] (figure 1.19).



Di-ruthenium polypyridyl complex



Ruthenium polypyridyl complex

Figure 1.19. Structures exemplary di-ruthenium and mono-ruthenium G4 binding complexes

The enantiomeric feature of the octahedral ruthenium(II) complexes yields one isomer which has preferential binding properties. For the di-ruthenium polypyridyl complex shown in figure 1.19, the $\Lambda\Lambda$ isomer displayed a binding affinity for HTelo G4 DNA 40 times greater than the $\Delta\Delta$ isomer. This was due to the $\Lambda\Lambda$ isomer having the correct geometry to bind to both G-quartet faces of the G4, whereas the $\Delta\Delta$ isomer could only access one G-quartet face [246]. The presence of the TAP ligand within the

ruthenium(II) complexes also resulted in G4 binding selectivity over duplex DNA. The inclusion of the TAP ligand resulted in complexes only showing toxicity upon irradiation as a result of photo-induced electron transfer (PET) quenching oxidising the G4 [245]. Since then, there have been many further modifications to the ruthenium(II) polypyridyl complexes yielding complexes which can induce senescence, reduce *c-myc* gene expression, and inhibit telomerase activity [247–249].

Metallophthalocyanines complexes have displayed better G4 binding abilities than the metalloporphyrin complexes [250, 251]. This is a result of the extended aromatic system and increased number of nitrogen atoms which makes the ligand more complementary to the spatial and electronic requirements for end stacking with G-quartets [221] (figure 1.20).

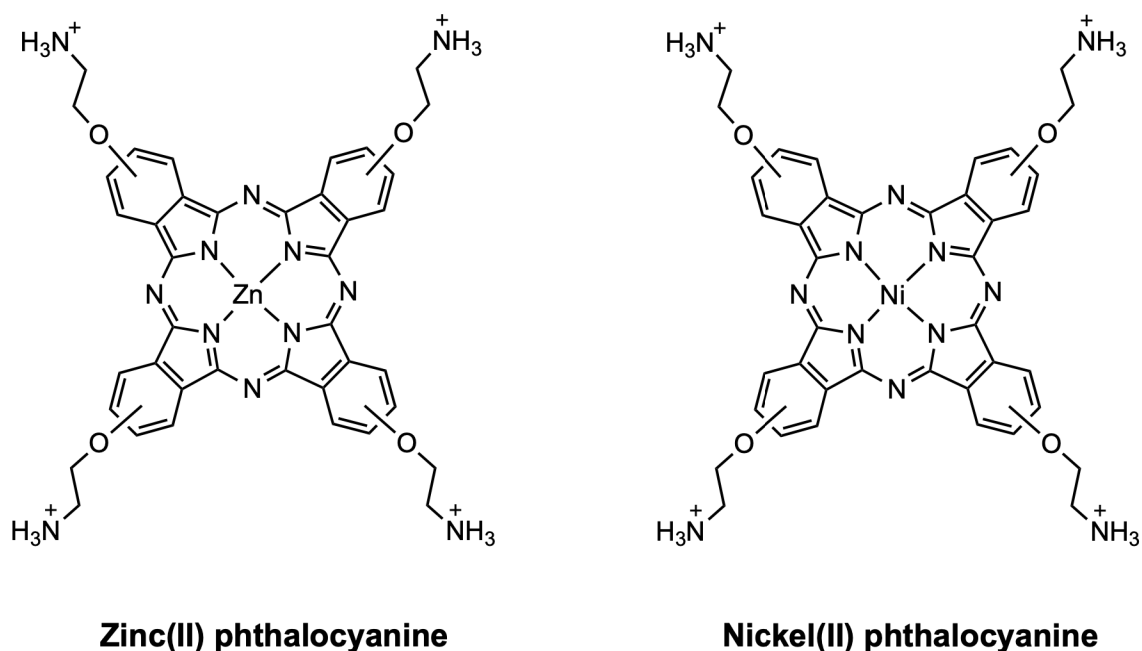


Figure 1.20. Structures of exemplary metallophthalocyanines G4 binding complexes

Polydentate polypyridyl ligands also show good G4 binding properties. Metal-terpyridine complexes are able to bind selectively to a variety of G4 topologies over duplex DNA, performing best with square planar platinum(II) or copper(II) centres [252]. Some derivatives of platinum(II) terpyridine complexes were found to outperform cisplatin within cytotoxicity assays, and were able to overcome cisplatin resistance within ovarian cancer cell lines [253]. Platinum(II) NNC polypyridyl complexes have

also demonstrate high affinity and selectivity for G4s (figure 1.21).

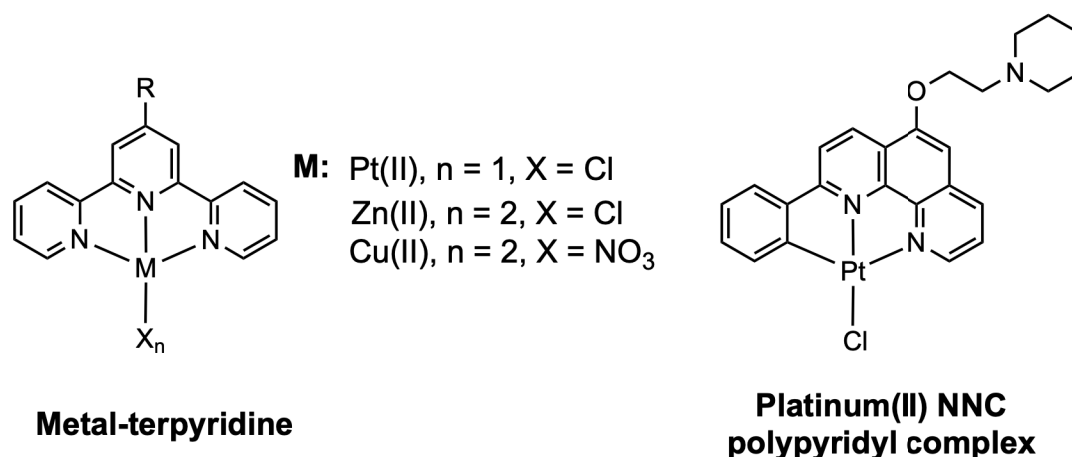


Figure 1.21. Structures of exemplary metal-terpyridine and NNC polypyridyl G4 binding complexes

Metal-phenanthrolines have been shown to inhibit telomerase via G4 binding [231, 254]. With enhanced activity when containing two sets of the bidentate phenanthroline ligands [154]. Derivatives of platinum phenanthrolines with extended aromatic systems such as phenanthroimidazole and dipyrrophenazine have also been explored and resulted in a significant improvement to G4 binding abilities. This observation further highlights the importance of aromatic surfaces and the planar geometries (enforced by the platinum(II) centre) being essential for G4 binding [255, 256] (figure 1.22).

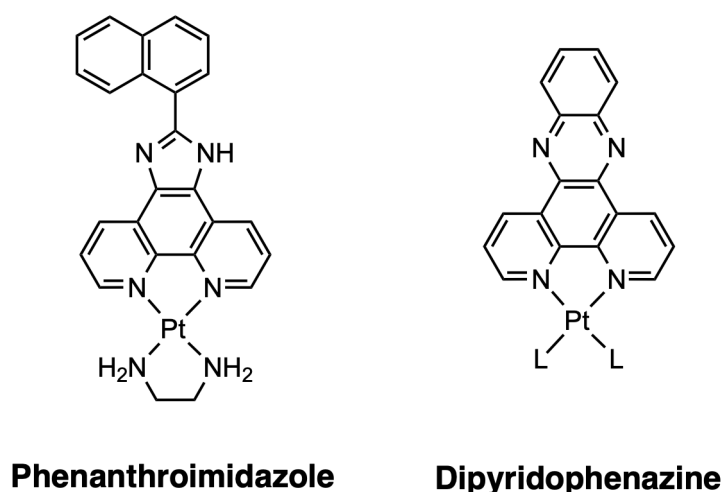


Figure 1.22. Extended aromatic derivatives on the metal-phenanthroline ligand scaffold with identified anti-cancer activity.

Within this thesis, platinum(II) salphen complexes were the main G4 ligand of interest due to their exploitable photoactivity, in terms of phosphorescence and ROS generation, in addition to their previously established strong G4 binding ability in the context of anti-cancer therapies [257, 258].

1.3.2 Machine learning in the context of G4s

To date, all applications of ML in the context of G4s have been in the prediction of G4 formation within the genome [259–264]. A key example includes DeepG4, a neural net trained on cellularly visualised G4 regions, which also accounts for G4 chromatin accessibility, resulting in a prediction accuracy of > 98% [264].

The first instance of a G4 ligand prediction ML model was released as a pre-print in November 2024 by L. D’anna et al., [265]. This work showcased “the first AI in-silico workflow for G4 binding affinity prediction”: AGAPE (computational G4 Affinity Prediction). AGAPE is a classification model which predicts if a ligand is active or inactive as a G4 binder, with an accuracy of 90%. The model makes its decision based on relationships between a range of quantum chemical descriptors (molecular descriptors are described in more detail in chapter 5) and compounds which are known to be active or inactive G4 binders, obtained from the G4 ligand database G4LDB. Despite the training of this model being biased towards organic compounds (training set contained 1073 organic compounds and 144 metal complexes), the model was still able to correctly classify previously unseen metal complexes from an unseen validation set with an accuracy of 75%. This publication demonstrates that with as few as 1217 data points, a reasonable classification prediction accuracy can be achieved, showing promise for the development of AI tools within domains which are data poor. Classification models like AGAPE can be used to prioritise the synthesis of compounds which are most likely to be active for a desired property, thereby reducing downstream attrition during drug development.

1.4 Project overview and objectives

This thesis presents a case study of the application of automation and ML in the development of platinum(II) salphen DNA binding photosensitisers as potential

anti-cancer therapeutics. This was performed in three stages: synthesis, screening and ML data analysis. Within the synthetic stage of this project, the goal was to develop a large library of platinum(II) salphen complexes using HT synthesis. These complexes were to contain a range of electron donating and withdrawing substituents, in addition to substituents which provided a solubilising effect to the complex. To achieve the overall goal of synthesising 96 complexes, the synthetic procedure was first optimised in batch, ensuring that the selected building blocks were compatible with the platinum(II) salphen synthetic route.

Following the synthesis of a selection of complexes, the synthetic route was then miniaturised to fit the 96 well plate format of the HT synthesis protocol. A Opentron2 (OT2) liquid handling robot protocol was developed to set up the reactions, and prepare NMR samples to assess conversion to the final platinum(II) salphen complex product. Assuming reactions had proceeded to 100% conversion as desired, the reaction mixtures would then have been screened directly for their DNA binding and ROS generating abilities within the HT screening stage of the project.

To undertake the HT screening stage of the project, first, assays to measure the biological properties of interest were developed which could be prepared using the OT2 liquid handling robot in a multi-well format. These included a DNA binding assay which takes advantage of the phosphorescent properties of the platinum(II) salphens: emission titrations to determine binding affinity. In addition to an assay to measure the ROS generation of the complexes.

The second part of the screening stage involved hand prepared cellular assays (due to the OT2 not being in a sterile environment appropriate for cell work), which assessed the cytotoxicity and photo-induced cytotoxicity (phototoxicity) of the platinum(II) salphen complexes. These assays were performed in a multi-well format using multi-channel pipettes to mimic HT screening methods. Finally, the cellular localisation of the complexes was screening by fluorescent imaging microscopy using a HT microscope: Nikon Ti2 Eclipse which was automated to image 96 well plates. The results of the HT screening assays determined which complexes were taking forward for DNA damage experiments and confocal microscopy with co-stains to determine organelle specific localisation.

The final stage of this project involved the use of ML to perform complex data analysis to better understand structure activity relationships (SAR) between the platinum(II) salphen complexes and their DNA binding photosensitising activity. However, as it was not possible to achieve the HT synthesis stage of this project, computational modelling techniques were implemented to supplement the experimental datasets. This included density functional theory (DFT) and time dependent density functional theory (TDDFT) to calculate energy-based properties of the complexes, and docking simulations to determine complex binding affinities to DNA.

Project objectives:

- Synthesise a diverse library of platinum(II) salphen complexes using HT synthesis.
- Develop HT screening protocols to measure the DNA binding affinity, and ROS generating ability of the platinum(II) salphen complexes.
- Assess the relationship between DNA binding and ROS generation in the context of DNA damage.
- Perform in-depth cellular studies on a selection of complexes which performed best in the HT screens.
- Identify a lead complex for further *in vivo* testing.
- Use computational modelling, including DFT, TDDFT and molecular docking, to generate larger datasets of the platinum(II) salphen complex properties.
- Use the computational datasets to train ML models to predict desirable complex properties.

Chapter 2

Synthesis of platinum(II) salphen complexes

2.1 Platinum(II) salphen complex library design

Metal salphen complexes are of increasing interest in medicinal applications due to their interesting photophysical properties and ability to target DNA, as mentioned in chapter 1. In addition to these desirable properties, they are also synthetically versatile, containing three key regions which can be systematically modified to introduce a range of diverse electronic and steric properties [242, 257]. Points of modification include the metal centre (**M**), diamine backbone (**DAB**) and lower carboxaldehyde (**CBA**) rings, which originate from the respective diamine and carboxaldehyde starting materials.

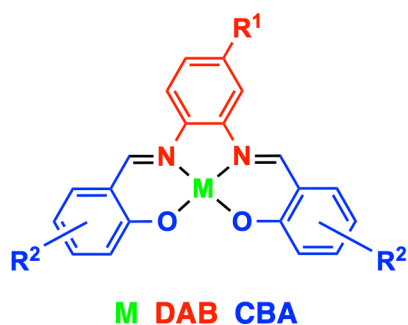


Figure 2.1. Breakdown of the metal salphen complex scaffold components including the metal centre (**M**) (green), the diamine backbone (**DAB**) (red), and the lower carboxaldehyde rings (**CBA**) (blue).

This work focuses on the photoactive square planar platinum(II) salphen complexes due to their emission and ROS generating abilities, in addition to their selectivity for G4s over duplex DNA, making them strong candidates for PDT agents. By introducing

different functionalities within the **CBA** rings, the electronic and structural properties of the complexes can be modified to tune their photophysical and DNA binding properties. Within this work, a range of commercially available electron donating and electron withdrawing functionalised **CBAs** were selected to build a library of platinum(II) salphen complexes (figure 2.2).

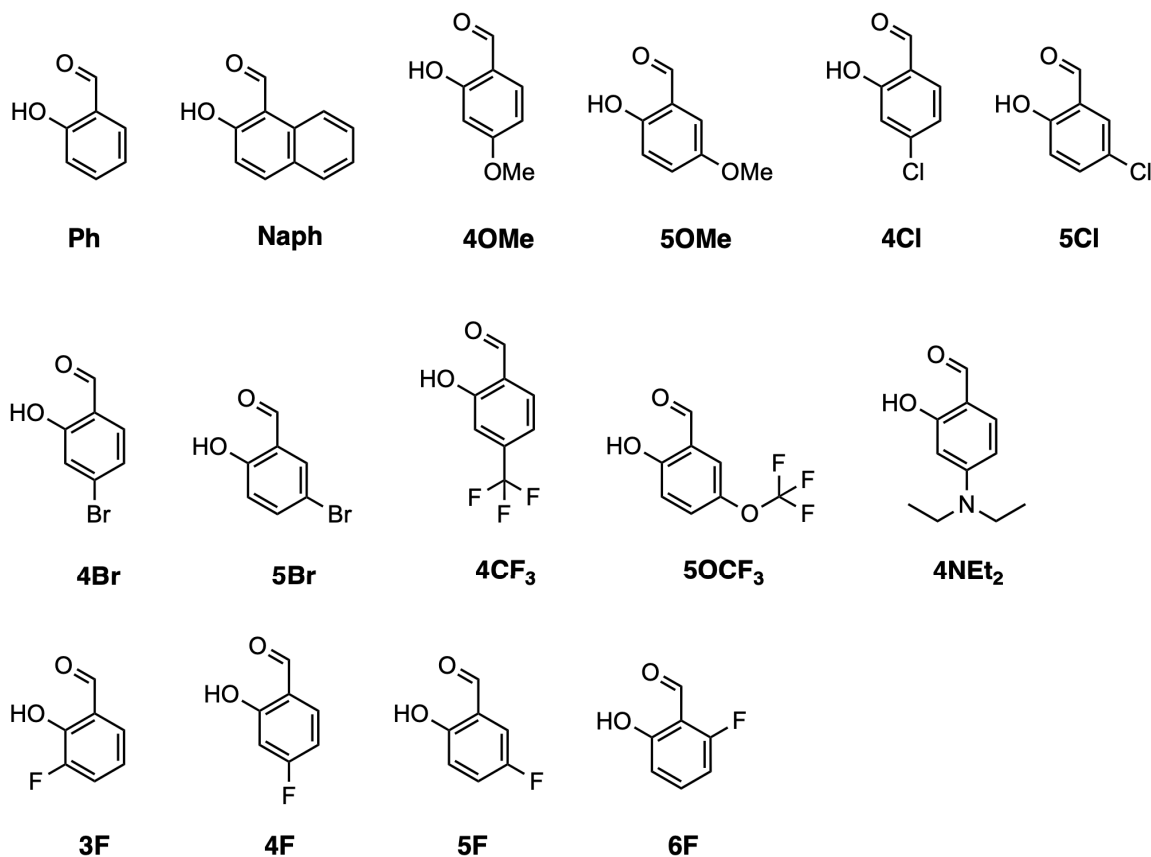


Figure 2.2. Structures of the **CBA** building blocks used within the library synthesis. Labelled by the name of the component within the final complex.

One key challenge of these polyaromatic planar platinum(II) complexes is their poor water solubility. This is due to the large planar aromatic surfaces which can engage in π - π stacking interactions, and the d^8 platinum(II) centre which can participate in metallophilic platinum-platinum interactions, which together promote aggregation [266]. This is further promoted by the rigid, planar structure, and lack of polar groups within the unsubstituted complex. To overcome this issue, a solubilisation group was installed into the **DAB** of the platinum(II) complexes. Based on previous work where the addition of protonatable amines resulted in an increase in the solubility of platinum(II) complexes, the substituents in figure 2.3 were investigated. The amine functionalities

were attached to **DAB** by either an ether or an amide linker, which are electron donating and withdrawing groups respectively. Therefore the **DAB** also introduces an additional point to modify the electronic properties of the core complex.

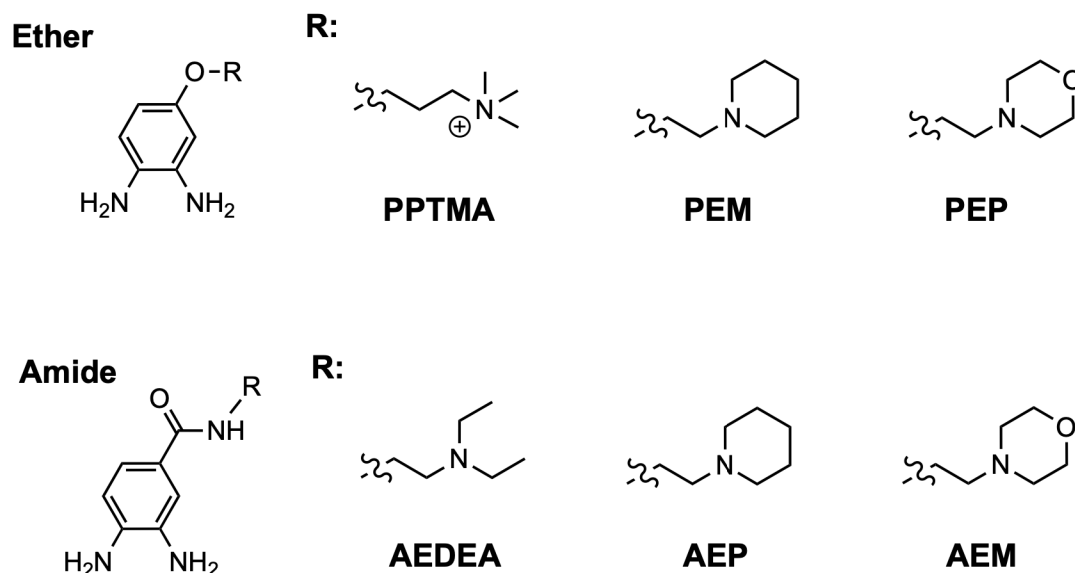
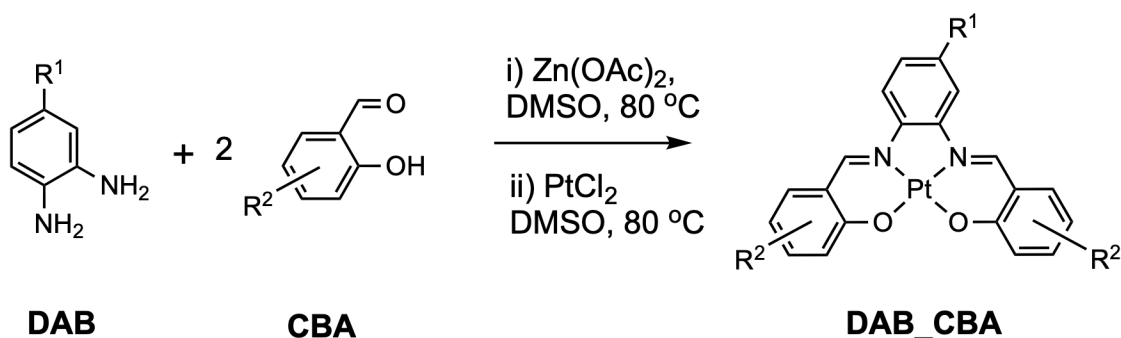


Figure 2.3. Structures of the **DAB** building blocks with a range of solubilising amine groups.

Complexes were then generated via a one-pot two-step synthesis, starting with a zinc(II) templating reaction followed by a platinum(II) transmetalation to give the final complex (scheme 1).



Scheme 2.1. General synthetic route for the formation of platinum salphen complexes. Where R^1 is the functionalised diamine backbones (**DAB**), and R^2 is the functionalised carboxaldehydes (**CBA**). Reaction times: i) 24 h, ii) 72 h.

The zinc templating route was selected over direct formation of the platinum(II) salphen to avoid the generation of a mixture of products, including the mono-substituted complex

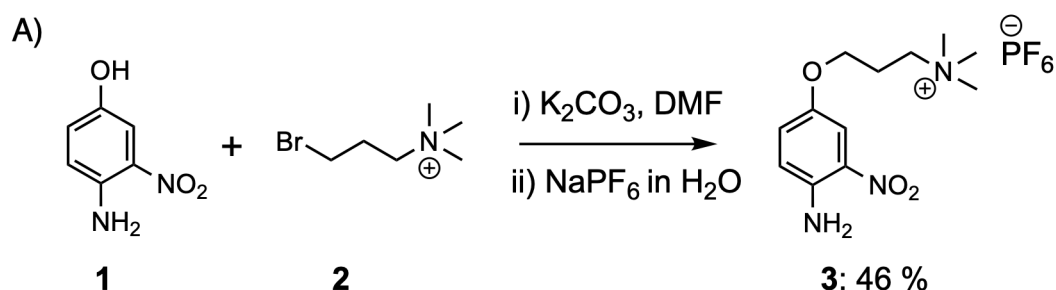
[267]. This route was also selected over the formation of the salphen ligand and subsequent insertion of the metal centre to increase the yield, and reduce the synthetic effort required to access these platinum(II) salphen complexes [233].

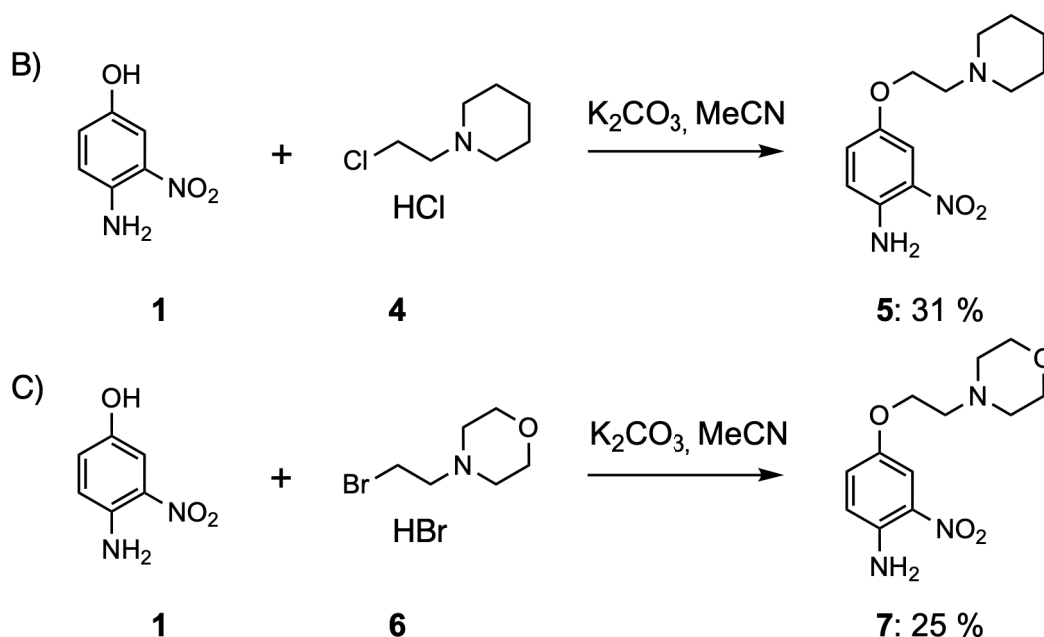
The modular and one-pot synthetic route was also a stronger candidate for HT synthesis, enabling the rapid production of diverse compound libraries. Paired with HT screening, the HT pipeline could be used to generate large volumes of photophysical and biological activity data for ML applications. However, for the reaction to be suitable for HT synthesis, it is desirable for 100 % conversion to be achieved to avoid lengthy purification procedures, which are complex or not feasible when using HT techniques [268]. In addition, all synthetic building blocks needed to be confirmed to be compatible with the experimental conditions required for the quantitative synthesis of the final complexes. Therefore, reactions were first investigated using batch conditions.

2.2 Platinum(II) salphen complex synthesis

2.2.1 Ether linked DAB platinum(II) salphen complexes

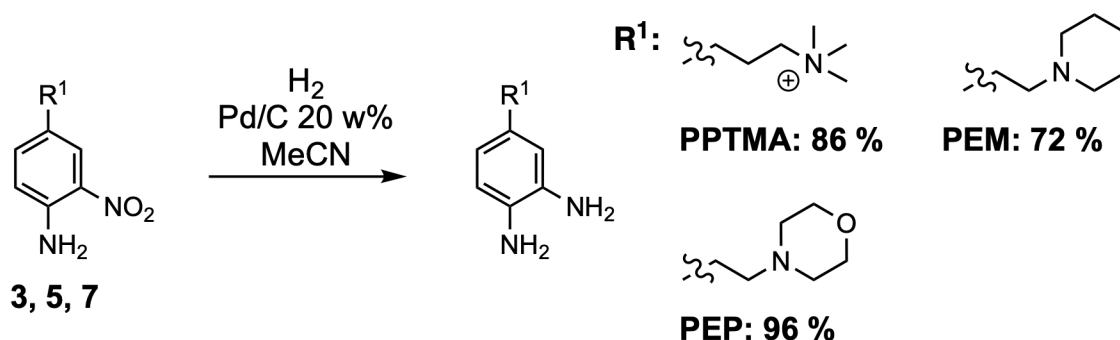
The ether linked solubilised backbones were synthesised via an S_N2 reaction between the nitro-amino precursor (**1**) of the **DAB** and a halide functionalised solubilised alkyl chain. The nitro-amino precursor (**1**) was selected due to the di-nitro precursor having a reduced reactivity from the nitro groups, and the diamino phenol starting material being known to oxidise in the presence of air [233, 269]. Scheme 2.2 shows the synthesis of each **DAB** precursor.





Scheme 2.2. Synthetic routes for the nitro-amino precursor of the **PPTMA** (A: **3**), **PEP** (B: **5**), and the **PEM** (C: **7**) diamine backbones. All reactions were performed at 60 °C for 18 h.

Following the syntheses of the **DAB** precursors **3**, **5** and **7**, the amino-nitrophenol compounds were converted to the diamino **DAB** building blocks via a hydrogenation reaction with hydrogen gas (H_2) and a palladium on carbon (Pd/C) catalyst at room temperature (scheme 2.3).



Scheme 2.3. Synthetic route for the hydrogenation of nitro-amine precursors to the diamine of the solubilising backbone for salphen formation. The reaction was ran for 18 h at room temperature.

The hydrogenation reactions generally went to 100 % conversion after 24 hours, with limited material lost during the workup of the reactions. Successful conversion was indicated by ^1H NMR spectroscopy which showed an up-field shift in the aromatic

peaks from around 7.3 ppm to 6.2 ppm. This was due to the loss of the highly electron withdrawing nitro substituent which had a de-shielding effect on the aromatic protons. This shift was also coupled with a change in the order which the aromatic peaks appeared. These changes can be seen in figure 2.4, which shows the ^1H NMR spectra for **3**, and **PPTMA** following hydrogenation.

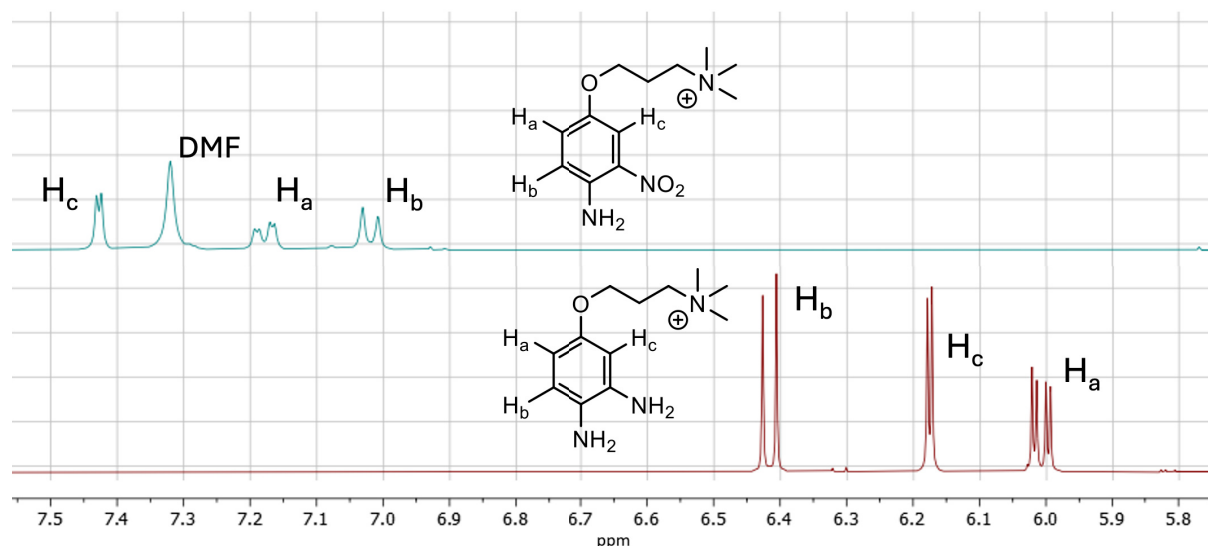


Figure 2.4. ^1H NMR spectra for **3** (top) and **PPTMA** (bottom) showing the shift of aromatic environments after reduction of the nitro group.

It was found that when storing the functionalised **DAB** compounds **PPTMA**, **PEP** and **PEM**, they would degrade over time. This was attributed to the air instability of the diamino phenol starting materials, suggesting that functionalisation of the alcohol did not prevent degradation of the diamine. Figure 2.5 shows the ^1H NMR spectra of the ether **DABs** after 1 h in a DMSO-d_6 solution (A) compared to the ^1H NMR spectra of the freshly isolated **DAB** (B).

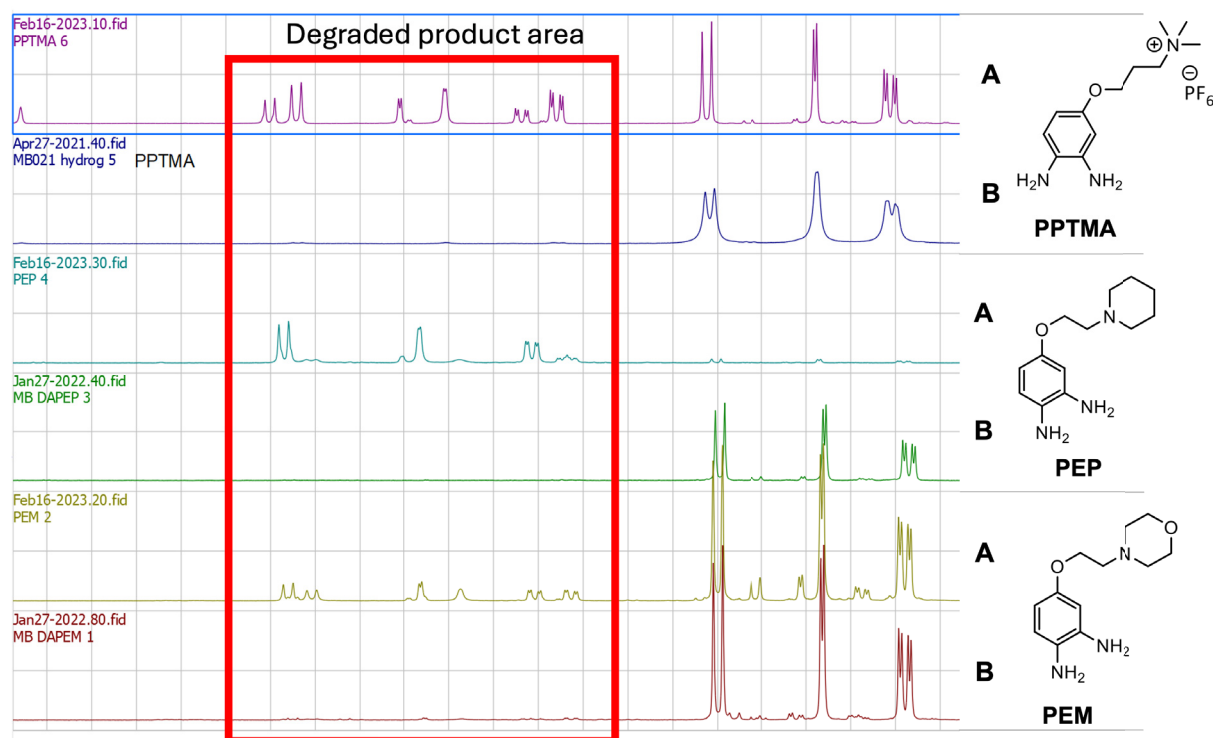


Figure 2.5. ^1H NMR spectra of the **DAB** DMSO-d_6 solutions freshly after preparation (B), and 1 hour later (A) for **PPTMA** (top), **PEP** (middle), and **PEM** (bottom).

This degradation arises from the 3,4-diamino phenols susceptibility to oxidation, which had been previously reported for the unsubstituted 3,4-diaminophenol analogue [270]. The final product of degradation was not determined but was likely a mixture of hydroxyl amine products ($-\text{N}[\text{H}]\text{OH}$), hence the presence of two major impurities for the **PPTMA** and **PEM DABs**. **PEP** appeared to have degraded to one major product.

Following this observation, complexation reactions with the ether functionalised **DABs** were carried forward immediately after their isolation, using a nitrogen environment for the initial templating step. The zinc(II) templating step appeared to occur very rapidly indicated by the reaction mixture almost instantly turning from colourless to bright yellow upon the combination of colourless starting materials. This reaction mixture was then left overnight to ensure full complexation before the addition of PtCl_2 . The zinc(II) salphen intermediates were not isolated and were used directly without further purification to prepare the final platinum(II) complexes.

Upon addition of the PtCl_2 , the reaction mixture turned a deep red colour. Despite this instant colour change the transmetalation occurred slowly, requiring 72 hours on

average before complete conversion was achieved. ^1H NMR spectroscopy was used to monitor the reaction as seen in figure 2.6 for the synthesis of **Pt_PPTMA_Ph**. Spectrum A shows the **DAB** starting material. Spectrum B shows the complete formation of the zinc(II) salphen after 24 hours, indicated by the presence of the imine peaks at 9 ppm, and absence of the peaks associated with the **DAB**. Spectrum C shows the partial conversion of the zinc(II) salphen to the platinum(II) salphen 48 after the initiation of the transmetalation reaction. Small peaks assigned to the imines of the zinc(II) salphen can still be seen at 9 ppm, with new major imine peaks associated with the platinum(II) salphen appearing around 9.4 ppm, showing an overall downfield shift of all complex peaks. This downfield shift can be attributed to the increased relativistic effect of the platinum(II) centre due to its larger size compared to the zinc(II) centre.

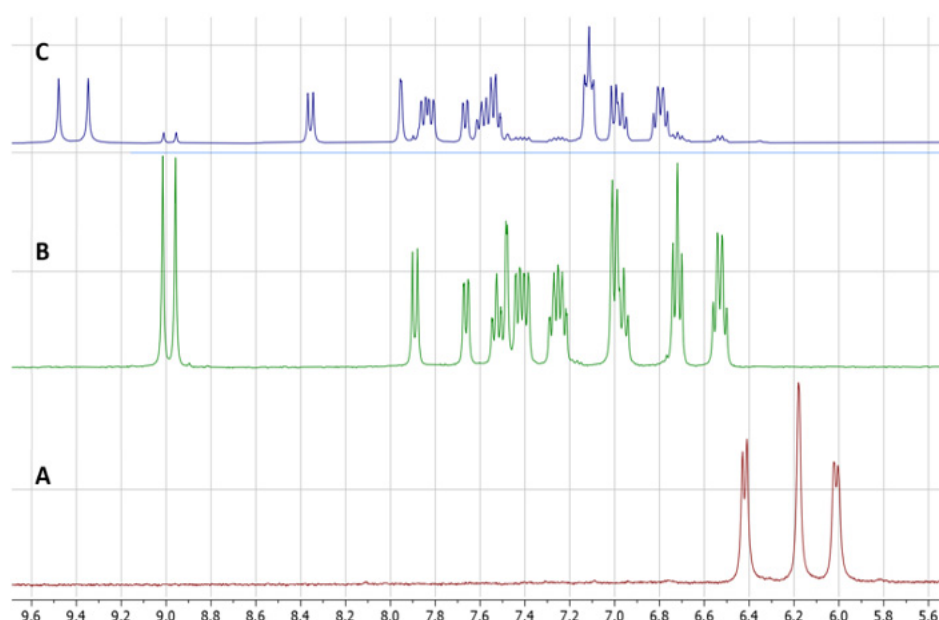


Figure 2.6. ^1H NMR spectra monitoring the formation of **Pt_PPTMA_Ph**. A) **PPTMA** diamine starting material. B) Zinc(II) salphen complex after 24 hours, with imine peaks at 9 ppm. C) platinum salphen complex product after 48 sequential hours, showing new platinum(II) salphen imine peaks at 9.4 ppm.

Following the transmetalation step, the **PPTMA** containing platinum(II) salphen complexes were precipitated out of solution using a saturation aqueous NaPF_6 solution, which resulted in complexes with purities of 99 % and above as assessed by liquid chromatography mass spectrometry (LCMS). As both the zinc templating and platinum transmetalation steps occurred readily with the **PPTMA DABs**, it was possible to

synthesise complexes with all 15 **CBA** groups.

The synthesis of complexes containing **PEM** were also generally high-yielding, with successful complex formations for **CBAs**: **Ph**, **Naph**, **4OMe**, **5Br** and **5OCF₃**. All confirmed by a distinct imine formation in their ¹H NMR spectra between 9 and 10 ppm, and purity confirmed by LCMS (see experimental details in section 7). As an example, figure 2.19 shows the ¹H NMR spectrum for **Pt_PEM_Ph** with a clean aromatic region and distinct imine peaks highlighted in green. The overall integration is 2 hydrogens less than expected due to the masking of the second aliphatic morpholine peak by the water peak at 3.6 ppm.

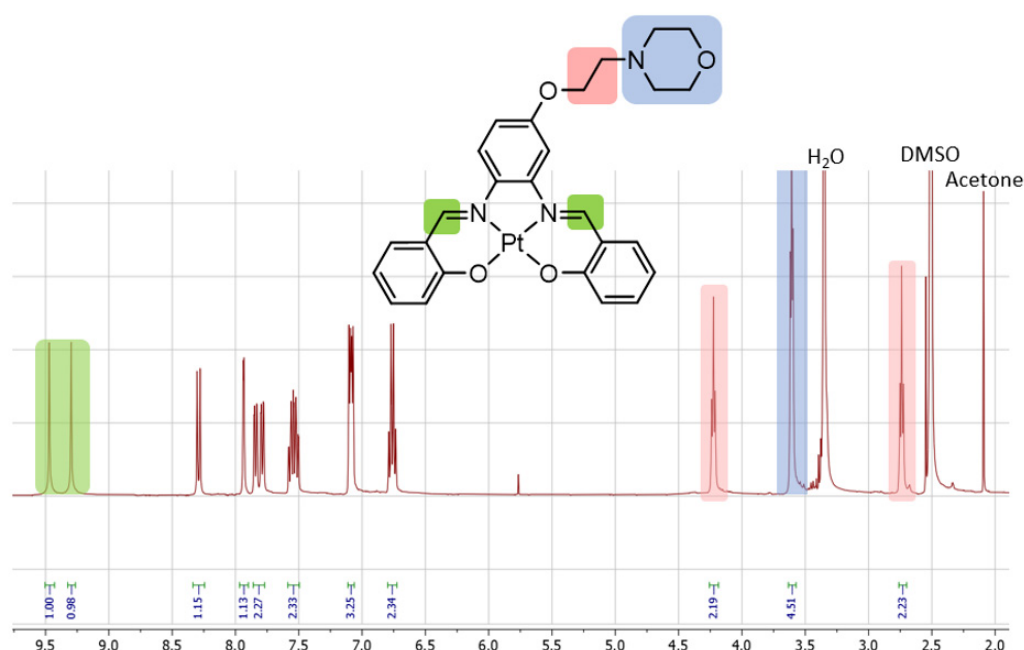


Figure 2.7. ¹H NMR spectrum of **Pt_PEM_Ph** highlighting distinct peaks for complex formation identification: di-imine (green). Ethyl linker (red). Morpholine (blue).

Complexes containing the **PEP** backbone did not form as readily. It was seen by ¹H NMR spectroscopy that following the same conditions used for the complexes with the **PEM DAB**, a mixture of products were formed. However, addition of 1 drop of 0.1 M HCl into the DMSO reaction mixture led to a slightly cleaner ¹H NMR spectrum. This can be seen below for the formation of **Pt_PEP_Ph**. Purple arrows indicate impurities in the reaction without HCl which are no longer present in the reaction with HCl (figure 2.8).

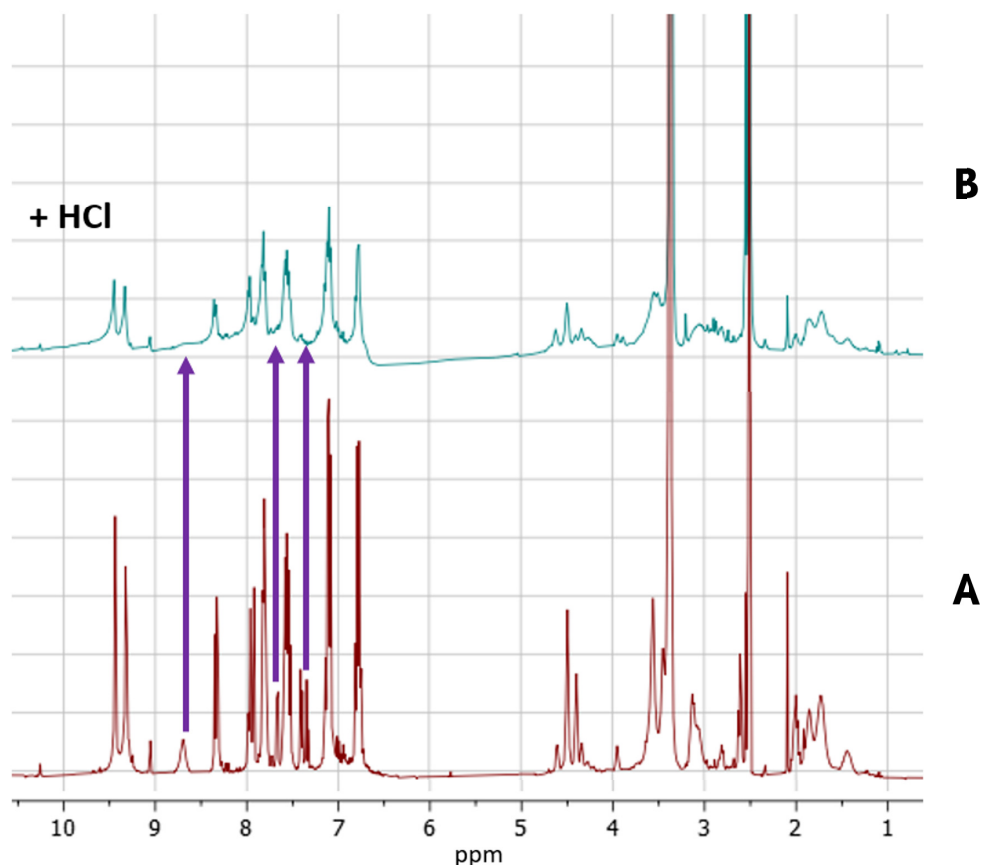
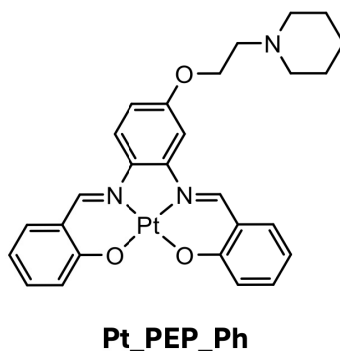
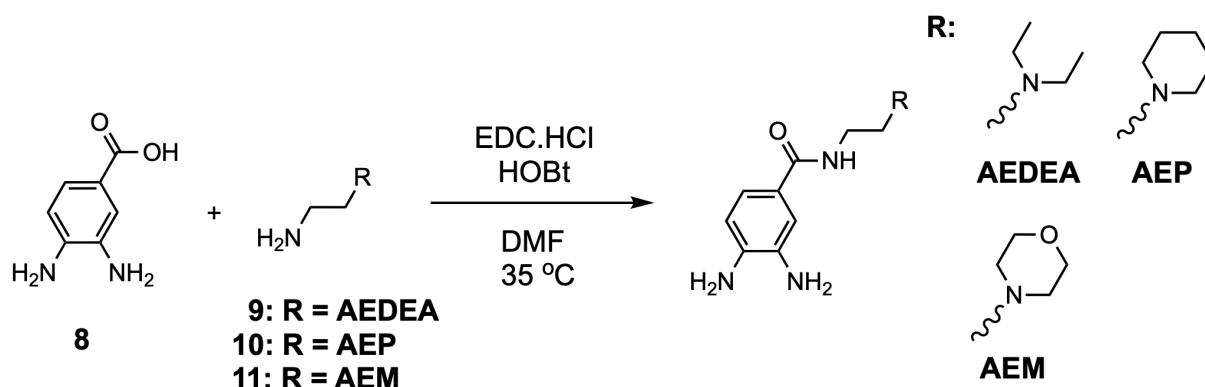


Figure 2.8. ^1H NMR spectra of **Pt_PEP_Ph** showing loss of some impurity peaks when HCl is added. A = before HCl addition. B = after HCl addition.

The synthetic procedure required further optimisation for a clean formation of the **PEP** complexes. However, due to the synthetic challenges associated with the unstable ether **DABs** it was decided to move on to explore the alternate functionalised **DABs**. Within this section, 20 pure platinum(II) salphen complexes were obtained and purity was confirmed by ^1H and ^{13}C NMR spectroscopy, LCMS, and high resolution mass spectrometry (HRMS) (see experimental details in section 7). The structures of all the isolated complexes can be found within the conclusion of this chapter (section 2.22).

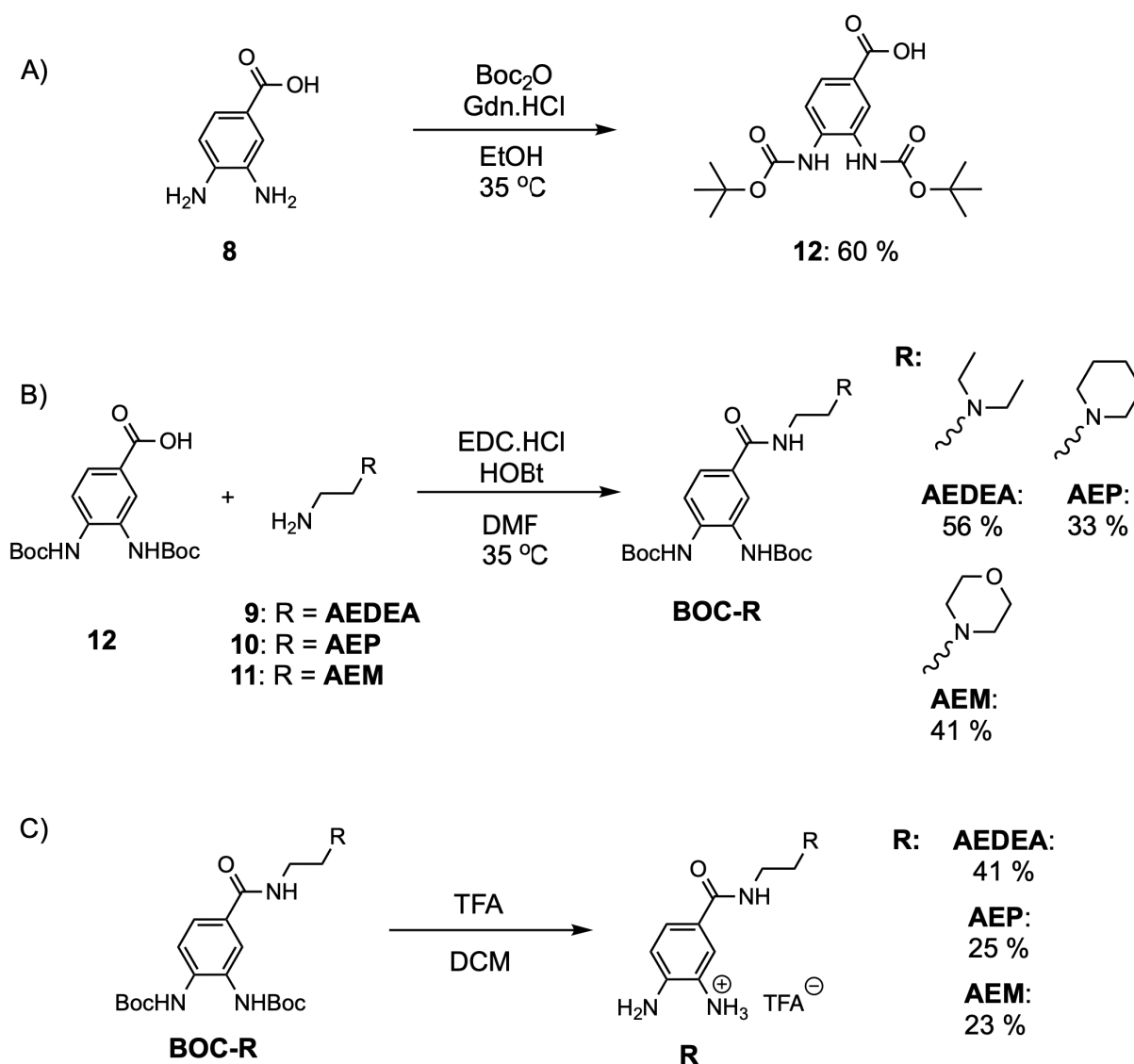
2.2.2 Amide linked DAB platinum(II) salphen complexes

The next set of complexes investigated contained amide-functionalised **DABs** (i.e. **AEDEA**, **AEP** and **AEM** – see figure 2.3). The amide coupling was initially attempted on the diamino benzoic acid starting material (**8**) (scheme 2.4).



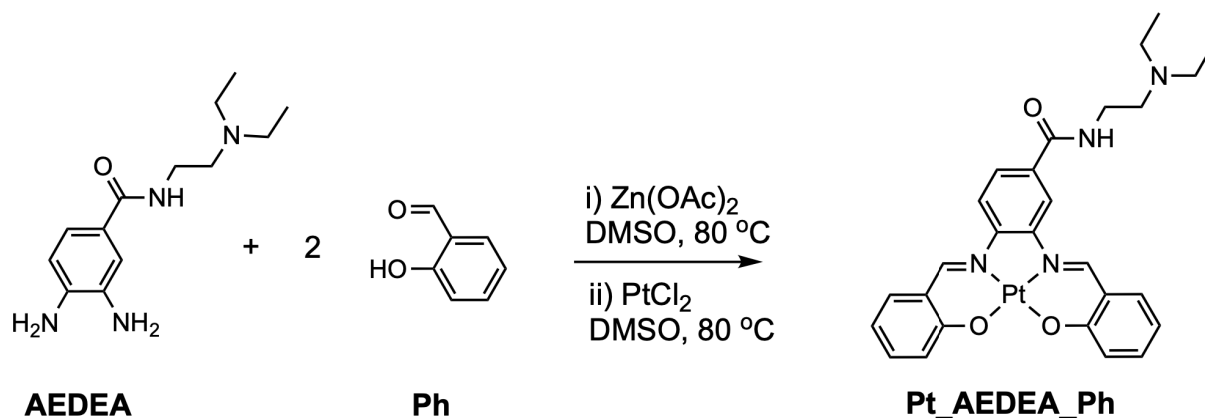
Scheme 2.4. Initial reaction scheme for the diamino benzoic acid amide formation. The final **DAB** is named after the amide linked solubilising group. All reactions were ran for 18 h.

However, it was found that the diamino benzamide products were extremely water soluble and therefore, not compatible with the water quench of the amide formation reaction. Very little product was obtained from the organic washes of the aqueous quench. In addition to further difficulties with the crude product co-eluting with the coupling agent when attempting further purification by column chromatography, as evidenced by thin layer chromatography (TLC). To overcome this issue, the amines in scheme 2.4 were protected prior to the amide coupling reaction (scheme 2.5).



Scheme 2.5. Synthetic route of the amide linked **DABs**: A) Boc protection reaction of diamino benzoic acid (12 h). B) Amide formation reaction of the Boc protected benzoic acid (All reactions were ran for 18 h). C) Deprotection of the diamine amide pre-cursor (Deprotection reactions were ran for 18 h at room temperature).

For the amide-substituted diamines, the synthesis of the platinum complexes were first attempted using the same conditions described in the previous section for the ether-substituted **DABs**. It was later found that higher temperatures and dry reaction conditions were required to ensure the full formation of the di-imine with the amide substituted **DABs**. The first amide functionalised complex which was attempted to be synthesised was **Pt_AEDEA_Ph** (scheme 2.6).



Scheme 2.6. Complex formation reaction between **DAB AEDEA** and **CBA Ph**. Reaction times: i) 24 h. ii) 72 h

The ^1H NMR spectrum of the final reaction mixture showed the formation of a single product; however, the integration was not consistent with the expected product. A product was isolated by precipitation and its ^1H NMR spectrum changed drastically as compared to the reaction mixture, showing new peaks at 13.43 and 12.87 ppm (figure 2.9). The precipitate was analysed by LCMS and was found to contain a major product of the mono-substituted ligand, with an m/z of 353.3.

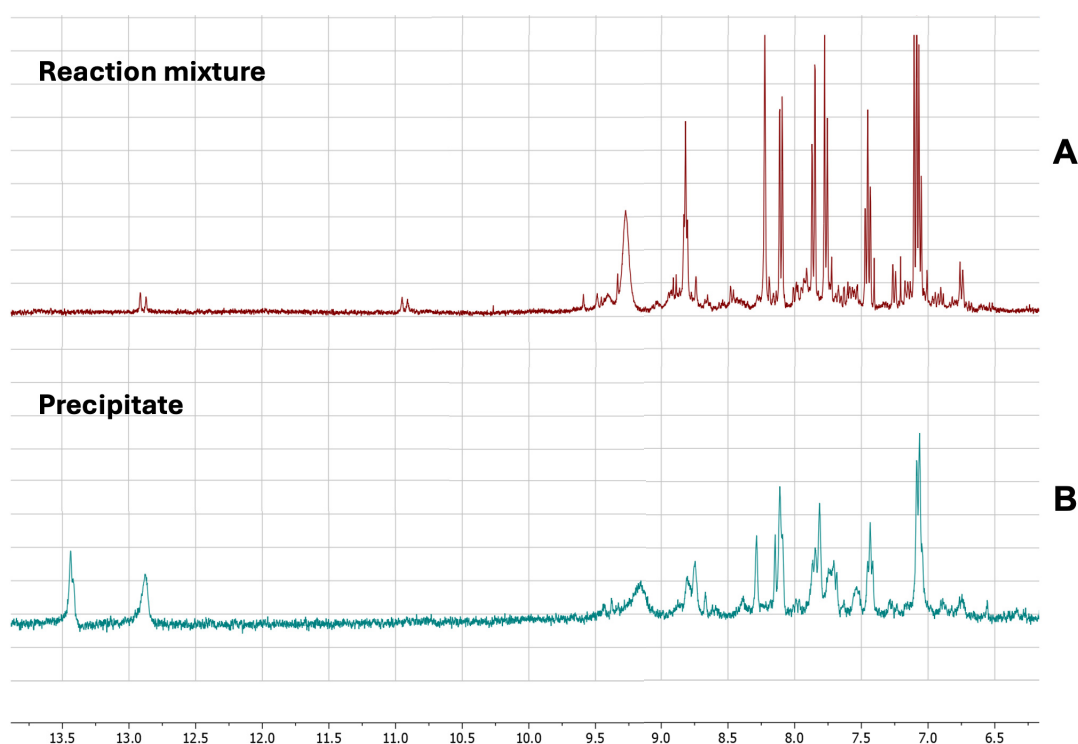


Figure 2.9. ^1H NMR spectra of **Pt_AEDEA_Ph** reaction mixture and precipitate from the reaction mixture.

To investigate the issues with this reaction, the zinc templating reaction step was repeated and the zinc(II) intermediate was isolated. The resultant ^1H NMR spectrum of the reaction mixture showed more clearly that the mono-substituted ligand was the major product (figure 2.10).

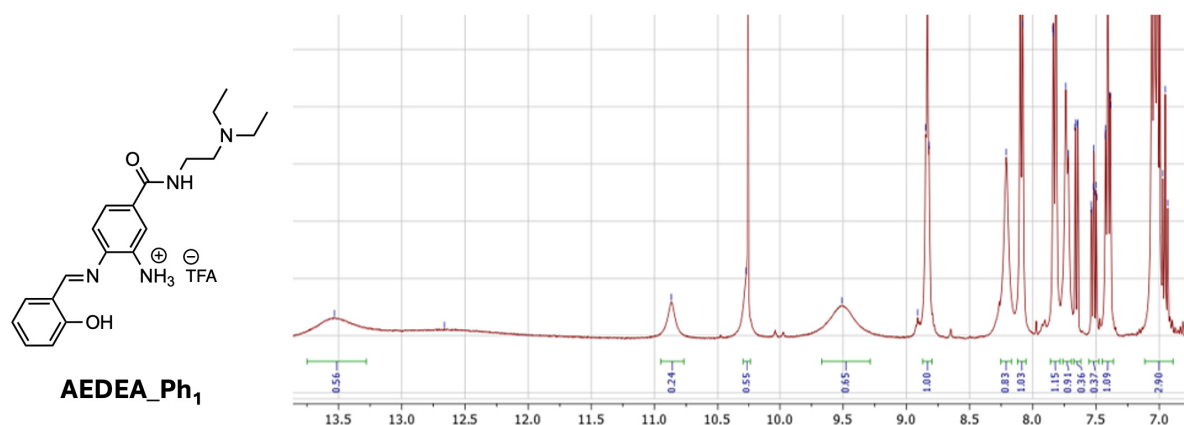


Figure 2.10. Aromatic region of the ^1H NMR spectrum for the mono-substituted salphen ligands from the failed formation of the **Zn_AEDEA_Ph₁**.

The aromatic region integrated to only 9 protons which was consistent with the mono-substituted ligand. Further evidence of this includes the presence of a 0.5:1 ratio of **CBA** to complex shown by the aldehyde singlet peak at 10.25 ppm. In addition, the broad peak observed at 13.5 ppm indicates the nitrogen is protonated, which could lead to the preferable formation of the mono-substituted complex. It was likely that zinc metal was not complexed to the ligand, as only the mass of the mono-substituted ligand was detected by LCMS (353 m/z).

It was suspected that the TFA salt of the **DAB** was preventing both amines from reacting, therefore the **DABs** were deprotonated using a potassium carbonate plug prior to reaction. Attempting the zinc templating step using the deprotonated diamine resulted in formation of one major product in addition to some minor products indicated by the cluttered aromatic region of the reaction mixtures ^1H NMR spectrum (figure 2.11).

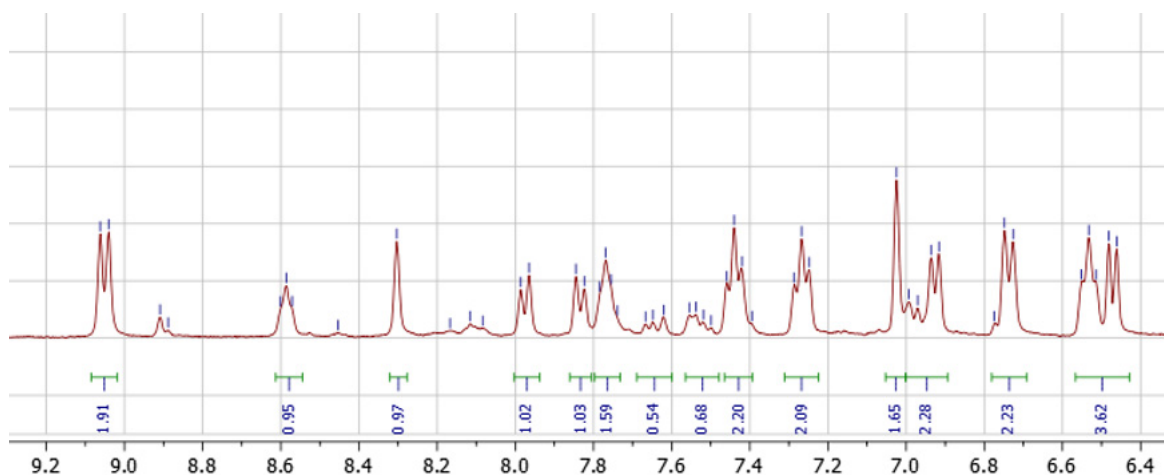


Figure 2.11. ^1H NMR spectrum of Zn_AEDEA_Ph reaction mixture using AEDEA which has been deprotonated using K_2CO_3 .

As complexation appeared to occur less readily for the amide-containing **DABs** (likely to be due to the electron-withdrawing nature of the amide), the templating reaction was repeated at 100°C for 24 hours. The resulting ^1H NMR spectrum showed a more clean aromatic region which integrated to the desired 14 protons. However, side products were still present, in addition to the presence of the **CBA Ph** starting material indicated by the aldehyde singlet at 10.25 ppm (figure 2.12).

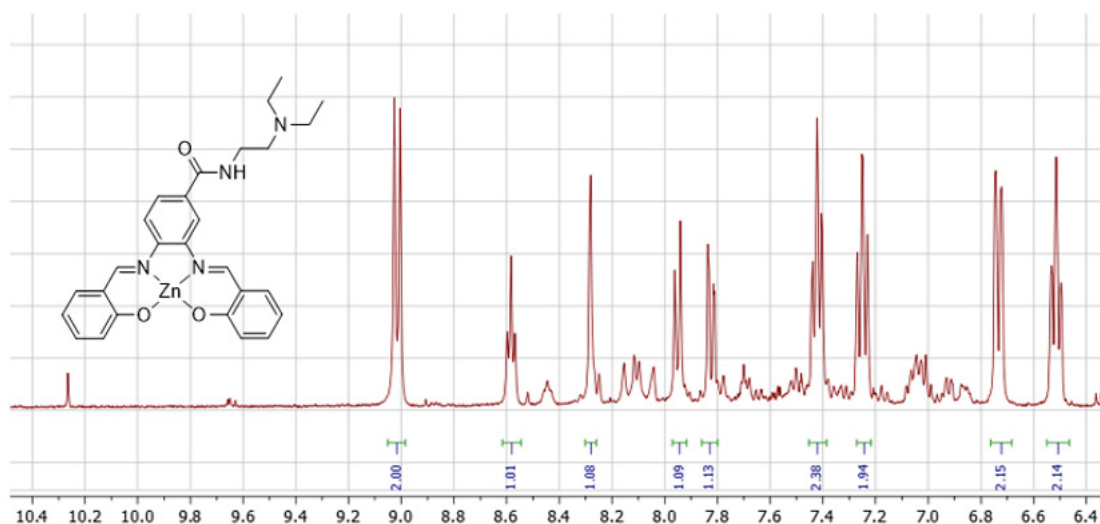


Figure 2.12. Aromatic region of the ^1H NMR spectrum for Zn_AEDEA_Ph reaction mixture formed at 100°C .

The reaction was heated to 100°C for a further 96 hours, but no change was observed

in the ^1H NMR spectrum of the reaction mixture. PtCl_2 was then added to the reaction mixture and heated for a further 12 hours. However, there was no significant change in the chemical shift to be attributed to the formation of the platinum(II) salphen complex within the new ^1H NMR spectrum shown in blue in figure 2.13, alongside the original spectrum in red, indicating that the transmetalation step was not occurring.

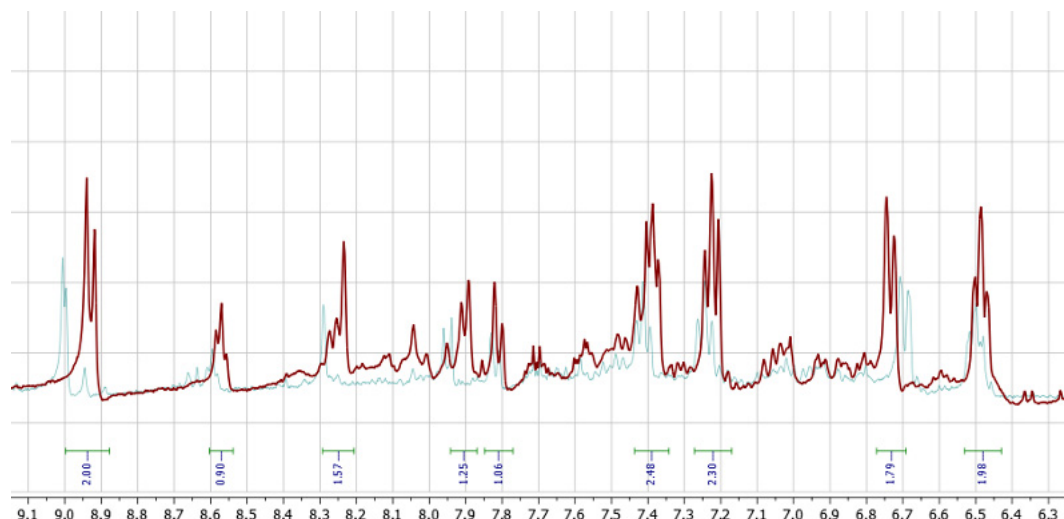


Figure 2.13. Aromatic region of the ^1H NMR spectra for **Zn_AEDEA_Ph** (red) and attempted transmetalation (blue).

The transmetalation reaction mixture was then heated for an additional 48 hours, but no further changes were observed in the ^1H NMR spectrum. This lack of reaction may have been due to the incomplete conversion of starting materials to the full zinc(II) salphen complex and the presence of mono-substituted side-products. Therefore, to promote full conversion to the tetradentate ligand (and zinc(II) complex) the reaction was repeated under a nitrogen atmosphere with molecular sieves to investigate if the presence of water was preventing the full conversion of the diamine to the di-imine. After 24 hours the zinc-templated reaction mixture was filtered, and water was added to precipitate out a yellow solid. A sample of this solid was then redissolved in DMSO-d_6 , and the ^1H NMR spectrum showed clean complex formation with the desired 14 aromatic protons, and almost no presence of side products (figure 2.14).

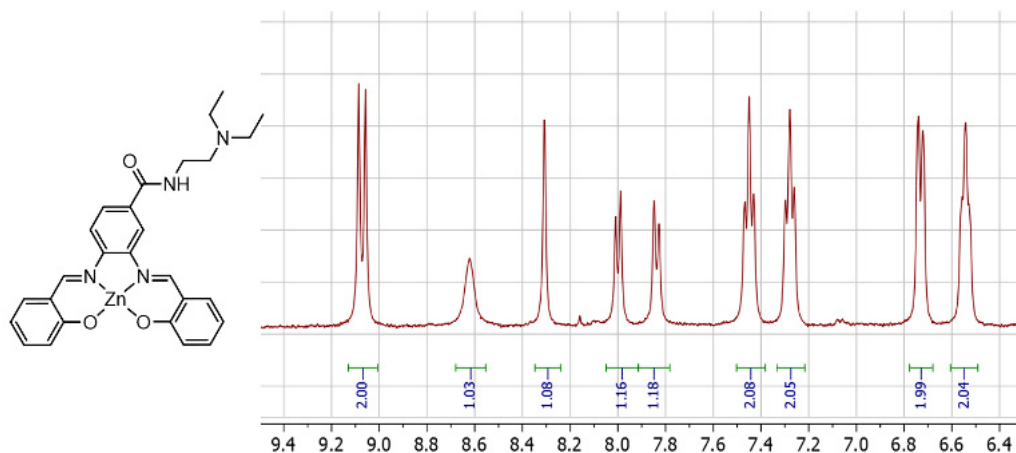


Figure 2.14. ^1H NMR spectrum for isolated **Zn_AEDEA_Ph**.

The isolated **Zn_AEDEA_Ph** was then redissolved in DMSO-d_6 and PtCl_2 was added to attempt the transmetalation step. After 42 hours the reaction mixture was filtered and an ^1H NMR spectrum was obtained. The spectrum showed a proton shift of the imine peaks from 9.07 ppm to 9.64 ppm, indicating that the zinc had been displaced from the complex by the platinum (figure 2.15).

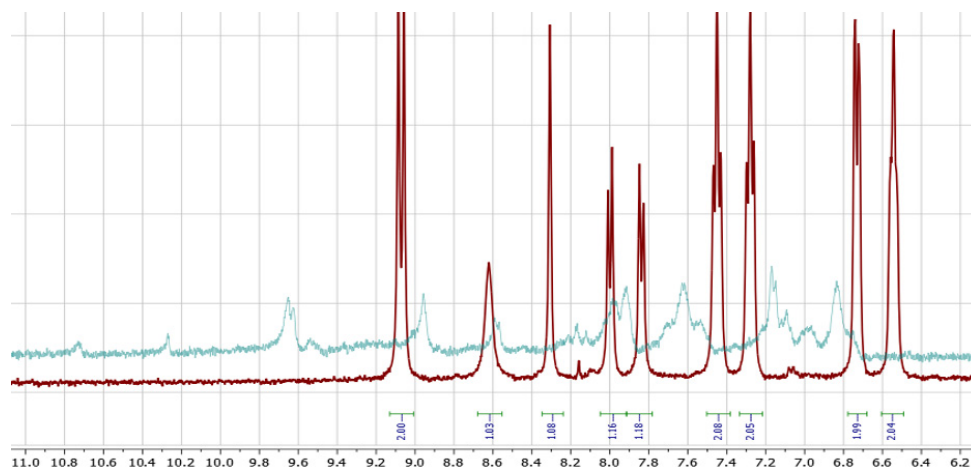


Figure 2.15. ^1H NMR spectra of the **Zn_AEDEA_Ph** (red) transmetalation to **Pt_AEDEA_Ph** (blue).

However, the intensity of the protons in the aromatic region were low and integrated to more protons than the desired 14 of the final **Pt_AEDEA_Ph** complex. The reaction mixture was left for a further 96 hours in an attempt to achieve full conversion, however, no change in the spectrum was observed. As the reaction mixture was cloudy it was

suspected that the **AEDEA** backbone was not providing a sufficient solubilising effect and the final complex was precipitating from the reaction mixture.

It was decided to explore the addition of a functionalised **CBAs** with the **AEDEA** backbone to assess if the highly planar aromatic surface was leading to the insolubility of the final **Pt_AEDEA_Ph** complex. The zinc transmetalation step was repeated for the **AEDEA DAB** with the **4OMe CBA**. The goal of the addition of the **OMe** functionality was to reduce the planarity of the complex, and provide additional solubilisation by the introduction of hetero atoms. Zinc templating appeared to be successful for **Zn_AEDEA_4OMe** resulting in the ^1H NMR spectrum of the reaction mixture to contain the desired 12 aromatic hydrogens (figure 2.16).

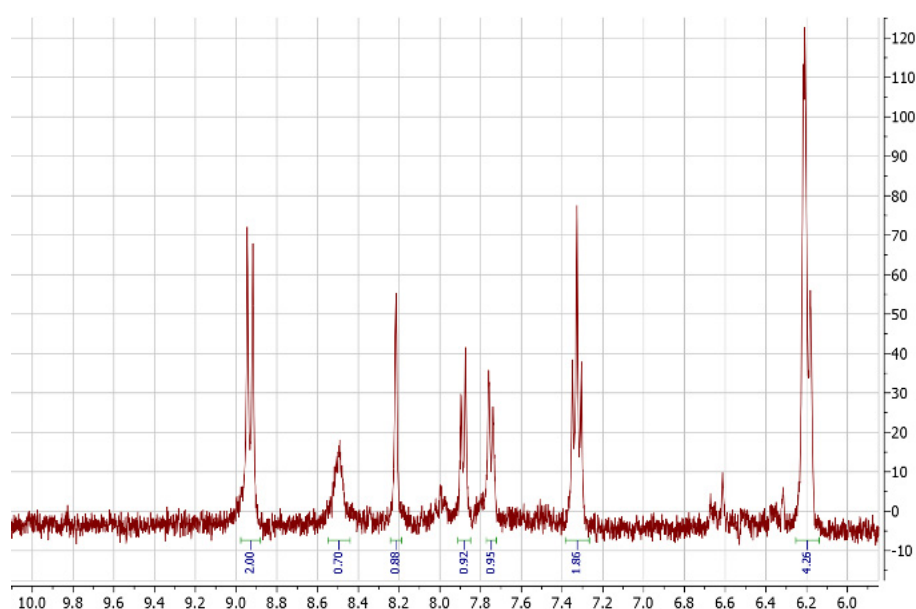


Figure 2.16. Aromatic ^1H NMR spectrum of the **Zn_AEDEA_4OMe** reaction mixture.

The **Pt_AEDEA_4OMe** reaction mixture was reacted further with PtCl_2 and heated at $100\text{ }^\circ\text{C}$ for 72 h. The resulting product was then precipitated from the reaction mixture with water, filtered, and dried under vacuum. The resulting ^1H NMR spectrum for the target platinum complex showed a non-distinct broad aromatic region with very low intensity (figure 2.17).

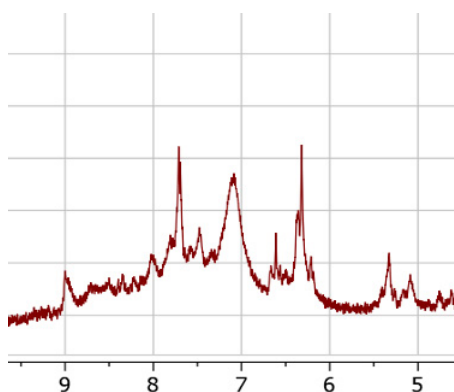


Figure 2.17. Aromatic region of the ^1H NMR spectrum of the isolated product following the attempted **Zn_AEDEA_4OMe** transmetalation reaction.

The inability to isolate the final platinum complex with the **AEDEA DAB** confirmed its unsuitability in providing a reasonable solubilising effect to the complex. To test the eligibility of the other amide functionalised **DABs**, complexes with the **AEM** and **AEP** amide **DABs** were synthesised with the **5Br** and **4OMe CBAs** (figure 2.18).

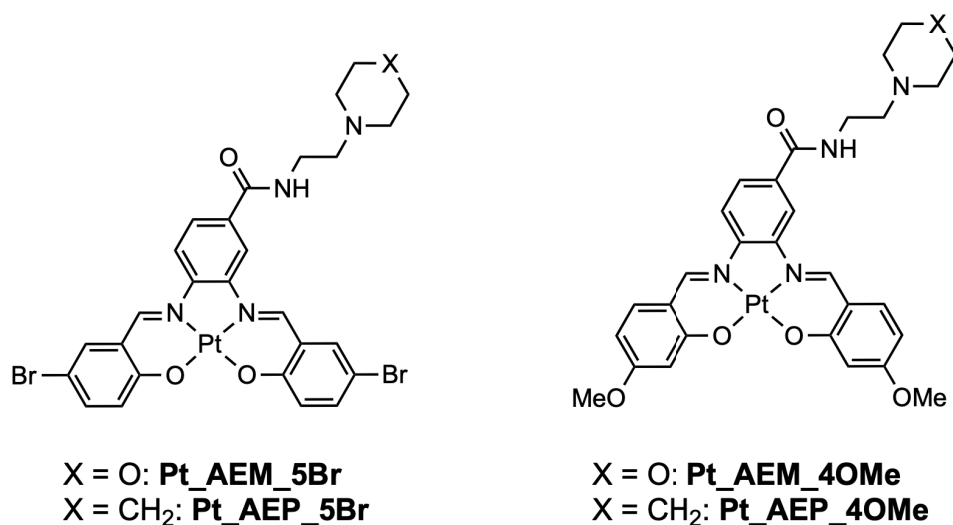


Figure 2.18. Structures of the **AEP** and **AEM** complexes explored synthetically.

Following the templation step, reaction mixtures were filtered to remove the molecular sieves. A saturated aqueous NH_4PF_6 solution was added to each of the bright yellow filtrates resulting in precipitation of the yellow zinc intermediate complexes. The formation of the zinc(II) salphen complexes was confirmed by ^1H NMR spectroscopy with each spectrum containing the desired 12 aromatic protons. The formation and purity of the compounds was further confirmed by LCMS data (table 2.1).

Complex	Molecular weight	LC area (%)	m/z $-[H^+]$ (m/z)	m/z $-[Zn^{2+}]$ (m/z)
Zn_AEM_5Br	693.70	96	692.10	631.20
Zn_AEM_4OMe	595.96	89	594.10	533.20
Zn_AEP_5Br	691.73	100	690.90	629.60
Zn_AEP_4OMe	593.99	97	593.40	531.40

Table 2.1. Table showing the area of the UV peak and the masses detected for each isolated zinc(II) salphen intermediate complex by LCMS.

The isolated zinc(II) salphen complexes were subsequently redissolved in DMSO- d_6 , combined with 1.1 equivalents of $PtCl_2$ and heated to 100 °C for 4 days. The **4OMe** complexes transmetalation yields were low (< 17 %) with < 3 mg of each complex being isolated. Despite the low yield, the **Pt_AEP_4OMe** complex was isolated with a purity of 95 % by LCMS, with broad complex peaks seen within 1H NMR spectroscopy. NH_4PF_6 was also seen to be present within the 1H NMR spectrum at 7.1 ppm and was not possible to remove by further washing (figure 2.19).

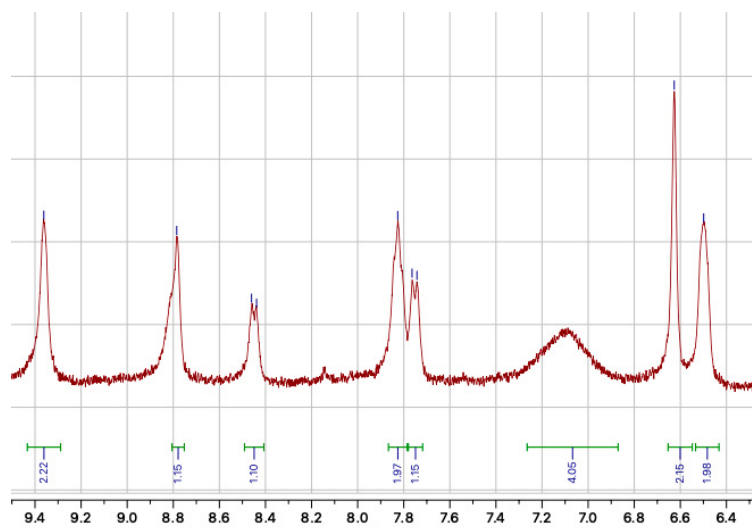


Figure 2.19. Aromatic region of the 1H NMR spectrum of **Pt_AEP_4OMe** in DMSO- d_6 .

The **5Br** complexes additionally contained broad peaks for the complex and the NH_4^+ ion within their 1H NMR spectra and were found to contain impurities by LCMS. Repurification was attempted by redissolving the crude solid in DMSO followed by precipitation with water. However, this did not improve their purity. The LCMS

measurements of the platinum complexes are shown in table 2.2.

Complex	Molecular weight	LC area (%)	m/z +[H ⁺] (m/z)
Pt_AEM_5Br	823.4	52	824.9
Pt_AEM_4OMe	725.7	75	726.1
Pt_AEP_5Br	821.4	92	821.3
Pt_AEP_4OMe	723.7	95	724.3

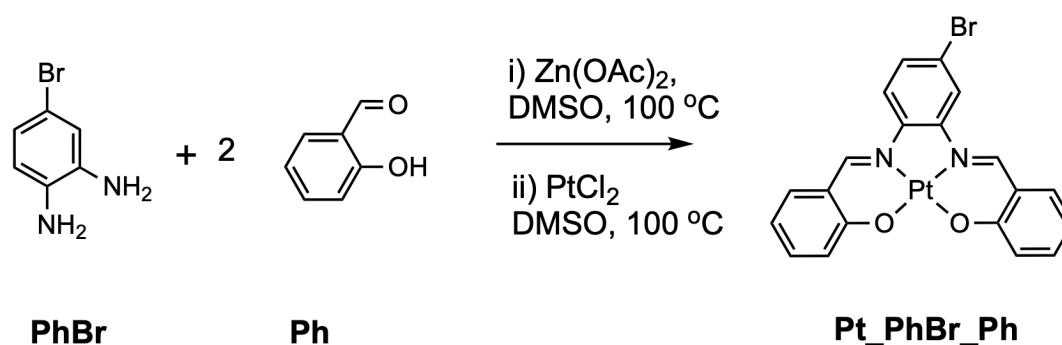
Table 2.2. LCMS data for the amide functionalised **AEP** and **AEM** platinum(II) salphen complexes.

The LCMS traces of the less pure amide functionalised platinum(II) complexes contained impurities which displayed platinum splitting patterns within the MS associated with the signals under the LC trace. This indicated other platinum complexes were being formed. The masses of these signals were assigned to the mono-substituted complexes.

The key differences between the reactivity of the ether-substituted **DABs** and the amide-substituted **DABs** were likely due to the electron withdrawing nature of the amide functionality, in contrast to the ether functionality which was electron donating. As such the ether **DAB** has a more electron-rich aromatic centre, promoting the dual imine formation reaction. Additionally, as imine formation is reversible, the amide **DABs** required slightly elevated temperatures (80 °C to 100 °C), the removal of water the by-product, and extended reaction times to push reactions to completion. The amide complexes also appeared to perform the transmetalation step less favourably, only occurring after isolation of the zinc(II) salphen pre-cursor. Due to the need to isolate the zinc(II) salphen intermediate before transmetalation, and the challenges seen within the purification of these complexes, it was decided to not continue their development. From this section the purest complex, **Pt_AEP_4OMe**, was taken forwards for photophysical and biological screening.

2.2.3 Alternate DAB systems

Due to the challenges of generating complexes with the functionalised **DABs**, it was decided to explore **DABs** which could be further functionalised following complex formation. Common reactive functional groups include halogen-substituted diamino benzenes which can be subsequently reacted through reaction types including nucleophilic aromatic substitutions, Suzuki couplings, Sonogashira couplings and Buchwald-Hartwig aminations [61, 62, 271]. The first halogen functionalised platinum(II) salphen complex generated was **Pt_PhBr_Ph** (scheme 2.7).



Scheme 2.7. **Pt_PhBr_Ph** synthesis: 1) zinc templating (24 h) and 2) platinum transmetalation (72 h).

Complexation was achieved under the standard one-pot zinc-templated procedure at 100 °C in DMSO, resulting in a pure complex after isolation by aqueous precipitation. This was confirmed by ^1H NMR spectroscopy where 13 aromatic peaks were observed with the distinct imine hydrogen environments within the expected 9.5 ppm region for the platinum complex (figure 2.20).

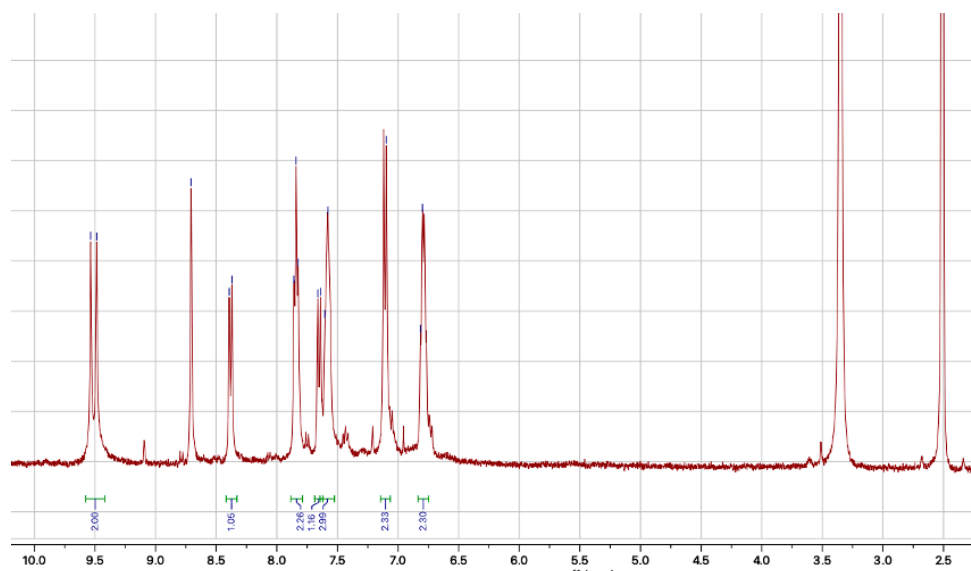


Figure 2.20. ^1H NMR spectrum for **Pt_PhBr_Ph**. Peaks at 3.5 ppm: DMSO, 2.5 ppm: H_2O .

This was supported by LCMS showing a 96 % purity (m/z of 589 + $[\text{H}^+]$). This complex was subsequently reacted with NaN_3 to install an azide group for further ‘click’ cycloaddition reactions which would be ideally suited for HT synthesis (scheme 2.21).

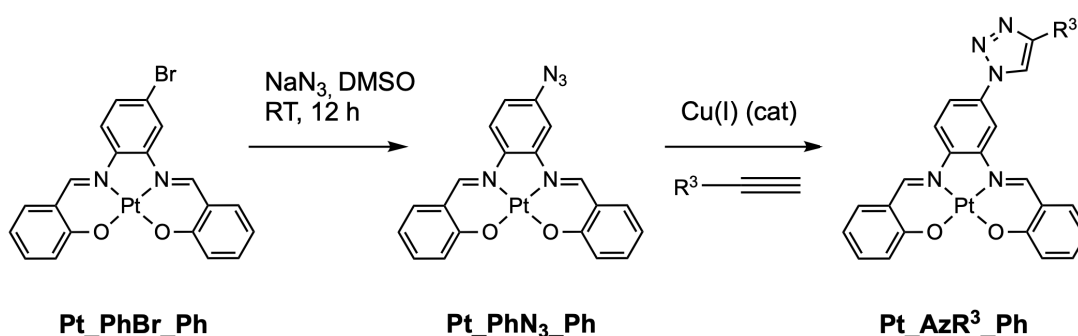
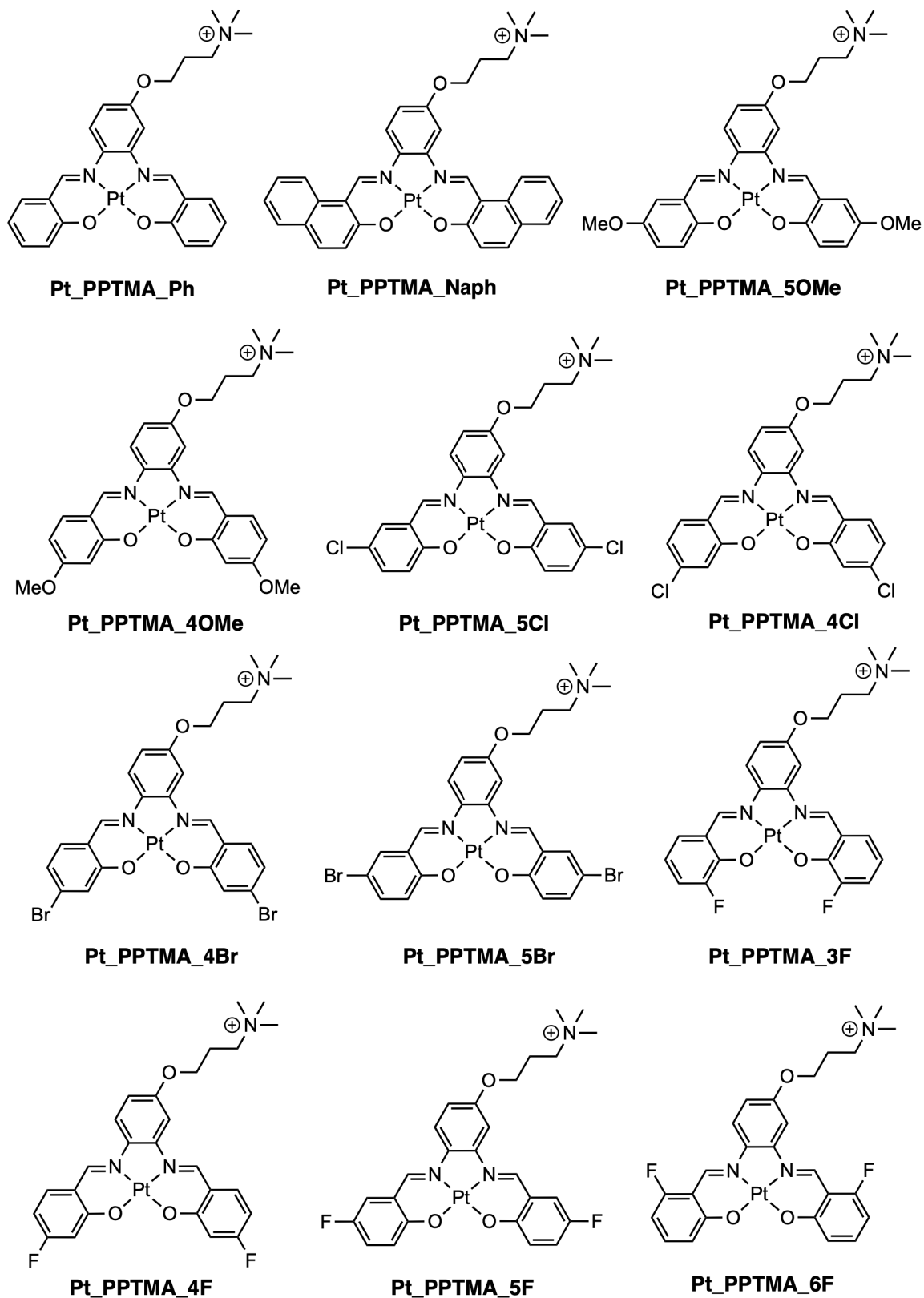


Figure 2.21. Diversification reaction of **Pt_PhBr_Ph** via azide formation and click chemistry.

However, when attempting to generate the azide, it was observed that after 2 days, no conversion had occurred at the original room temperature conditions. The ^1H NMR spectrum of the isolated solid showed it was the unreacted **Pt_PhBr_Ph** complex, as confirmed by LCMS. The reaction was then repeated at 60 °C over 120 hours, but no conversion was observed. **Pt_PhBr_Ph** was taken forward for biological testing to provide an example of a complex without a linked solubilising group on the **DAB**.

In total, 22 platinum(II) salphen complexes were isolated with a > 95 % purity as

evidenced by ^1H and ^{13}C NMR spectroscopy, LCMS, and HRMS. The structures for the final complexes taken forward for photophysical and biological screening are shown in figure 2.22:



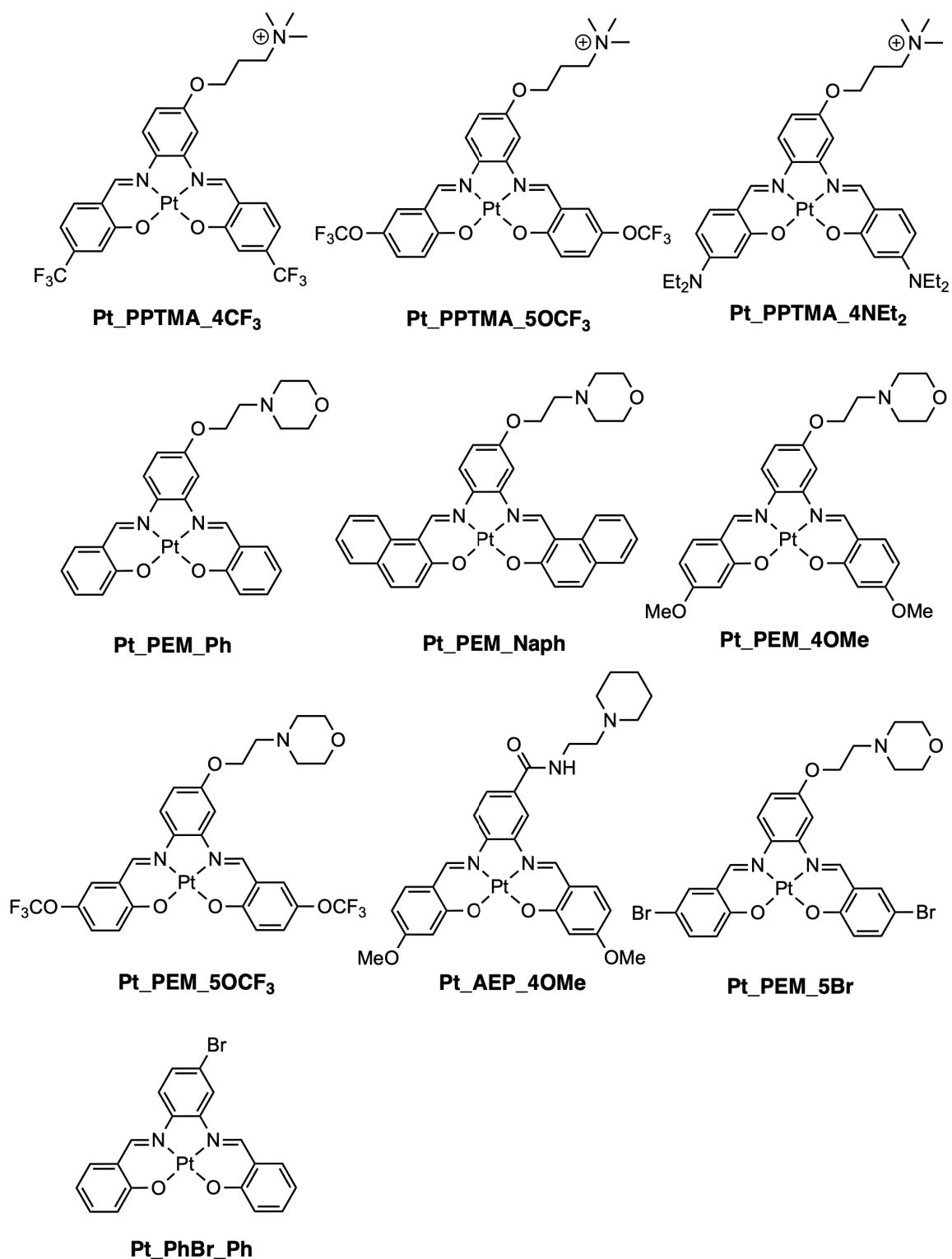


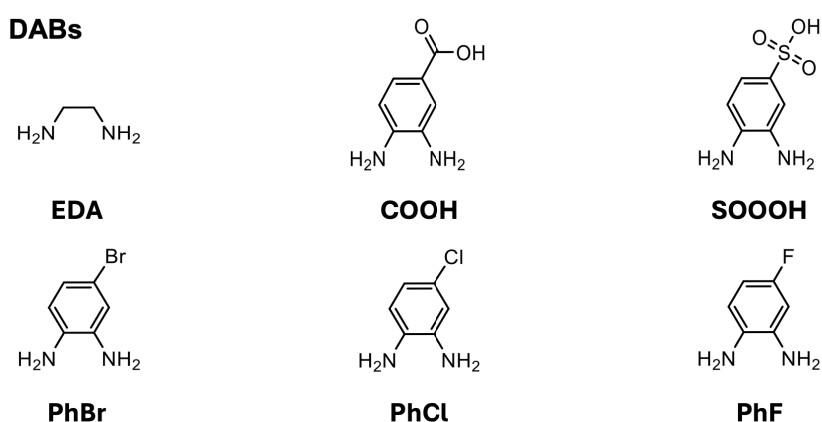
Figure 2.22. Figure showing the structure and systematic names of the 22 pure isolated platinum(II) salphen complexes.

2.3 High throughput synthesis

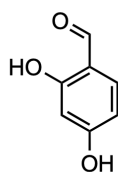
The one-pot synthetic route for the platinum(II) salphen complexes was selected for its potential suitability for HT synthesis. The objective of this work was therefore to extend the library of platinum(II) salphen complexes using HT synthesis, which would be followed by HT screening of photophysical and biological properties. The data obtained from this HT pipeline would then be used within ML SAR analysis, and for the development of predictive ML models to aid future complex discovery.

When monitoring complexation reactions in batch by ^1H NMR spectroscopy it was seen that they were proceeding to 100 % conversion. However, when attempting to use the ether-functionalised **DABs** within HT synthesis it was found that they degraded over time, even when contained within a glove box. This was likely due to the presence of water within the DMSO solution resulting in the amine oxidation as discussed in section 2.2.1. Therefore, due to the extended residence time of the **DABs** in solution for the set-up of the HT synthesis, these **DABs** were determined to be unsuitable. The amide-functionalised **DABs** were also deemed unsuitable for HT synthesis due to the required isolation of the zinc(II) salphen intermediate before the transmetalation step.

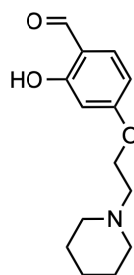
As a final recourse it was decided to select 6 commercially available **DABs** (figure 2.23) to attempt to generate a library of 96 complexes with 16 **CBAs**. These include the 14 in figure 2.2 (minus **4NEt₂**), plus an additional 2 shown in figure 2.23: **4OH** and **PhPip**.



Additional CBA's



4OH



PhPip

Figure 2.23. Structures for the commercially available **DAB** starting materials selected for HT synthesis. In addition to the structures of the additional **CBA** used within the HT synthesis.

The concentrations of the stock solutions were calibrated using a known concentrated solution of ethylene carbonate. Zinc templating reactions were performed at a complex concentration of 8 μM in 300 μL of DMSO-d_6 . Following the addition of 200 μL of PtCl_2 (1.1 equivalent) the final reaction concentration was 5 μM in 500 μL . Figure 2.24 shows the layout of the reaction plate.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Ph	4OH	Ph	4OH	Ph	4OH	Ph	4OH	Ph	4OH	Ph	4OH
B	4OMe	5OMe	4OMe	5OMe	4OMe	5OMe	4OMe	5OMe	4OMe	5OMe	4OMe	5OMe
C	4Cl	5Cl	4Cl	5Cl	4Cl	5Cl	4Cl	5Cl	4Cl	5Cl	4Cl	5Cl
D	4Br	5Br	4Br	5Br	4Br	5Br	4Br	5Br	4Br	5Br	4Br	5Br
E	3F	4F	3F	4F	3F	4F	3F	4F	3F	4F	3F	4F
F	5F	6F	5F	6F	5F	6F	5F	6F	5F	6F	5F	6F
G	4CF ₃	5CF ₃	4CF ₃	5CF ₃	4CF ₃	5CF ₃	4CF ₃	5CF ₃	4CF ₃	5CF ₃	4CF ₃	5CF ₃
H	5OCF ₃	PhPip	5OCF ₃	PhPip	5OCF ₃	PhPip	5OCF ₃	PhPip	5OCF ₃	PhPip	5OCF ₃	PhPip

EDA

COOH

SOOOH

PhBr

PhCl

PhF

Figure 2.24. 96 well reaction plate layout showing the **DAB** blocks defined by their colours, and the subsequent **CBA** group combined with the **DAB**.

An OT2 liquid handling robot was used to prepare the reaction plates, as shown in figure 2.25. The protocol consisted of three steps: 1) OT2 transferred the starting material

stock solutions to a 96-well reaction plate. The reaction plate was sealed with rubber caps and heated at 100 °C for 24 hours using the Unchained Lab robot's heater stirrer deck. 2) The plate was then unsealed, and the PtCl_2 stock solution was added by the OT2, before being resealed and heated at 100 °C for a further 72 hours. 3) The OT2 was used to prepare NMR samples for each reaction mixture.

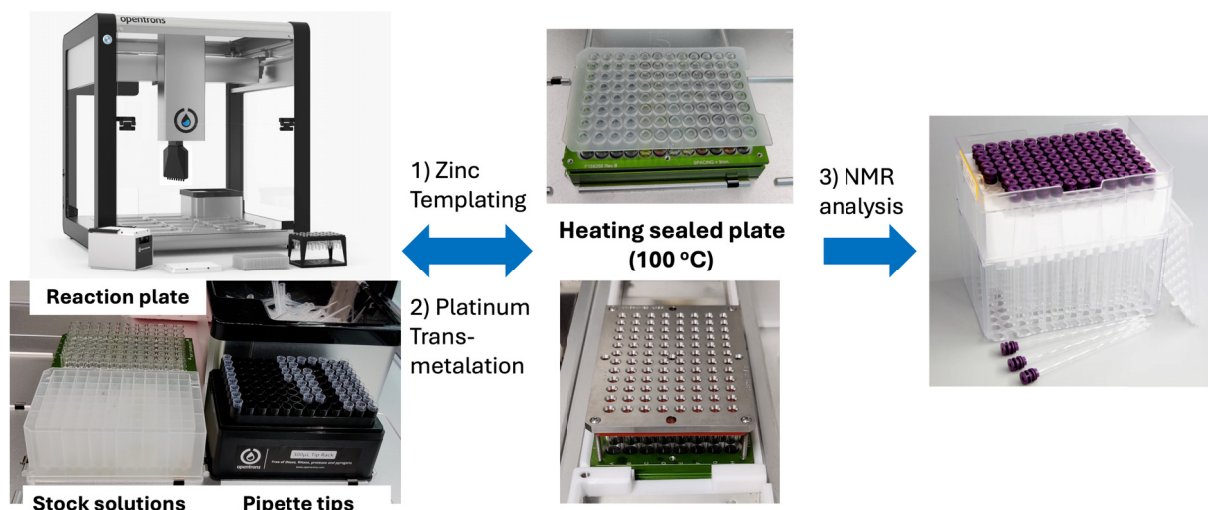


Figure 2.25. Schematic showing the OT2 (top left) and its deck set up (bottom left) containing the reaction plate, stock solution plate, and pipette tips for the 8-channel pipette. 1) First step of the protocol, adding all the **CBA** and **DAB** components across the plate, and finally the zinc acetate reagent. The plate was then sealed and heated using the Unchained Lab robots at 100 °C for 24 hours. 2) The plate was unsealed and placed back on the OT2 deck which was then used to add the PtCl_2 across the plate. The plate was then resealed and placed back on the Unchained Lab robot and heated at 100 °C for 72 hours. 3) The plate was then unsealed and transferred back to the OT2 which then prepared NMR samples of the reaction mixtures.

To assess the conversion, ^1H NMR spectroscopy was implemented, enabled by the Bruker SampleJet, which allowed 96 NMR samples to be queued using a 96 well plate array. However, no significant quantities of platinum(II) salphen complexes were detected upon analysis of the ^1H NMR spectra of the reaction mixtures (searching for the distinct di-imine peaks around 9.5 ppm). Exemplary figures 2.26 and 2.27 show the ^1H NMR spectra of the reaction mixtures for the complexation reactions with the **COOH** and **PhBr DABs** respectively, with each **CBA**. The top spectrum within each figure is the **DAB** stock solution which was used within each reaction.

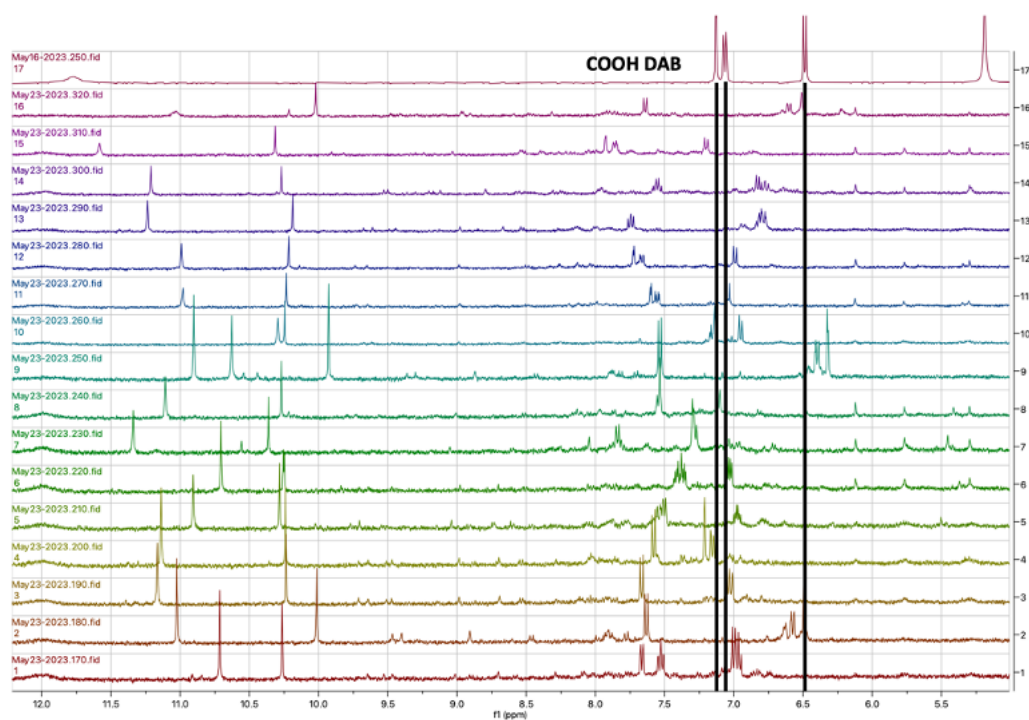


Figure 2.26. ^1H NMR spectra from the HT synthesis for each **CBA** with the **COOH DAB**.

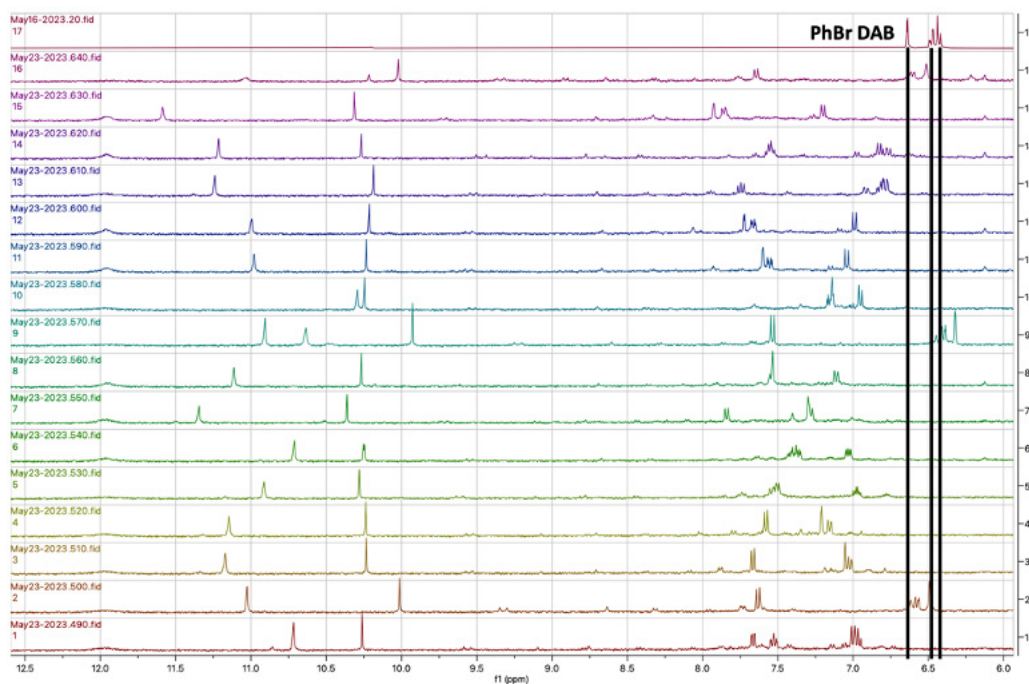


Figure 2.27. ^1H NMR spectra from the HT synthesis for each **CBA** with the **PhBr DAB**.

In the ^1H NMR spectra of the final reaction mixtures no **DAB** was present. NMR peaks could still be seen for the aldehyde **CBA** starting material (aldehyde peak around 10.25 ppm). In some reaction mixtures, small peaks associated with di-imines

from complexation could be seen between 10 - 9 ppm. When visually inspecting the reaction plates, there was a clear colour change of the reaction mixtures following the zinc templating step, and a further distinct colour change following the platinum transmetalation step, suggesting that the reactions may have been proceeding to a small extent (figure 2.28).

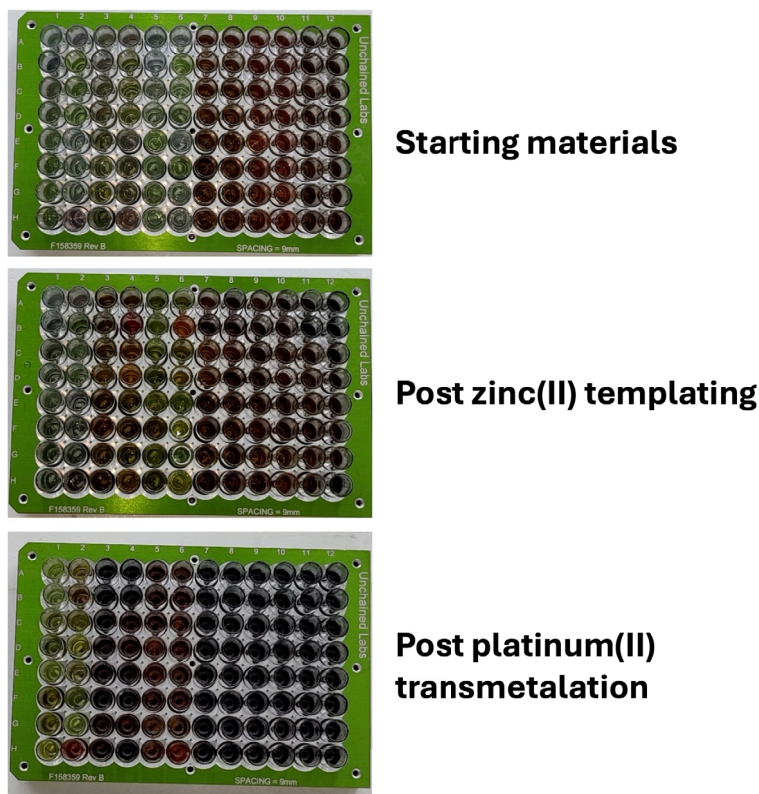


Figure 2.28. Three images showing the reaction plates containing the starting materials, the plate after the zinc templating reaction, and the plate following the transmetalation reaction.

The reaction mixtures were prepared using an OT2 that was not under an inert atmosphere, and the plastic seals on the reaction plates were not perfectly airtight. Therefore, it was possible that the **DAB** reagents were degrading before they were able to fully react with the **CBA** components. The HT synthesis was attempted a further two times, and the same results were obtained. To attempt to provide further time before complete decomposition of the **DABs**, the reactions were attempted a fourth time at a higher final complex concentration of 50 μM . The rationale behind this was to ensure that there was sufficient **DAB** remaining within the reaction mixture to enable reasonable complex formation. The final complexes could then be precipitated

out of the reaction mixtures, filtered using a 96-well filtration plate, and redissolved in DMSO-d₆, followed by determination of their final concentrations using ethylene carbonate. However, excessive precipitation was seen following the zinc templating step, likely due to the increased concentration, and the lack any solubilising groups in the **DAB**. Due to the presence of solid within the reaction mixtures it was not possible to assess conversion by ¹H NMR spectroscopy.

Unfortunately, it was not possible to achieve the HT synthesis within the conditions explored. The challenges which arose were that complexes were not forming at low reaction mixture concentrations, and at higher concentrations complexes were precipitating out of solution. Future work would include performing further optimisation of reaction conditions using the HT platforms, in addition to using robots which are fully contained within a glovebox. However, due to time constraints it was decided to proceed with the activity screening of the 22 isolated pure platinum(II) salphen complexes obtained during the condition optimisation in batch.

2.4 Synthesis conclusions

Observations from the successful synthesis of the 22 platinum(II) salphen complexes were as follows: complexes with an ether-functionalised **DAB** formed more readily than with the amide-functionalised **DAB**, likely due to the different electronic properties of each **DAB** linker group. This influence on the reactivity of the **DABs** amine groups resulted in the amide functionalised complexes forming mono-substituted side products, which were not observed for the ether functionalised complexes.

Unfortunately, neither family of solubilising **DABs** were found to be suitable for HT synthesis due to the ethers air-instability, and the amides low reactivity. As a result, commercially available **DABs** were explored for the HT synthesis. However, they appeared to degrade before complete complexation at low concentrations. At higher concentrations the **DABs** were unable to provide the required solubilising effect to the complexes, resulting in zinc intermediates precipitating out of solution.

Alternative DAB linker functionalities were explored including a sulfonamide linker, however, functionalisation of the 3,4-Diaminobenzenesulfonic acid was not possible

due to an observed lack of reactivity of the sulfonic acid. Further investigation for more reactive and stable solubilising **DABs** should therefore be undertaken before reattempting the HT synthesis.

However, considering the observed **DAB** building block sensitivity of the platinum(II) salphen multi-step complexation reaction, this system may not be suitable for HT synthesis. The requirement to adjust the reaction conditions depending on the reactivity of the **DABs**, in addition to the systems sensitivity to water and oxygen with some **DABs**, makes this reaction challenging to translate to HT. As discussed in section 1.1, for an ideal HT reaction it is desirable for the reaction conditions to be tolerated across a range of building blocks, and for the reactions to be air and water stable.

Chapter 3

Photophysical and biological properties of platinum(II) salphen complexes

The platinum(II) salphen complexes have been previously reported to have desirable photophysical and biological properties which make them candidates as DNA binding photosensitisers [230, 233, 235]. Within this section, a selection of photophysical and biological properties were tested for the platinum(II) salphen complexes synthesised in chapter 2. These include absorbance and emission, ROS generation, DNA binding, and cellular localisation and toxicity. The combined ROS and DNA binding properties were also investigated to identify compounds that could produce targeted DNA damage and in turn photo-induced cytotoxicity (phototoxicity). Figure 3.1 shows a schematic overview of the work performed within this chapter. The full experimental details for each experimental method within this chapter are detailed within chapter 7. Additionally, the findings of this chapter were published in Chemistry - A European Journal in 2024.

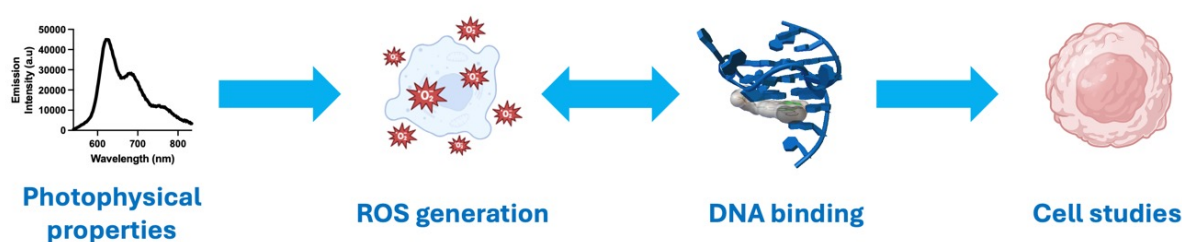


Figure 3.1. Schematic showing the photophysical, biophysical and cellular properties studied for the platinum(II) salphen series, including their absorbance and emission profiles, ROS generation, DNA binding, the interaction between ROS and DNA binding in inducing DNA damage, and finally cellular studies exploring cytotoxicity, ROS induced cytotoxicity and cellular permeability and localisation.

Platinum(II) salphen complexes absorb energy when irradiated, promoting an electron from the S_0 energy level to the S_1 level (see Jablonski diagram in figure 3.2) [137, 272]. These complexes then readily undergo intersystem crossing transitions, accessing the T_1 energy level. From there, the excited state can either decay radiatively, to give rise to emission via phosphorescence, or non-radiatively through an energy (type 2) or electron (type 1) transfer to molecular oxygen, in turn producing ROS. The excited state can also decay via other non-radiative processes such as collision and vibrational interactions.

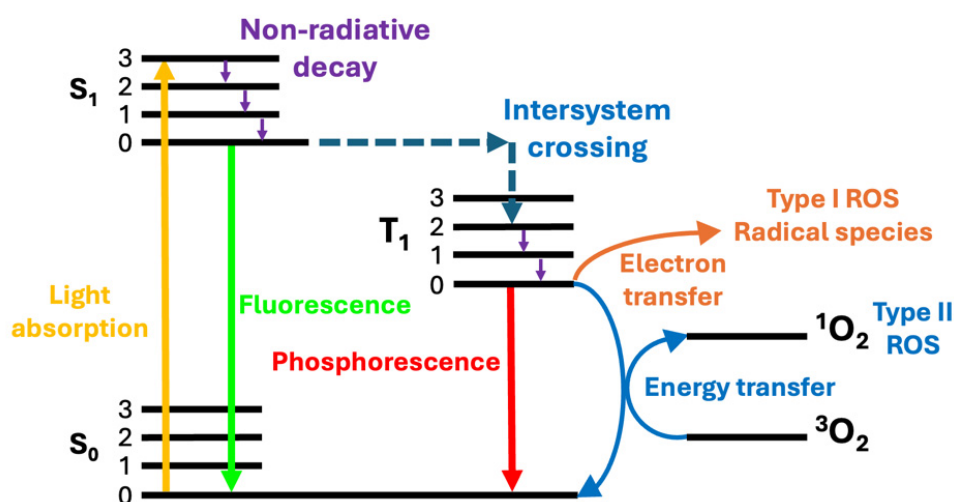


Figure 3.2. Jablonski diagram showing the energy levels of molecules involved in light absorbance from the S_0 to S_1 energy level, fluorescence from the S_1 to S_0 energy level, intersystem crossings from the S_1 to the T_1 energy level, and the three methods of energy decay from T_1 to the ground state including phosphorescence and type 1 and type 2 ROS generation. The diagram also shows the vibrational energy states within each energy level.

There are two forms of ROS generation: type 1 involves the formation of free radical species including superoxide ($O_2^{\cdot-}$) and peroxide (O_2^{2-}) anions, and hydroxyl radicals ($OH\cdot$). Type 2 involves the generation of singlet oxygen (1O_2) by exciting an electron from the ground triplet state of an oxygen molecule (O_2) to its singlet state. Singlet oxygen is extremely reactive due to its violation of Hund's rule by having spin-paired electrons in one π^* orbital, whilst another π^* orbital is unoccupied. Both radical and singlet oxygen ROS species are powerful oxidising agents which readily react with proteins, lipids and nucleic acids, leading to broad cytotoxicity [137–139].

3.1 Photophysical properties

To gain an understanding of the photophysical and photochemical properties of the platinum(II) salphen complexes, their absorbance and emission spectra were first recorded. The absorbance of all the complexes in DMSO displayed an intense peak at wavelengths below 400 nm. Complexes also showed a second absorbance peak around 500 nm. Generally, the complex modifications did not appear to have a significant impact on the peak absorbance wavelength. However, some had an influence on the shape of the absorbance spectra, as demonstrated by a selection of complexes in figure 3.3. The absorbance spectra for all complexes are shown in appendix figure A.145.

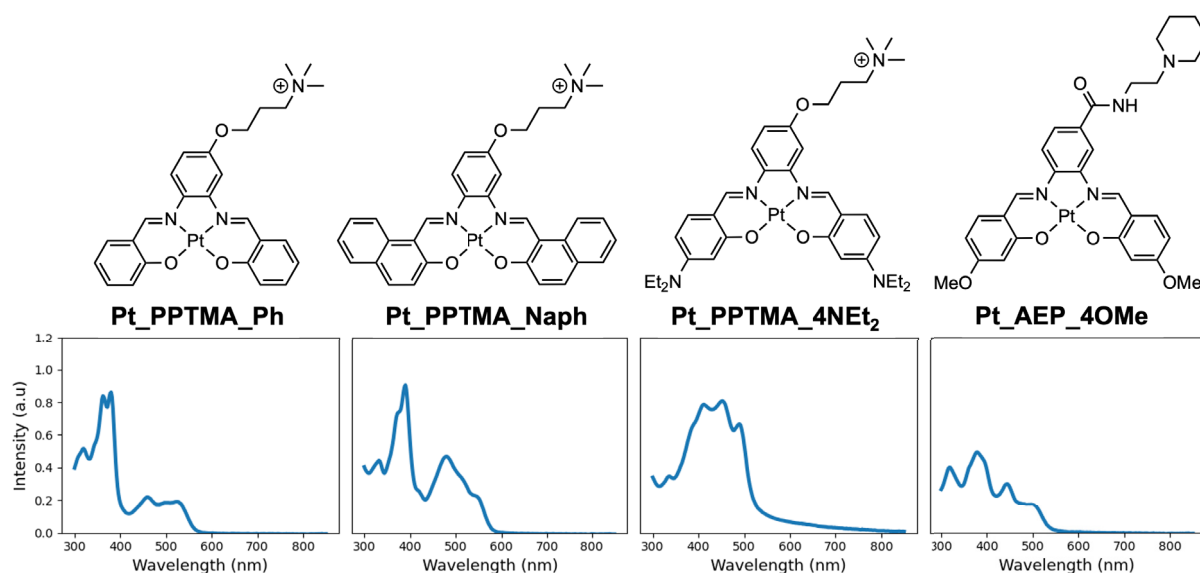


Figure 3.3. Absorbance spectra for a selection of platinum(II) salphen complexes between 300 – 850 nm at 100 μ M in DMSO. Complexes include **Pt_PPTMA_Ph**, **Pt_PPTMA_Naph**, **Pt_PPTMA_4NEt₂** and **Pt_AEP_4OMe**.

The shape of the absorbance spectra for the synthesised platinum(II) salphen complexes align with absorbance profiles for platinum(II) salphen complexes reported in the literature. Sharing the key characteristics of a major absorbance peak around 380 nm, which corresponds to the intraligand $\pi \rightarrow \pi^*$ transition, and second peak around 500 nm which originates from the metal-to-ligand $^1[\text{Pt}(5d) \rightarrow \pi^*(\text{L})]$ and intraligand $^1[\text{L}(\text{phenoxide}) \rightarrow \pi^*(\text{imine})]$ charge transfer transitions [273]. It was challenging to obtain the absorbance spectra of the complexes in water due to the high concentrations

required for a readable absorbance intensity and the low solubility of the complexes.

The complexes with the most unique spectra were **Pt_AEP_4OMe** and **Pt_PPTMA_4NEt₂**. The latter, has two additional nitrogen atoms bound directly to the lower aromatic rings of the complex which causes a bathochromic shift of the peak absorbance wavelength. **Pt_AEP_4OMe** contains an amide bond linking the aromatic core to a solubilising group which has an electron withdrawing effect on the aromatic system, relative to the electron donating ether linked complexes.

The emission of the complexes were measured in DMSO following excitation at their measured peak absorbance wavelengths (figure 3.4). The emission spectra for all complexes are shown in appendix figure A.146.

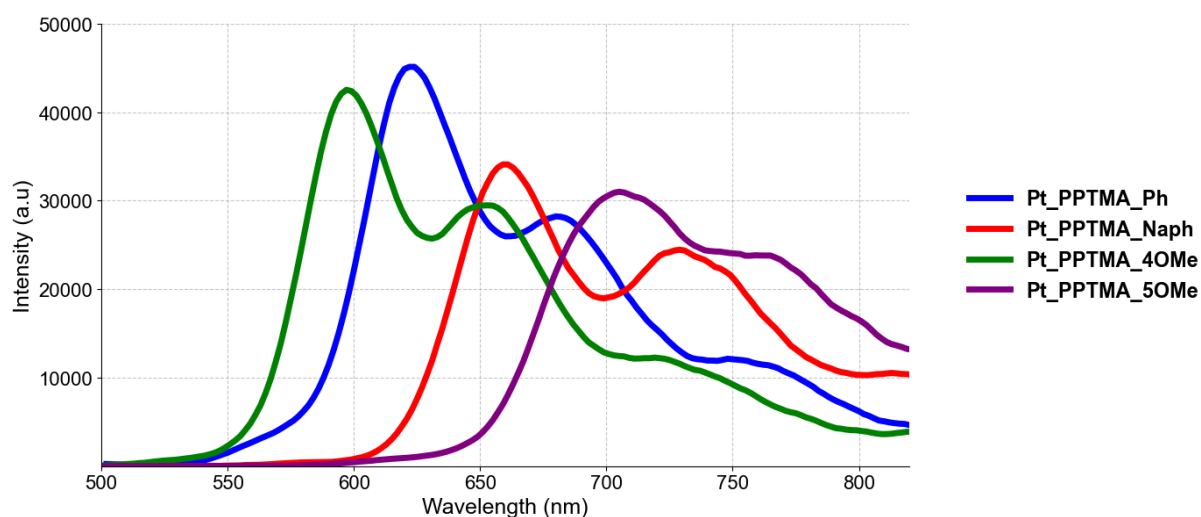


Figure 3.4. Emission spectra of the platinum(II) salphen complexes in DMSO at 2 μ M, following excitation at their recorded peak absorbance wavelengths.

The shape of the emission spectra is reasonably consistent across the series of complexes, showing a single major peak followed by one or two less intense shoulder peaks. This aligns with emission spectra of platinum(II) salphen complexes reported within the literature. The emission peaks have been found to originate from transitions between the T_1 excited state, which has mixed [L(phenoxide) $\rightarrow\pi^*(\text{imine})$] and [Pt(5d) $\rightarrow\pi^*(\text{Schiff-base})$] charge transfer character, and the S_0 ground state [273–275]. In water however, the emission of the complexes was quenched, as seen in appendix figure A.147. This is due to a process known as resonant energy

transfer, where the excitation energy of the complex is transferred to the high-frequency vibrational stretching modes of water, in addition to loss via collisions between the water molecules and the surface of the complexes, thereby losing energy by non-radiative decay [276].

The nature and position of **CBA** functionalities had a strong influence on the complexes peak emission wavelength. This was seen most strongly for **Pt_PPTMA_4OMe** and **Pt_PPTMA_5OMe**, which displayed hypsochromic (lower wavelength) and bathochromic (higher wavelength) shifts of the peak emission wavelength respectively. The differences in the peak emission wavelengths were also accompanied with a change in the physical appearance of the complexes, resulting in a wide spectrum of coloured solutions ranging from yellow to dark red, demonstrated by the **Pt_PPTMA** series in figure 3.5.

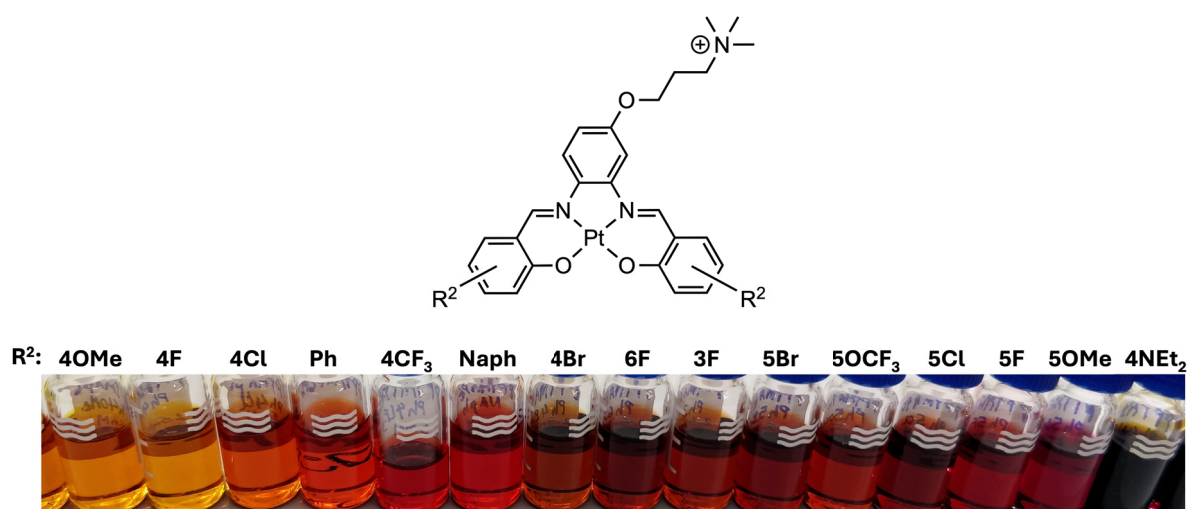


Figure 3.5. 5 mM DMSO stock solution vials of the **PPTMA** complexes.

The spectrum of colours displayed by the complexes correlated to the shift in their peak emission wavelengths, with the lighter coloured complexes having a lower peak emission wavelength. The spread of the peak absorbance and peak emission wavelengths for each complex are shown in figure 3.6, which visualises the different **CBA** and **DAB** combinations of the complexes with respect to the measured photophysical properties.

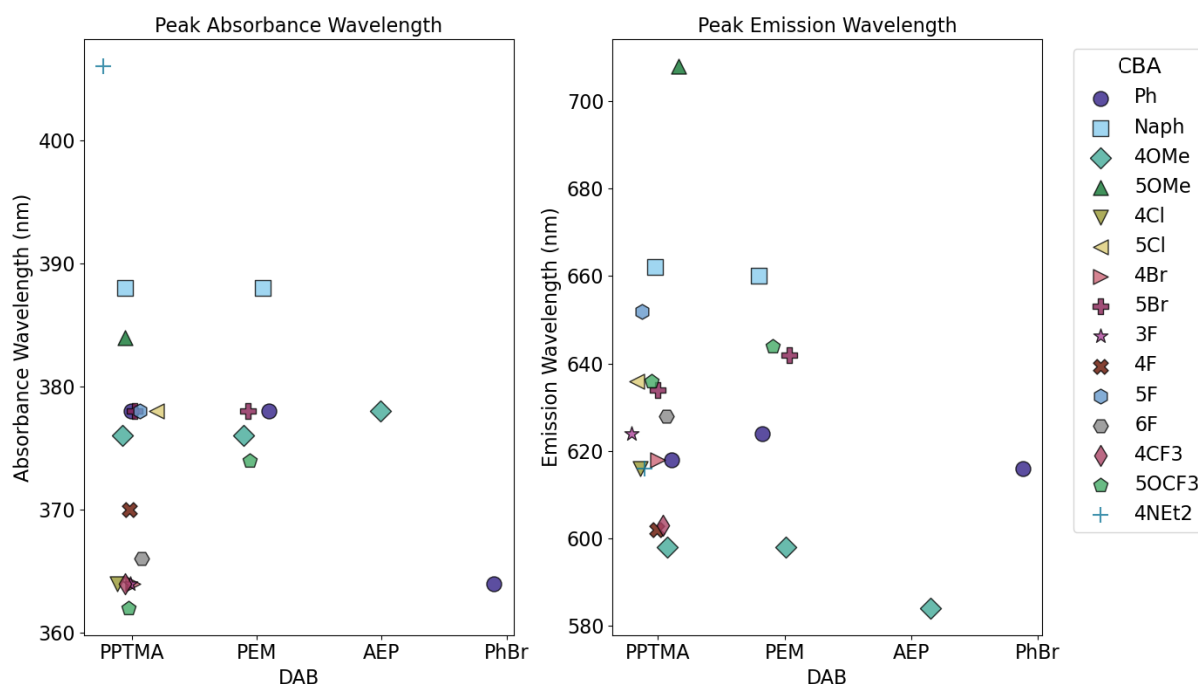


Figure 3.6. Strip plot showing the spread of absorbance (left) and emission (right) wavelengths for the platinum(II) salphen complexes, split by their **DAB** and coloured by their **CBA** components.

It can be seen that the **CBA** component of the complex has a dominant effect on the absorbance and emission wavelengths, demonstrated by complexes with the same **CBAs** and different **DABs** sharing very similar peak emission wavelengths (e.g 4OMe with **PPTMA**, **PEM** and **AEP**). The spread of the peak absorbance wavelength is much smaller than the peak emission wavelength which range 362 nm to 452 nm and 587 nm to 705 nm respectively.

Complexes containing **4OMe** and **5OMe** functionalities displayed the most extreme peak emission wavelengths, as demonstrated by the **DAB** pair: **Pt_PPTMA_4OMe** (λ_{em} = 601 nm) and **Pt_PPTMA_5OMe** (λ_{em} = 705 nm). This similar pairwise trend can be seen for **Pt_PPTMA_4F** (λ_{em} = 602 nm), **Pt_PPTMA_5F** (λ_{em} = 652 nm), **Pt_PPTMA_4Cl** (λ_{em} = 616 nm), **Pt_PPTMA_5Cl** (λ_{em} = 636 nm), **Pt_PPTMA_4Br** (λ_{em} = 618 nm), and **Pt_PPTMA_5Br** (λ_{em} = 634 nm). The difference between the pairs of peak emission wavelengths can be attributed to the electronic influence of the substituents, which can be represented by Hammett constants. Hammett constants were originally developed to quantify the effect of substituents on the linear free

energy of reactions with benzoic acid derivatives [277]. This approach has since been more broadly applied to assess the electronic contributions of substituents in other aromatic systems [278–282]. However, it should be noted that Hammett constants are limited in the context of emission wavelength as they do not fully capture the complex relationship between other contributing factors such as orbital overlap. Within this context 4-substituted complexes are considered meta substituted due to their position with respect to the oxygen of the salphen which is bound directly to the platinum, and 5-substituted complexes are considered para. Figure 3.7 shows the Hammett constants for the **CBA** within each complex against the peak emission wavelength, with the points colour and shape representing the **CBA** group and position respectively.

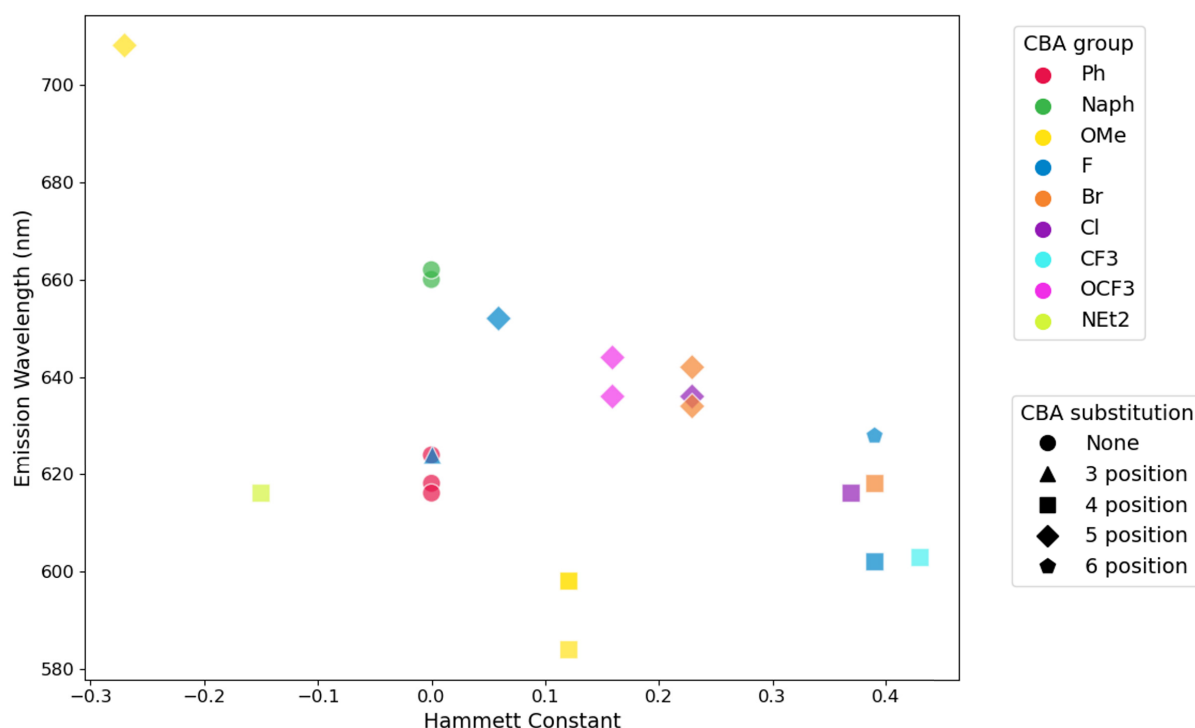


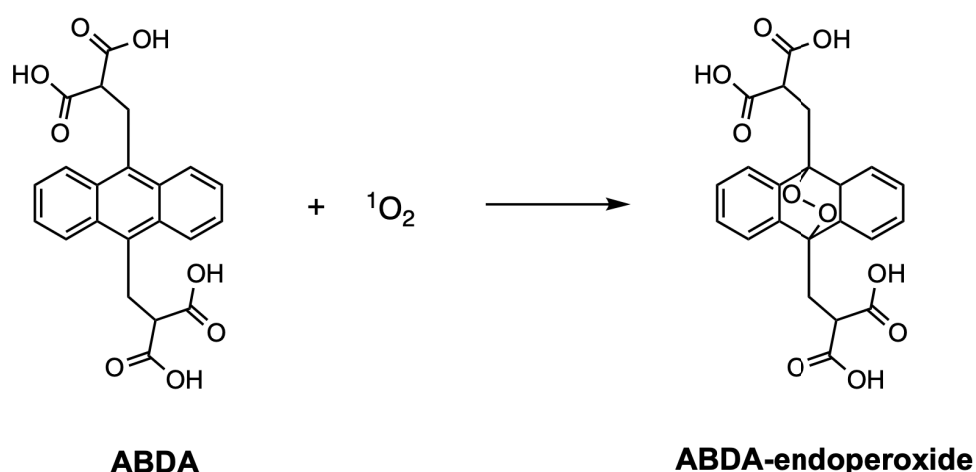
Figure 3.7. Scatter plot showing the Hammett constants for the **CBA** components against the emission wavelength of the platinum(II) salphen complexes. Data points colour and shape based on the complexes **CBA** group and position respectively. This plot does not describe the **DABs** of the complexes.

The weak negative correlation suggests that larger Hammett constants result in higher peak emission wavelengths. Complexes which display the greatest difference between their pairwise peak emission wavelengths also had the largest gaps between their Hammett constants, demonstrated by the **Pt_PPTMA** series: **4OMe** (0.12) and **5OMe**

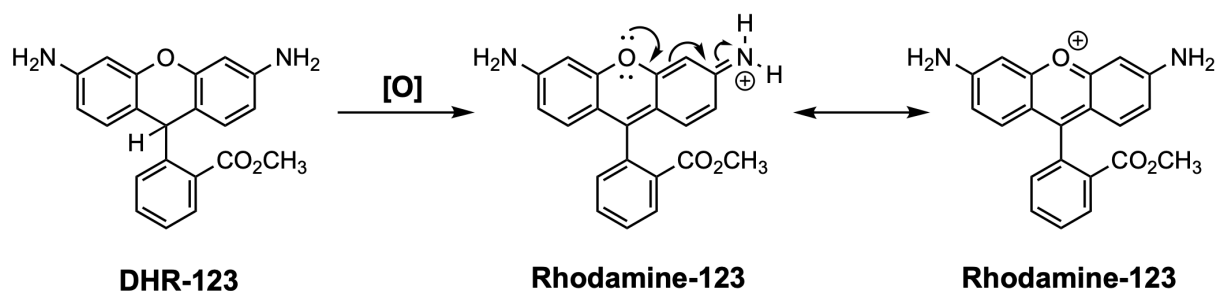
(-0.27) have the greatest difference of 0.39, followed by **4F** (0.39) and **5F** (0.06) with a difference of 0.33. More extreme Hammett constants indicate greater influence from the electron donating or electron withdrawing nature of the substituents. Interestingly the **4OMe** substitution doesn't perfectly fit the trend, showing a modest Hammett constant of 0.12, but regardless, the substituent has a large effect on the peak emission wavelength, yielding complexes with the lowest wavelengths. This is an example of where the Hammett constant alone is unable to explain the full story. This discrepancy is explained more clearly using the HOMO and LUMO energies and orbital diagrams of the complexes in chapter 4.

3.2 ROS generation

The second pathway by which excited complexes can return to ground state is via ROS generation, as shown in figure 3.2. The two forms of ROS generated can be measured using ROS specific commercially available dyes. To detect type 2 ROS, ABDA (9,10-Anthracenediyl-bis(methylene)dimalonic acid) reacts with singlet oxygen via a Diels-Alder reaction to form an endoperoxide [283] (scheme 3.1). This reaction breaks the conjugation of the molecule, resulting in a reduction in absorbance measured by UV-vis. To measure type 1 ROS, DHR-123 (Dihydrorhodamine 123), which is a non-fluorescent dye, reacts with radical ROS species to undergo an oxidation reaction to form rhodamine-123, a fluorescent molecule [284] (scheme 3.2).



Scheme 3.1. Reaction of ABDA with type 2 ROS singlet oxygen to form the endoperoxide.



Scheme 3.2. Oxidation of DHR-123 in the presence of type 1 ROS species to generate rhodamine 123.

The ROS detection assays are compatible with HT screening as they were performed in 96 well plates which can be prepared using liquid handling robots [64]. Within each well there was a mixture of the dye and the platinum(II) salphen complex, which were then irradiated using a 457 nm 96-LED array (8 mA, 2 mW/cm²) at 30 second intervals between absorbance or emission measurements using a plate-reader. This LED wavelength was selected as all complexes displayed a broad absorbance peak from 400 nm – 550 nm.

Within the ROS assays [Ru(bpy)₃]²⁺, a known photosensitiser, was used as a ROS generation standard [285–288]. The ROS generation of the platinum(II) salphen complexes were compared to the ROS generation of [Ru(bpy)₃]²⁺. ROS generation was quantified by the gradient of absorbance reduction (ABDA), or switch-on of emission (DHR-123) for each assay, divided by the gradient for [Ru(bpy)₃]²⁺, to give a relative ROS generation percentage (complexes which generated more ROS than [Ru(bpy)₃]²⁺ obtained > 100 % relative ROS generation). Figure 3.8 shows each complexes type 1 (green) and type 2 (pink) ROS generating abilities, ordered by the complexes combined performance.

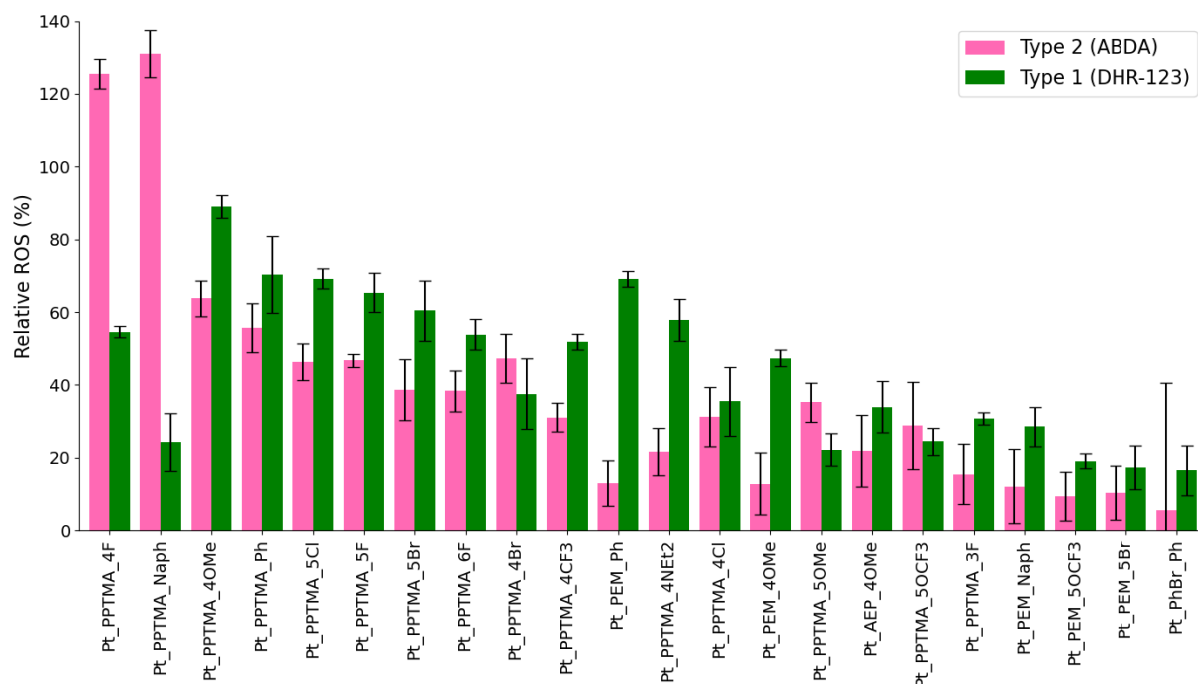


Figure 3.8. Bar chart of the percentage ROS generation of each platinum(II) salphen complex relative to $[\text{Ru}(\text{bpy})_3]^{2+}$. Pink bars: Singlet ROS generating ability measured using ABDA ($n = 3$). Green bars: Superoxide ROS generating ability measured using DHR-123 ($n = 3$). Full ROS generation percentages are detailed within appendix tables A.3 and A.4.

Multiple complexes demonstrated contrasting performances across the type 1 and type 2 ROS generating assays. The top type 2 ROS generating complexes showed a much lower type 1 ROS generation. **Pt_PPTMA_Naph** scored an impressive 131 % in the type 2 ROS assay, but only 24 % in the type 1 ROS assay. The complex with the highest combined overall ROS generation ability was **Pt_PPTMA_4F** which scored 125 % in the type 2 assay, and 54 % in the type 1 assay. For the type 1 assays **Pt_PPTMA_4OMe** scored 89 %, the highest of the series, as well as 64 % in the type 2 assay, yielding the 3rd overall best ROS generating complex, with the most balanced performance across both assays.

To assess the presence of trends between the complex structures and their ROS generating abilities, figure 3.9 contains strip plots showing the ROS generation of each complex with the **CBA** and **DAB** components visualised.

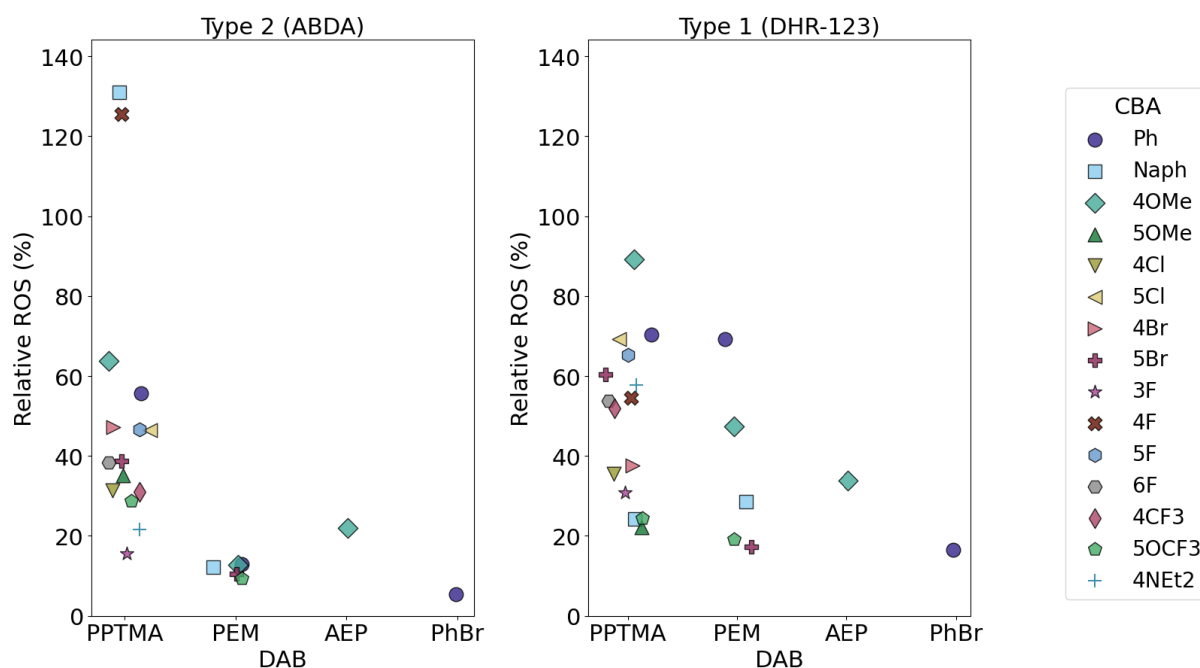


Figure 3.9. Strip plots showing the type 2 (ABDA, left) and type 1 (DHR right) ROS generation ability of the platinum(II) salphen complexes relative to $[\text{Ru}(\text{bpy})_3]^{2+}$, split by their **DAB** and coloured by their **CBA** components.

From these plots it can be seen that the **DAB** of the complexes had a strong influence on ROS generation. Complexes containing the **PEM DAB** appeared to quench type 2 ROS generation, with the highest scoring complex, **Pt_PEM_Ph**, scoring 13 %, and the lowest, **Pt_PEM_5OCF₃**, scoring 9.4 %. In contrast, these complexes performed more evenly within the type 1 assay with **Pt_PEM_Ph** scoring 69 % and **Pt_PEM_5OCF₃** scoring 19 %. The worst overall complex was **Pt_PhBr_Ph** which performed poorly across both assays. This could have been due to its poor solubility within the aqueous media of the experiment, and therefore it is likely that its ROS-generating ability was not measured accurately. When assessing the influence of the **DAB** in the type 1 ROS assay, the performance of the **4OMe** complexes with varying **DABs** displayed a trend of **PPTMA > PEM > AEP**. The **DAB** trends are less clear with other **CBA** components due to a lack of representation.

The effect of the **CBA** components of the complex on ROS generation are also less clear. Within the **PPTMA** series, the trend of the best performing complexes within the type 2 assay were **Naph > 4F » 4OMe > Ph**, and within the type 1 assay were

4OMe > Ph > 5Cl > 5F. Interestingly, complexes which perform best in one ROS assay, appeared to perform very poorly in the other, suggesting that particular complex structures may prefer one ROS generation pathway over another. However, it is not clear which structural elements influenced this. It was expected that identifying structural trends in ROS generation would be challenging, as ROS generation is a highly complex process which has previously been shown to not be easily predicted based on the nature of substituents [64].

3.3 DNA binding

As discussed in the introduction, the polyaromatic planar structure of the platinum(II) salphen complexes make them well suited for binding to DNA. However, within cells there are many potential DNA targets, including a diverse range of G4 structures. It is therefore important for complexes to have specific binding for the target of interest to prevent off target side effects. Therefore, the screening of multiple DNA targets is required to identify the most selective complexes.

The assay used to measure the platinum(II) salphen complexes DNA binding takes advantage of the complexes emissive properties. The emission of platinum(II) salphen complexes is quenched in aqueous media, however, when bound to DNA the surface of the complex is protected from water. This prevents non-radiative decay and leads to an emission switch-on. Therefore, by measuring the change in emission with increasing concentrations of DNA, the binding affinity can be calculated. To calculate the binding affinity the change in emission intensity is plotted against the concentration of DNA and fitted to the GraphPad equation for one site specific binding [29, 34, 233, 289] (equation: 3.1).

$$Y = \frac{B_{max} \cdot X}{K_d + X} \quad (3.1)$$

- Y: Emission intensity
- X: DNA concentration
- B_{max} : Maximum emission response
- K_d : Equilibrium dissociation constant

When X is equal to K_d , Y is equal to half of B_{max} , allowing the determination of binding affinity (K_a) as shown in equation 3.2:

$$K_a = \frac{1}{K_d} \quad (3.2)$$

Figure 3.10 below shows a representative emission titration and binding curve for **Pt_PPTMA_4OMe** against four unique G4 sequences and duplex DNA.

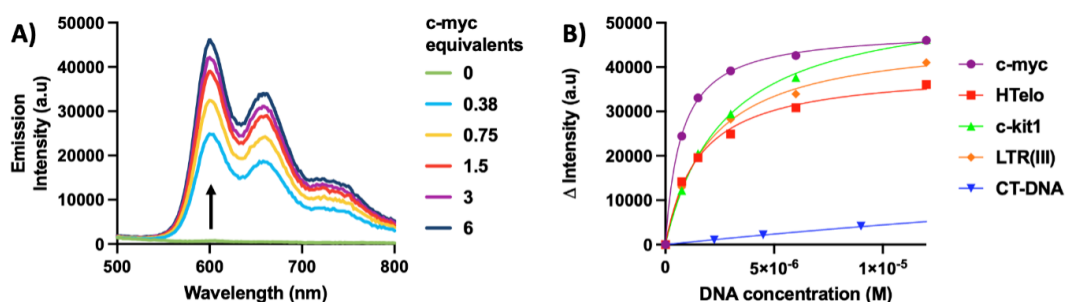


Figure 3.10. A) Emission spectra showing the switch-on of emission for **Pt_PPTMA_4OMe** with the addition of increase equivalents of *c-myc* G4 DNA. B) Plot showing the change in emission intensity with increasing DNA concentrations for four G4 DNA sequences; *c-myc*, HTelo, *c-kit1*, and *LTR(III)*, and duplex DNA (CT-DNA). Curves are fit to the one site specific binding equation (3.1).

The G4 sequences selected for study represent a range of potential G4 topologies. Promoter G4s *c-myc* and *c-kit1* occur as parallel and anti-parallel topologies respectively. Telomeric G4s HTG4 and HTelo occur as hybrid 3+1 and hybrid-2 topologies respectively. Additionally, LTR(III), a G4 found within the long terminal repeat of the HIV virus, was investigated for its unique combined topology of G4 and duplex hairpin DNA topology [290]. The biological relevance of these sequences was described in chapter 1, and the precise sequences and topologies can be found within chapter 7.

Preliminary emission titration assay optimisation was performed for the platinum(II) salphen complexes using HTG4 G4 DNA. The optimised protocol was then translated into a HT assay using liquid handling robots to screen the four additional G4 DNA sequences: *c-myc*, HTelo, *c-kit1*, and LTR(III), as well as duplex DNA (CT-DNA).

3.3.1 Emission titration assay optimisation

To ensure the one site specific binding equation (3.1) measures binding affinity accurately, the switch-on of emission must reach saturation indicating that all complex is bound to DNA (B_{max}). It is also important to have many data points around the point of inflection: the DNA concentration at which half of all available binding sites are occupied, to enable an accurate calculation of the K_d . The focus of the initial protocol optimisation was to select the appropriate concentrations of DNA to use within the assay.

The complex concentration used within the assay was 2 μM as it provided a reasonable emission signal when bound to DNA. Therefore 12 μM of DNA (6 equivalents) was initially selected as the top DNA concentration of the assay. Assays were designed to use one row per 96 well plate per test, screening 11 DNA concentrations and one well containing only the complex for a zero-point emission measurement. The plate was prepared using serial dilution to efficiently access a range of DNA concentrations. However, starting with a single DNA concentration and diluting down resulted in large gaps between the lower concentrations, which coincidentally were the concentrations closest to the point of inflection. Therefore, two concentration starting points, 12 μM and 4 μM were used to facilitated the inclusion of more data point around the point of inflection, which was found to typically be between 0.5 – 2 μM . Table 3.1 illustrates the plate layout for the serial dilution with two DNA starting point concentrations shown in blue (starting at 12 μM) and green (starting at 4 μM).

Well	1	2	3	4	5	6	7	8	9	10	11	12
[DNA] (μM)	0	0.13	0.25	0.50	0.75	1.0	1.5	2.0	3.0	4.0	6.0	12

Table 3.1. Map of 96 well plate columns for the emission titration protocol and the final concentration of DNA within each well.

The preliminary HTG4 emission titrations were performed with a sample of the platinum(II) salphen complexes, and showed only a small variation in binding affinities across the complexes, with all complexes achieving K_a values within the 10^5 - 10^6 range.

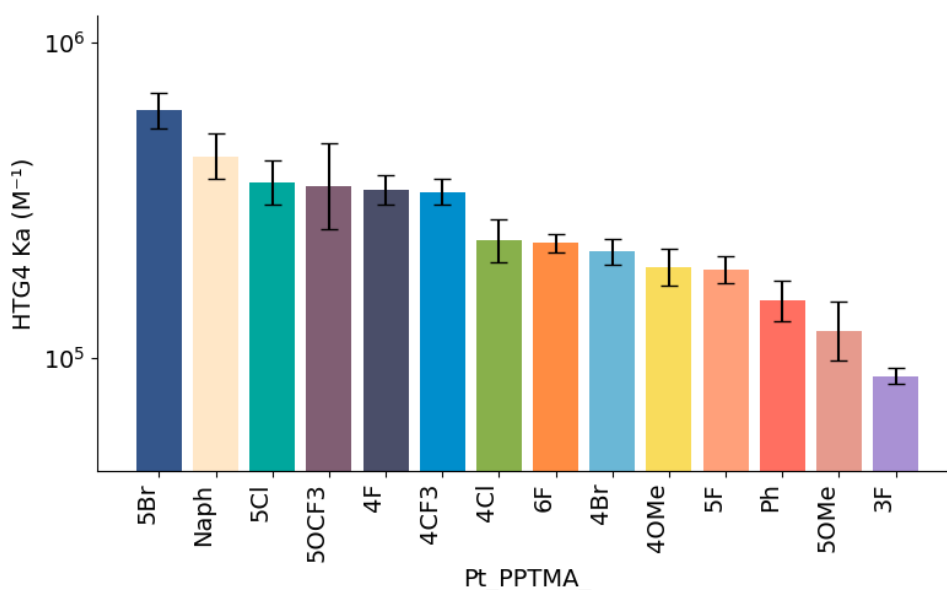


Figure 3.11. Bar chart showing the HTG4 binding affinity (K_a) for the **PPTMA** solubilised platinum(II) salphen complexes obtained using hand prepared assays. $n = 3$. Full K_a values are detailed within appendix table A.5.

Pt_PPTMA_5Br had the highest K_a at $6.09 \times 10^5 M^{-1}$. The binding affinity of complexes with halogens in the 5 position decreased with their increasing electronegativity (**Br** > **Cl** > **F**). The complex with an electron donating group in the 5 position, **Pt_PPTMA_5OMe**, had one of the lowest binding affinities of the series at $1.22 \times 10^5 M^{-1}$. This would suggest that more favourable binding occurs with weak electron withdrawing groups in the 5 position of the **CBA**. This trend was further supported with the binding affinity of **Pt_PPTMA_5OCF₃** being three times greater than **Pt_PPTMA_5OMe** due to the weaker electron donating effect resulting from the electronegative **CF₃** group bound to the oxygen. Additionally, it appeared that the presence of a larger aromatic surface within the **CBA** component resulted in an increased binding affinity, demonstrated by **Pt_PPTMA_Naph** having a binding affinity more than 3 times greater than **Pt_PPTMA_Ph**. This observation could be attributed to the increased surface area for π - π stacking interactions.

3.3.2 Measuring DNA binding in high throughput

With the emission titration assay optimised, a protocol using the OT2 liquid handling robot was developed. Using this protocol, full plates containing 8 experiments could

be prepared in under 12 minutes. Before performing the HT screen, the accuracy of the serial dilution performed by the OT2 was tested using the absorbance spectra of Acid Violet 17. The expected linear increase in absorbance with increasing dye concentration was observed across all rows with an R^2 of 0.99 (figure 3.12).

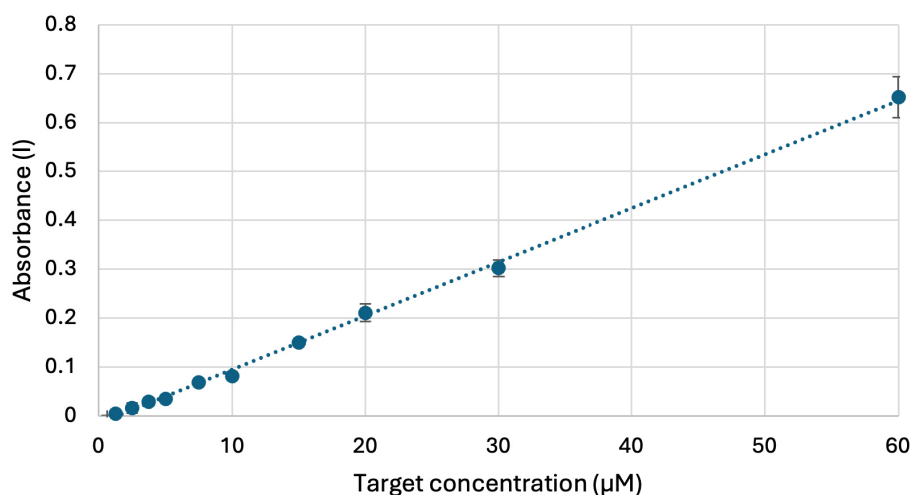


Figure 3.12. Graph showing the expected linear pattern of absorbance intensity of Red Violet 17 at the target concentrations which were prepared by the OT2. $n = 3$, $R^2 = 0.99$.

Following confirmation that the OT2 was reliably pipetting the target volumes, the emission titration protocol was performed using HTG4 DNA with the full platinum(II) salphen series (figure 3.13).

When comparing the measured binding affinities from the plates prepared using the OT2 to assays prepared by hand, it was seen that the order of complex binding affinity was preserved, with **Pt_PPTMA_3F** being the weakest binder of the **PPTMA** series (of those which displayed any binding affinity), and **Pt_PPTMA_5Br** being the strongest binder of the series. **Pt_PPTMA_4NEt₂**, **Pt_PEM_5OCF₃** and **Pt_PEM_Naph** displayed no switch-on of emission when binding to HTG4 DNA and therefore had no measurable binding affinity.

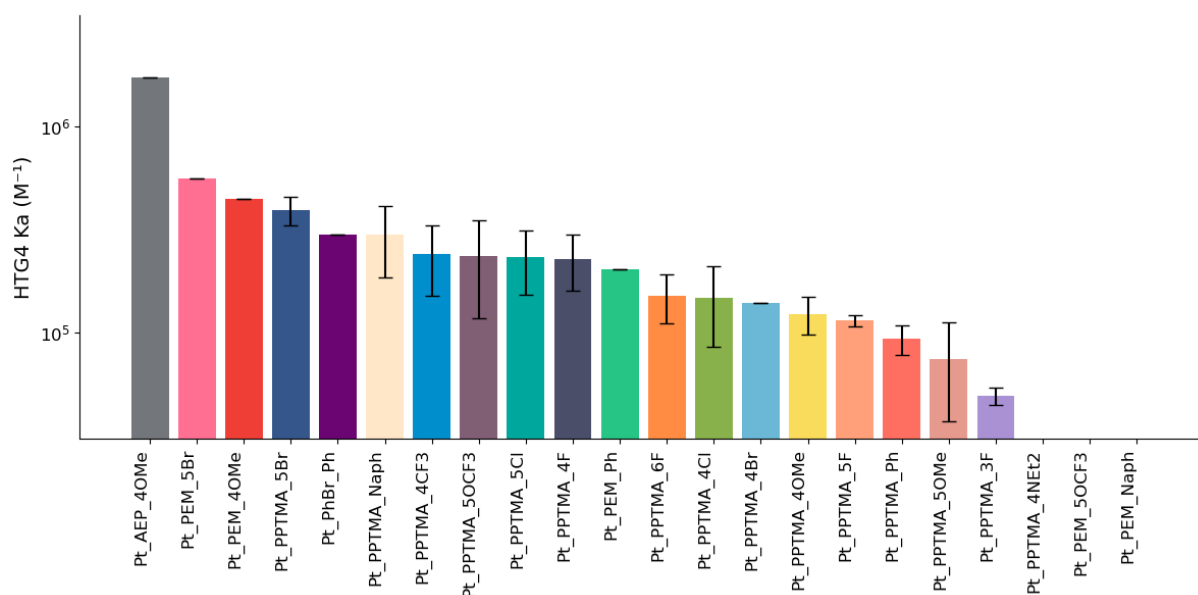


Figure 3.13. Bar chart showing the HTG4 binding affinity (K_a) for the platinum(II) salphen complexes obtained using OT2 prepared assays. $n = 3$. Full K_a values are detailed within appendix table A.6.

Having established the emission titration protocol using the OT2, the DNA binding affinities were measured for the full series of 22 platinum(II) salphen complexes against CT-DNA (duplex) and the four G4 topologies: *c-myc* (parallel), HTelo (hybrid), *c-kit1* (anti-parallel), and LTR(III) (Hybrid 3+1 with duplex stem-loop). However, using the initial protocol design in triplicate would have required 3780 data points, and therefore forty 96 well plates. To ensure optimal resource usage, the assay was redesigned to duplicate measurements using 6 data points per experiment, reducing the full screen down to 1320 data points, and only required fourteen 96 well plates. The binding affinities for all 22 platinum(II) salphen complexes with all 6 DNA structures are shown in figure 3.14. The final protocol was additionally used within institutional and international collaborations, resulting in contributions to two publications within the European Journal of Inorganic Chemistry (2023), and the Journal of the American Chemistry Society (2023).

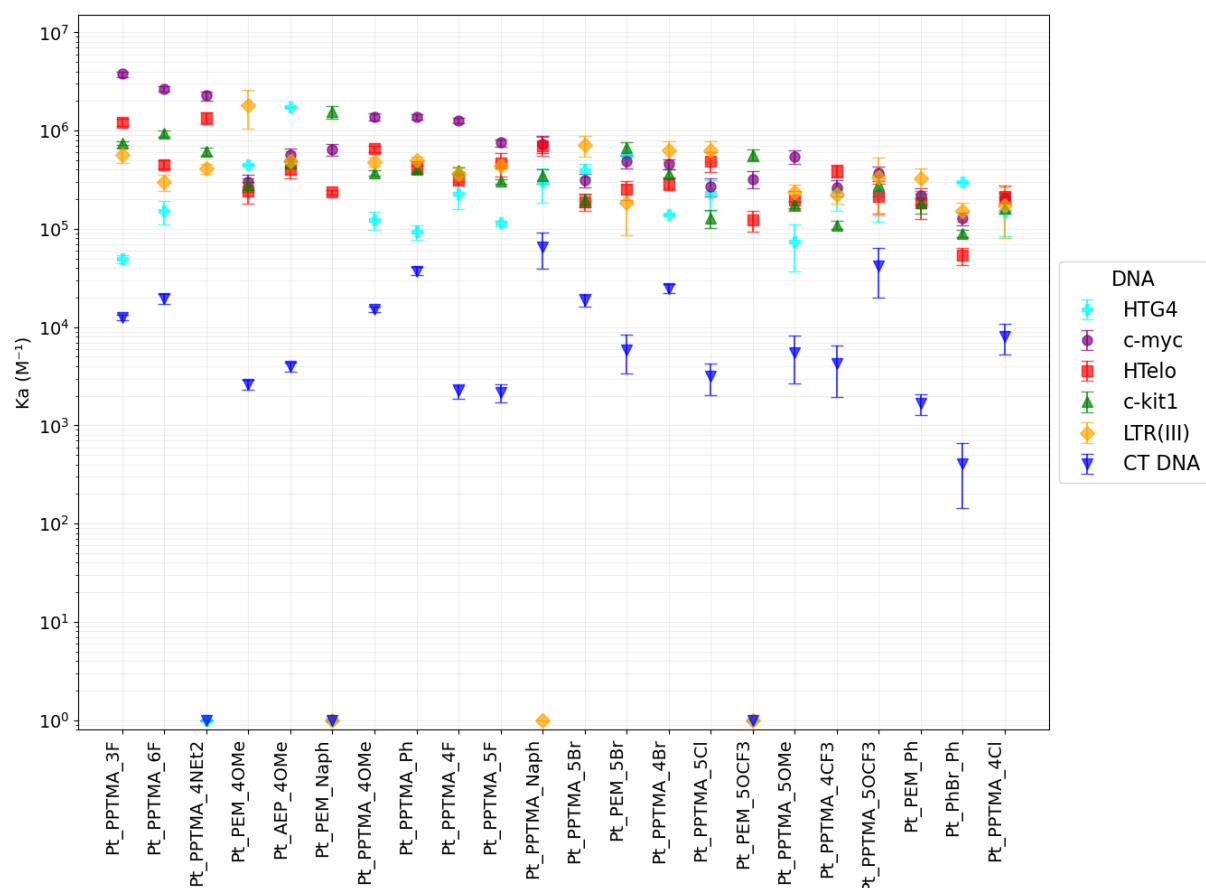


Figure 3.14. Multi series scatter plot showing the binding affinity (K_a) for the platinum(II) salphen complexes against six DNA targets: HTG4, *c-myc*, HTelo, *c-kit1*, LTR(III) and CT-DNA, obtained using OT2 prepared assays. $n = 2$.

The most notable result of this screen was that each complex had a unique DNA binding preference across the 6 DNA topologies. All complexes displayed a binding preference for G4 DNA over duplex DNA, this is likely due to the complexes preferring π - π stacking interactions with the G4 quartets over intercalation or groove binding interactions with the CT-DNA. Of the G4 sequences there was a general preference for *c-myc* DNA, likely due to its more accessible G-tetrad in comparison to the others which have flanking bases protruding into the space above the quartet where complexes bind. This is visualised further within section 3.3.3 where the X-ray crystal structures of the DNA sequences are used within docking simulations.

The structural factors which influence the complexes binding affinity were unique to each G4 topology. This can be seen more clearly in figure 3.15 which shows the binding affinities for each DNA topology, with complexes separated by their **DAB** components

and coloured by the **CBA** components.

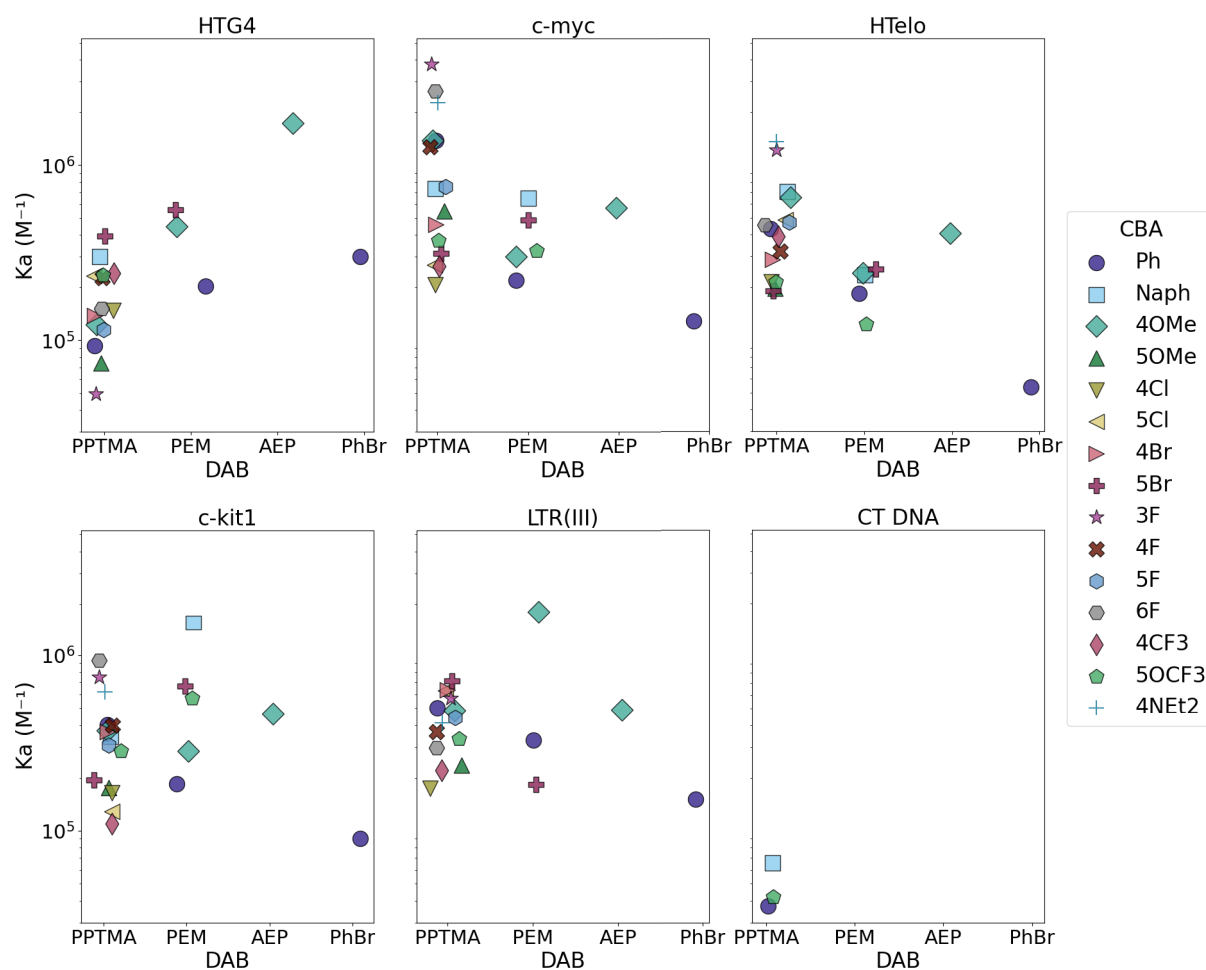


Figure 3.15. Strip plots showing the binding affinities (K_a) for each DNA target (title) for the platinum(II) salphen complexes, split by their **DAB** and coloured by their **CBA** components. Binding affinities below 3.0×10^4 were not included within these plots.

For HTG4 binding, **PPTMA DABs** exhibit a lower binding affinity compared to complexes with the same **CBAs**. This can be seen most distinctly with **Pt_AEP_4OMe** which has the highest binding affinity for HTG4, an order of magnitude greater than **Pt_PPTMA_4OMe**. In contrast, **PPTMA** appeared to increase the binding affinity to HTelo. The effect of **PPTMA** is less clear with *c-myc* as although multiple **PPTMA** complexes had the highest binding affinity, the cross-**DAB** derivatives appeared to have similar binding affinities.

The **CBA** which resulted in the highest binding affinity was unique for each DNA topology. Overall, **Naph** consistently provided high binding affinities, likely due to its

increased aromatic surface resulting in greater π - π interactions between the complex and the exposed G-tetrads of the G4s. The **CBAs** which resulted in the best binding affinities for HTG4 DNA were **5OCF₃** and **4OMe**, suggesting that adding electron donating hydrogen bond acceptor groups to the complex favoured binding to HTG4. For *c-myc* DNA the best **CBA** groups were **3F** and **6F**, which have very little influence on the sterics of the complex, but add strong electron withdrawing groups close to the metal centre. Conversely, the next best **CBA** was **4NEt₂** which was the largest functional group of the series, a hydrogen bond acceptor, and an electron donor in the 4 position. These were followed by **4OMe**, an additional electron donating hydrogen acceptor in the 4 position, **Ph**, the unsubstituted complex, and **4F**, a small electron withdrawing group. Overall, this suggested that the presence of hydrogen bond acceptors increased the binding affinity to *c-myc* G4s, with a preference for smaller groups, unless they are electron donating and in the 4 position. This observation was repeated for HTelo DNA with **4NEt₂**, **3F**, **Naph** and **4OMe** providing the best binding affinities. Similarly, *c-kit1* DNA had the best binding affinities with **Naph**, **6F**, **3F**, followed by **5Br** which was also seen to result in more favourable binding to HTG4. The spread of binding affinities were much smaller for LTR(III) DNA with **Pt_PEM_4OMe** having the best binding affinity of the series. The other **4OMe** complexes also demonstrate a reasonable binding affinity, but not definitively better than other complexes with the same **DABs**, making the trends of binding affinities for LTR(III) less clear.

As previously mentioned, some complexes demonstrated no switch-on of emission when titrated with select DNA sequences. **Pt_PPTMA_4NEt₂** did not appear to bind to HTG4 nor CT-DNA. **Pt_PEM_Naph** did not appear to bind to HTG4, LTR(III) nor CT-DNA. **Pt_PPTMA_Naph** did not appear to bind to LTR(III). Finally, **Pt_PEM_5OCF₃** did not appear to bind to HTG4, LTR(III) nor CT-DNA. These combinations of complexes and DNA sequences contain complexes with bulkier substituents, and DNA binding sites which are less accessible. This could explain why no switch-on of emission intensity was detected. However, this does not necessarily mean that the complexes do not bind to the DNA sequences, they could instead be participating in groove binding which does not shield the aromatic surface from water, resulting in non-radiative decay. This could be confirmed by performing a groove binder displacement assay [291].

3.3.3 Docking simulations of the platinum(II) salphen complexes with G4 DNA targets

To further investigate the platinum(II) salphen complexes DNA binding, docking simulations were performed with the DFT geometry optimised platinum(II) salphen complexes and x-ray crystal structures of previously explored DNA targets, using AutoDock4 to predict the binding modes [292, 293]. Docking predicts binding interactions and binding poses which are then scored to assess their relative strength. There are a number of factors which simulations need to consider to enable accurate prediction of binding interactions. These include the dynamic properties of the ligand and receptor, the different types of binding interactions (van der Waals, hydrogen bonding, π -stacking interactions, and ionic interactions), and the effects of water within the system [294].

Due to this complexity, docking programs are simplified to assess ensembles of different binding poses to account for dynamic systems, use potential energy functions and force fields to assess binding in place of computationally expensive quantum mechanical calculations, and do not account for water in the system. Following the initial pose generation, the docking program then assigns a score to rank each binding pose. The key scoring functions consider force-fields of direct surface interactions between the ligand and receptor, knowledge based statistical functions of known ligand-receptor interactions, and empirical scoring functions for the electronic configuration of surfaces and their bonding, in addition to rotational and translational degrees of freedom in balancing equations [295, 296].

Docking simulations were performed for a diverse selection of the platinum(II) salphen complexes to assess the effect that the **CBA** and **DAB** components had on the complexes interactions with DNA. The DNA sequences investigated were *c-myc* (parallel), HTelo (hybrid-2), *c-kit1* (dimeric anti-parallel), HTG4 (hybrid 3+1), LTR(III) (hybrid 3+1 with stem-loop), and duplex DNA.

The general binding motifs of complexes with duplex DNA was intercalation, and with G4s was π - π stacking interactions with the terminal G-tetrads. These predicted motifs are consistent with binding poses determined by X-ray crystallography [208, 297].

Except in the case of HTG4 and *c-kit1* where the G-tetrads had limited accessibility and so docking predicted that complexes were participating in more groove binding interactions. This may have led to complexes having lower recorded binding affinities as groove binding does not protect the complexes from water, and therefore non-radiative decay. The results of the docking simulations further supported what was seen within the experimental results: the nature and position of the **DAB** and **CBA** groups influence the binding interactions between the complex and DNA targets. Figure 3.16 shows the complex **Pt_PPTMA_Ph** docked with each DNA target of interest.

Within the docking simulations a range of interactions were predicted. For HTelo G4 DNA a bonding interaction was predicted between the linker groups ether oxygen and the phosphate backbone which was not observed with any other DNA targets. Whereas for HTG4, *c-kit1*, and LTR(III), simulations predicted bonding interactions between the platinum coordinated oxygen group and the phosphate backbone. For *c-myc* and duplex DNA no additional bonding interactions were predicted. **DABs** which included a protonated amine ring at physiological pH (7.4), such as those solubilised with tertiary amine groups (**PEM** and **PEP**), were predicted to participate in hydrogen bonding interactions with the DNA targets, shown below with *c-myc* G4 DNA. Whereas the **PPTMA** complexes with the quaternary triethylamine were unable to make the same interactions due to the lack of a protonated nitrogen or free lone pairs.

The nature of the linker group also influenced the potential interactions between the complex and the DNA targets. For example, the oxygen within the ether linker was predicted to behave as a hydrogen bond acceptor (**Pt_PPTMA_Ph** with HTelo in figure 3.16), and the NH group in the amide linker was predicted to behave as a hydrogen bond donor (figure 3.17). The key differences between the **CBA** components was that **CBAs** which contained oxygens, such as **OMe** and **OCF₃**, were predicted to participate in additional bonding interaction with the phosphate backbone of HTG4 (figure 3.18).

Docking also helped to explain why no binding was detected for the **Naph CBA** complexes with LTR(III). Generally, the complexes bound to LTR(III) within the cavity between the G4 and the duplex loop of the sequence (shown by **Pt_PPTMA_Ph** in figure 3.16). Whereas with the **Naph CBA**, the complexes could no longer fit within the LTR(III) binding cavity as shown in figure 3.19 with **Pt_PPTMA_Naph**. This resulted

in complexes favouring groove binding modes, which do not protect the complex from water, leading to non-radiative decay pathways and no detectable switch-on of emission.

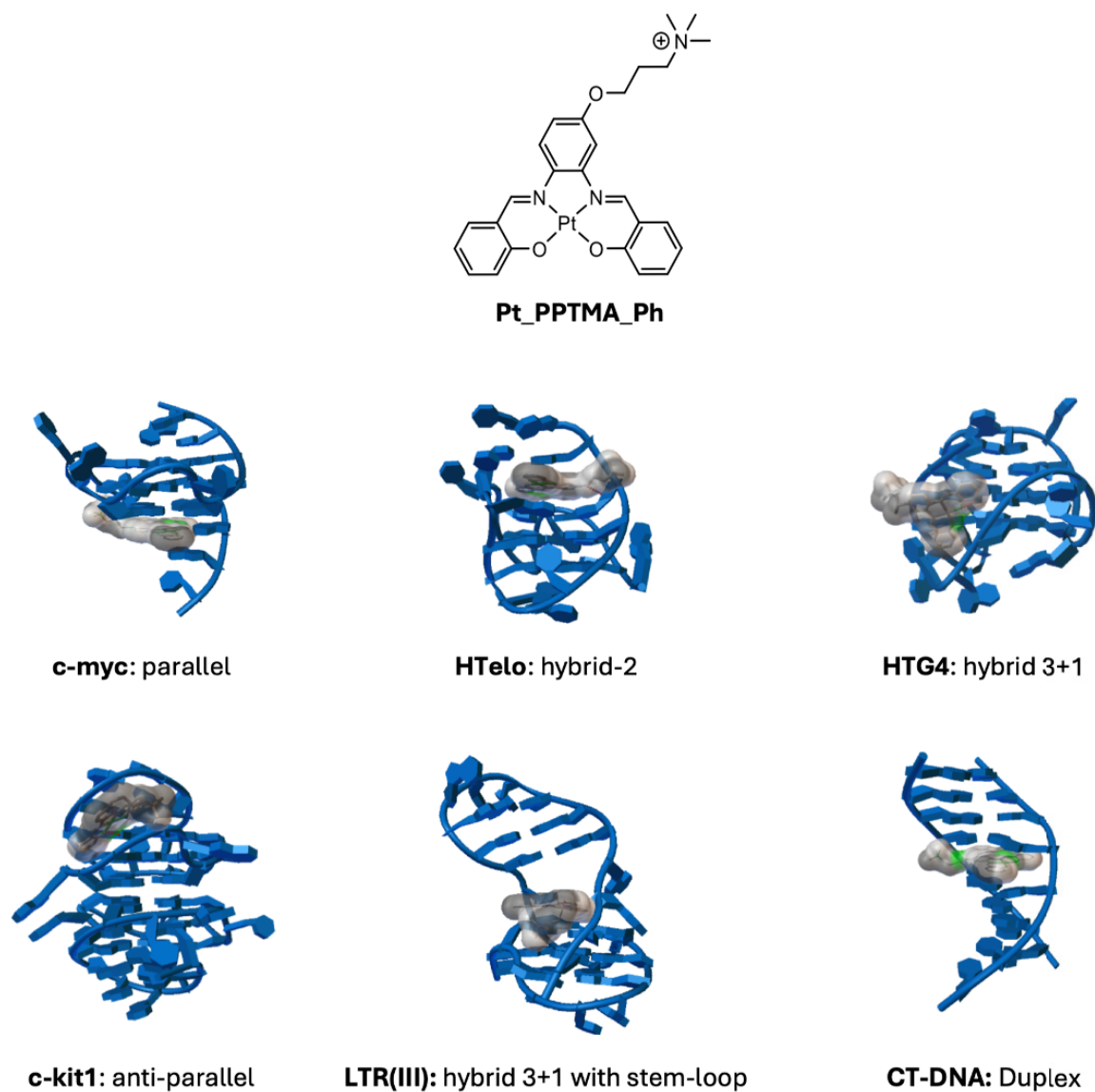


Figure 3.16. Top docking poses for the crystal structure of each DNA sequence used for the binding assays with Pt_PPTMA_Ph. Generated using AutoDock4. Predicted interactions are shown by red dotted lines.

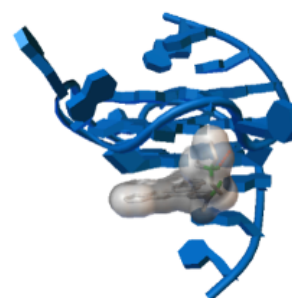


Figure 3.17. Docking poses for the crystal structure of *c-myc* with **Pt_X_Ph** where X is each of the **DABs** proposed for synthesis in chapter 2: **PEP**, **PEM**, **PPTMA**, **AEP**, **AEM** and **AEDEA**. Generated using AutoDock4.

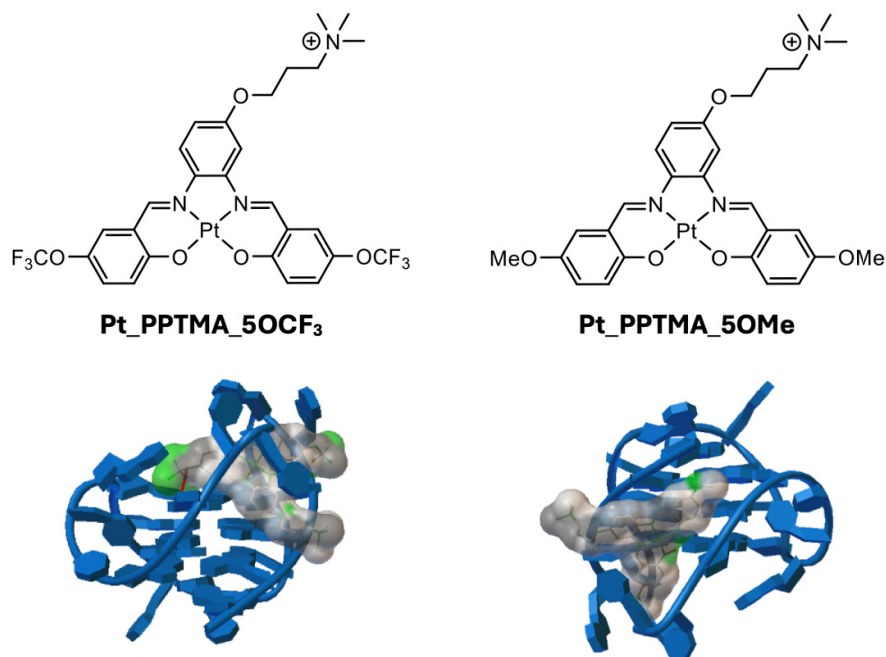


Figure 3.18. Docking poses for the crystal structure of HTG4 with **Pt_PPTMA_5OCF₃** and **Pt_PPTMA_5OMe**. Generated using AutoDock4. Predicted interactions are shown by red dotted lines.

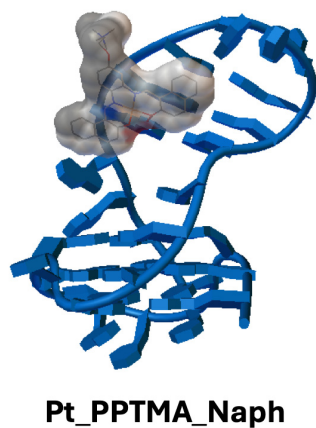


Figure 3.19. Docking pose for the crystal structure of LTR(III) with **Pt_PPTMA_Naph**. Generated using AutoDock4.

The possibility of using binding scores from docking in place of experimental measurements was also explored with the aim of building a larger computational

dataset to train ML models to predict G4 binding of untested complexes. However, the order of best to worst predicted binders were found to be inconsistent with the experimental binding results (table 3.2) and therefore docking was instead only used as a qualitative assessment of binding modes.

Pt_PPTMA_R²	Measured K_a (M^{-1})	Docking score (kcal/mol)	Docking rank order
5Br	6.09×10^5	-9.86	3
Naph	4.36×10^5	-10.2	1
5Cl	3.59×10^5	-9.32	5
5OCF₃	3.51×10^5	-8.54	14
4F	3.40×10^5	-9.02	7
4CF₃	3.36×10^5	-8.90	8
6F	2.31×10^5	-8.66	12
4Br	2.17×10^5	-9.99	2
4Cl	2.35×10^5	-9.65	4
4OMe	1.93×10^5	-8.81	10
5F	1.90×10^5	-8.58	13
Ph	1.52×10^5	-8.87	9
5OMe	1.22×10^5	-9.18	6
3F	8.76×10^4	-8.74	11

Table 3.2. Table showing the measured HTG4 K_a (in descending order) for **PPTMA** series, compared to the computationally calculated binding score and the ranked docking score.

The results in table 3.2 show that docking was unable to preserve the true order of binding affinities, likely due to the fact that the differences between the binding affinities of many of the complexes was very small, and docking only provides approximate values. Docking simulations were also unable to provide binding scores which were significantly different (> 1 kcal/mol) for complexes which were determined to have binding affinities with an order of magnitude difference (**5Br** vs **3F**). Furthermore, the differences between the calculated docking scores were within the standard method error of 2.5 kcal/mol, meaning that the difference in binding scores between the complexes was indistinguishable [298]. This confirmed that docking simulations were

not an appropriate alternative for the experimental measurement of binding affinities.

Docking is limited in its predictive ability due to assumptions made during the calculations: the DNA target is represented as rigid, meaning that there is no flexibility to the position or orientation of flanking bases. In solution, DNA is flexible and flanking bases can move to accommodate ligand binding. In an attempt to minimise the influence of this on the binding predictions, DNA structures were taken from PDB which had been calculated with a bound ligand. The ligand was removed to create a cavity for binding. However, this was not possible for *c-kit1* as PDB structures with bound ligands were not available. Additionally, the absence of water within the simulation could have had a negative effect on the prediction. This is because water often plays a role within binding, through water mediated hydrogen bonding between the ligand and the target [295, 296]. Therefore, docking provides more of a qualitative assessment of binding poses rather than strong quantitative results.

3.4 ROS induced DNA damage

Having established the ability of platinum(II) salphen complexes to generate ROS, and their ability to bind to G4 DNA, it was of interest to see if these properties could work complementary to one another and be used to induce G4 DNA damage.

3.4.1 Plasmid DNA ROS induced damage

To assess if photocleavage of DNA was possible with the platinum(II) salphen complexes, an assay using plasmid DNA (pUC19) was performed. This assay used gel electrophoresis to measure conversion of the supercoiled topology of plasmid DNA to the nicked circular DNA topology induced by DNA damage [299, 300] (figure 3.20).

In the supercoiled form (sc), DNA moves more quickly through the gel. However, upon single strand breakage, as is typical of ROS DNA damage, the DNA takes an open circular (oc) topology which migrates more slowly through the gel. When both strands of the DNA are broken linear (ln) duplex DNA is formed, which moves at an intermediate speed, appearing between the supercoiled and open circular topologies of the DNA within the gel. As a control, an endonuclease enzyme (Eco31I)

was used to cut the plasmid across both strands of DNA at a single point (BsaI sequence) to generate linear DNA. The variability in gel migration of the different DNA topologies enabled the quantification of conversion upon irradiation of the DNA in the presence of the ROS generating platinum(II) salphen complexes. The complexes selected for this assay included **Pt_PPTMA_4OMe**, **Pt_PEM_4OMe**, **Pt_AEP_4OMe** and **Pt_PEM_5Br**. These complexes demonstrate a range of ROS generating abilities which are shown in table 3.3.

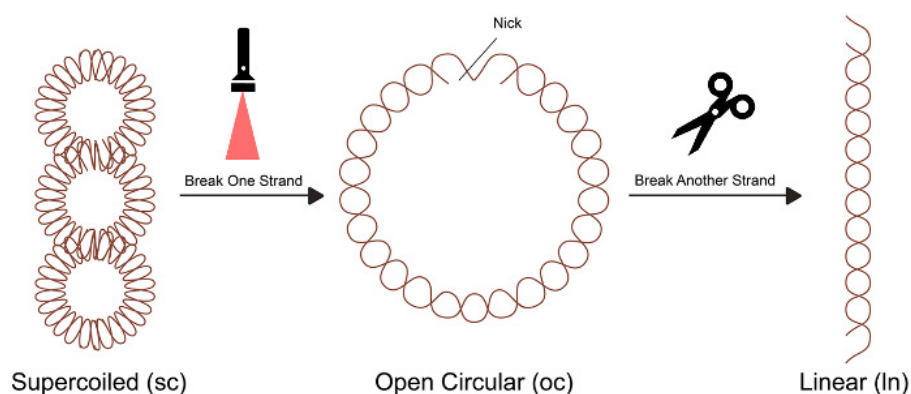


Figure 3.20. Schematic showing the different topologies of pUC19 plasmid DNA in its native supercoiled form (sc), its open circular form (oc) after single strand breakage, and the linear duplex form (ln) following double strand breakage.

Complex	ABDA: Type 2 ROS (%)	DHR: Type 1 ROS (%)
Pt_PPTMA_4OMe	64	90
Pt_PEM_4OMe	13	50
Pt_AEP_4OMe	22	35
Pt_PEM_5Br	10	12

Table 3.3. Table showing the relative type 2 (ABDA) and type 1 (DHR) ROS generating abilities of the complexes selected for the plasmid DNA damage assay.

The main focus was on the **4OMe** complexes due to their consistent performance within the DNA binding assays across each **DAB**. Complexes were combined with the plasmid DNA and irradiated at 457 nm for 1 hour before being loaded on to the agarose gel. As

controls, DNA was irradiated in the absence of complex and incubated with complexes kept in the dark. Figure 3.21 shows the resulting gel.

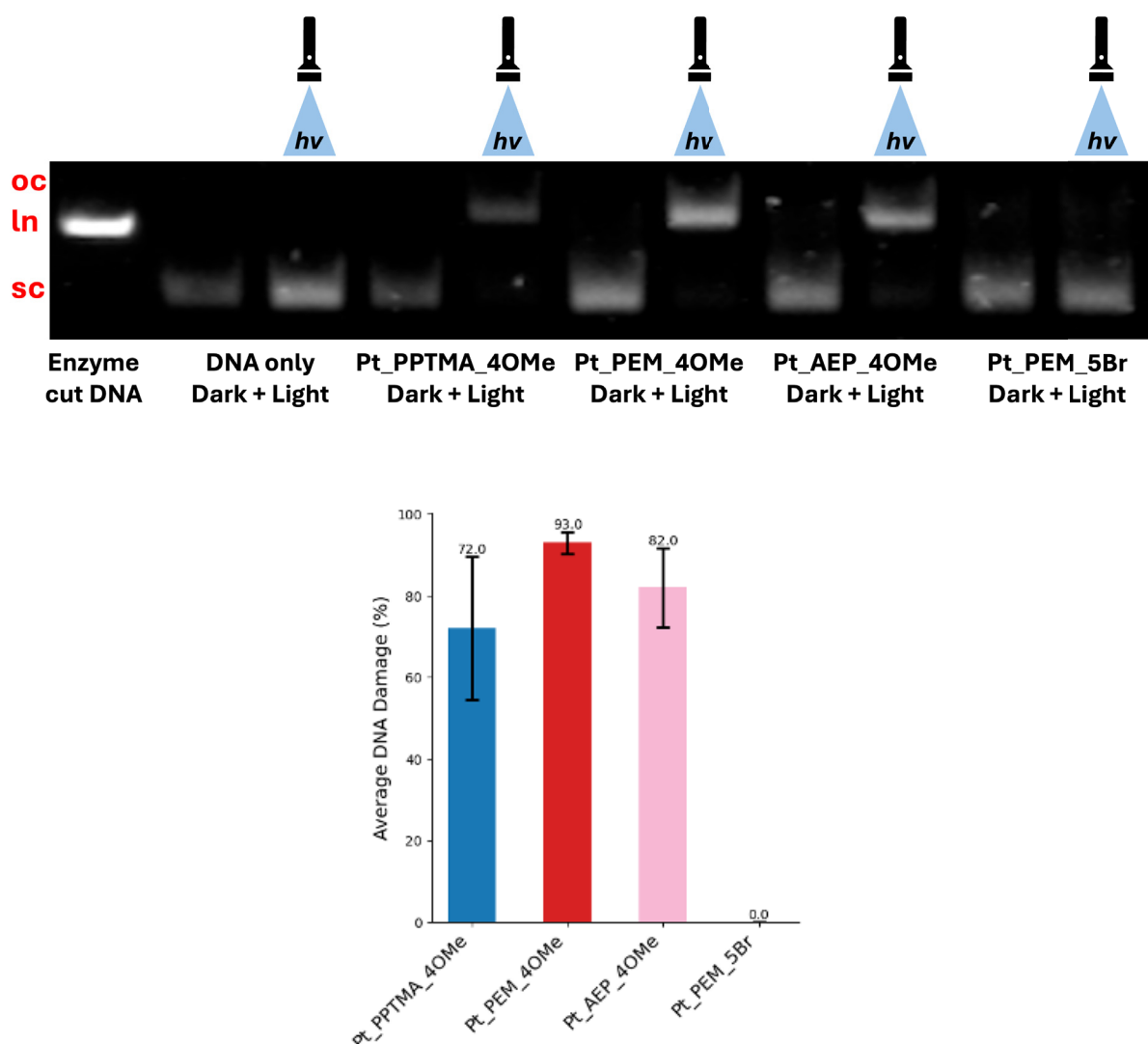
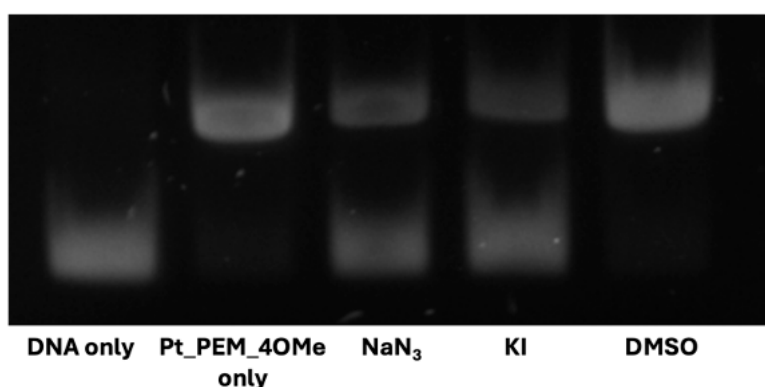


Figure 3.21. Photocleavage of DNA induced by platinum(II) salphen complexes. Top: Image of the gel electrophoresis plasmid damage assay. The contents of each lane are: 1) Eco31I. 2) pUC19 dark. 3) pUC19 irradiated (with no platinum complex added). 4) pUC19 + **Pt_PPTMA_4OMe** dark. 5) pUC19 + **Pt_PPTMA_4OMe** irradiated. 6) pUC19 + **Pt_PEM_4OMe** dark. 7) pUC19 + **Pt_PEM_4OMe** irradiated. 8) pUC19 + **Pt_AEP_4OMe** dark. 9) pUC19 + **Pt_AEP_4OMe** irradiated. 10) pUC19 + **Pt_PEM_5Br** dark. 11) pUC19 + **Pt_PEM_5Br** irradiated. Light irradiation was performed with a Lumidox 457 nm 96-LED array (8 mA, 2 mW/cm²) for 1 hour. Bottom: The percentage conversion of pUC19 in the presence of each complex is: **Pt_PPTMA_4OMe** 72 %. **Pt_PEM_4OMe** 93 %. **Pt_AEP_4OMe** 82 %. **Pt_PEM_5Br** 0 %. n = 3. oc = open circular pUC19 topology. ln = linear pUC19 topology. sc = supercoiled pUC19 topology.

It was seen that all **4OMe** containing platinum(II) salphen complexes induced DNA cleavage upon light irradiation (lanes 5, 7 and 9), but DNA was not cleaved when incubated with complexes in the dark (lanes 4, 6 and 8). In comparison **Pt_PEM_5Br** did not appear to induce any DNA damage upon irradiation (lane 11), this was expected due to the low ROS generating ability of this complex.

Pt_PEM_4OMe appeared to result in the best DNA cleavage as shown by its very bright band of open circular DNA. This was surprising due to its moderate performance within the type 1 DHR assay (50 %) and very poor performance within the type 2 ABDA assay (13 %). This result, alongside the inability of **Pt_PEM_5Br** to induce DNA damage with poor type 1 and type 2 ROS generation (12 % and 10 % respectively) suggested that type 1 ROS played a key role in inducing DNA damage.

To further explore which ROS species were responsible for DNA cleavage, the experiments were repeated for **Pt_PEM_4OMe** in the presence of ROS scavenger species: NaN_3 , KI and DMSO, which scavenge singlet oxygen (type 2), superoxide (type 1), and hydroxyl radicals (type 1) respectively. Figure 3.22 shows the resultant agarose gel and a bar chart with the average DNA damage in the presence of each scavenger.



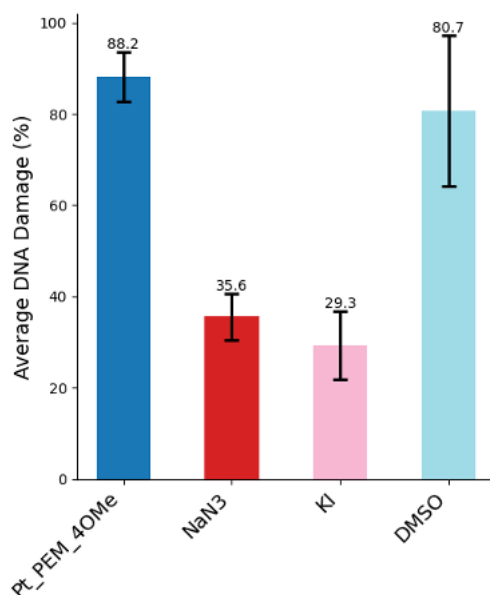


Figure 3.22. Top: Gel contain ROS quenchers: The contents of each lane were irradiated and contain: 1) pUC19. 2) **Pt_PEM_4OMe**. 3) **Pt_PEM_4OMe** + NaN₃. 4) **Pt_PEM_4OMe** + KI. 5) **Pt_PEM_4OMe** + DMSO. n = 3. Bottom: bar chart showing the average DNA damage upon addition of each ROS quencher compared to without a quencher.

Unexpectedly, both the singlet and superoxide scavengers reduced DNA cleavage, suggesting both type 1 and type 2 ROS are equivalently important. Overall this indicated that the degree of DNA damage induced by a complex was as a result of more complex factors than just ROS generation, likely including how well the complex bound to the DNA before cleavage. Interestingly, the hydroxyl radical scavenger did not significantly affect the amount of DNA damage, suggesting this species is not responsible for DNA cleavage. In future work, a complex with high type 2 and low type 1 ROS, such as **Pt_PPTMA_Naph**, should be investigated.

3.4.2 G4 DNA ROS induced damage

With the knowledge that the ROS generation of the platinum(II) salphen complexes could cleave plasmid DNA, it was next determined if the complexes could also induce G4 DNA damage. Due to the chiral nature of G4 DNA structures, their relative concentration can be measured using circular dichroism (CD) spectroscopy. This is the measure of absorbance of left and right-handed circularly polarised light. The G4 sequence used within this experiment was *c-myc* G4 DNA due to its parallel topology

giving rise to a distinct CD spectrum with an intense positive ellipticity at 265 nm, and a weaker negative ellipticity at 245 nm [301]. Should the G4 DNA become damaged, there would be a reduction in the CD signal intensity [138, 302]. ROS can damage DNA through a variety of processes, two of the mechanisms of interest for G4 DNA damage include cleavage and guanine oxidation [245]. For DNA cleavage the CD spectra is expected to flatline after sufficient irradiation where no more G4 is present. For guanine oxidation the CD spectra will reduce slightly and reach a plateau where no further oxidation from the complex is possible, and a perturbed G4 structure remains.

The complexes taken forward for the G4 damage assay included **Pt_PPTMA_4OMe**, **Pt_PEM_4OMe** and **Pt_AEP_4OMe**. The *c-myc* DNA was combined with 3 equivalents of each complex and irradiated at 457 nm in 10-minute intervals, with CD spectra being measured between each irradiation. An extended irradiation experiment of 3 hours was performed with **Pt_PPTMA_4OMe** to confirm whether the G4 DNA damage was a result of cleavage or oxidation. TMPyP4, a G4 ligand which causes irradiation-induced oxidation of guanines was also tested as a positive control [303]. Figure 3.23 shows the CD spectra for *c-myc* DNA with **Pt_PPTMA_4OMe**, and the change in ellipticity for the extended 3-hour time period.

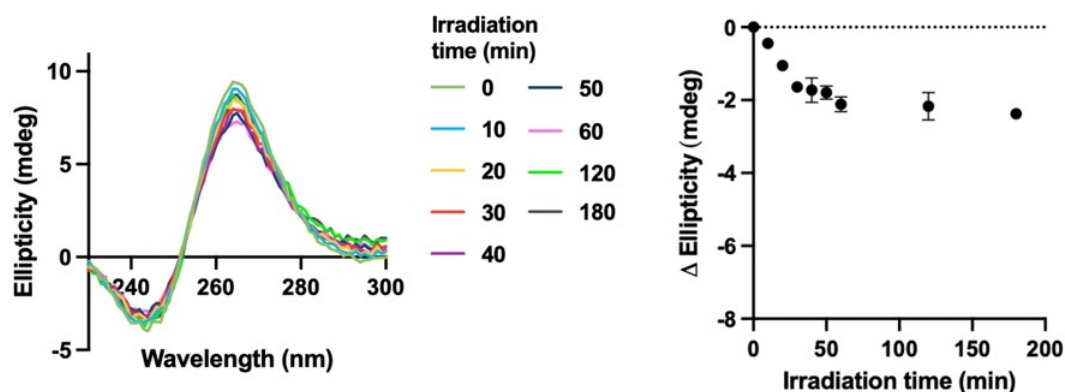


Figure 3.23. left) CD spectra of *c-myc* DNA with **Pt_PPTMA_4OMe**, showing a reduction in ellipticity with increasing intervals of irradiation up to three hours. Right) plot showing the change in ellipticity at 265 nm from the CD plot with increasing irradiation time up to three hours. $n = 3$.

The plateau after 1 hour of irradiation suggested that the G4 DNA was being damaged by oxidation as opposed to cleaved. This is further supported when looking at the

performance of each of the platinum(II) salphen complexes in comparison to TMPyP4, which shows a similar plateau in the reduction in ellipticity (figure 3.24). These plots show that TMPyP4 reduces ellipticity by 6 mdeg after 1 hour, compared to **Pt_PPTMA_Ph** by 2 mdeg, **Pt_PEM_4OMe** by 1.5 mdeg and **Pt_AEP_4OMe** by 1 mdeg. This suggests that the complexes are not as efficient as TMPyP4 at oxidising the G4 structures. The order of G4 DNA CD signal perturbation followed the same trend as the performance of the complexes at generating type 1 ROS (see section 3.2).

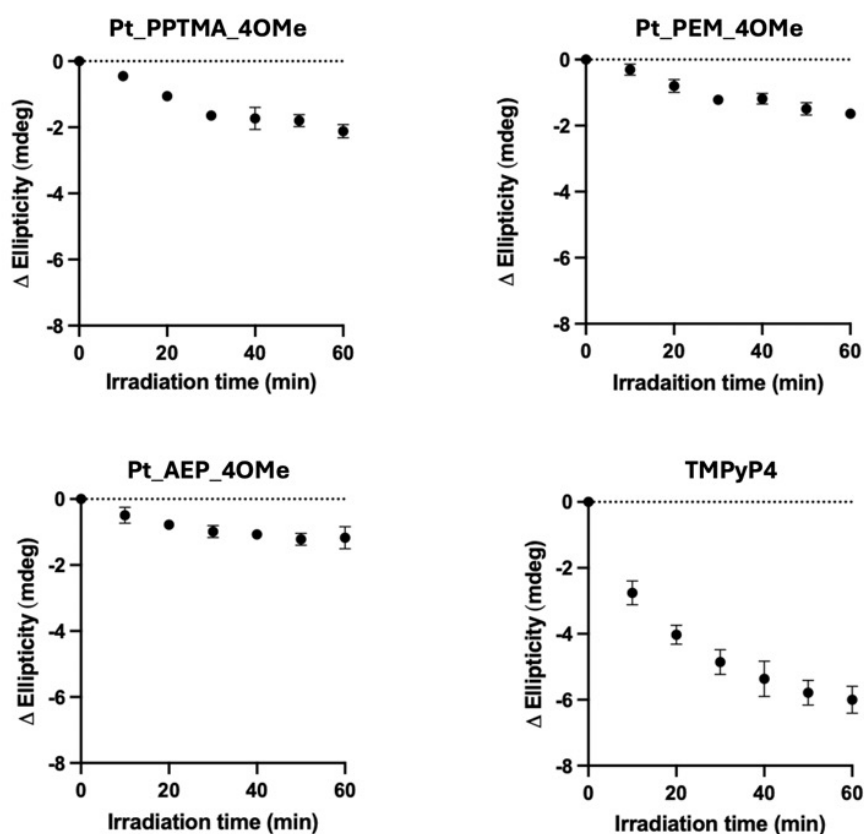


Figure 3.24. Plots showing the change in ellipticity for *c-myc* DNA at 265 nm with increasing intervals of irradiation with **Pt_PPTMA_4OMe**, **Pt_PEM_4OMe**, **Pt_AEP_4OMe** and TMPyP4 up to 1 hour of irradiation. $n = 3$.

To confirm that the change in ellipticity for these complexes was a result of ROS generation upon irradiation, experiments were repeated for TMPyP4 and **Pt_PPTMA_4OMe** under dark conditions (figure 3.25). When kept in the dark the G4 CD spectra showed a very small increase in ellipticity overtime, suggesting minor stabilisation of the G4 DNA from binding of the complexes. As the intensity of the

spectra did not decrease, this confirmed that the damage observed was a result of ROS generation from irradiation of the complexes.

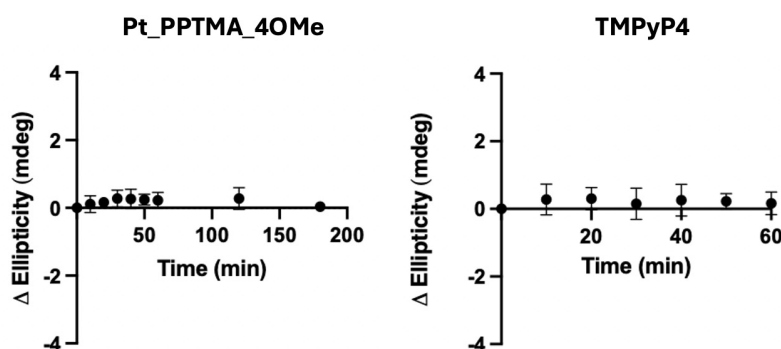


Figure 3.25. Plots showing the change in ellipticity for *c-myc* DNA with Pt_PPTMA_4OMe and TMPyP4 for 3 hours and 1 hour respectively when incubated in the dark. $n = 3$.

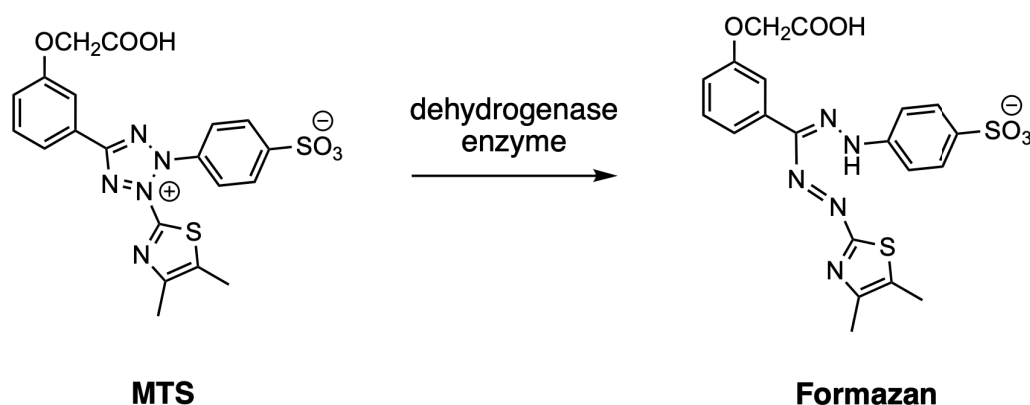
3.5 Cellular behaviour

Next, the cellular behaviour of the platinum(II) salphen complexes was investigated. Knowing that the complexes were capable of inducing ROS-mediated DNA damage it was of interest to assess their performance as phototoxic agents. To be good phototoxic agents, complexes are required to be non-toxic under dark conditions, and toxic upon irradiation. The results of the *in vitro* DNA damage assays of the complexes were promising for this goal as they were found to not damage DNA under dark conditions and induce damage upon irradiation. Additionally, their G4 DNA binding affinities showed modest activity (between 10^6 to 10^5 M^{-1}) indicating that binding may be labile, and therefore not induce direct cell death as is seen for more potent G4 binding complexes. Within this work HeLa cells, immortalised cervical cancer cells, were used as models of cancer cells.

3.5.1 Cytotoxicity of platinum(II) salphen complexes

To measure the cytotoxicity of the platinum(II) salphen complexes, a well established MTS assay was implemented. This assay is a colorimetric method which assess cell viability and proliferation using the MTS dye: 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium).

The dye's tetrazolium group is oxidised to formazan by dehydrogenase enzymes within metabolically active cells [304] (scheme 3.3).



Scheme 3.3. Scheme showing the conversion of the MTS dye to Formazan via the dehydrogenase enzyme within metabolically active cells.

This dye conversion results in a reduction in absorbance, enabling a quantitative measurement of cell viability in comparison to positive controls of 100 % cell viability. To measure the platinum(II) salphen complexes cytotoxicity, the IC_{50} (concentration of compound for 50 % of cell death) was calculated using a dose response curve ranging from 25 μ M to 0.78 μ M with 24 hour incubation. Under these conditions no significant cytotoxicity was detectable for the platinum(II) salphen complexes, with no cell viabilities dropping below 70 % (figure 3.26).

The experiment was repeated in quadruplicate across two biological repeats, showing the same results. This lack of toxicity displayed by the platinum(II) salphen complexes was also reported in the literature for similar complexes [235]. This result was very promising for their potential application as photosensitisers as they displayed the desirable property of being non-toxic in the absence of irradiation. Therefore, the next step was to determine their cytotoxicity upon irradiation (phototoxicity).

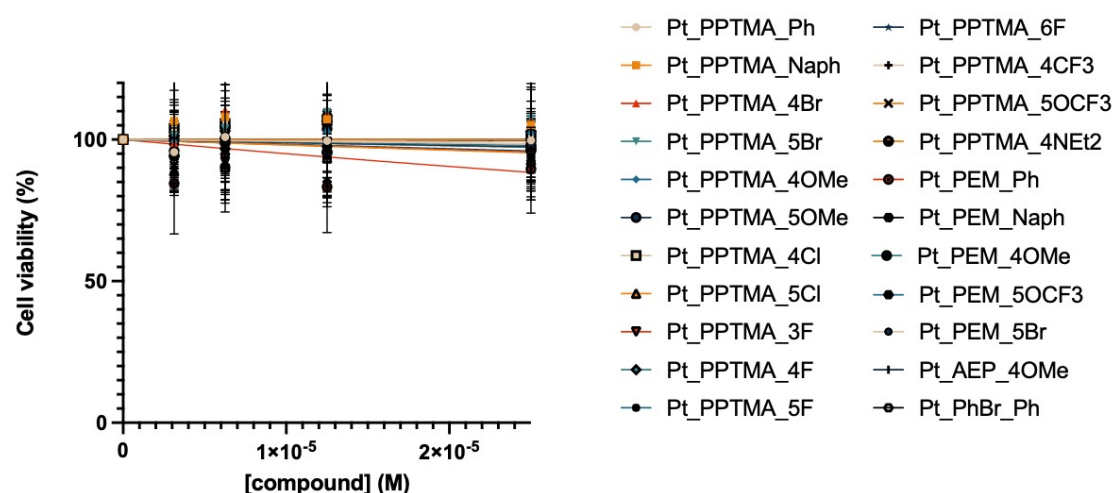


Figure 3.26. Plot showing the concentration dependent cytotoxicity of the platinum(II) salphen complex series against HeLa cells under dark conditions. Showing a maximum concentration of 25 μ M of complex. $n = 4$.

3.5.2 Phototoxicity of platinum(II) salphen complexes

Initially a single point concentration experiment was performed in one 96-well plate. The plate was split in two halves, one side of the plate was kept in the dark, and the other side was irradiated for 12 minutes using the 457 nm Lumidox 96 LED plate, following a previously established cell irradiation protocol by Dr Tim Kench within the Vilar group. Black plates were used to avoid light diffusion between samples. Cells were incubated with the complexes for 3 hours before irradiation. Following a further 21 hours incubation, the resultant cell viability for each complex under light (irradiation) and dark conditions was measured (table 3.4).

It was seen from this preliminary screen that many of the complexes which were non-toxic under dark conditions become very cytotoxic following irradiation, making them phototoxic. This initial screen was used to determine the top concentration for each complex within the dose response phototoxicity investigation. Complexes which resulted in less than 50 % cell death were screened from 25 μ M to 0.78 μ M, whilst complexes which resulted in more than 50 % cell death were screened from 10 μ M to 0.31 μ M. From these experiments **Pt_PPTMA_Naph**, **Pt_PEM_4OMe** and **Pt_PEM_Naph** were identified as being highly phototoxic and so their measurements

were repeated from 2.5 μM to 0.078 μM . The dose response curves for each experiment parameter, and a bar chart of all complex IC_{50} values (within the range of the assay) are shown in figures 3.27 and 3.28.

Complex at 5 μM	Light % cell viability	Dark % cell viability
Pt_PPTMA_Ph	79	95
Pt_PPTMA_Naph	9	89
Pt_PPTMA_4Br	60	120
Pt_PPTMA_5Br	64	90
Pt_PPTMA_4OMe	68	91
Pt_PPTMA_5OMe	85	95
Pt_PPTMA_4Cl	56	115
Pt_PPTMA_5Cl	55	76
Pt_PPTMA_3F	84	81
Pt_PPTMA_4F	84	122
Pt_PPTMA_5F	96	110
Pt_PPTMA_6F	75	122
Pt_PPTMA_4CF ₃	39	106
Pt_PPTMA_5OCF ₃	23	120
Pt_PPTMA_4NEt ₂	93	125
Pt_PEM_Ph	2	78
Pt_PEM_4OMe	2	95
Pt_PEM_Naph	3	120
Pt_PEM_5OCF ₃	24	100
Pt_PEM_5Br	34	118
Pt_AEP_4OMe	15	134
Pt_PhBr_Ph	3	85
Nickel Salphen	5	6

Table 3.4. Dark and light cell viability of HeLa cells with the platinum (II) salphen complexes at 5 μM . Low cell viability is shown as red and high cell viability as green. Phototoxicity was induced following irradiation of cells at 457 nm for 12 minutes at 8 mA.

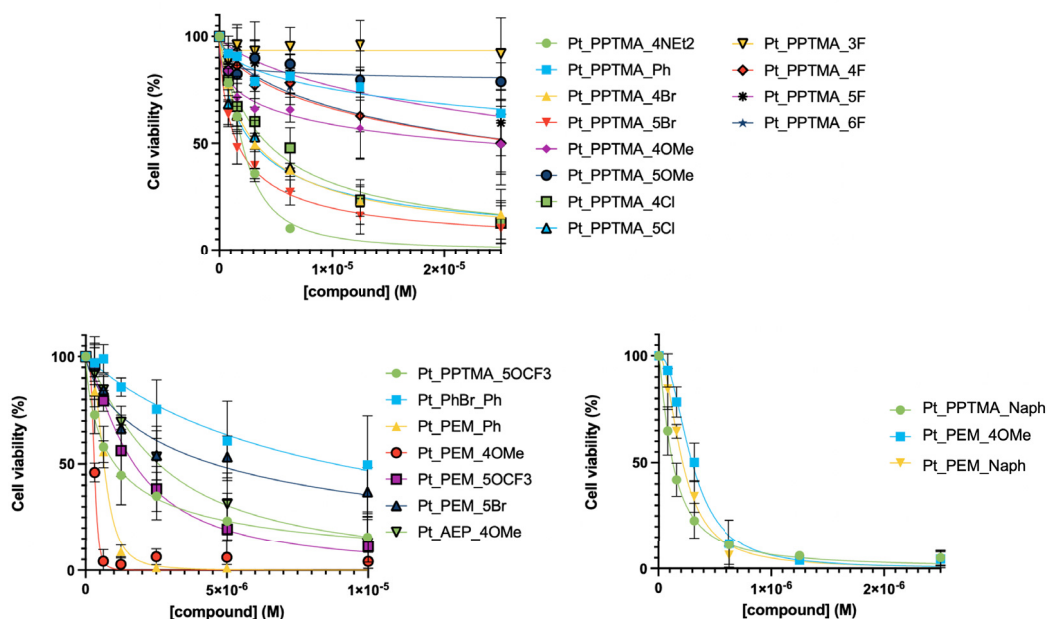


Figure 3.27. Phototoxicity dose response curves for the platinum(II) salphen complexes screened from concentrations of: 25 μM (top), 10 μM (bottom left), and 2.5 μM (bottom right). $n = 3$.

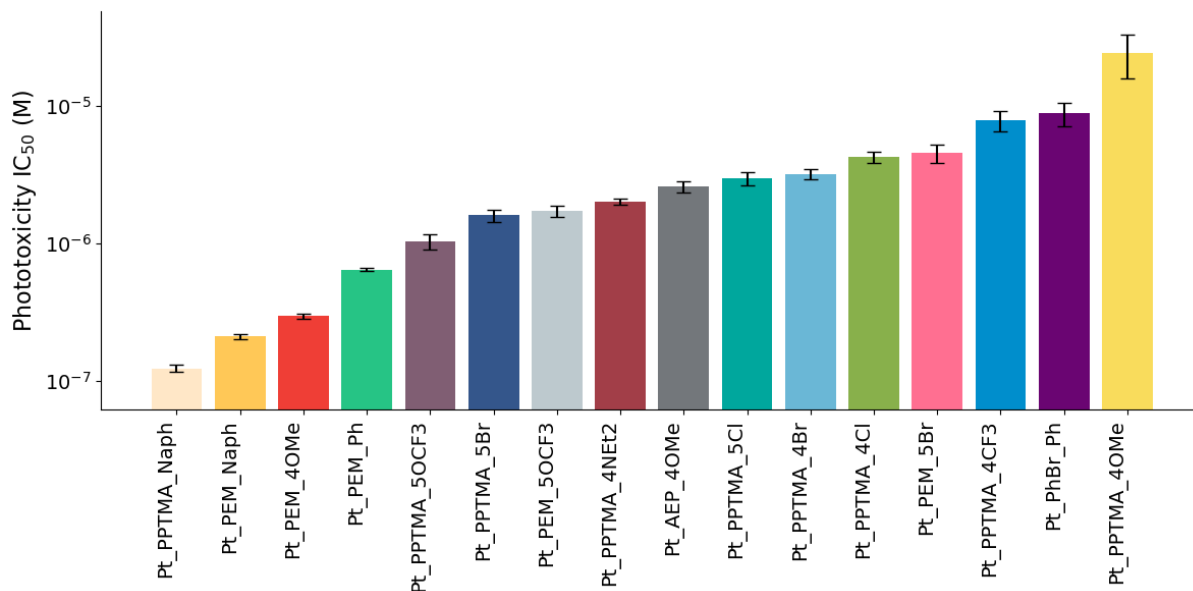


Figure 3.28. Phototoxicity IC_{50} values for the platinum(II) salphen complexes. Phototoxicity was induced following irradiation of cells at 457 nm for 12 minutes at 8 mA. $n = 3$. The full IC_{50} values are detailed in appendix table A.8.

The complex which displayed the highest phototoxicity was **Pt_PPTMA_Naph** with an

IC₅₀ value of 123 nM, which is ten times more potent than the common anti-cancer drug cisplatin (22.4 µM after 24 hours in HeLa cells [36]). **Pt_PEM_Naph** and **Pt_PEM_4OMe** also demonstrated excellent IC₅₀ values of 210 nM and 294 nM respectively. Table 3.5 shows the ROS generation and relevant G4 binding abilities of the most and least phototoxic complexes.

Complex	IC ₅₀ (µM)	Type 2 ROS (%)	Type 1 ROS (%)	<i>c-myc</i> K _a (M ⁻¹)	HTelo K _a (M ⁻¹)	<i>c-kit1</i> K _a (M ⁻¹)	HTG4 K _a (M ⁻¹)
Pt_PPTMA_Naph	0.123	131	24.3	7.35 x 10 ⁵	7.04 x 10 ⁵	3.71 x 10 ⁵	3.00 x 10 ⁵
Pt_PEM_Naph	0.210	12.1	28.5	5.45 x 10 ⁵	2.35 x 10 ⁵	1.55 x 10 ⁶	Non-binding
Pt_PEM_4OMe	0.294	12.8	47.3	3.01 x 10 ⁵	2.41 x 10 ⁵	2.84 x 10 ⁵	4.45 x 10 ⁵
Pt_PPTMA_Ph	>25.0	55.7	70.4	1.38 x 10 ⁶	4.33 x 10 ⁵	4.03 x 10 ⁵	9.29 x 10 ⁴
Pt_PPTMA_6F	>25.0	38.4	53.8	2.66 x 10 ⁶	4.53 x 10 ⁶	9.36 x 10 ⁵	1.51 x 10 ⁵

Table 3.5. Table summarising the phototoxicity IC₅₀ values, type 1 and 2 ROS, and G4 DNA binding for the most phototoxic complexes **Pt_PPTMA_Naph**, **Pt_PEM_Naph** and **Pt_PEM_4OMe**, and the least phototoxic complexes **Pt_PPTMA_Ph** and **Pt_PPTMA_6F**.

Interestingly, the most phototoxic complexes did not have the best ROS generating or DNA binding abilities. The least phototoxic complexes **Pt_PPTMA_Ph** and **Pt_PPTMA_6F** demonstrated good ROS generating ability (> 50 % type 1 and type 2) and strong G4 binding (values in the 10⁶ range) yet did not induce phototoxicity within the range of the assay. This suggested that there is not a clear correlation between *in vitro* ROS generation and G4 binding with phototoxicity. This is likely due to the additional considerations for cytotoxicity measurements including the complexes ability to generate ROS within a cellular environment, in addition to the complexes cell permeability and localisation.

3.5.3 Cellular imaging

As a result of the phosphorescent nature of the platinum(II) salphen complexes, fluorescent imaging microscopy could be used to determine the cellular uptake and localisation. This would provide further rationale for the mismatch between the ROS generation and G4 binding of complexes and phototoxicity.

3.5.3.1 High throughput fluorescent microscopy screening

To obtain the optimal images of the platinum(II) salphen complexes within HeLa cells several conditions were screened including concentration (2.5 μ M, 5 μ M, 6 μ M, 10 μ M, 12.5 μ M, and 25 μ M), and incubation time (3, 6 and 24 hours). Due to the large number of complexes and conditions selected for screening it would be very time consuming to perform on a traditional confocal microscope. Fortunately, using a Nikon Eclipse Ti2 microscope it was possible to scan and image 96 well plates. The intensity of each image was adjusted independently from one another to provide clear images of each complex across the range of cell fluorescent intensities. A larger distinction between the background and the fluorescent signal within cell images indicates a more strongly emissive complex (the more blue the background the weaker the complex intensity). To quantitatively assess the brightness of the cellular fluorescence the corrected total cell fluorescence (CTCF) was calculated [305, 306] (equation 3.3):

$$\text{CTCF} = \text{Integrated Density} - (\text{Area of Cell} \times \text{Mean Background Fluorescence}) \quad (3.3)$$

The integrated density is the sum of all pixel values within the cell area, representing the total fluorescent signal. The signal is then corrected based on the size of the cell, and the background fluorescent signal from an area without any cells present.

6 hours was determined to be the optimal incubation time as after 3 hours incubation most concentrations resulted in low fluorescent intensity images, and 24 hours incubation showed no significant change in fluorescent intensity to incubation for 6 hours. The optimal complex concentration was determined to be 12.5 μ M, as at 25 μ M the fluorescent intensity of the brighter complexes saturated the detector, and at 10 μ M the fluorescent intensity of the least bright complexes was too low to detect. Figure 3.29 shows the fluorescent microscopy images for each complex within HeLa cells after 6 hours incubation at 12.5 μ M.

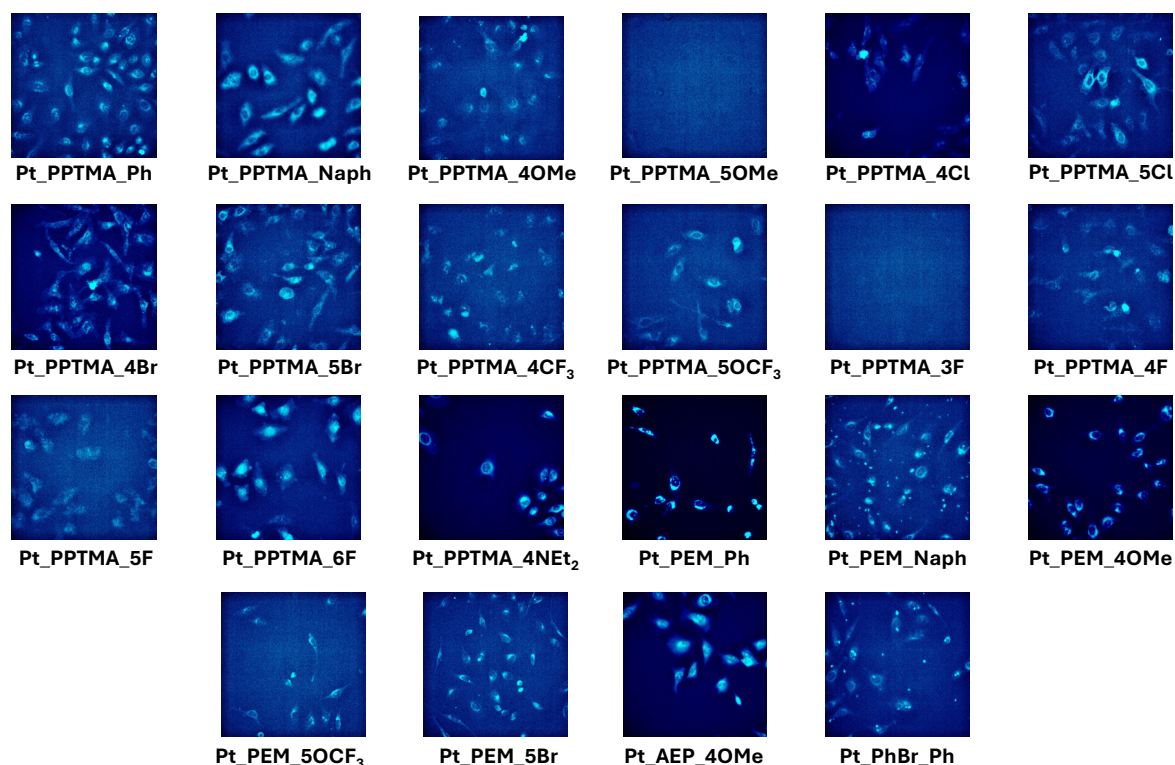


Figure 3.29. Fluorescent microscopy images of the platinum(II) salphen complexes at 12.5 μ M after 6 hours incubation in HeLa cells. Following excitation at 405 nm and emission measured from 600 to 650 nm. The contrast of the images was adjusted using a contrast adaptive equalisation formula applied to each individual image. Image size: 250x250 μ m.

The HT microscopy images showed that the majority of the platinum(II) salphen complexes generated an emissive signal within cells, with the exception of **Pt_PPTMA_5OMe** and **Pt_PPTMA_3F**. Neither of these complexes achieved measurable phototoxicity IC_{50} values within the range of the assay, which could now be attributed to their poor cellular uptake. Small bright flecks were seen within the images of **Pt_PEM_Naph** and **Pt_PhBr_Ph**, suggesting complex aggregation. Their aggregation was likely due to their low solubility from the increased aromatic system of **Pt_PEM_Naph**, and the lack of solubilising groups in **Pt_PhBr_Ph**. The brightest complexes of the series were **Pt_PEM_Ph**, **Pt_PEM_4OMe**, **Pt_AEP_4OMe**, and **Pt_PPTMA_4NEt₂**. Figure 3.30 below shows a scatter plot exploring trends between CTCF and the complexes **CBA** and **DAB** components.

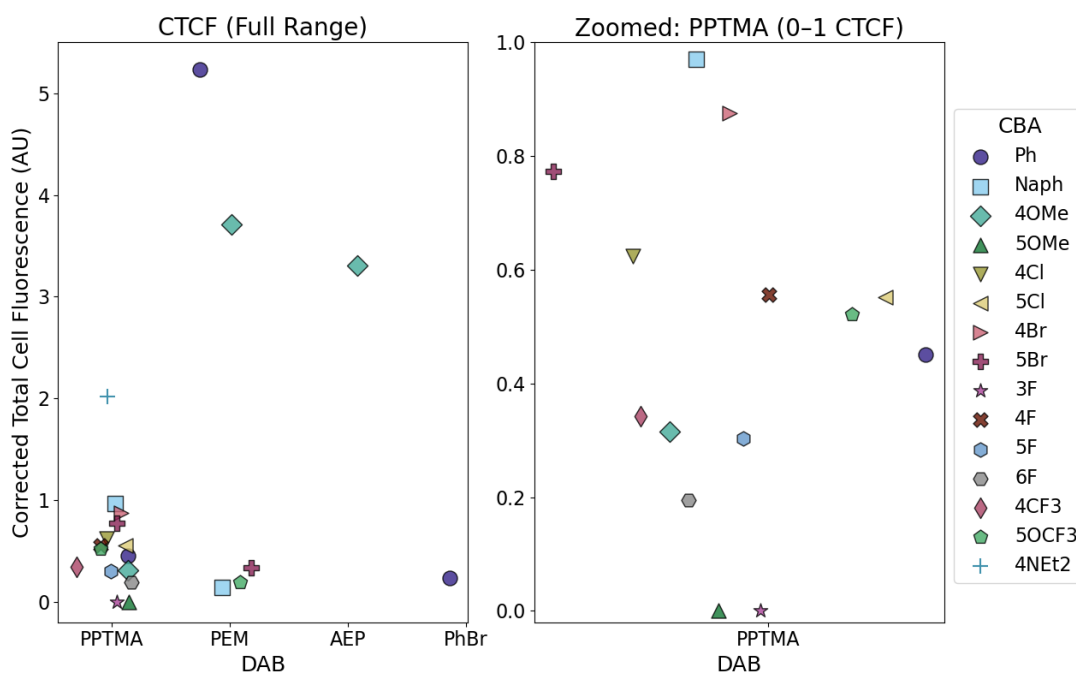


Figure 3.30. Strip plots showing the corrected total cell fluorescence (CTCF) determined from the HT microscopy imaging for the platinum (II) salphen complexes, split by their **DAB** and coloured by their **CBA** components. Left: all complexes across all CTCF values. Right: Zoomed in CTCF plot between 0-1 for **Pt_PPTMA_R²** complexes.

Notably, no single component appeared to result in consistently bright complexes. For example, the **Ph CBA** yields the brightest complex with **Pt_PEM_Ph**, however, the other **Ph** containing complexes displayed a low CTCF. The **PPTMA DAB** resulted in the lowest brightness across complexes with shared **CBA** derivatives, most distinctly shown for **4OMe** which demonstrated very high CTCF with **PEM** and **AEP**, but low CTCF with **PPTMA**. The inability to distinguish clear trends within this data is likely due to the complexity of factors effecting the CTCF measurement which include cell permeability, localisation, and their ability to bind to components within the cellular environment which provide protection from water, preventing non-radiative decay.

The key limitations of the Nikon Eclipse Ti2 were that the auto focus feature would not always reach an optimal z-dimension, resulting in some of the images being at a lower resolution. Additionally, during imaging protocols it was not possible to relocate the camera position within each well, resulting in some images having low cell density. Due to these limitations, it was not possible to confidently decipher the cellular

localisation of each complex. To enable more definitive conclusions of the cellular localisation of the platinum(II) salphen complexes, more in-depth confocal microscopy was performed with a selection of complexes which appeared to have unique cellular localisation and contained a variety of **CBA** and **DAB** components. The complexes selected were **Pt_PPTMA_Naph**, **Pt_PPTMA_4NEt₂**, **Pt_PPTMA_4Br**, **Pt_PEM_Ph**, **Pt_PEM_4OMe** and **Pt_AEP_4OMe**.

3.5.3.2 Confocal imaging for in-depth cellular localisation studies

The confocal images were taken by Dr Petr Sherin using a confocal dual laser-scanning microscope: Leica TCS SP5 II. The experiments undertaken included imaging of the complexes within HeLa cells at various time points. In addition to images with co-stains MitoTracker™ deep red (MitoTracker) to investigate mitochondrial staining, and LysoTracker™ deep red (LysoTracker) to investigate lysosome staining. A beneficial feature of the Leica TCS SP5 II is the ability to define the wavelength range for excitation and emission. This enabled the careful selection of wavelengths to prevent the overlap of emission signals from the platinum(II) salphen complexes and the co-stains, even when their emissions spectra showed some overlap. Additionally, the dual laser nature of the microscope enabled very low energy imaging (< 0.5 μW) which prevented photodamage in all cases except **Pt_PPTMA_Naph**, which was the most phototoxic complex of the series.

Figure 3.31 shows the confocal fluorescent microscopy images of the complexes at 10 μM within HeLa cells after 1 hour of incubation.

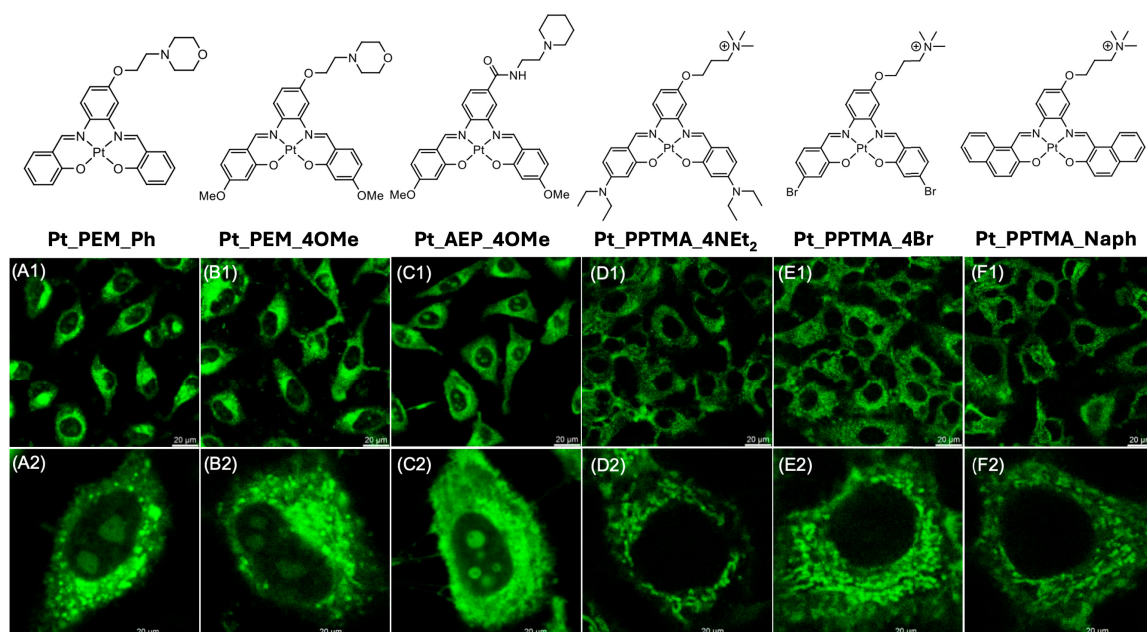


Figure 3.31. Platinum(II) salphen complexes and their confocal images in HeLa cells at low magnification (A1-F1): 155×155 μm image size and high magnification (A2-F2): 40×40 μm image size, after one hour of incubation with 10 μM of complexes; scale bars are 20 μM. Images taken by Dr Petr Sherin.

Within the images it can be clearly seen that **Pt_PEM_Ph**, **Pt_PEM_4OMe** and **Pt_AEP_4OMe** stained the nucleoli of the HeLa cells, in addition to staining the cytoplasm in granular patterns, with very little non-nucleolar nucleus staining. For the **PPTMA** solubilised complexes **Pt_PPTMA_4NEt₂**, **Pt_PPTMA_4Br** and **Pt_PPTMA_Naph**, there was no nuclear or nucleoli staining, with complexes only staining within the cytoplasm, this indicated that the presence of the charged **PPTMA** backbone may have been preventing complexes from entering the nucleus of the HeLa cells. An additional observation to note was that the staining within the cytoplasm was not uniform across the cell, showing fine interconnected structured patterns, unique to the cytoplasmic staining patterns observed for **Pt_PEM_Ph**, **Pt_PEM_4OMe** and **Pt_AEP_4OMe**.

To determine if the distinct staining patterns of the complexes within the cytoplasm were a result of organelle specific staining, MitoTracker and LysoTracker dyes were added to the HeLa cells alongside the platinum(II) salphen complexes. However, due to the very

broad emission profiles of the platinum(II) salphen complexes it was not possible to obtain images with both, or in some cases either stain with all complexes. MitoTracker has a peak emission wavelength of 665 nm, and LysoTracker has a peak emission wavelength of at 663 nm. These dyes have the most red-shifted emission profiles of the available MitoTracker and LysoTracker co-stains. An alternate approach of a blue shifted dye was not possible as MitoTracker Green has a peak emission wavelength of 516 nm, and LysoTrack Green had a peak emission wavelength of 511 nm, both of which would directly clash with the peak emission wavelengths of all of the platinum(II) salphen complexes, which generally ranged from 500 nm to 650 nm. LysoTracker Blue with a peak emission wavelength of 400 nm would have clashed with the excitation wavelength of the experiments.

Unfortunately, it was not possible to achieve co-staining with **Pt_PPTMA_5Br** or **Pt_PPTMA_Naph** due to their very red shifted emission profiles which directly overlapped with, or tailed across the emission spectra of the co-stains as shown in figure 3.32.

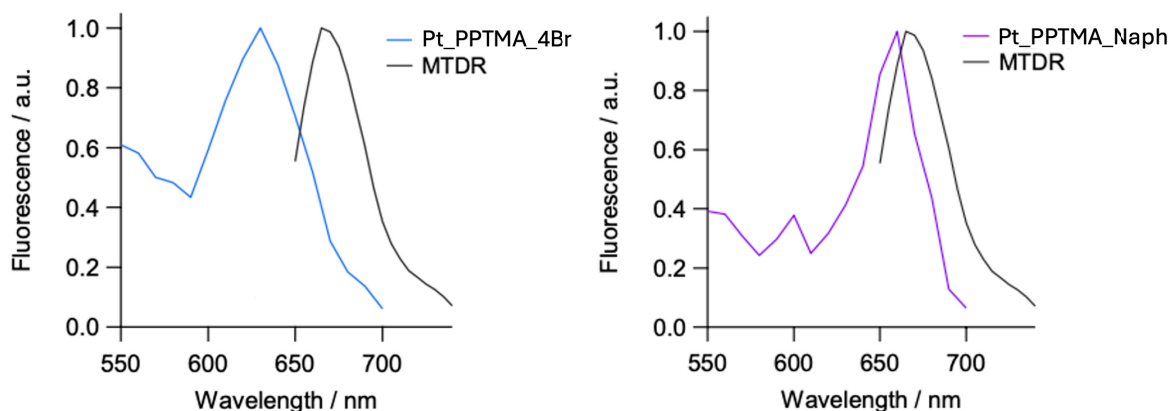


Figure 3.32. Emission spectra for **Pt_PPTMA_4Br** (left) and **Pt_PPTMA_Naph** (right) with MitoTracker showing that there was too much overlap between the spectra to obtain co-stained images.

Figure 3.33 shows the co-stain images of **Pt_AEP_4OMe** with MitoTracker (A) and LysoTracker (B). It was possible to obtain these despite a substantial overlap between the co-stains and the complex emission spectra by setting the emission detection of the microscope between 680 – 720 nm, which only contained a very small signal overlap

from the complex emission profile to the co-stains emission profiles.

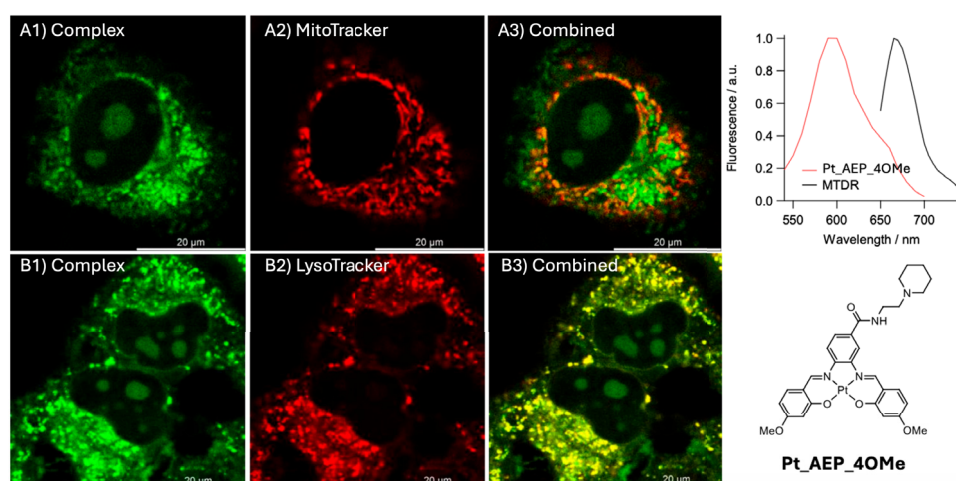


Figure 3.33. HeLa cells stained with 10 μM **Pt_AEP_4OMe** – 6 hours (A1, B1), 50 nM MitoTracker – 5 mins (A2), and 50 nM LysoTracker – 5 mins (B2), and the overlaid images for **Pt_AEP_4OMe** with MitoTracker (A3) and **Pt_AEP_4OMe** with LysoTracker (B3). **Pt_AEP_4OMe** detection: excitation 514 nm, emission 540 – 630 nm. MitoTracker detection: excitation 514 nm, emission 680 – 720 nm. LysoTracker detection: excitation 633 nm, emission 680 – 720 nm. Images taken by Dr Petr Sherin.

Within these images it can be seen that there was no overlap between the complex fluorescent signal and the MitoTracker signal across images. Whereas, there was substantial overlap between the images of the complex and the LysoTracker signal (shown as gold from the overlap of the green and red images), indicating that the complex was localising within the nucleoli and lysosomes of the HeLa cells.

Similar observations could be seen for **Pt_PEM_4OMe** with the MitoTracker and LysoTracker co-stains (figure 3.34).

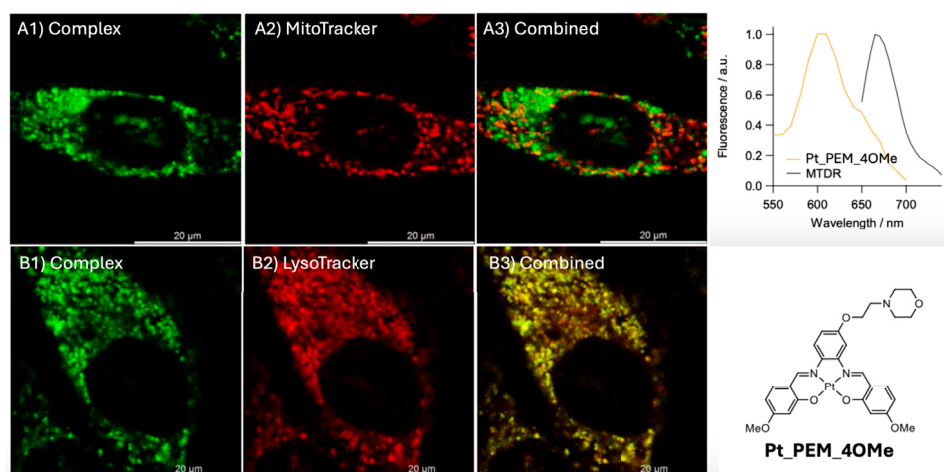


Figure 3.34. HeLa cells stained with 10 μM **Pt_PEM_4OMe** - 6 hours (A1, B1), 50 nM MitoTracker – 5 mins (A2), and 50 nM LysoTracker – 5 mins (B2), and the overlaid images for **Pt_PEM_4OMe** with MitoTracker (A3) and **Pt_PEM_4OMe** with LysoTracker (B3). **Pt_PEM_4OMe** detection: excitation 514 nm, emission 590 - 660 nm. MitoTracker detection: excitation 514 nm, emission 680 – 720 nm. LysoTracker detection: excitation 633 nm, emission 680 – 720 nm. Images taken by Dr Petr Sherin.

From the images it was clear that **Pt_PEM_4OMe** was also localising within the lysosomes of the HeLa cells. Interestingly there was much less nucleoli staining within this image than was seen previously for **Pt_PEM_4OMe** in figure 3.31 (1 hour incubation). This could be explained by the observed time-dependent localisation behaviour of the platinum(II) salphen complexes which is shown in figure 3.36.

Co-stained images for **Pt_PEM_Ph** were only obtained with MitoTracker (figure 3.35).

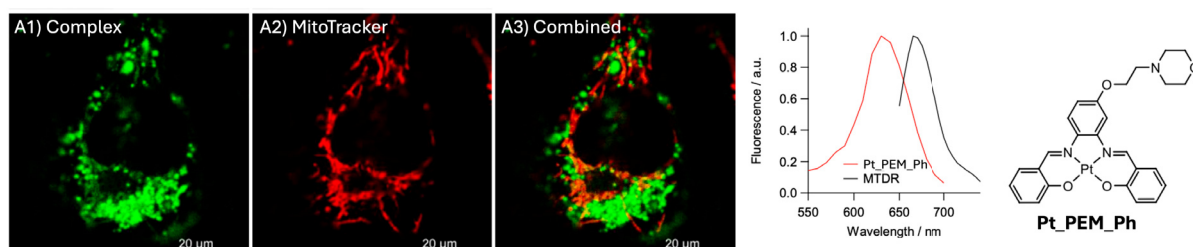


Figure 3.35. HeLa cells stained with 10 μM **Pt_PEM_Ph** - 6 hours (A1), 50 nM MitoTracker – 5 mins (A2), and the overlaid image for **Pt_PEM_Ph** with MitoTracker (A3). **Pt_PEM_Ph** detection: excitation 514 nm, emission 590 - 660 nm. MitoTracker detection: excitation 514 nm, emission 680 – 720 nm. Images taken by Dr Petr Sherin.

This confirmed that the complex was not localising within the mitochondria of the HeLa cells. As the staining appeared to have a very similar pattern to **Pt_AEP_4OMe** and **Pt_PEM_4OMe** it could be deduced that this complex was likely also localising within the lysosomes.

Finally, co-stained images were obtained for **Pt_PPTMA_4NEt₂** with MitoTracker (figure 3.36).

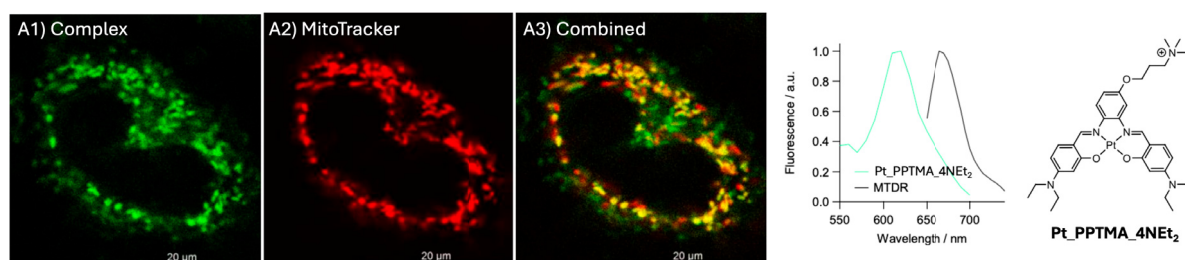


Figure 3.36. HeLa cells stained with 10 μ M **Pt_PPTMA_4NEt₂** - 6 hours (A1), 50 nM MitoTracker – 5 mins (A2), and the overlaid image for **Pt_PPTMA_4NEt₂** with MitoTracker (A3). **Pt_PPTMA_4NEt₂** detection: excitation 514 nm, emission 590 - 660 nm. MitoTracker detection: excitation 514 nm, emission 680 – 720 nm. Images taken by Dr Petr Sherin.

Within these images it appeared that the positively charged complexes were localising within the mitochondria of the HeLa cells. This was likely due to the negative charge of the mitochondrial matrix, which attracts the positively charged **PPTMA DAB**. It has also reported in the literature that the presence of charged groups, such as NMe_3^+ and PPh_3^+ , result in mitochondria localisation [307–310]. As the staining of **Pt_PPTMA_4Br** and **Pt_PPTMA_Naph** share very similar patterns to **Pt_PPTMA_4NEt₂** which has been confirmed to localise within the mitochondria, it could be inferred that the other **PPTMA** complexes share a similar localisation. It was not possible to conclude if the **CBA** had any effect on the cellular localisation of these complexes.

During these imaging experiments it was noted that the staining of the complexes was changing overtime, therefore a time dependent imaging experiment was performed which involved imaging the complexes at incubation times of approximately 30 minutes, 1 hour, 6 hours, and for **Pt_AEP_4OMe**, 22 hours. Figure 3.36 shows the fluorescent images for each complex and the precise time point of imaging.

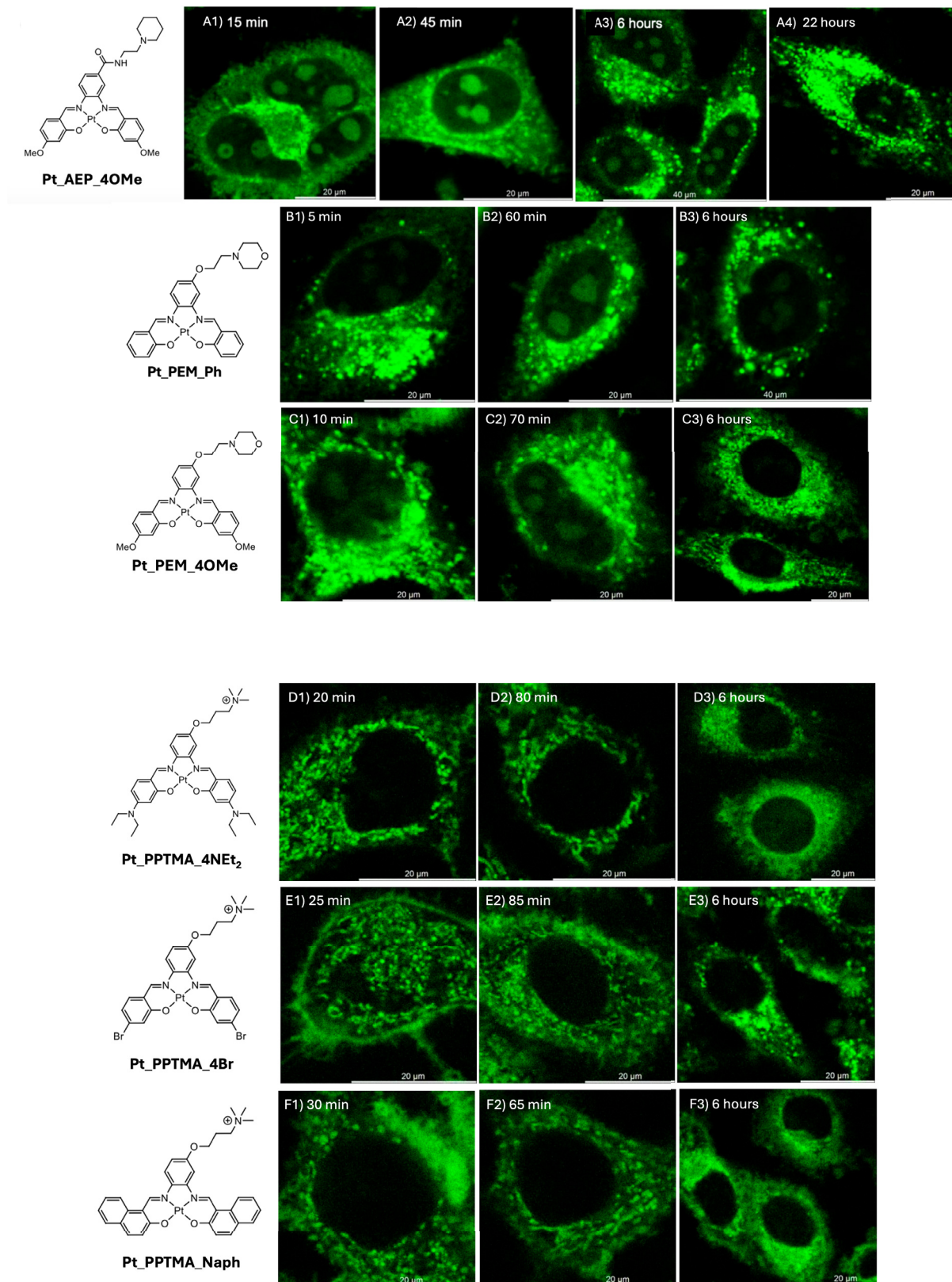


Figure 3.37. Time dependent images of the platinum(II) salphen complexes all excited at 514 nm. A) **Pt_AEP_4OMe** at 1) 15 mins 2) 45 mins 3) 6 hours 4) 22 hours. Emission measured from 540 – 660 nm. B) **Pt_PEM_Ph** at 1) 5 mins 2) 60 mins 3) 6 hours. Emission measured from 590 – 660 nm. C) **Pt_PEM_4OMe** at 1) 10 mins 2) 70 mins 3) 6 hours. Emission measured from 590 – 660 nm. D) **Pt_PPTMA_4NEt₂** at 1) 20 mins 2) 80 mins 3) 6 hours. Emission measured from 590 – 660 nm. E) **Pt_PPTMA_4Br** at 1) 25 mins 2) 85 mins 3) 6 hours. Emission measured from 590 – 680 nm. F) **Pt_PPTMA_Naph** at 1) 30 mins 2) 65 mins 3) 6 hours. Emission measured from 630 – 700 nm. Images taken by Dr Petr Sherin.

The complexes **Pt_AEP_4OMe** (A), **Pt_PEM_Ph** (B) and **Pt_PEM_4OMe** (C) demonstrated a fast penetration of the cells, followed by accumulation within the nucleoli and lysosomes overtime. **Pt_PEM_Ph** and **Pt_PEM_4OMe** showed maximum nucleoli staining after approximately 1 hour (B2 and C2). After extended incubation time it appeared that the complexes were changing localisation from the nucleoli to the lysosomes (B3 and C3). Contrastingly, **Pt_AEP_4OMe** demonstrated rapid localisation within the nucleoli (A1) and remained there for an extended period (A2 and A3), only appearing to leave the nucleoli after overnight incubation (A4).

Complexes containing the **PPTMA** containing ligands (i.e. **Pt_PPTMA_4NEt₂** (D), **Pt_PPTMA_4Br** (E) and **Pt_PPTMA_Naph** (F)) demonstrated a rapid penetration into the mitochondria of the HeLa cells (D1, E1 and F1). After extended incubation periods the pattern of staining appeared more granular (D3, E3 and F3), becoming similar to the cytoplasmic staining patterns of **Pt_AEP_4OMe**, **Pt_PEM_Ph** and **Pt_PEM_4OMe**. This suggested that overtime the complexes were being actively removed from the mitochondria into lysosomes. These observations are promising in the context of photodynamic therapies as they imply that the cells are able to uptake, and then remove the complexes overtime, preventing bioaccumulation.

3.6 Correlation analysis of photophysical and biological properties

As previously mentioned, photosensitisers should be non-toxic under dark conditions. Upon irradiation they should generate ROS efficiently, and result in a high phototoxicity. Complexes need to be cell permeable and localise in areas of the cell which enable

phototoxicity. If the cytotoxicity of the complex is mediated by DNA damage it would be desirable for complexes to have selectivity for G4 DNA over duplex DNA, leading to preferential localisation within the nucleus of cancer cells, which are enriched with G4 DNA. However, the cellular data has shown that the most phototoxic complexes do not localise within the nucleus, and therefore DNA binding may not be important for the platinum(II) salphen complexes mechanism of phototoxicity.

Designing a complex to have all of these properties is no simple feat, requiring an intricate balance of properties which are tuneable through structural modifications which are not always complementary. Therefore, to assess how the experimental properties relate to one another Pearson's correlation (PC), a statistical correlation assessment, was performed on all the experimental data. PC is described in more detail in section 4.3. This statistical analysis uncovered some expected, and unexpected property relationships (figure 3.38).

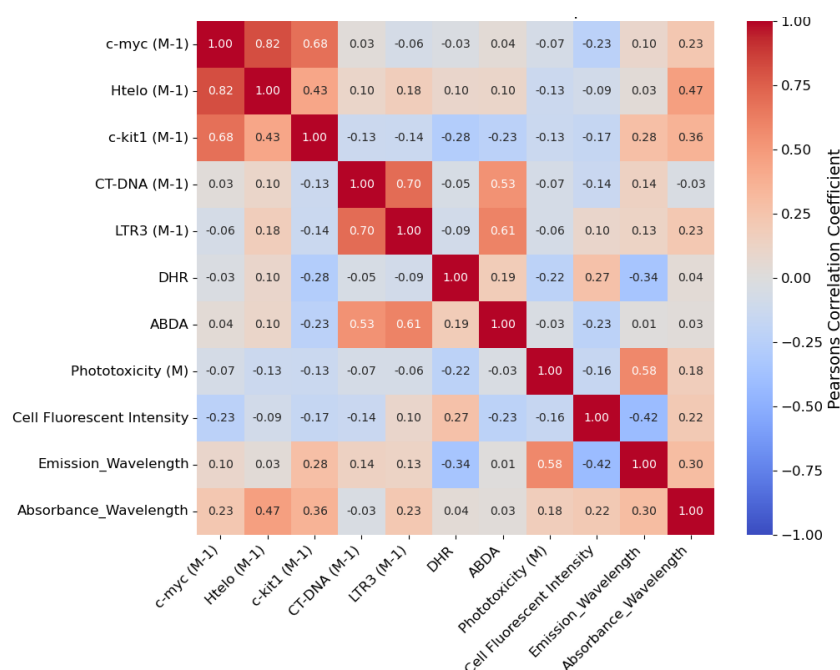


Figure 3.38. PC coefficient heat map assessing correlations between experimental properties.

As expected, there was a correlation between the binding affinities of the three G4 DNA structures: *c-myc* (parallel), Htelo (hybrid-2) and *c-kit1* (anti-parallel). The similarities were likely due to the similar binding mode of π - π stacking interactions with the terminal g-quartets. G4 binding did not correlate to the binding affinities of LTR(III) (hybrid 3+1

with stem-loop) and CT-DNA (duplex). LTR(III) has a much more restricted G4 binding site in which complexes must fit into the cavity between the G4 and the duplex loop, and CT-DNA binding is likely occurring via a mixture of groove binding and intercalation. The strong correlation between the binding affinities for LTR(II) and CT-DNA suggests they may be binding by similar mechanisms. LTR(III) groove binding was postulated in section 3.3.3 for the platinum(II) salphen complexes which were unable to fit in the binding cavity, such as the **Naph** containing complexes, which resulted in poor switch-on of emission upon binding.

Surprisingly there was no correlation detected between the two ROS generating assays (i.e. DHR and ABDA). This suggests that the different types of ROS occur independently and are determined by separate factors of the complex.

Phototoxicity and cell fluorescence intensity were not correlated to ROS generation or DNA binding properties. In the context of phototoxicity, this is likely due to the complexity of the measurement having several dependencies including cell permeability, cell localisation, *in cellulo* ROS generation and the resultant damage to the cell. The same conditions apply for cell fluorescent intensity, with the addition of how well these complexes bind to cellular components which provide protection from the aqueous environment of the cell to prevent non-radiative decay. This was somewhat surprising as it would be expected that complexes which have better cellular uptake, indicated by greater cellular fluorescence intensity, would result in higher phototoxicity,

Most surprisingly was the correlation between emission wavelength, DHR (ROS type 1), phototoxicity, and cell fluorescence intensity. These may be correlated due to their relation to the electronic properties of the complex and how the complexes interact with light. Although with that in mind it is surprising that the peak absorbance wavelength was not correlated to any of these features.

Overall, the relationships are highly complex and difficult to consider within the design of a complex using typical medicinal chemistry heuristics. This does however pose an interesting challenge for the application of AI, which is known to handle problems of high dimensional space well through pattern recognition. By mapping molecular descriptors to target properties in multi-dimensional space it would be possible to

identify compounds which are predicted to have an optimal balance of target properties. As mentioned in chapter 1, AI models are already being used within medicinal chemistry for predicting drug-like properties [97, 112–116].

3.7 Screening conclusions

To conclude, within this chapter the photophysical properties, ROS generation, DNA binding, DNA damage, and the cellular behaviour of 22 platinum(II) salphen complexes were investigated. The **CBA** and **DAB** complex modifications resulted in varied performances across all properties. The key observations from the data presented in this chapter are as follows:

- Emission energy: Functionalities in the 4 and 5 **CBA** position resulted in a hypsochromic and bathochromic shift of the peak emission energy respectively. Electron donating groups, such as **OMe** had the strongest influence on this observation. The **DAB** had very little influence on the peak emission wavelength.
- ROS generation: The **DAB** had a significant influence on ROS generation, with the **PEM DAB** quenching type 2 ROS. It was not possible to deduce clear trends between the **CBA** functionalities and ROS generation.
- DNA binding: All complexes displayed preferential binding to G4 DNA over duplex DNA. Each G4 topology had distinct binding preferences with different **CBA** components. **Naph** containing complexes performed well across all G4 topologies, likely due to the increased aromatic surface (except with LTR(III)). The presence of functionalities on the **CBA** component of the complexes appeared increase G4 binding affinity, particularly in those which could participate in hydrogen bonding interactions (**Ph CBA** containing complexes performed generally poorly). Furthermore, although docking was unable to provide distinguishable docking scores between the complexes, it was still useful for providing insights into the binding modes. For example, it was seen that the nature of the **DAB** could influence binding through additional bonding interactions. In addition to highlighting the size incompatibility of the **Naph** containing complexes and the LTR(III) binding pocket, which resulted in no switch on of emission.

- DNA damage: Only moderate ROS activity was required for DNA damage (**Pt_PEM_4OMe** type 1: 50 %, type 2: 13 %). Complexes which had poor generation of both type 1 and type 2 ROS were unable to damage DNA (**Pt_PEM_5Br** type 1: 12 %, type 2: 10 %). Complexes which had good type 1 ROS generation, but poor type 2 generation were still able to damage DNA (**Pt_PEM_4OMe** type 1: 50 %, type 2: 13 %). ROS-quenching assays revealed that DNA damage was induced by both type 1 and type 2 ROS. Finally, the platinum(II) salphen complexes induced G4 DNA damage via oxidation rather than strand cleavage. This observed difference between the cleavage of plasmid DNA and the oxidation of G4 DNA could be a result of the unique DNA topologies and binding interactions with the platinum(II) salphen complexes. Within the plasmid DNA, the phosphate backbone is more exposed, and likely the site of complex binding. Whereas in G4 DNA, complexes bind to the G-quartet and therefore the ROS are more likely to react with the guanine residues over the DNA phosphate backbone, resulting in oxidation as opposed to DNA cleavage.
- Cellular toxicity: Complexes were found to be non-cytotoxic in the dark up to 25 μ M. Upon light irradiation, 68 % of complexes became highly phototoxic (IC_{50} values below 8.8 μ M) as a result of ROS generation. However, there was not a direct correlation between ROS generation and phototoxicity. Except in the case of **Pt_PPTMA_Naph** (type 1 ROS: 24 %, type 2 ROS: 131%) which displayed the highest type 2 ROS generating ability, and the highest phototoxicity IC_{50} value of 123 nM. However, the next most phototoxic complex was **Pt_PEM_Naph** (type 1 ROS: 29 %, type 2 ROS: 12 %) which displayed a moderate to low ROS generation and had an IC_{50} value of 210 nM. The G4 binding ability of the most phototoxic complexes were worse than that of the least phototoxic complexes, suggesting G4 binding may not play a role in the phototoxicity mechanism of the complexes.
- Cellular localisation: The HT imaging screen unveiled that most complexes were cell permeable, except for **Pt_PPTMA_3F** and **Pt_PPTMA_5OMe**. Further confocal imaging is required to confirm if all other complexes also localise within the discussed organelles, as this was unclear from the lower resolution

HT imaging screen. Interestingly, the cellular localisation did not appear to significantly influence the phototoxicity, with the most phototoxic complexes **Pt_PPTMA_Naph** localising within the mitochondria, and the next most phototoxic complex, **Pt_PEM_Naph**, localising within the nucleoli and lysosomes. However, other complexes which were also determined to localised within the discussed organelles, had much higher IC₅₀ values (less toxic): **Pt_AEP_4OMe** (type 1 ROS: 34 %, type 2 ROS: 22 %) localised within the nucleoli and lysosomes, and only achieved a phototoxicity IC₅₀ value of 2.57 µM. **Pt_PPTMA_4Br** (type 1 ROS: 38 %, type 2 ROS: 47 %) localised within the mitochondria, and only achieved phototoxicity IC₅₀ value of 3.18 µM.

Overall, it is challenging to definitively say which structural modifications best promote each individual property, and how these properties can complement one another, which as mentioned above, could be an interesting challenge for AI approaches if there were greater volumes of experimental data available.

With the results at hand, **Pt_PEM_4OMe** and **Pt_AEP_4OMe** were identified as lead candidates for further biological and *in vivo* testing. Both complexes display good G4 binding affinities, and selectivity over duplex DNA, with strong photo-cleavage abilities. In addition to good cellular uptake in HeLa cells, where they were non-toxic until irradiated, after which they achieve phototoxicity IC₅₀ values of 294 nM and 2.57 µM respectively. Figure 3.39 shows the properties of these lead complexes compared to the highest achieved measurements across all the platinum(II) salphen complexes.

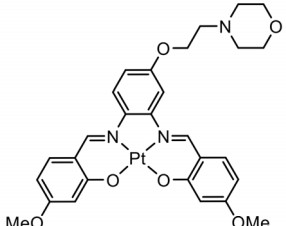
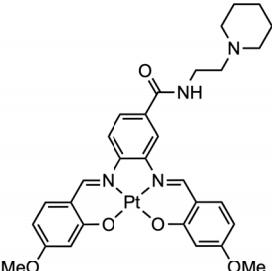
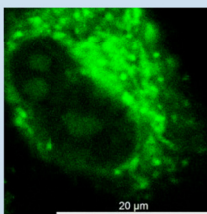
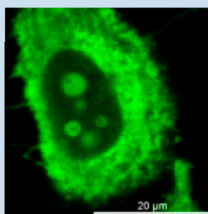
	 Pt_PEM_4OMe	 Pt_AEP_4OMe	Best in class
c-myc binding affinity (M⁻¹)	2.41 x 10 ⁵	5.73 x 10 ⁵	3.96 x 10⁶
ROS: type 1, type 2 (%)	47, 13	34, 22	90, 131
Phototoxicity IC₅₀ (M)	2.94 x 10 ⁻⁷	2.57 x 10 ⁻⁶	8.87 x 10⁻⁸
Cell localisation	Nucleoli + Lysosomes 	Nucleoli + Lysosomes 	

Figure 3.39. Measured properties of the lead complexes **Pt_PEM_4OMe** and **Pt_AEP_4OMe** compared to the best in class of each property.

Chapter 4

Computational modelling of platinum(II) salphen geometries and photophysical properties

To explore relationships between the electronic system of the platinum(II) salphen complexes and their photophysical and biological properties (chapter 3), density functional theory (DFT) and time dependent DFT (TDDFT) calculations were performed. From DFT the HOMO and LUMO orbital diagrams and energies were obtained. In addition to the S_1 to S_0 and T_1 to S_0 transition energies from TDDFT.

4.1 Introduction to DFT and TDDFT

DFT is a quantum mechanical method which can be used to model the geometry and electronic structure of molecules. DFT provides an alternate framework for solving the Schrödinger equation (4.1): a quantum mechanics equation which describes the behaviour of many electron quantum systems [311].

$$H\psi = E\psi \quad (4.1)$$

The Hamiltonian (H) represents the total energy of a quantum system: the kinetic and potential energy. To determine the energy of the system the Schrödinger equation (4.1) must be solved to find the many-electron wavefunction and eigenvalue for the energy of the system. The Hamiltonian is described in equation 4.2.

$$H = \left[-\frac{\hbar^2}{2m} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^N V(r_i) + \sum_{i=1}^N \sum_{j<i} U(r_i, r_j) \right] \quad (4.2)$$

- $\hbar$: Reduced Planck's constant.
- m : The mass of each particle.
- N : The number of particles (electrons or other interacting particles in the system).
- ∇_i^2 : The Laplacian operator representing the kinetic energy of particle i .
- r_i : The position vector of the i -th particle.
- $V(r_i)$: The external potential acting on the i -th particle.
- $U(r_i, r_j)$: The Coulomb repulsion interaction potential between particles i and j .

The many-electron wavefunction (ψ) is an approximated product of each single electron wavefunction as seen in equation 4.3.

$$\psi(r_1, r_2, \dots, r_n) \approx \psi_1 \psi_2 \dots \psi_n \quad (4.3)$$

As the number of spatial degrees of freedom ($3N$: x, y, z) in a system of N particles increases, the Schrödinger equation becomes increasingly more difficult to solve due to the growing dimensionality of the wavefunction (ψ). As a result, for many body systems $3N$ increases exponentially. DFT addresses this problem by using electron density ($n(r)$) in place of the many-body wavefunction. This fundamental principle of DFT is based on the Hohenberg-Kohn theorem, which proves that the ground state electronic density of a molecule determines the external potential of electron-electron interactions and consequentially, all properties of the system. This led to the Kohn-Sham (KS) equation (equation 4.4) which maps the system into non-interacting electrons, significantly simplifying the calculations and providing an approximate solution to the Schrödinger equation. The KS equation is a set of one-electron equations for non-interacting particles [312, 313].

$$\left\{ -\frac{1}{2} \nabla^2 + v_s[n](r) \right\} \phi_i(r) = \epsilon_i \phi_i \quad (4.4)$$

- ∇^2 : Laplacian operator (kinetic energy of non-interacting particles).
- $v_s[n](r)$: KS potential (functional of electron density ($n(r)$)).

- $\phi_i(r)$: KS orbitals (wavefunction).
- ϵ_i : Eigenvalues (energies) of KS orbitals.
- r : Spatial coordinate.

The goal of the KS equation is to determine the KS orbitals and associated energies. KS orbitals are mathematical functions which describe electron density as non-interacting particles, simplifying the many-electron system to a single electron framework, reducing the $3N$ dimensions of the Schrödinger equation to $1N$. The electron density of the system is obtained as the sum of the squared KS orbitals: the core output of DFT calculations.

KS orbitals are dependent on the KS potential which consists of three components. The Hartree potential describes the electrostatic repulsion of electrons, and the external potential describes the attraction between nuclei and electrons. The third component is the exchange-correlation potential which is a functional of the ground state density, which provides the exact properties of the interacting system, encapsulating all the complex many-body effects that are not captured by other terms. This potential is the only term that requires approximation, making it a crucial part of DFT development.

As the KS potential is the functional density, and the functional density is the sum of the KS orbitals, the KS equation is self-consistent, and so iteration methods are required to find the solution. This occurs by starting with an initial guess of the electron density, calculating the effective KS potential, solving the KS equation to obtain new orbitals, then calculating a new electron density from these orbitals. This cycle is repeated until the electron density value converges.

The accuracy of DFT depends critically on the choice of exchange-correlation functional. As a result, various approximation methods for the exchange-correlation functional have been developed to improve the accuracy of DFT calculations, with different functionals best suited for different tasks. There are three key classes of functionals: Local Density Approximation (LDA) which works well on simple systems where electron density does not vary significantly. Generalised gradient approximation (GGA) which

uses information about the gradient of electron density to account for rapid changes in electron density. Along with hybrid functionals which combine exact Hartree-Fock exchange with DFT approximations. Hartree-Fock exchange is used to approximate the wavefunction and energy of a many-body quantum system. The hybrid B3LYP functional was used for the platinum(II) salphen complexes.

The final component of the DFT calculation to consider is the basis sets, which are mathematical functions which represent the molecular orbitals (ψ_i). The final molecular wavefunction is constructed as a linear combination of these basis functions (ϕ_μ), defining the electronic structure of the system (equation 4.5).

$$\psi_i(r) \approx \sum_{\mu} c_{\mu i} \phi_{\mu}(r) \quad (4.5)$$

- $\psi_i(r)$: Molecular orbital at position r
- $\phi_{\mu}(r)$ Basis function
- $c_{\mu i}$: Expansion coefficient (determines contribution of each basis function to the molecular orbital).

Basis sets can vary in size and complexity through changing the number of functions used to describe each atoms valence electrons, orbital shape, and polarizability. Within this work the basis set 6-31G(d) was used for all non-metal components of the complex. This basis set uses Gaussian-type orbitals which model the spatial distributions of electron density around an atom based on the characteristic bell-shaped Gaussian distribution curve. This basis set uses 6 Gaussian functions for the core electrons and 3 Gaussian functions for the valence electrons in a split-valence configuration. In addition to a polarization function (d) which helps to describe distortion during chemical bonding. This is particularly important for the context of metal complexes as the metal to ligand bond lengths can vary. For the platinum centre the basis set LANL2DZ was used. This basis set uses a double-zeta basis, consisting of two basis functions to represent each atomic orbital, one for valence electrons and the other for the effective core potential in place of core electrons. This is important due to platinum's large size and therefore large number of core electrons. The combination of the described basis

sets and functionals have previously been used for DFT calculations with platinum(II) salphen complexes, and have been validated further by comparison to bond lengths and angles of x-ray crystal structures of reported platinum(II) salphen complexes [289, 314, 315]. Complexes were modelled in DMSO using a conductor-like polarizable continuum model (cPCM) which simulates the solvent environment by applying the dielectric constant of the solvent to the system. The purpose of this additional calculation feature was to ensure that the computational results would be directly comparable to the experimental photophysical measurements, which were performed in DMSO. Furthermore, comparison studies of a range of alternative functionals and basis sets were performed by MSci Joe Burke, validating the performance of B3LYP with 6-31G(d) and LANL2DZ for platinum(II) salphen complexes.

When solving for the energy of the entire system, DFT determines the molecular orbitals and their corresponding energies from the KS equation. The orbitals are then sorted by energy and electron occupancy. The HOMO is identified as the highest energy orbital which contains electrons, and the LUMO is the lowest energy orbital without electrons.

TDDFT is an extension of DFT which incorporates time-dependent external potentials to calculate excited state properties including excitation energies and oscillation strengths. These properties enable calculation of the spectroscopic properties of molecules. TDDFT is based on the time dependent Schrödinger equation which describes the changes in a quantum system under time-varying perturbations (equation 4.6).

$$\hat{H}(t)\psi(t) = i\frac{\partial}{\partial t}\psi(t) \quad (4.6)$$

- $\hat{H}(t)$: Time-dependent Hamiltonian operator.
- $\psi(t)$: The wavefunction of the quantum system at a given time t .
- i : The imaginary unit ($i^2 = -1$).
- $\frac{\partial}{\partial t}$: The time derivative operator, describing how the wavefunction $\psi(t)$ changes with respect to time (t).

In which the Hamiltonian consists of three components (4.7):

$$\hat{H} = \hat{T} + \hat{V}_{ext}(t) + \hat{W} \quad (4.7)$$

- $\hat{T}$: Kinetic energy operation.
- $\hat{V}_{ext}(t)$: Time-dependent external potential.
- $\hat{W}$: Electron-electron interaction operator.

TDDFT translates the time dependent Schrodinger equation to a more manageable calculation using the principles established by Hohenberg and Kohn in a time (t) dependent context. This is known as the Runge-Gross theorem which establishes a one-to-one correspondence between the time-dependent external potential and the time-dependent electron density for a given initial state. Meaning, the time-dependent many-body wavefunction is equivalent to the time-dependent electron density, resulting in the time-dependent KS equation [316, 317] (equation 4.8).

$$\left(-\frac{1}{2}\nabla^2 + v_s(r, t) \right) \phi_j(r, t) = i \frac{\partial}{\partial t} \phi_j(r, t) \quad (4.8)$$

- ∇^2 : Laplacian operator, representing the second spatial derivative with respect to position r , accounting for kinetic energy.
- $v_s(r, t)$: Time-dependent KS potential, including the external potential and the self-consistent interaction with the electron density at position r and time t .
- $\phi_j(r, t)$: Time-dependent KS orbital for the j -th electron, which describes the quantum state of the electron at position (r) and time (t).

TDDFT evolves the initial ground-state orbitals, which are calculated using DFT, over time under the influence of an external perturbation, such as the oscillating electric field of a light pulse. This perturbation excites the electrons from the ground state to higher-energy states, such as the singlet excited state (S_1) or the triplet excited state (T_1). The excitation energy is then determined by the energy difference between the excited state and ground state. The transition energies calculated by TDDFT (i.e. S_1 to S_0 if the molecule is fluorescent, and T_1 to S_0 if the molecule is phosphorescent) correspond to the molecule's emission energy and hence peak emission wavelength, as shown in the Jablonski diagram in figure 4.1 [318]. Using Planck's equation (equation

4.9) the emission energy can be converted to wavelength to provide the peak emission wavelength and vice versa.

$$E = \frac{hc}{\lambda} \quad (4.9)$$

- E : Energy (eV).
- h : Planck's constant (4.14×10^{-15} eV·s).
- c : Speed of light in a vacuum (3.00×10^8 m·s⁻¹).
- λ : Wavelength of the electromagnetic wave (m).

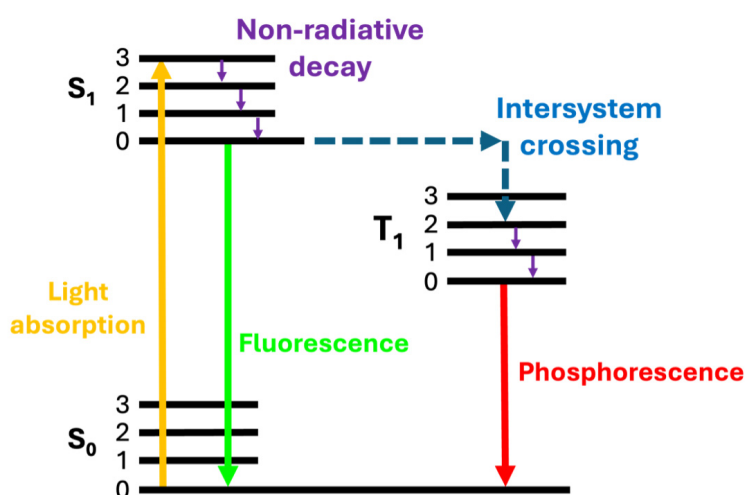


Figure 4.1. Simple Jablonski diagram showing the energy levels from which fluorescence and phosphorescence originate. The levels 0, 1, 2, and 3 represent vibrational energy states within each energy level. Energy moves between vibrational levels via intersystem crossings.

4.2 Relationships between DFT and photophysical properties

DFT and TDDFT calculations can be used to rationalise the effect of substitution on the photophysical properties of the platinum(II) salphen complexes. As was discussed in chapter 3, the nature and position of substituents on the **CBA** rings of the complexes have a distinct effect on the peak emission wavelength. This same trend was observed for the T₁ to S₀ transition energies of the complexes. This can be seen within a strong positive correlation between T₁ to S₀ transition energies and experimental emission

wavelengths (plots A and B in figure 4.2). The HOMO LUMO gap is not as strongly correlated to the emission wavelengths, but a general trend can be seen between larger HOMO LUMO gaps and lower emission wavelengths (plots C and D in figure 4.2). Plots are coloured by the complexes **DAB** (A and C) and **CBA** (B and D) components.

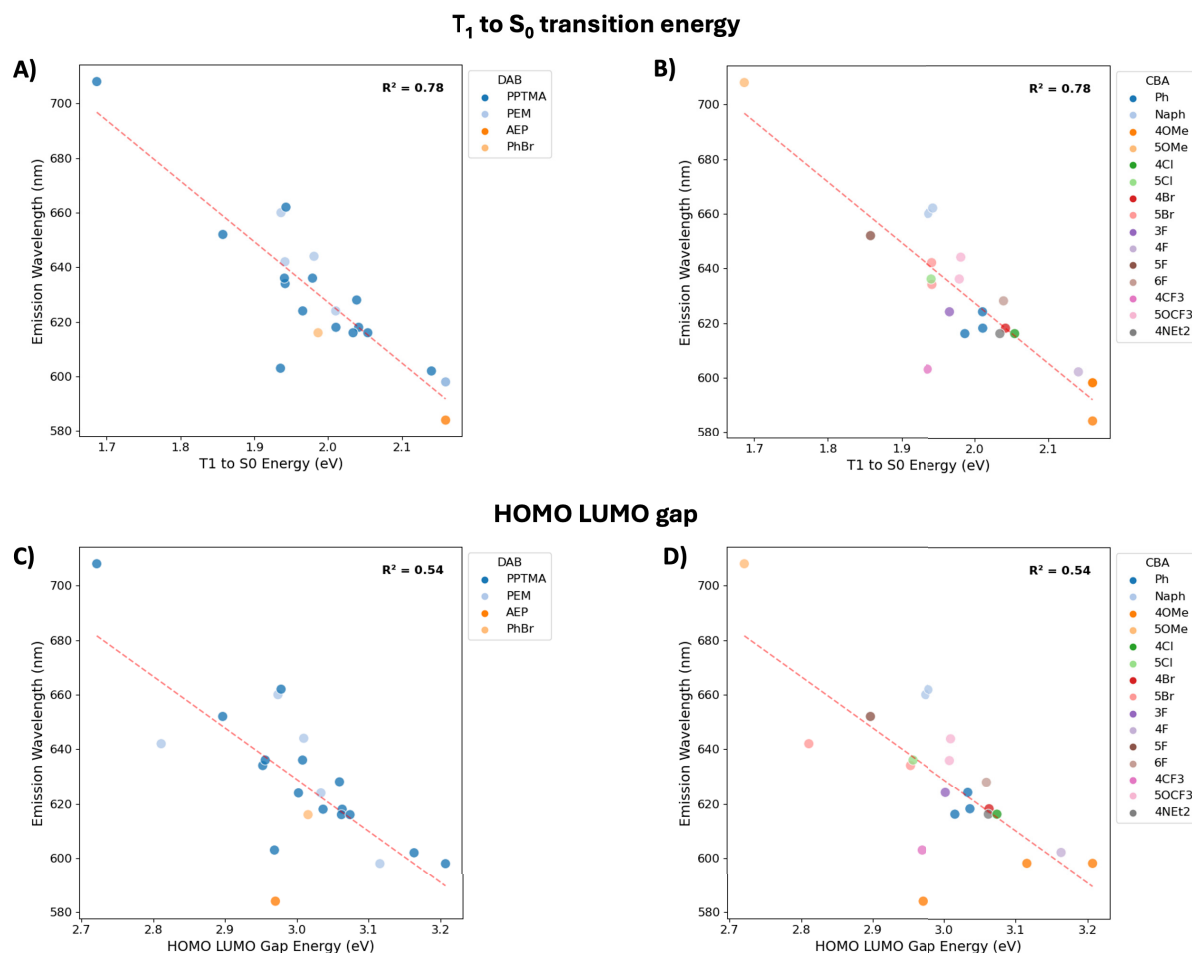
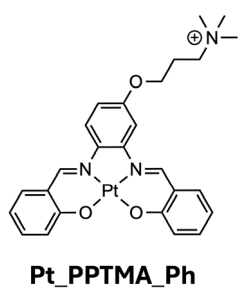
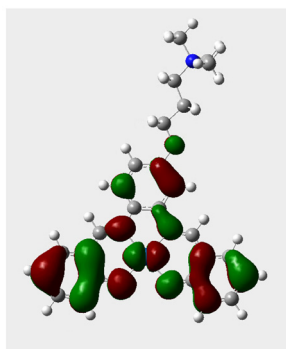


Figure 4.2. Scatter plots showing the T₁ to S₀ transition energy vs emission wavelength (plots A and B) and the HOMO LUMO gap vs emission energy (plots C and D), coloured by the **DAB** (plots A and C) and **CBA** (plots B and D) complex components.

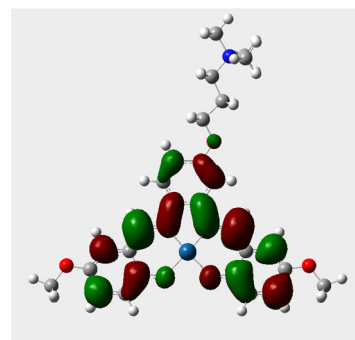
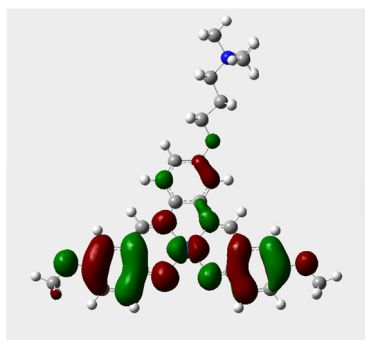
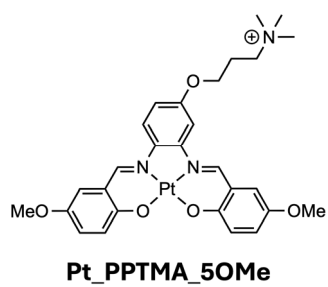
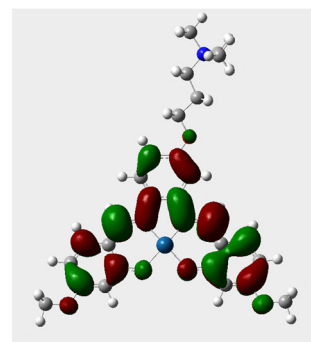
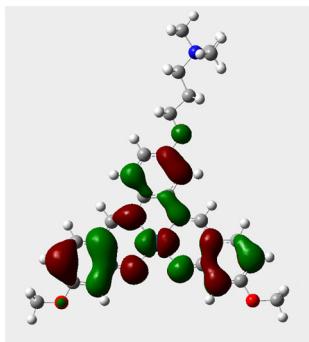
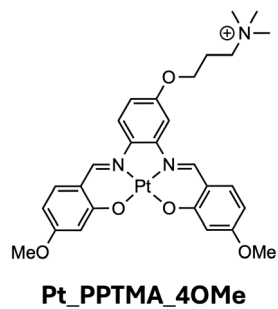
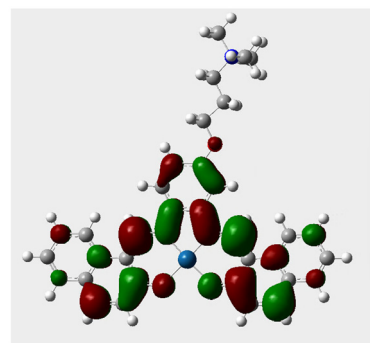
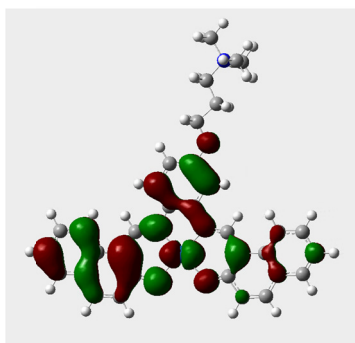
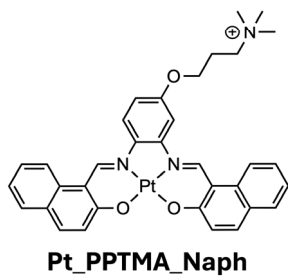
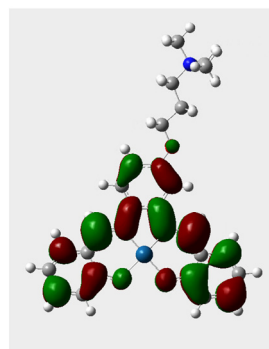
Within the plots it can be seen that the different **CBA** functionalities not only influence the emission wavelengths, but also the HOMO LUMO gap and the T₁ to S₀ transition energy. This can be seen by distinct groupings of complexes with the same **CBA** groups (plots B and D). These trends can be explained further by visualising the HOMO and LUMO orbitals of the complexes: The position of the substituent determines its contribution to the HOMO or the LUMO orbitals, causing either stabilisation (- eV) or destabilisation (+ eV) of the orbital energies (figure 4.3).



HOMO



LUMO



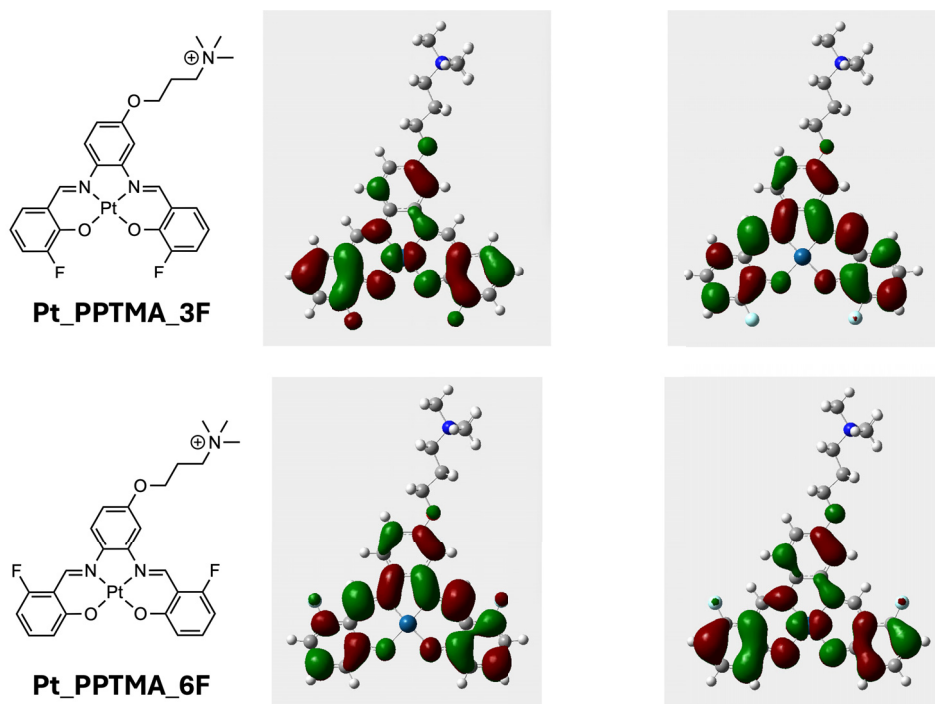


Figure 4.3. HOMO (right) and LUMO (left) of platinum(II) salphen complexes.

Substituents in the 5-position contribute to the HOMO of the complex, whereas substituents in the 4-position contribute to the LUMO. This effect can be seen most explicitly in **Pt_PPTMA_4OMe** and **Pt_PPTMA_5OMe** which have the lowest and highest emission wavelengths of the series respectively. Of the series, **Pt_PPTMA_4OMe** also has the largest T_1 to S_0 transition energy and HOMO LUMO gap. Conversely, **Pt_PPTMA_5OMe** has the smallest T_1 to S_0 transition energy and HOMO LUMO gap. This is due to the methoxy in the 5-position (para) having significant anti-bonding contribution to the HOMO, resulting in a relative destabilisation of +0.29 eV. Whereas the methoxy in the 4-position has a significant anti-bonding contribution to the LUMO, causing a relative destabilization of +0.21 eV. As a result, the HOMO LUMO gap for the **Pt_PPTMA_5OMe** complex is reduced (2.69 eV), resulting in a higher emission wavelength. Whilst the HOMO LUMO gap for the **Pt_PPTMA_4OMe** complex increases (3.19 eV), resulting in a lower emission wavelength.

The same trend is seen for the other functionalities where substitutions in the 4 or 5 position cause LUMO or HOMO destabilisation, and therefore a decrease or increase in emission wavelength respectively. The energies of the **6F** and **3F** complexes lie in between the energies of the **4F** and **5F** substituted complexes. This is because

the substituents in the 6 and 3 positions contribute to both the HOMO and the LUMO. Substitution in the 6 position results in a main contribution into the LUMO with a slightly larger contribution to the HOMO than substitution in the 3 position, resulting in a smaller difference between the HOMO LUMO gap of the **3/6F** substituted complexes than compared to the 4/5 substituted **F** complexes.

For the **Naph** substituted complexes, the HOMO and LUMO are delocalised across a larger conjugated system resulting in a small relative stabilisation of both the HOMO (-0.06 eV) and the LUMO (-0.005 eV). The HOMO of the Naph complexes are asymmetrical, with greater electron density on the aromatic ring opposite to the linking group on back of the **DAB**. Overall, there is no strong influence from the **DAB** on the HOMO or LUMO orbital systems, explaining why the **DAB** does not significantly impact the emission wavelength or energy values calculated by DFT for complexes with varying **DAB** and constant **CBA** components.

4.3 Correlation analysis of DFT energies and experimental properties

Following the generation of DFT and TDDFT data for the platinum(II) salphen complexes, statistical analysis was performed to investigate their relationships with the experimental results obtained within chapter 3. The goal of this section is to uncover trends between complex activity and calculable properties, as well as assess the viability of using computational data in place of experimental data to perform ML prediction tasks.

The key statistical tool implemented was the Pearson's correlation (PC) coefficient which measures linear relationships between 2 variables as seen in equation 4.10 [319]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4.10)$$

- n : Number of data points.
- x_i and y_i : Individual data points indexed with i .

- $\bar{x}$ and $\bar{y}$: Sample means of x (equation 4.11) and y (equation 4.12):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.11)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4.12)$$

For the PC coefficient to be considered significant: $r > 0.3$. PC was performed on the experimental data and the TDDFT data, figure 4.4 shows the resulting heat map.

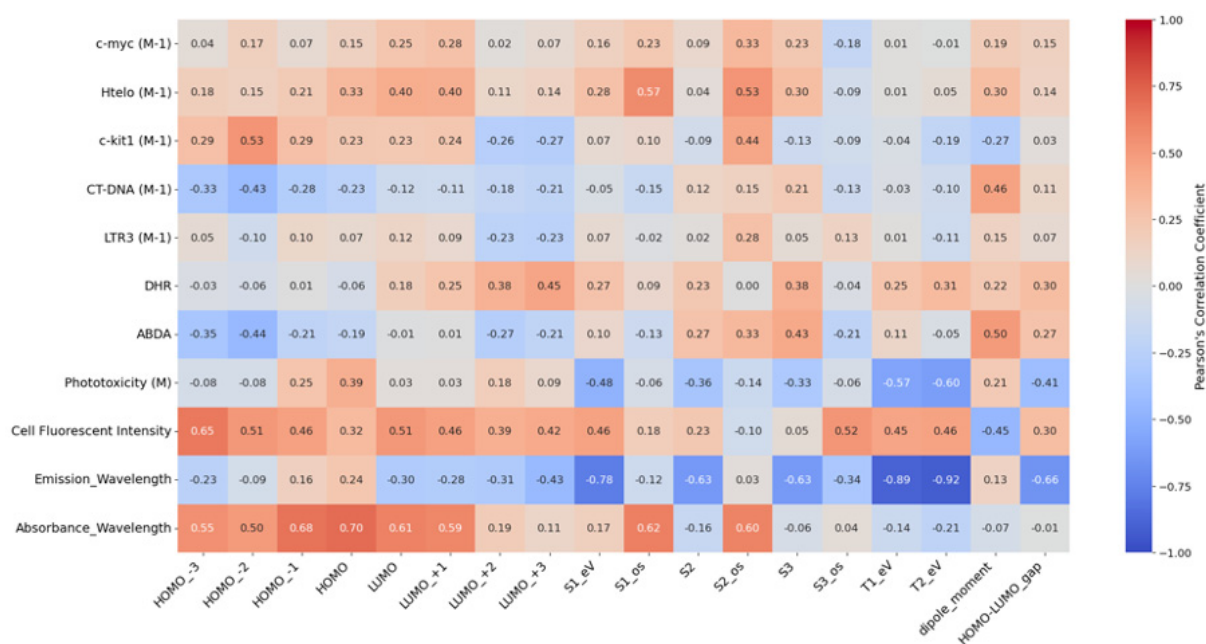


Figure 4.4. PC heat map assessing correlations between experimental properties against DFT and TDDFT calculated energies. A strong positive correlation is represented as dark red (+1). A strong negative correlation is represented as dark blue (-1). No correlation is represented as grey.

When assessing the binding affinities for the G4 sequences there were some moderate to strong correlations to the HOMO and LUMO energies (HTelo correlated by 0.40 to LUMO+1 and 0.53 to the S_2 oscillator strength). *c-kit1* correlated by 0.53 to HOMO-2. Suggesting that the electronic and orbital properties of complexes influence G4 binding affinity: it is known that electron deficient aromatic systems bind more favourably to the electron rich aromatic systems of the G-quartets [174]. In contrast, CT-DNA and LTR(III) showed weak correlations to the TDDFT properties, suggesting binding is less

influenced by the energy of the complex, and therefore may be influenced more by steric factors.

The ROS generations assays also demonstrated moderate correlation to the DFT and TDDFT energies. Type 1 ROS (DHR) showed moderate correlation to LUMO+2 and LUMO+3 (0.38 and 0.45). The type 2 ROS assay (ABDA) was most strongly correlated with the HOMO-2, dipole moment and the S_1 to S_0 transition energy of the complexes. Interestingly the results of the ROS assays were not correlated to the expected T_1 to S_0 transition energy which was shown to have the closest relationship to ROS generation within the Jablonski diagram (figure 3.2).

Phototoxicity appeared to correlate to many of the calculated energies including the HOMO, HOMO LUMO gap, and the S_1 to S_0 and T_1 to S_0 transition energies. This was interesting as the phototoxicity is determined by many more factors besides the electronic properties of the complexes, including cell permeability and localisation. Additionally, the phototoxicity did not correlate to ROS generation, which results in the mechanism of phototoxicity. This could imply that the *in-vitro* ROS assays were not truly representing the ROS generating abilities of the complexes within cells. The cell fluorescent intensity showed a moderate correlation to many of the TDDFT energies suggesting a very strong correlation to the electronic properties of the complexes. This makes sense as CTCF is a secondary measurement of the photophysical quantum yield of the complexes within cells.

For the absorbance wavelength there was a general correlation between the HOMO and LUMO energies of the complex, with little correlation to the transition energies. This was expected as the absorbance is dependent on accessibility of electrons to excite, therefore the strong correlation to the HOMO (0.70) is reasonable. The correlation to the oscillation strengths of the S_1 and S_2 to S_0 transition energies was also expected as this determines how readily the transition occurs, and therefore how readily the complex absorbs energy.

Finally, for emission wavelength there was a very strong correlation between the peak energy of emission and the T_1 to S_0 (-0.89) and T_2 to S_0 (-0.92) transition energies. It is known that the T_1 to S_0 transition energy is related to the peak emission energy as

demonstrated by Planck's equation (4.9). The emission wavelength also shows strong correlation to the S_1 , S_2 and S_3 to S_0 transition energies, in addition to the HOMO LUMO gap. This is likely due to the dependency of an electron to be excited to allow for an emission via radiative decay.

4.4 DFT and TDDFT conclusions

Using DFT and TDDFT it was possible to correlate a larger T_1 to S_0 transition energy and a larger HOMO LUMO gap to a lower peak emission wavelength. Using the HOMO LUMO orbital diagrams it was possible to attribute the changes in energy to the **CBA** functionalisation of the platinum(II) salphen complexes. Substituents in the 4 position had a significant anti-bonding contribution to the LUMO, increasing the HOMO LUMO gap, resulting in a decrease in the peak emission wavelength (increase in energy). Substituents in the 5 position had a significant anti-bonding contribution to the HOMO, reducing the HOMO LUMO gap, and resulting in an increase in the peak emission wavelength (decrease in energy).

From the DFT and TDDFT calculations, the T_1 to S_0 transition energy was deemed to be a suitable representative of experimental peak emission wavelengths for ML analysis in chapter 5. The rationale behind this decision was a result of the known relationship between the T_1 to S_0 transition energy and the peak emission wavelengths, as shown in the Jablonski diagram in figure 4.1. Additionally, when comparing the calculated T_1 to S_0 transition energy to the experimental peak emission energies (following conversion of wavelength to energy using Planck's equation 4.9) there was a mean absolute error (MAE) of only 0.045 eV. This was further supported by the high PC coefficient value between the complexes emission wavelengths and T_1 to S_0 transition energy.

The objective was therefore to perform HT DFT and TDDFT calculations to generate a large dataset of platinum(II) salphen energy properties to train ML models. The final models could then be used in place of performing further DFT and TDDFT calculations to determine the peak emission wavelengths of new complexes.

Chapter 5

Machine learning analysis on the photophysical properties of metal complexes

To generate larger datasets for the platinum(II) salphen complexes, a semi-automated HT computational pipeline was developed. This was developed in collaboration with Wilson Wu (PhD student in the Vilar group) and Joseph Burke (UG project student in the Vilar group) who developed the bash scripts for Gaussian input file generation and semi-automated HPC submission. This pipeline was used to generate an *in silico* library of 263 platinum(II) salphen complexes, perform DFT and TDDFT calculations, and extract the output data. The *in silico* library was also used to generate computational descriptors, enabling the interpretation of complexes within ML applications (see section 5.2).

The selected descriptors were analysed using unsupervised ML techniques to explore their representation of the complexes. Next, supervised ML regression algorithms were optimised to develop models which could predict the energies obtained from DFT and TDDFT calculations, with a focus on T_1 to S_0 energy to represent the experimental peak emission wavelength. The objectives of the final models were to generate energy data for new complexes without running further DFT and TDDFT calculations, which are more computationally expensive than using a trained ML model.

Next, it was investigated whether it was possible to predict the peak emission energies of alternate metal complex systems using the optimised platinum(II) salphen DFT ML models. The metal complexes investigated included zinc(II) salphen complexes and platinum(II) polypyridyl (NNC) complexes, using experimental emission energies

obtained from the literature [289, 320–326] (see figure 5.11 and appendix section A.3.5.1). To improve the predictive power of the models, transfer learning (TL) and domain adaptation (DA) techniques were explored.

Finally, SAR of the platinum(II) salphen complexes and DFT and TDDFT energies were explored using chemically interpretable molecular descriptors. From this, the features determined to be important to predict target properties (orbital and transition energies) were visualised. Figure 5.1 shows a schematic overview of chapter 5.

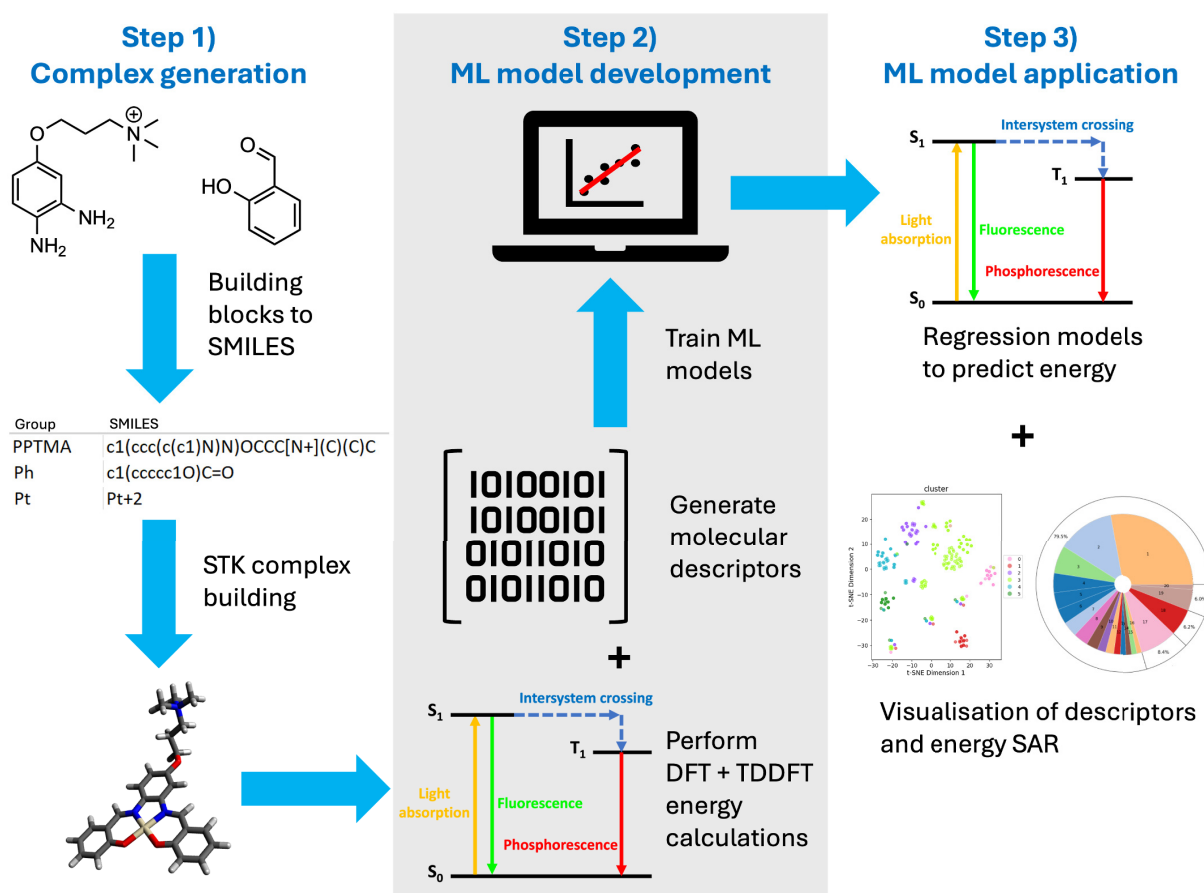


Figure 5.1. Schematic of the workflow within chapter 5. Step 1) Computational generation of complexes from the SMILES of the building blocks. Step 2) Development of ML models using data from DFT and TDDFT with molecular descriptors to train regression models. Step 3) Application of ML models to predict energy properties for platinum(II) salphen complexes, zinc(II) salphen complexes and platinum(II) NNC complexes. In addition to visualisation of the structure activity relationships (SAR) between molecular descriptors and complex energies.

5.1 Computational generation of complexes

A set of 263 platinum(II) complexes were generated by combining 12 **DAB** building blocks with 22 **CBA** building blocks. The selected building blocks introduced a range of electron donating and electron withdrawing functionalities to platinum(II) salphen core, with the objective to more broadly explore the effects of substitution on the complexes electronic properties. As seen previously, the nomenclature for the generated complexes is based on the identity of the building blocks (figure 5.2).

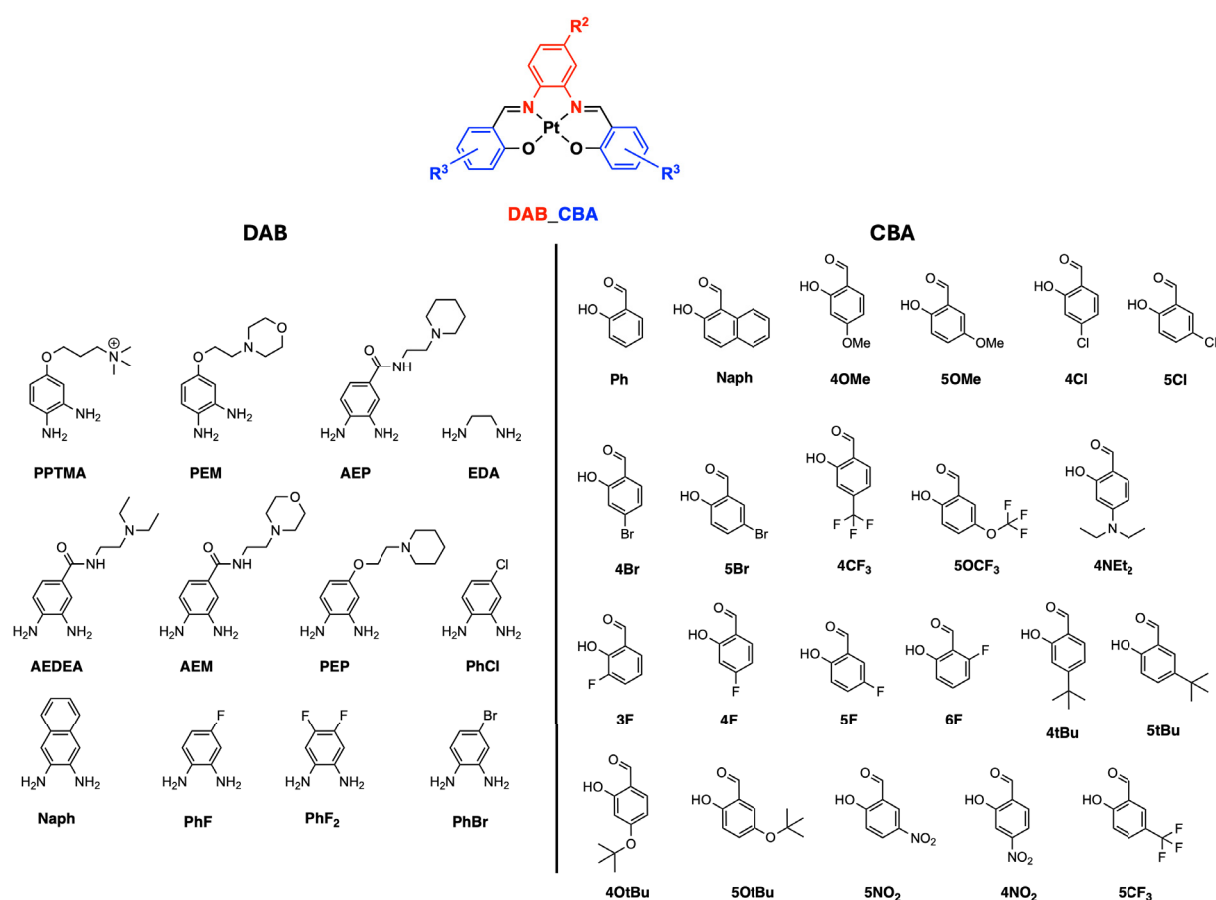


Figure 5.2. Figure showing all of the **DAB** and **CBA** building blocks used within the generation of the 263 platinum(II) salphen complexes.

To develop a large *in silico* library, a systematic and efficient way of generating the complexes was developed. To generate metal complexes computationally, it is not as simple as drawing a molecule in ChemDraw. Metal complexes have more complicated geometries to consider which are not represented by 2D inputs. Therefore, complexes were generated using the STK and STKO python toolkits [327, 328]. For

these toolkits, the initial inputs are SMILES strings of the complex building blocks: **DAB**, **CAB**, and metal centre (Pt). SMILES stands for simplified molecular input line entry system, and represents the atoms and bonds of a molecule in a text format [329]. The reactive groups of the SMILES building blocks are defined and “reacted” together using templates to create a constructed molecule of the ligand. The templates define which atoms within the reactive groups are lost and retained following bond formation. The ligand bond lengths are then optimized (1), the metal is inserted (2), and several rounds of geometry optimizations are performed using pipelines of optimization software including OpenBabel, GULP and xTB (3) [330–332]. Figure 5.3 shows the computationally generated structure of complex **Ph_PPTMA_Ph** at each point described above.

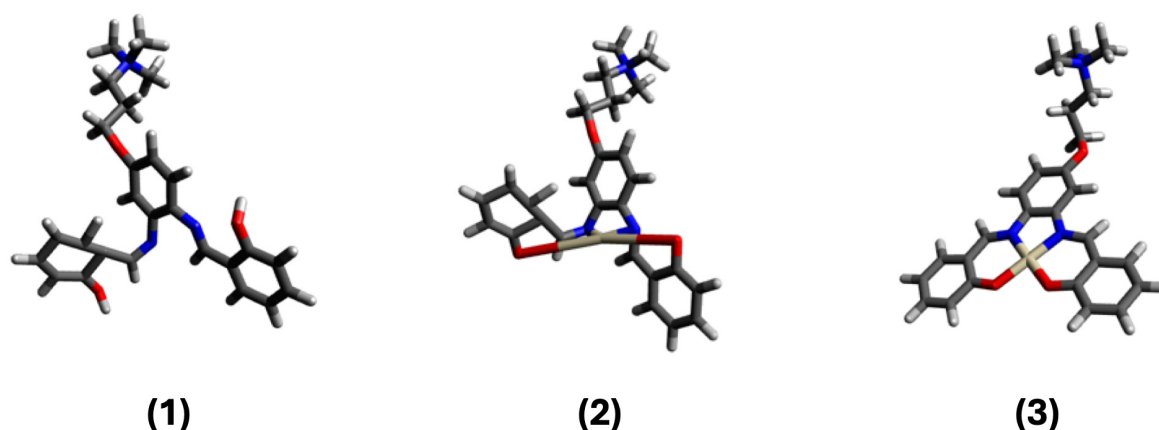


Figure 5.3. Figure showing the platinum(II) salen complexes in each steps of the STK complex generation. 1) Organic ligand generated and optimized. 2) Insertion of the metal centre. 3) Multiple rounds of geometry optimization using OpenBabel, GULP and XtB.

OpenBabel and GULP use Universal Force Fields (UFF), an alternate energy optimisation method to DFT. UFF’s optimise molecular geometries by minimising the potential energy of the system, which includes bonding (bond length and angle) and non-bonding (electrostatic and van der Waals) interactions. The force on each atom is determined by calculating the gradient of the potential energy, adjusting atomic positions to minimise these forces, and thereby the overall system. This results in a molecular geometry with a local minimum potential energy [330, 331].

xTB is a semi-empirical quantum mechanical method which uses quantum mechanical

approaches based on the Hartree-Fock method, with empirical parameters to model the electronic energy of the system. This means the calculations use only the valence electrons of the system, minimal basis sets, and neglect or approximate certain integrals [332]. The final optimised complexes were exported into rdkit, a cheminformatics python toolkit, which was used to generate molecular descriptors. As well as used to generate DFT and TDDFT input files. When comparing the optimised complex geometries from STK to DFT, the average difference in bond length was found to be $< 0.05 \text{ \AA}$.

5.2 Molecular descriptors

A determining factor of the performance of ML models is how accurately the molecular descriptors are able to represent the SAR. Without adequate representation, ML models struggle to find reliable correlations between complex representations and the target properties. This is particularly important within SAR analyses as subtle structural changes can dramatically alter molecular properties [333–336]. Two classes of descriptors were investigated within this work: Mordred, a collection of over 1800 chemically interpretable 2D and 3D descriptors [337]. In addition to more structure focused Morgan's Fingerprints (fingerprints): numerical vectors which encode local and extended structural features of a molecule [338].

The generation of Mordred descriptors comes from the open-source Mordred software which can calculate descriptors efficiently and accurately with very few lines of code. Mordred descriptors represent electronic and steric properties of molecules, and are widely applied for correlating these chemically interpretable descriptors with biological activity or chemical properties. Mordred has been used across fields including drug discovery, material science, and environmental chemistry [339–346]. Within Mordred descriptors there are multiple classes of descriptor types. Each class contains a range of molecular features which capture unique properties, based on the same underlying principles or mathematical approach for their determination. Appendix section A.3.1 contains the details of the descriptor classes within Mordred. The full list of each descriptor and its description are available within the Mordred descriptor github [347].

Fingerprints encode the structural features of molecules into binary vector formats. This

is performed via a circular fingerprinting approach which captures the connectivity of atoms within a molecule. Fingerprints are generated through a systematic process that involves several key steps (figure 5.4):

1. Atom-centric feature identification: the molecular structure is converted to a graph representation. Each atom is numbered and analysed in terms of its immediate neighbours, considering features such as atom type, hybridization state, and the presence of functional groups.
2. Circular fragmentation: a radius is defined that determines how many bonds from the central atom the algorithm will explore within each iteration.
3. Hashing and bit vector representation: after identifying relevant fragments within the specified radius, each fragment is hashed into a fixed-length binary vector. This vector encodes the presence (1) or absence (0) of specific structural features. The resulting fingerprint is a compact representation of the molecular structure that can be efficiently used for similarity comparisons.

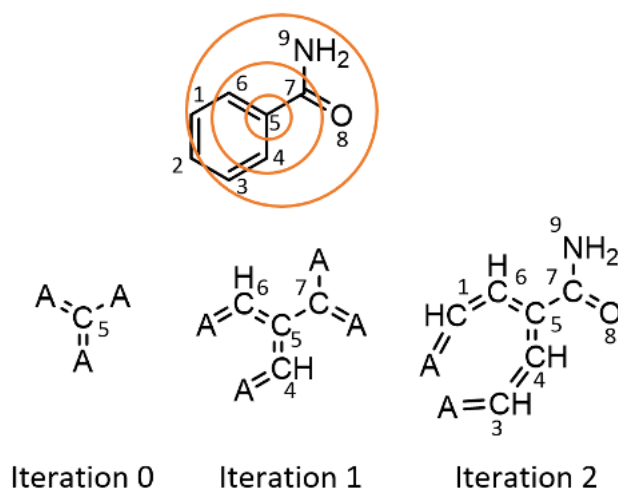


Figure 5.4. Fingerprint generation showing the atom numbering and circular fragmentation iterations.

Fingerprints are used extensively within cheminformatics and drug discovery. Examples of their use include screening large libraries of compounds against target proteins by performing similarity searches between potential ligands and known active compounds [348–350]. As well as applications with ML for predicting molecular

properties, and visualising chemical space [113, 351–355].

Mordred and fingerprint descriptors were selected due to their prevalence within the literature, how accessible they were to generate, and to provide descriptors which have unique molecular representation with respect to one another.

5.2.1 Dimensionality reductions and clustering

Due to the high dimensionality of Mordred and fingerprint descriptors, it is challenging to visualise and interpret how they represent the platinum(II) salphen complexes within chemical space. Dimensionality reduction techniques can be employed to translate high dimensional data into low dimensional space (2D) whilst preserving the meaning of the dataset, thereby enabling interpretation [356].

Dimensionality reduction techniques are designed for specific uses. For the primary use of visualisation there is t-SNE: a nonlinear technique which translates pairwise similarities between data points in high-dimensional space to low-dimensional space, grouping similar compounds close together whilst dissimilar ones are spaced apart [357, 358]. t-SNE performs this task through complex mathematical equations which compute the pairwise similarities between points in high dimensional space. Then, recompute the pairwise similarities in a lower dimensional space. t-SNE then minimises the Kullback-Leibler divergence which measures the difference between high-dimensional and low-dimensional pairwise similarity distributions to ensure probability distributions are similar in high and low dimensional space. This is achieved by adjusting the positions of points in low dimensional space using a gradient descent. The full equations for each step are detailed within appendix section A.3.2.1. By applying t-SNE to the Mordred and fingerprint descriptors it was possible to visualise how complexes occupy chemical space. It also enabled visualisation of how well these descriptors are able to differentiate between the complexes which are chemically very similar due to the shared core salphen structure.

An additional method for assessing how well descriptors are differentiating between complexes is k-means clustering: an unsupervised ML algorithm which partitions data into clusters based on their similarity [359, 360]. Cluster assignment occurs via an initial

random assignment of cluster centres (k centroids), followed by the assignment of data points to the nearest centroid. An update step then calculates the mean of all points assigned to each cluster and updates cluster assignments until the mean stabilises or the centroids stop moving. The optimal number of clusters is determined by two key metrics. The first is an objective function which minimises the total within-cluster variance (inertia). Along with the silhouette score which measures how similar objects are within a cluster. The final k value for each clustering task below is determined by balancing the optimal inertia and silhouette scores with a reasonable number of clusters, which for this dataset is less than 20. The k-means equations are detailed within appendix section A.3.2.1.

t-SNE and k-means clustering methods are typically employed together to assess the similarity in high-dimensional space with k-means, and how well this is translated within a lower dimensional space following dimensionality reduction. All plots within this section show the 2D t-SNE dimensionality reduction of the Mordred and fingerprint descriptors. Plots have been annotated using colours to represent complex structural features including; **DAB** and **CBA** components, nature of the **DAB** linker, position of the **CBA** functionality, the electron donating (ED) or electron withdrawing (EW) nature of the **CBA**, and the Hammett constant for the **CBA** substituent.

Applying k-means clustering to fingerprints resulted in an optimal 6 k-means clusters, and sparse groupings of complexes indicating clear distinction between complexes (figure 5.5).

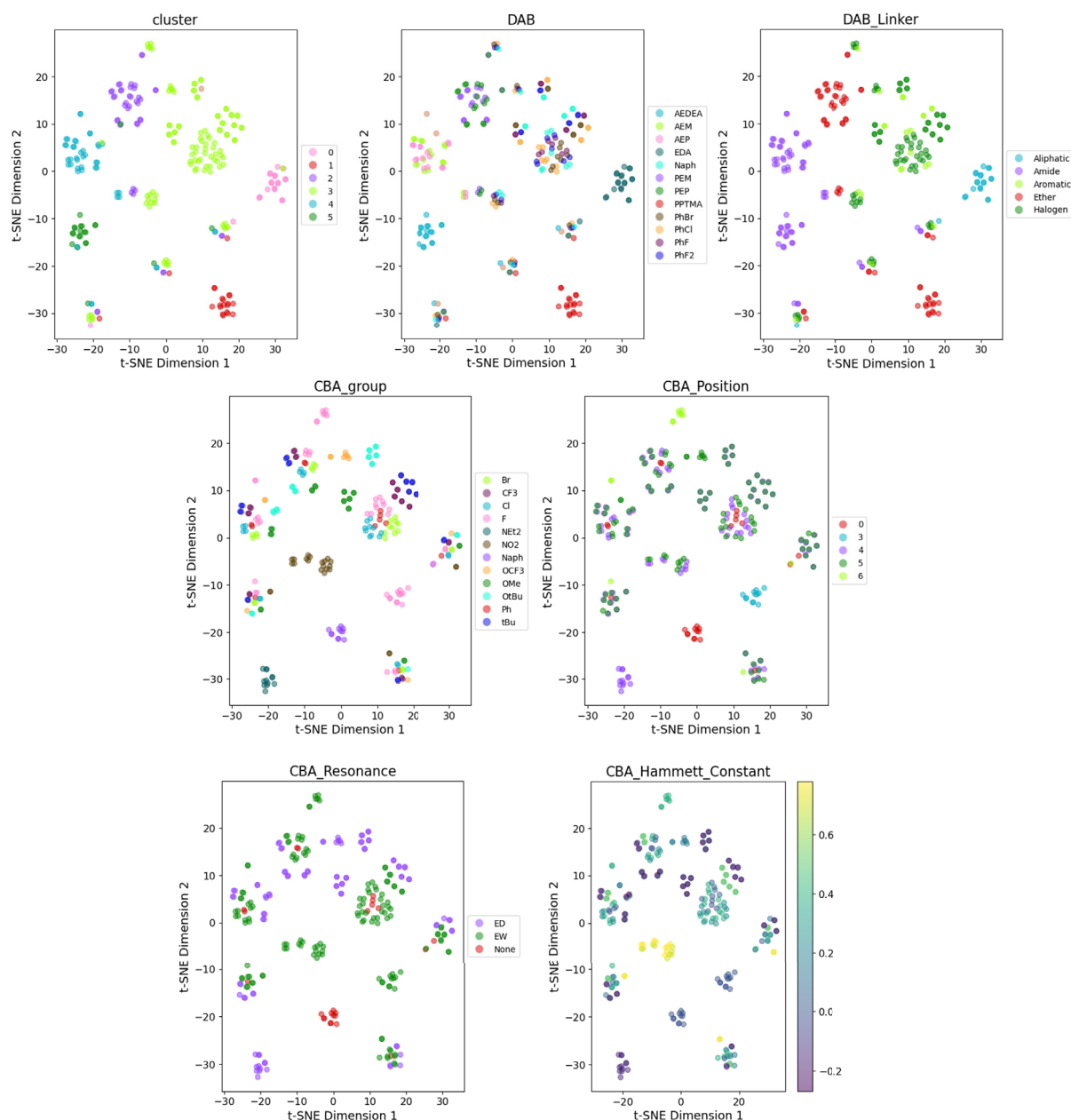


Figure 5.5. 2D t-SNE plots of fingerprint descriptors coloured from right to left as designated: cluster) assigned k-means cluster. DAB) **DAB** group within the complex. DAB_Linker) nature of the atom connecting the functionality to the **DAB**. CBA_group) **CBA** functionality atom type without position information. CBA_Position) position of the **CBA** functionality. CBA_Resonance) electron donating or electron withdrawing **CBA** resonance property. CBA_Hammett_Constants) relative Hammett constant of the substituent on the **CBA** functionality with meta/para positions measured from the platinum bound oxygen.

The clustering of the fingerprint complexes appeared to be dominated by the nature of the **DAB**, seen by consistent clustering of complexes with shared **DABs** (table 5.1).

Cluster	DAB
0	EDA
1	PPTMA
2	PEM + PEP
3	PhBr, PhCl, PhF, PhF₂, Naph
4	AEM + AEP
5	AEDEA

Table 5.1. Table containing the **DAB** functionalities within each k-means cluster of the fingerprint descriptors.

The combination of **DABs** within clusters 2 and 4 were expected as both **PEM/PEP** and **AEM/AEP** are very similar in structure, differing only by 1 atom in their morpholine or piperidine rings respectively. For cluster 3 the combination of the halide functionalised DAB was expected due to their shared electronegative character which has a small inductive electron withdrawing effect on the π -system of the complex, in addition to the electron donating resonance effect from the lone pair electrons. However, the inclusion of **Naph** within cluster 3 was unexpected. Its inclusion is potentially due to both these groups increasing the electron density of the system. For the **Naph** this is through extending the aromatic system, and for the halogenated groups this is through resonance with their lone pair electrons.

The t-SNE 2D visualisation groups the fingerprint descriptors based on the **DAB** components linker type. This can be seen in the DAB_Linker plot within figure 5.5. The **DABs** which are not grouped with their cluster in the t-SNE plot are those which share distinct **CBA** functionalities. The **CBA** functionalities with a strong influence on the t-SNE positioning include **4NEt₂** (plot axis: -20,-30), **Naph** (plot axis: 0, -20), and **3F** (plot axis: 5, -12). Substituent pairs in the 4 and 5 positioned were also grouped very closely together for functionalities including **4/5NO₂**, **4/5tBu**, **4/5OtBu**, **4/5Cl**, **4/5Br**, **4/5OMe**. In some cases, there was so much overlap between the 4 and 5 position points that it was difficult to see the pair (**4/5OtBu** - plot axis: 10, 18).

When applying t-SNE dimensionality reductions to the Mordred descriptors, the

resulting groups were much closer in space with less well-defined edges. There was also a greater mix of k-means clusters within the t-SNE groups (figure 5.6).

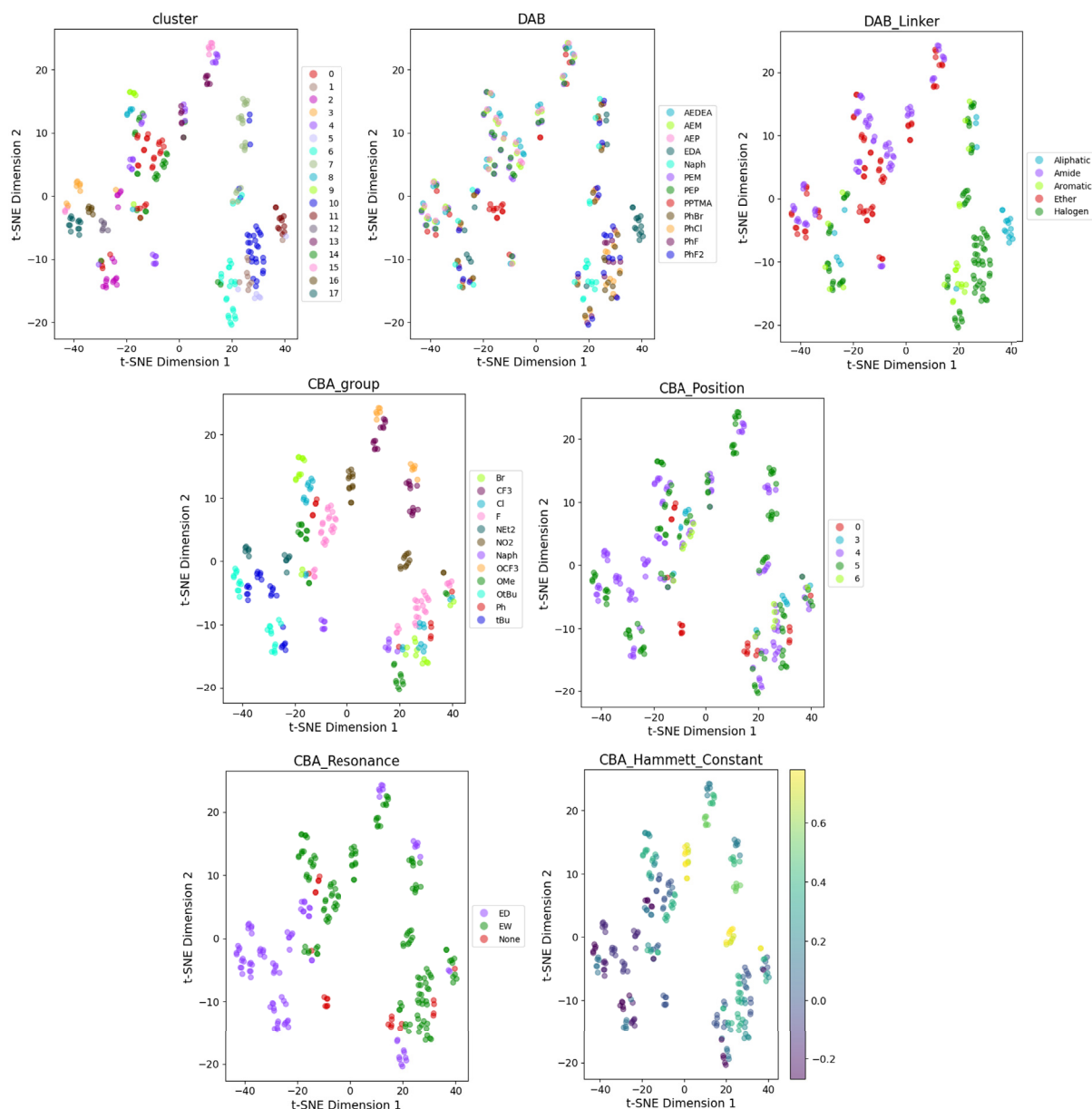


Figure 5.6. 2D t-SNE plots of Mordred descriptors coloured from right to left as designated: cluster) assigned k-means cluster. DAB) **DAB** group within the complex. DAB_Linker) nature of the atom connecting the functionality to the **DAB**. CBA_group) **CBA** functionality atom type without position information. CBA_Position) position of the **CBA** functionality. CBA_Resonance) electron donating or electron withdrawing **CBA** resonance property. CBA_Hammett_Constants) relative Hammett constant of the substituent on the **CBA** functionality with meta/para positions measured from the platinum bound oxygen.

The complexes contained within each k-mean cluster were more mixed for the Mordred descriptors: clusters all contained mixtures of the **DABs**. Some clusters contained collections of similar **CBAs** (table 5.2).

Cluster	DAB	CBA	Cluster	DAB	CBA
0	Mixed	Mixed	9	Ether/Amide	5Br
1	Mixed	4Br, 4Cl, 4F, 6F	10	Halogenated	Halogenated
2	Mixed	4NEt₂, 4tOBu, 5tBu	11	EDA	Mixed
3	Mixed	4NEt₂, 5OtBu	12	Mixed (small)	4tBu
4	Ether/Amide	Mixed	13	Ether/Amide	5NO₂, 5CF₃
5	Mixed	5Br	14	Ether/Amide	Mixed
6	Mixed	Mixed	15	Ether/Amide	5OCF₃, 5OtBu
7	Halogenated + Naph	4CF₃, 5CF₃, 5OCF₃, 4NO₂, 5NO₂	16		4tBu
8	Ether/Amide	4Br	17	Ether/Amide	5tBu, 4OtBu

Table 5.2. Table containing the **DAB** and **CBA** functionalities within each k-means cluster of the Mordred descriptors.

Overall, k-means clustering does not appear to have been able to detect any clear patterns of the **CBAs**, indicating that Mordred may struggle to distinguish between the subtle differences across the platinum(II) salphen **CBA** functionalities. Despite this, some structure can be seen within the t-SNE groups of the Mordred descriptors: **DABs** containing halogens, and the **Naph DABs** remain close in space seen in the DAB_Linkers plot in figure 5.6. The aliphatic **EDA DABs** were also grouped together. t-SNE was unable to distinguish between the amide and ether linked **DABs**, which were mixed together closely in space.

The most obvious trend that could be seen for the t-SNE grouping of Mordred descriptors was within the Hammett's constants (CBA_Hammett_Constant plot in figure 5.6), showing groupings of **CBAs** with similar constants, and a trend of lower constants towards the bottom left of the plot. This suggested that the Mordred descriptors were picking up on the electronic properties of the complexes more than the specific structural features. This was expected as Mordred descriptors are primarily made up of

property-based features in contrast to structural features.

k-means clustering and t-SNE dimensionality reduction may have been unable to capture clear relationships within the Mordred dataset due to the high density and collinearity of the dataset. Collinearity arises from the presence of many features which represent very similar properties, and therefore share trends within the data. As t-SNE attempts to preserve all local relationships, the importance of features with collinearity are exaggerated. The presence of redundant features can also lead to negative performance of ML models due to the creation of noise, over-fitting, and slow computational times. Therefore, there is an interest in reducing the dimensions of data for purposes beyond visualisation. Some common methods include Principle Component Analysis (PCA) and Uniformed Manifold Approximation and Projection (UMAP).

PCA is a linear dimensionality reduction technique which transforms features into orthogonal variables known as the principal components which preserve the variation of the original dataset [361, 362]. The number of principal component dimensions is determined using the PCA elbow: the number of dimensions required to explain 95 % of the dataset's variance. Once an optimal number of principal components has been determined, data is standardised using each individual features mean and standard deviation to ensure all features have equal variance. The covariance matrix is then computed which assesses how the features vary together across the dataset, calculating eigenvalues and eigenvectors to represent this covariance. The eigenvectors become the PCA axes, and the eigenvalues represent the amount of variance explained by each component. The equations used in PCA are detailed within appendix section A.3.2.1. PCA is designed to preserve global structures by capturing variance across all dimensions. However, PCA has challenges in maintaining local relationships in complex datasets and data which displays non-linear relationships.

UMAP is a non-linear dimensionality reduction technique which is able to preserve both local and global structures [363]. UMAP works by computing pairwise distances between data points to generate a fuzzy topological representation of the data. Meaning that relationships between points are represented as continuous (fuzzy), as opposed to binary (crisp), using Riemannian manifolds (smooth surfaces in high dimensions).

This allows for more complex relationships to be represented, resulting in a good representation of both local and global relationships. This is then used to initialise a low-dimensional embedding which is optimised using stochastic gradient descent to minimise the cross-entropy, the difference between probability distributions between the high and low-dimensional representations. The UMAP equations are detailed within appendix section A.3.2.1.

Using these additional dimensionality reduction techniques on the Mordred and fingerprint descriptors will provide insights into the local and global representations of complexes. It is expected that PCA will result in descriptors which represent global structures, such as the nature of the **CBA** functionalities. Whereas, UMAP will likely provide more representation of local structures, such as substituent positions and the nature of solubilising linkers. Following the implementation of these dimensionality reduction techniques, k-means clustering was performed, followed by t-SNE for visualisation.

Using the PCA elbow method on the fingerprint descriptors revealed that 23 dimensions were required to explain 95 % of the variance from the original 2048 features. Performing k-means clustering on the PCA reduced dataset revealed, with high confidence, that the optimal number of clusters was 6. Visualisation using t-SNE resulted in 6 distinct outer groups in a hexagonal shape, and a larger collection of less well-defined groups within the centre of space (figure 5.7).

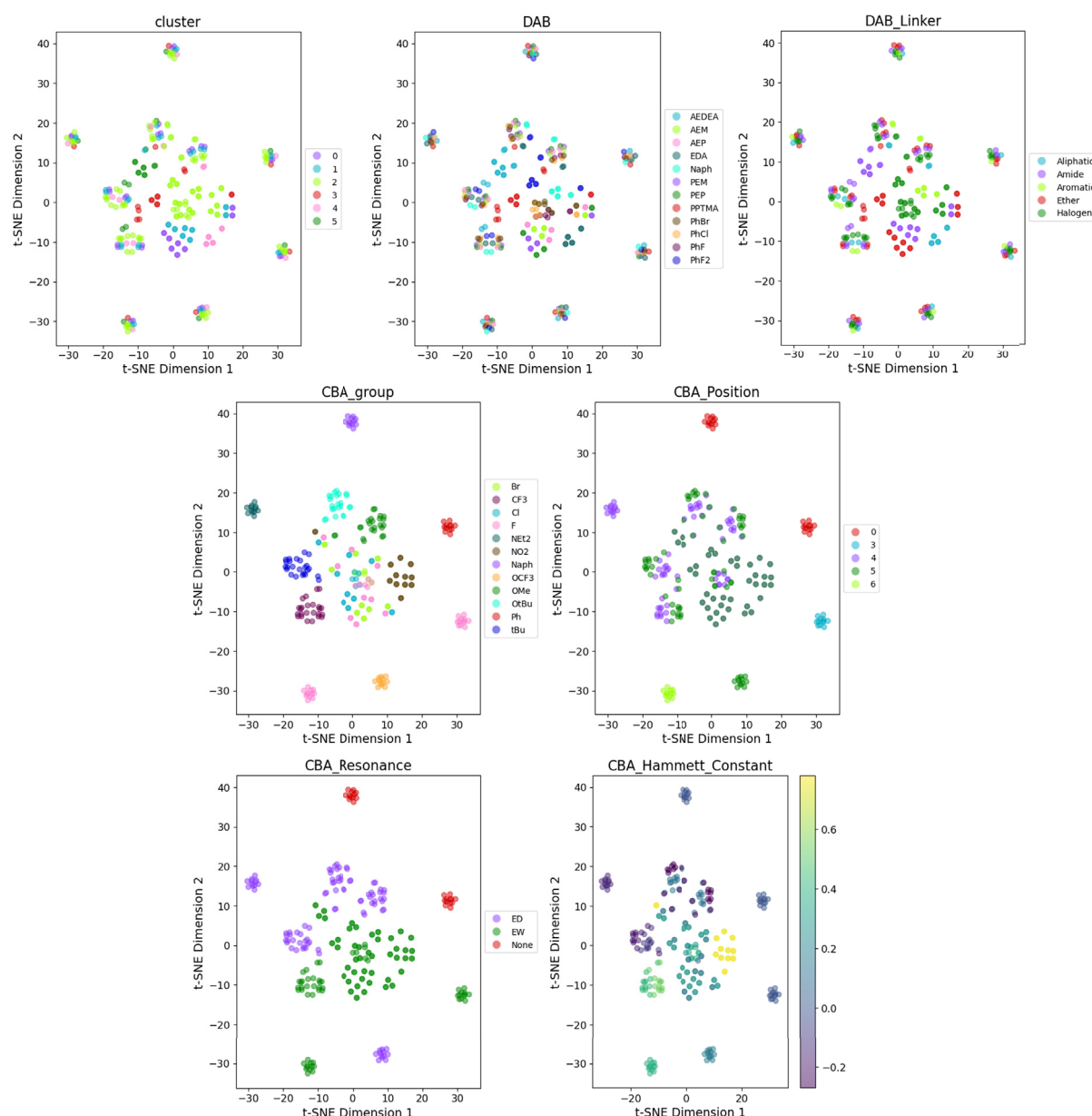


Figure 5.7. 2D t-SNE plots of PCA reduced fingerprint descriptors ($n = 23$) coloured from right to left as designated: cluster) assigned k-means cluster of PCA data. DAB) **DAB** group within the complex. DAB_Linker) nature of the atom connecting the functionality to the **DAB**. CBA_group) **CBA** functionality atom type without position information. CBA_Position) position of the **CBA** functionality. CBA_Resonance) electron donating or electron withdrawing **CBA** resonance property. CBA_Hammett_Constants) relative Hammett constant of the substituent on the **CBA** functionality with meta/para positions measured from the platinum bound oxygen.

k-means clustering once again exclusively clustered the complexes based on the nature of the **DAB**. This indicated that no information regarding the structure was lost during the dimensionality reduction. Table 5.3 shows a comparison of the k-means clusters

within the original and PCA reduced fingerprint descriptors.

Cluster	Original Fingerprint: DAB	PCA Fingerprint: DAB
0	EDA	PEM + PEP
1	PPTMA	AEM + PEP
2	PEM + PEP	PhF, PhCl, PhBr, PhF₂, Naph
3	PhBr, PhCl, PhF, PhF₂, Naph	PPTMA
4	AEM + AEP	EDA
5	AEDEA	AEDEA

Table 5.3. Table comparing the **DAB** functionalities within each k-means cluster of the fingerprint descriptors before and after PCA dimensionality reduction.

One key difference with the PCA reduced fingerprints was that the t-SNE groupings were now dominated by the **CBA** functionalities, with the outer 6 groups in the plots of figure 5.7 containing purely **Naph**, **Ph**, **5OtBu**, **5OCF₃**, **6F** and **4NEt₂** (clockwise from the top group axis position: 0, 40) **CBAs**. Pairwise groupings were still observed, in some cases with less of an extreme overlap (**4/5tBu**, **4/5OtBu**, **4/5CF₃** and **4/5OMe**). In other cases, the pairs had a greater overlap: **4/5Cl**, **4/5NO₂**, **4/5Br** and **4/5F**. This indicated that although global structural information (presence of a functionality) was retained after PCA dimensionality reduction, the more local relationships (positions of the functionalities) were lost. This was consistent with the commonly known limitations of PCA.

UMAP was initially employed in the same way as PCA, using the optimal PCA dimensions. When further dimensionality reduction using t-SNE was performed to visualise the data, the t-SNE grouping of the data points became very dense and sparse suggesting a loss of information. Therefore, UMAP dimensionality reduction was performed with 2 dimensions as a comparison to t-SNE. The 2D UMAP dimensionality reduction plot looked very similar to the t-SNE dimensionality reduction plot, providing no additional insights. The fingerprint UMAP plots can be found within appendix section A.3.2.2.

Applying the PCA elbow to the Mordred descriptors revealed that only 18 dimensions

were required from the original 1823 to explain 95 % of the variance. After applying k-means clustering and the t-SNE dimensionality reduction to the PCA data, there was a clear increase in the structure of t-SNE groupings (figure 5.8).

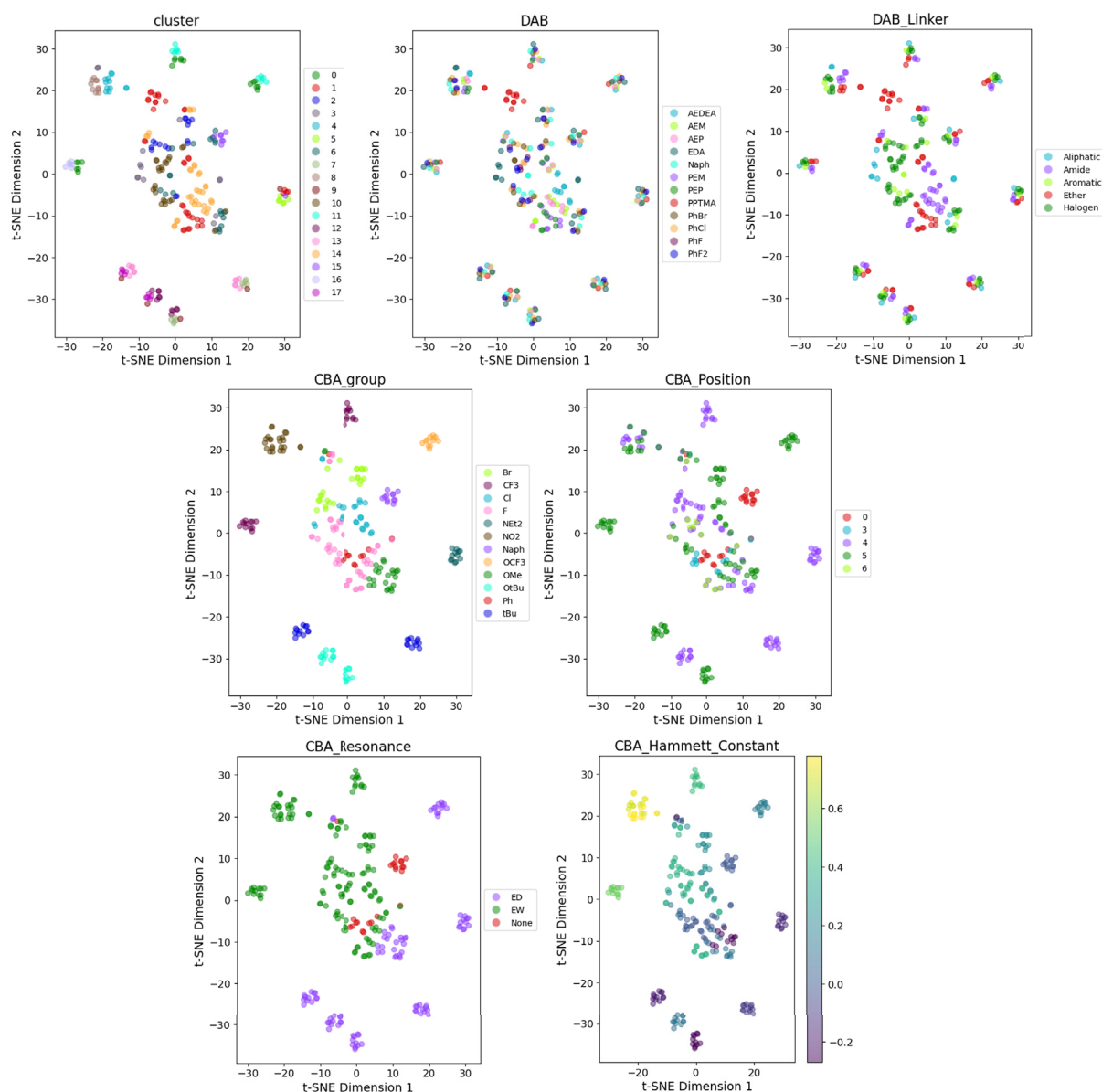


Figure 5.8. 2D t-SNE plots of PCA reduced Mordred descriptors (n = 23) coloured from right to left as designated: cluster) assigned k-means cluster of PCA data. DAB) **DAB** group within the complex. DAB_Linker) nature of the atom connecting the functionality to the **DAB**. CBA_group) **CBA** functionality atom type without position information. CBA_Position) position of the **CBA** functionality. CBA_Resonance) electron donating or electron withdrawing **CBA** resonance property. CBA_Hammett_Constants) relative Hammett constant of the substituent on the **CBA** functionality with meta/para positions measured from the platinum bound oxygen.

k-means clustering of the Mordred PCA data revealed clusters capturing both the nature of the **DAB** and the **CBA**. This was demonstrated particularly well by clusters 4 and 8 which both contained the **4/5NO₂ CBA** but were separated by the Ether/Amide linked **DABs** and the halogenated and **Naph DABs** respectively. The t-SNE grouping appeared to be dominated by the nature of the **CBA**, showing 9 pure **CBA** functionality groupings around the edge of the plot including **3F**, **5OCF₃**, **4NEt₂**, **4tBu**, **5OtBu**, **4OtBu**, **5tBu**, **5CF₃**, and **4/5NO₂** (appearing clockwise from the top group from the axis position 0, 30 on the plots of figure 5.8). The pure **CBA** groups were split between two k-means clusters based on the nature of the **DABs**.

Another feature picked up within the t-SNE visualisation of the PCA reduced Mordred descriptors was the distribution of EW and ED **CBA** functionalities seen in the CBA_resonance plot of figure 5.8. EW functionalities localised in the top left of the plot, and the ED functionalities localised in the bottom right of the plot. This was also seen in the distribution of the **CBA** Hammett constants in the CBA_Hammett_Constant plot of figure 5.8, where low Hammett constants were trending towards the bottom right of the plot, and higher Hammett constants were trending towards the top left of the plot. This impressive amount of added detail in the plots resulting from the PCA reduction of the Mordred descriptors confirmed that as a whole the Mordred descriptors contain a lot of redundant features which were producing noise within the dataset, leading to a loss of t-SNE structure.

As was seen with the fingerprint descriptors, when performing UMAP on the Mordred descriptors followed by t-SNE reduction for visualisation much of the descriptor's details were lost: resulting in sparse groupings which were densely populated. The 2D UMAP visualisation of the Mordred descriptors resulted in a slight improvement in the separation of the different groups of complexes within the data. However, there was no significant improvement in the separation of complexes than was seen within the t-SNE representation of the complexes. The Mordred UMAP plots can be found within appendix section A.3.2.2.

Overall, it was clear that both fingerprint and Mordred descriptors were able to differentiate between the structural differences of the platinum(II) salphen complexes. Representation was enhanced by the application of dimensionality

reduction techniques. Particularly in the case of Mordred where the abundance of redundant features had a negative effect on complex representation. A stronger understanding of the representation of complexes allows predictions of descriptors performances within ML models. Therefore, it was expected that the fingerprint descriptors would perform best overall for prediction models, followed by the PCA reduced Mordred descriptors.

5.3 Development of machine learning predictive models

To attempt the development of ML models which predicted properties of the platinum(II) salphen complexes, the two descriptors (Mordred and fingerprint) were tested with seven supervised regression algorithms: linear, ridge, lasso, random forest (RF), XGBoost (XGB), support vector machine regression (SVM) and neural networks (NN). Algorithms were implemented using the sklearn python toolkit [364].

Linear regression assumes a linear relationship between the dependent variable (y) and a single independent variable (x) (equation 5.1) [365]:

$$y = \beta_0 + \beta_{1x} + \dots + \beta_p x_p + \epsilon \quad (5.1)$$

- y : Dependent variable (response/property).
- x : Independent variable (predictor/descriptor).
- β_0 : y -intercept of the regression line (the value of y when $x = 0$).
- β_1 : Slope of the regression line (the change in y for a one-unit change in x).
- ϵ : Error term (the difference between the observed and predicted values).
- $+\dots + \beta_p x_p$: Multiple variables.

The regression is optimised by an ordinary least squares (OLS) objective function to find the best fitting line/hyperplane that minimises the sum of the squared differences between observed and predicted values (equation 5.2):

$$J(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5.2)$$

- $J(\beta)$: Cost function (sum of squared errors).
- n : Number of observations.
- y_i : Observed value for the i -th observation.
- $\hat{y}_i$: Predicted value for the i -th observation, given by the regression equation.

After β coefficients are estimated, predictions can be made for new data points using the linear regression equation (equation 5.1).

It is unlikely that linear regression could be used to predict molecular properties as there is an assumption that all relationships are linear. Linear regression also assumes there is no multicollinearity within the variables. However, as the most simplistic model, it can be used as a baseline to compare the more complex algorithms against.

Ridge regression builds on the linear regression model as an OLS objective function which also includes a penalty term (equation 5.3) [365, 366]:

$$\text{Minimise } J(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (5.3)$$

- p : Number of independent variables (excluding the intercept).
- β_j : Coefficient for the j -th independent variable.
- λ : L2 regularisation hyperparameter.

Lasso regression works similarly to ridge regression as an OLS model with an almost identical objective function (equation 5.4) [365, 367]:

$$\text{Minimise } J(\beta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (5.4)$$

- λ : L1 regularisation parameter.
- $|\beta_j|$: Absolute coefficient for the j -th independent variable.

The key difference between the OLS objective functions of ridge and lasso is the regularisation method: ridge uses L2 regularisation, the sum of the squared coefficients, and lasso uses L1 regularisation, the sum of absolute values of coefficients. L1 regularisation can reduce coefficients to zero, whereas L2 regularisation can only shrink coefficients towards zero.

Ridge and lasso model relationships between a dependent variable and multiple independent variables, accounting for multicollinearity through the L2/L1 regularisation hyperparameter λ respectively. Regularisation shrinks coefficients which are not deemed significant, reducing the variance of the estimates. This leads to improved prediction accuracy, especially when dealing with multicollinearity and high-dimensional data. This shrinkage is also useful for model interpretability, allowing the identification of features deemed to be important for predicting the target property.

However, these models also have limitations: both are sensitive to the regularisation hyperparameter α . When α is large, the penalty for large coefficients is higher, leading to more significant shrinkage of the coefficients. When α is small (close to zero), the ridge regression model behaves more like the linear regression OLS objective function, applying less regularisation. In cases of high multicollinearity, lasso can arbitrarily select one variable over another without a strong rationale, which can lead to models that are less stable, less interpretable, and introduce bias into the coefficient estimates. There is also a risk of over shrinkage where lasso shrinks coefficients too aggressively, leading to the exclusion of variables that could be important for the model [368, 369].

RF and XGB are ensemble models which use a combination of base models to make an average prediction based on non-linear relationships [370–374]. The RF regression algorithm consists of an ensemble of multiple decision trees built up of nodes which represent the decision point, and branches which represent the decision. Leaf nodes are the terminal nodes where the final prediction is made. The model trains multiple trees on different subsets (bootstrapping) of data to prevent over-fitting. To give the final prediction ($\hat{Y}(x)$) (equation 5.5):

$$\hat{Y}(x) = \frac{1}{M} \sum_{j=1}^M m_j(x, \Theta_j) \quad (5.5)$$

- M : Total number of trees in the Random Forest. Each tree contributes to the final prediction.
- $m_j(x)$: Prediction made by the j -th tree for a given input x .
- Θ_j : Random variable associated with the j -th tree, capturing the randomness introduced during the tree construction process, such as the selection of features and the bootstrapping of the training data.

The benefit of using RFs is that they are designed to reduce over-fitting by averaging the predictions over multiple trees. They generally perform well across datasets and are less sensitive to noise compared to other regression methods.

Similarly, XGB models are built by ensembles of decision trees consisting of nodes and branches. Each node represents a decision based on a specific feature and threshold, while branches represent the paths taken based on those decisions. The tree structures allow RF and XGB to capture complex non-linear relationships between features and the target variable, making them powerful and flexible algorithms for ML tasks. The key difference between RF and XGB is that the XGB algorithm is based on gradient boosting ML. Meaning, gradient boosting frameworks build models sequentially and error correct the previous models by minimizing the loss function. The objective function contains a loss function and a regularisation term (equation 5.6):

$$L(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{j=1}^m \Omega(f_j) \quad (5.6)$$

- $L(\theta)$: Overall objective function.
- $l(y_i, \hat{y}_i)$: Loss function for the i -th observation, measuring the difference between the true value y_i and the predicted value $\hat{y}_i$ (equation 5.7):

$$l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2 \quad (5.7)$$

- $\Omega(f_j)$: regularisation term for the j -th tree, which penalizes complexity to prevent over-fitting.

XGB is designed for efficiency, utilizing parallel processing and optimised memory usage, which allows it to handle large datasets effectively. XGB incorporates L1 (lasso) and L2 (ridge) regularisation techniques, helping to prevent over-fitting and improve model generalization. However, while it is efficient, XGB can require significant computational resources, especially when using large datasets or many boosting iterations. XGB can be sensitive to outliers since it tries to correct the errors of previous models, which may lead to poor performance.

SVM regression is a non-parametric technique which uses kernel functions. This means it makes no assumptions of the relationships between elements of the inputs and outputs [375, 376]. Unlike parametric models, such as linear regression, which assumes a fixed functional form. Kernel functions are mathematical functions which allow SVMs to operate in high dimensional space. There are three main SVM kernels which are shown in table 5.4:

Kernel Name	Kernel Function
Linear (dot product)	$G(x_i, x_j) = x_i \cdot x_j$
Radial Basis (RBF)	$G(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$ σ determines the width of the Gaussian function
Polynomial	$G(x_i, x_j) = (1 + x_i \cdot x_j)^q$ $q \in \{2, 3, \dots\}$

Table 5.4. Table containing the equations for the three kernel functions explored within the SVM regression algorithm.

These kernels are inserted into the Dual Formula for the non-linear SVM regression model (equation 5.8):

$$\text{Minimise } \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)G(x_i, x_j) + \varepsilon \sum_{i=1}^N (\alpha_i - \alpha_i^*) - \sum_{i=1}^N y_i(\alpha_i - \alpha_i^*) \quad (5.8)$$

- $G(x_i, x_j)$: Kernel function.
- α_i, α_i^* : Dual variables for training with each sample of features, representing the final weight of each sample.

- α_j, α_j^* : Dual variables for training with each pair of data points (i paired with j).
- ϵ : Width of the epsilon-insensitive tube. Defines the margin of error tolerance before the model is penalized. Support vectors are the data points that fall on or outside the epsilon-insensitive tube and define the decision boundary.
- y_i : Target output.

SVM is effective at handling high dimensional data, particularly where the number of dimensions is larger than the number of samples. Additionally, its flexibility from the different kernels makes it much more versatile for a range of tasks. However, SVM is computationally expensive. It is also very sensitive to the kernel selection, therefore requiring parameter optimisation searches with each kernel before application.

The final algorithm explored within this work was NNs. NNs are built up of three components: An input layer which receives the input features, hidden layers which perform data transformations and computations, and an output layer which provides the final prediction [377, 378]. Each layer is built up of neurones, and each neurone computes a weighted sum of its inputs (equation 5.9), applied to an activation function to give an output:

$$z = \sum_{i=1}^n \omega_i x_i + b \quad (5.9)$$

- z : Weighted sum (net input)
- ω_i : Weights associated with each input x_i
- b : Bias term
- n : Number of inputs

The output of the neurone is then passed through an activation function (a) (equation 5.10):

$$a = f(z) \quad (5.10)$$

The activation function used within this work is ReLu (equation 5.11):

$$f(z) = \max(0, z) \quad (5.11)$$

The loss function used within the NN regression tasks is mean squared error (equation 5.12):

$$MSE = \frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2 \quad (5.12)$$

- y_j : True value
- $\hat{y}_j$: Predicted value
- N : Number of observations

The NN uses propagation to pass information between the layers: forward propagation passes inputs through the network to obtain outputs. For multi-layer networks the output from one layer becomes the input to the next layer. Back propagation is the learning process where the network adjusts its weights to minimise loss by calculating the gradient of the loss function with respect to each weight, and updates the weight using gradient descent.

NN are highly customisable machine learning algorithms. Additional layers can be added to increase complexity, as well as the number of neurones within each layer. The learning rate can also be modified to control the change of weights on each neurone during training. Epochs can be changed to modulate the number of complete passes the model performs through the training data. NNs also apply L1 and L2 regularisation to prevent over-fitting by penalising large weights. Finally, neurones can be randomly dropped, controlled by the drop-out rate, which helps prevent over-fitting. NNs are able to learn complex non-linear relationships between dependent and independent variables. However, the limitations of NNs are that they are black box algorithms, which results in a lack of interpretability of the model. They are computationally expensive, often having to compromise complexity to manage the computational burden. NNs can also over-fit if not well regularised. Most importantly they tend to require large datasets for accurate training to prevent over-fitting.

To optimise the ML algorithms, grid search parameter optimisation was implemented. This involved setting restricted values of hyperparameters for the algorithms to implement, the final model used the hyperparameters which yielded the best

predictions. To find the optimal input descriptors, feature engineering was performed on the molecular descriptors, which included data scaling, PCA and UMAP dimensionality reductions. Interestingly the best performances were seen with the non-engineered molecular descriptors. This is likely due to many of the algorithms having built in feature engineering steps as part of their training.

To assess the models performances, the datasets were split into 20 random sets containing 80 % of the data for training, and 20 % for testing. This is known as Monte Carlo Cross Validation (CV) [379, 380]. This tested the model's sensitivity to different splits of data and helped to identify models which were over-fitting to the training data. Additionally, the models were designed to check for data leakage between the training and test set to prevent any false positive performances. The metrics used to quantify the model's performance were the R^2 score (equation 5.13) which measures the variance explained by the model, assuming a linear relationship:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \quad (5.13)$$

- SS_{res} : $\sum (y_i - \hat{y}_i)^2$ (sum of squared residuals).
- SS_{tot} : $\sum (y_i - \bar{y})^2$ (total sum of squares).
- y_i : Actual values.
- $\hat{y}_i$: Predicted values.
- $\bar{y}$: Mean of actual values.

R^2 can be sensitive to outliers within the prediction and can be more volatile with smaller datasets. The second metric selected was the MAE (equation 5.14). MAE measures the average absolute difference between predictions and actual values, and is less sensitive to outliers.

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (5.14)$$

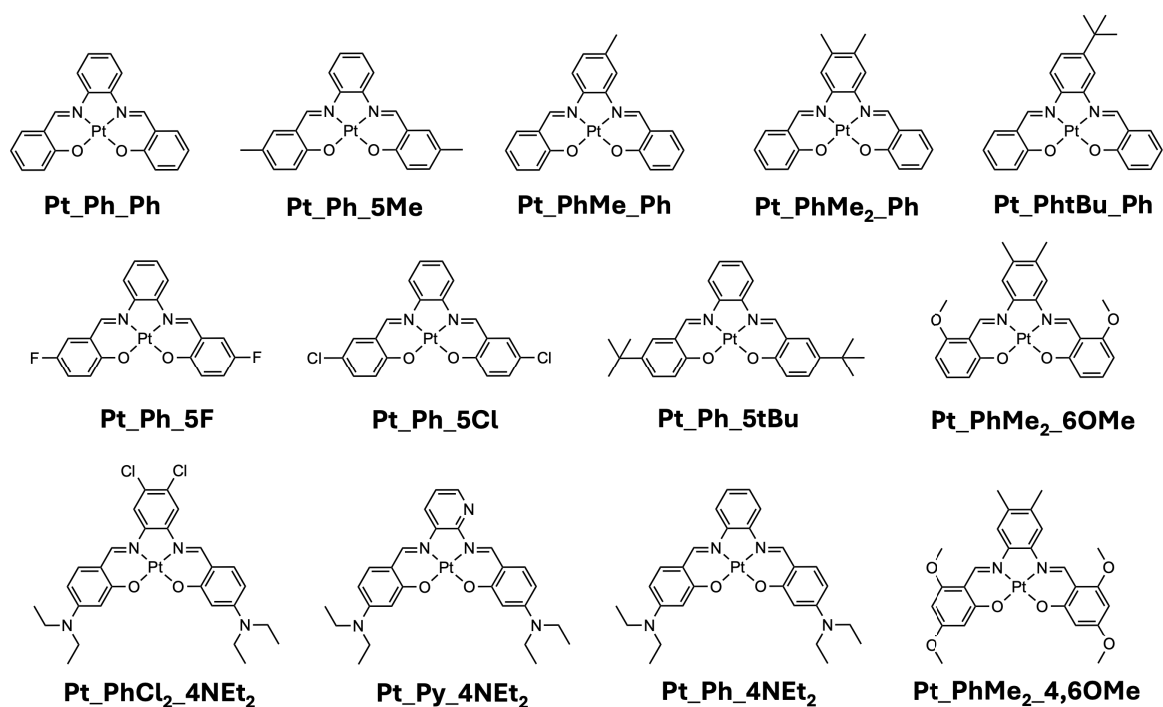
- n : Number of samples.
- y_i : Actual values.

- $\hat{y}_i$: Predicted values.

The performance of the final models were assessed by the average R^2 and MAE across the CV data splits [381, 382]. The inclusion of standard deviation ($\pm$) indicates the stability of the models across data splits, and therefore the reliability of the model.

5.3.1 ML models predicting platinum(II) salphen complexes orbital and transition energies

Although the key calculated energy of interest was the T_1 to S_0 transition energy (as a representation of experimental peak emission energy), four additional energy properties were investigated with predictive models: HOMO, LUMO, HOMO LUMO gap, and S_1 to S_0 energy transition. These were selected to explore how the ML algorithms performed across different prediction tasks. To validate the predictive abilities of the final models, an additional 21 platinum(II) salphen complexes were selected from the literature which had reported experimental peak emission wavelengths [34, 235, 273, 274, 314]. These complexes contained unique combinations of known building blocks, in addition to unseen building blocks to test the models' generalisability with unseen ligands (figure 5.9). DFT and TDDFT calculations were performed on the validation complexes to generate a full validation dataset for all models.



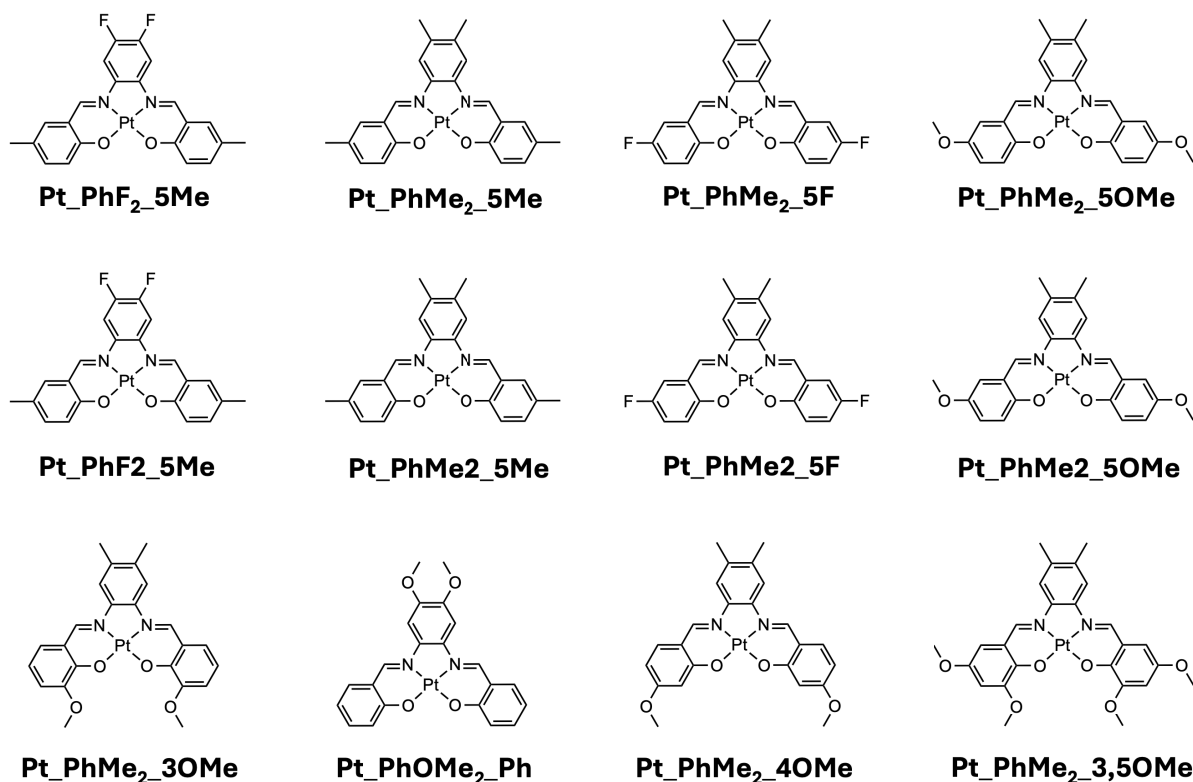


Figure 5.9. Structures of the literature platinum(II) complexes used to validate the regression models.

5.3.1.1 Prediction of T_1 to S_0 transition energy

Following the methods described in section 5.3, the platinum(II) salphen complexes T_1 to S_0 transition energies were used to train and test a series of predictive regression models. Table 5.5 details the R^2 scores and MAEs for each regression algorithm trained using the Mordred and fingerprint descriptors.

T ₁ to S ₀ Model	Mordred		Fingerprint	
CV	R ²	MAE (eV)	R ²	MAE (eV)
Linear	No fit	0.273 ± 0.109	No fit	3.82 × 10 ⁹ ± 4.19 × 10 ⁹
Ridge	0.474 ± 0.205	0.0786 ± 0.0130	0.804 ± 0.0985	0.0468 ± 0.00580
Lasso	0.762 ± 0.131	0.0557 ± 0.0098	-0.0531 ± 0.0418	0.122 ± 0.0165
RF	0.696 ± 0.0990	0.0580 ± 0.0088	0.694 ± 0.210	0.0356 ± 0.00820
XGB	0.681 ± 0.152	0.0598 ± 0.0111	0.806 ± 0.147	0.0347 ± 0.00660
SVM	0.747 ± 0.0801	0.0557 ± 0.0107	0.868 ± 0.107	0.0357 ± 0.00860
NN	No fit	49.0 ± 6.11	0.666 ± 0.115	0.0695 ± 0.011

Table 5.5. ML model performances for predicting T₁ to S₀ transition energy. The best performing model is highlighted in bold. The optimised model parameters are as follows. Mordred: Ridge – alpha: 10. Lasso – alpha: 0.1. RF – max depth: 10, n_estimators: 300. XGB – learning rate: 0.1, max depth: 3. SVM – C: 0.1, degree: 2, epsilon: 0.01, gamma: 'scale', kernel: 'linear'. NN – batch size: 32, epochs: 300, dropout rate: 0.2, learning rate: 0.01. Fingerprint: Ridge – alpha: 1. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 200. XGB – learning rate: 0.1, max depth: 3. SVM – C: 10, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. NN – batch size: 32, epochs: 100, dropout rate: 0.1, learning rate: 0.005. Unspecified parameters were default.

The best model performance was obtained with the fingerprint + SVM model which had an R² score of 0.87 and MAE of 0.036 eV. The performance of this model is visualised in figure 5.10 with plots for the actual vs predicted value of the T₁ to S₀ transition energies of complexes within the test set (plot A), as well as the R² score (plot B) and MAE (plot C) across the CV test splits. The models predictions varied across the CV splits indicating a sensitivity to the platinum(II) salphen complexes present within the training set. All models with an R² standard deviation (±) of > 0.1 were classified as being sensitive to complex representation within the training data.

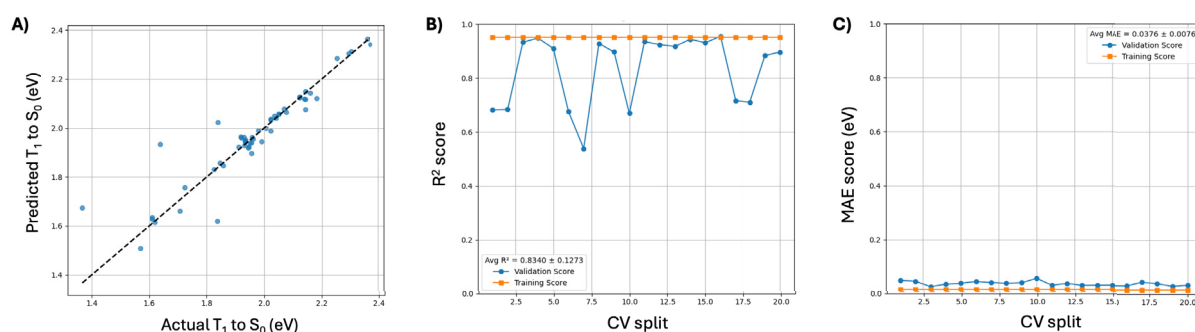


Figure 5.10. Plots showing the performance of the the fingerprint + SVM model for predicting T_1 to S_0 transition energies. Plots showing the actual (DFT) vs predicted (ML) values of the T_1 to S_0 transition energy on the test dataset (A), and the R^2 score (B) and MAE (C) across the CV training and test data splits.

The fingerprint + ridge model came in close second with an R^2 score of 0.80, with less instability than the XBG model ($R^2 = 0.87$), and an MAE of 0.047 eV. For the Mordred descriptors, the best model was obtained using the lasso algorithm which achieved an R^2 score of 0.76 and an MAE of 0.056 eV. Overall, fingerprint descriptors resulted in models with higher R^2 scores and lower MAEs. Except in the case of the fingerprint + lasso model which performed very poorly, resulting in a negligible R^2 score, indicating that the predictions were worse than fitting to the average of the dataset [383]. This was due to the incompatibility of the lasso model with the binary format of the fingerprint descriptors. The binary nature means there are many multi-collinearities within the dataset and a lack of continuous variability. As a result, lasso was unable to scale the coefficients smoothly, risking over-sparsification as lasso sets too many coefficients to zero. Which in turn results in much of the information within the descriptor to be lost.

The poor performance from the NN + Mordred model can be explained by the nature of the input data being high dimensional with few data points (>1800 features to 263 data points). Along with many of these features displaying multicollinearity and redundancies which are not compatible with how NNs handle input data (one neuron per feature). These factors combined led to the NN largely over-fitting to the training data as mentioned earlier. This problem is not observed for fingerprints due to their

binary format, which is compact, sparse, and without correlation, and can therefore be handled more easily by the NN.

To validate the models' performances, the T_1 to S_0 transition energies were predicted for the 21 literature platinum(II) salphen complexes and compared to their experimental peak emission energies (table 5.6):

T_1 to S_0 Model	Mordred		Fingerprint	
Validation	Validation R^2	Validation MAE (eV)	Validation R^2	Validation MAE (eV)
Linear	No fit	4.41	No fit	1.98×10^{11}
Ridge	No fit	0.263	0.817	0.0454
Lasso	0.411	0.0892	No fit	0.109
RF	0.626	0.0665	0.868	0.0403
XGB	0.667	0.0637	0.826	0.0408
SVM	No fit	0.251	0.795	0.0457
NN	No fit	1000	No fit	1.78

Table 5.6. ML model validation performances for predicting T_1 to S_0 transition energy. Parameters were the same as the original models. The best performing model is highlighted in bold.

The Mordred models generally performed worse for the literature complexes than in the training data. With the exception being XGB which performed similarly to the CV performance: the R^2 score only dropped from 0.68 to 0.67 and the MAE increased from 0.060 eV to 0.064 eV. The fingerprint models outperformed the Mordred descriptors across the board. The fingerprint + RF model performed the best with an R^2 score of 0.87, increasing from the CV R^2 score of 0.69, and the MAE increasing slightly from 0.036 eV to 0.040 eV. Despite the small increase, the MAE of the final model was lower than that between the experimentally determined emission energies and the TDDFT calculated T_1 to S_0 transition energies (0.045 eV). Therefore, the fingerprint + RF predictive model provides a more accurate and efficient method for determining the peak emission energies, and thereby peak emission wavelengths of novel platinum(II) salphen complexes in place of further computationally expensive DFT and TDDFT calculations.

5.3.1.2 Prediction of additional DFT and TDDFT energies

In addition to the T_1 to S_0 transition energy models, models were developed to predict energies of the HOMO and LUMO orbitals, HOMO-LUMO gap and the S_1 to S_0 transition. The full model performance tables are detailed within appendix section A.3.4. Across the prediction tasks, it was interesting to see that the models and descriptors did not behave in a one size fits all manor. In fact, different combinations of descriptors and regression algorithms appeared to perform best across almost every task (table 5.7):

Prediction Task	Best model (by validation)	R^2	MAE (eV)
HOMO	Mordred + SVM	0.729	0.0588
LUMO	Mordred + XGB	0.871	0.0472
HOMO LUMO gap	Fingerprint + XGB	0.806	0.0429
S_1 to S_0	Fingerprint + XGB	0.874	0.0397
T_1 to S_0	Fingerprint + RF	0.868	0.0403

Table 5.7. Summary of best models for each DFT and TDDFT energy prediction task. Determined by validation scores. Model hyperparameters are described in table 5.6 and appendix section A.3.4.

For the prediction of the HOMO and LUMO orbital energies, the Mordred descriptors performed best with the SVM and XGB regression algorithms respectively. Whereas for the more complex tasks of energy transitions (HOMO LUMO gap, S_1 to S_0 and T_1 to S_0 transitions), it was the fingerprint descriptors with ensemble learning models (RF and XGB) which performed best. When predicting the transition energies the Mordred descriptors consistently fell short, with no model exceeding an R^2 score of 0.70. Whereas the fingerprint descriptors consistently achieved an R^2 score 0.80 and above (except for those which reported negligible R^2 scores).

The strong performance of Mordred within predicting the orbital energies may have been due to the descriptors containing information which directly represented the individual orbital energies. With comparison to the transition energies being better represented by structural descriptors which are significant for electronic transitions, and

are not captured as well within the Mordred descriptors.

Overall, the algorithms which performed best were those which have more complexity than the linear regression models, likely due to the relationships between features and the energies not being linear. XGB was the best performing regression algorithm across the board, providing three out of five of the best final models. XGB is known for its ability to model non-linear relationships and handle complex features. Additionally, due to being an ensemble model it does not rely on a single prediction of the target property, instead it uses a combination of multiple decision trees, resulting in a model which is more accurate and robust, hence the consistent performance of the model.

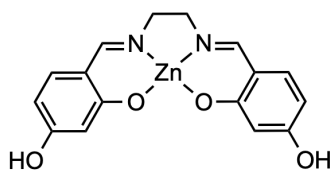
To test the bounds of the final models, their ability to extrapolate to other systems was investigated. This was achieved by assessing the models' ability to predict emission energies of alternate metal complex systems.

5.3.1.3 Prediction of emission energies for alternate metal complexes

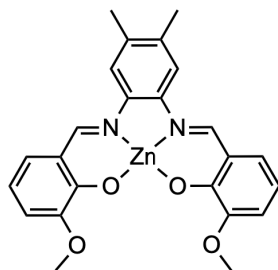
The systems selected for model extrapolation were emission energies of zinc(II) salphen complexes, and platinum(II) NNC complexes. Datasets were generated using experimental emission wavelengths from the literature converted to energy (eV) using Planck's equation (equation 4.9). Complexes were built using STK to generate Mordred and fingerprint descriptors.

These complexes were selected to explore the model's performance when changing the metal centre, the ligand, and a combination of both. To enable the comparison of models' performances when changing the metal independently, and when changing both the metal and the ligand (of the same class), additional TDDFT calculations were performed for the platinum(II) salphen derivatives of the zinc(II) salphen complexes. Figure 5.11 shows examples of the structures within the alternate metal complex datasets. The full list of complex structures can be seen within appendix section A.3.5.1.

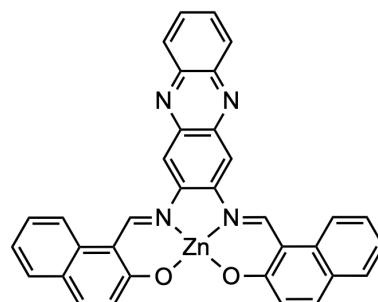
Zinc(II) salphen complexes



Zn_EDA_4OH

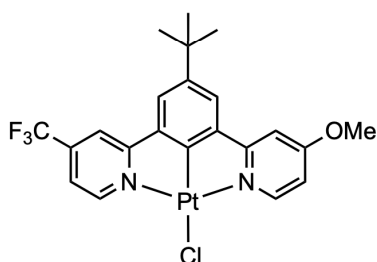


Zn_PhMe2_3OMe

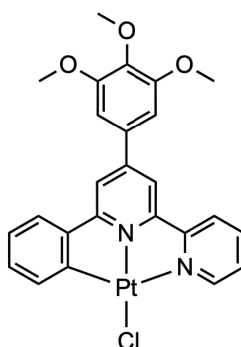


Zn_Phaz_Naph

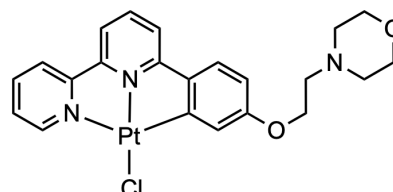
Platinum(II) NNC complexes



NCN_1CF₃_3OMe_2tBu



CNN_2Ph_{3,4,5}OMe₃



NNC_3PEM

Figure 5.11. Examples of complexes within the zinc(II) salphen and platinum(II) NNC complex libraries. The full list of complex structures are within appendix section A.3.5.1. Complex names were based on structural features, similarly to the platinum(II) salphen nomenclature.

When predicting the emission energy of the zinc(II) salphen complexes, most of the original platinum(II) salphen models reported negligible R^2 scores and MAEs of 0.50 eV and above. The only models which resulted in positive R^2 scores were those trained on platinum(II) salphen S_1 to S_0 transition energies (table 5.8). This is likely due to the fluorescent emission of the zinc(II) salphen complexes originating from the S_1 to S_0 energy transition as shown in the Jablonski diagram (figure 4.1). However, even the best performing models still had an MAE of 0.14 eV, which is 50 % of the standard deviation of the zinc(II) salphen peak emission energy dataset (0.23 eV). This suggests that the platinum(II) salphen models are unsuitable for directly predicting the emission energy of zinc(II) salphen complexes where both the metal centre and ligand system are unknown to the model.

S ₁ to S ₀ transition energy original models	R ²	MAE (eV)
Mordred + XGB	0.52	0.14
Mordred + RF	0.30	0.18
Fingerprint + Ridge	0.12	0.20

Table 5.8. Table showing the performance of the original ML models trained to predict the S₁ to S₀ energy of the 263 platinum(II) salphen complexes, applied to the task of predicting the emission energy of the zinc(II) salphen series.

Upon training new platinum(II) salphen models with the additional 49 platinum(II) salphen derivatives of the zinc complexes, the previous best performing models all had a performance boost with respect to their R² scores. Despite this, the MAEs remained high (table 5.9).

S ₁ to S ₀ transition energy model with additional platinum derivatives.	R ²	MAE (eV)
Mordred + XGB	0.62	0.13
Mordred + RF	0.68	0.11
Fingerprint + Ridge	0.32	0.18
Fingerprint + SVM	0.50	0.14

Table 5.9. Table showing the performance of the extended training set ML models trained to predict the S₁ to S₀ energy of the 302 platinum(II) salphen complexes, applied to the task of predicting the emission energy of the zinc(II) salphen series. The parameters and performance of the extended training models for the task of predicting the platinum(II) salphen S₁ to S₀ transition energy are as follows: Mordred + XGB – learning rate: 0.1, max depth: 3, Monte-Carlo R2: 0.682 ± 0.103 , MAE: 0.0791 ± 0.0081 eV. Mordred + RF: max depth: 10, n_estimators: 100, Monte-Carlo R2: 0.704 ± 0.145 , MAE: 0.0740 ± 0.0140 eV. Fingerprint + Ridge – alpha: 1, Monte-Carlo R2: 0.843 ± 0.0614 , MAE: 0.0555 ± 0.0056 eV. Fingerprint + SVM - C: 10, degree: 2, epsilon: 0.01, gamma: 'scale', kernel: 'linear', Monte-Carlo R2: 0.796 ± 0.0826 , MAE: 0.0494 ± 0.0075 eV. Unspecified parameters were default.

Figure 5.12 shows a comparison of the actual vs predicted emission energies for

the zinc(II) salphen complexes when predicted using the Mordred + XGB platinum(II) salphen S_1 to S_0 transition energy models trained on the original (A: unseen metal and ligand) and the extended (B: unseen metal) datasets.

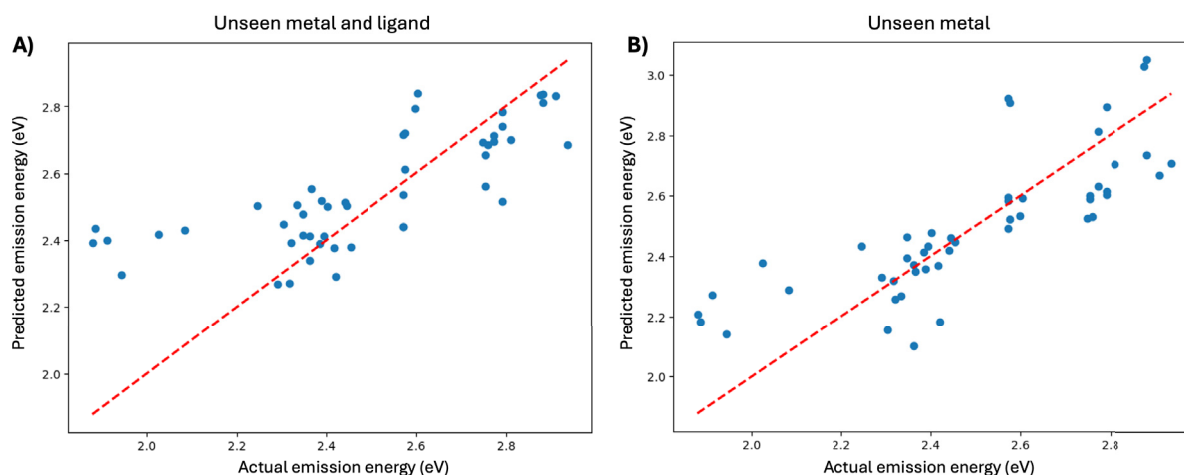


Figure 5.12. Prediction of zinc(II) salphen emission energies using the original (A) and extended (B) Mordred + XGB platinum(II) salphen S_1 to S_0 transition energy source domain models.

It could be seen that when predicting the emission energy of the zinc(II) salphen complexes with the original models (plot A) there was a moderate error between the actual and predicted values. However, the overall trend of the predictions followed a similar plane to the ideal prediction, resulting in the reasonable R^2 score (unseen metal and ligand: 0.52). Upon training new models with the platinum(II) salphen derivatives of the zinc(II) salphen complexes, prediction improved slightly, particularly for complexes which previously did not fit the trend of the data (data points on the left of each plot), improving the R^2 score further (unseen metal: 0.62).

When attempting to predict the emission energy of the platinum(II) NNC complexes with the original platinum(II) salphen models, no reasonable predictions were possible, resulting in negligible R^2 scores and large MAEs.

5.3.2 Transfer learning for the prediction of emission energies for alternate metal complexes

To improve the predictive power of the platinum(II) salphen models for the alternative metal complex systems, TL was explored. TL is an ML technique where models

which are trained on one task are adapted to perform a related task [384, 385]. TL model training starts from the pre-optimised hyperparameters of the models trained for the original task. The model is then re-trained with the new task data, using the previously learnt patterns between the input features (molecular descriptors) and the target variable of the original task, known as the source domain. Within the new models the learned feature weights of the original model are kept unchanged, and the decision boundary of the new model is adapted to fit the new input data. The algorithms explored in the context of TL were lasso, ridge, RF, XGB and SVM. TL was implemented to adapt the initial platinum(II) salphen models (see section 5.3.1) to the task of predicting the emission energies of zinc(II) salphen complexes and platinum(II) NNC complexes. To explore the effect of the source domain on the performance of the final TL models, models were re-trained from source domains including the HOMO LUMO energy gap, the S_1 to S_0 transition energy and the T_1 to S_0 transition energy, with both Mordred and fingerprint descriptors. Models were tested for their robustness using the R^2 and MAE values across CV data splits as detailed in section 5.3.

5.3.2.1 Transfer learning to predict zinc(II) salphen emission energy

The zinc(II) salphen emission energy TL models were re-trained on 80 % of the zinc(II) salphen emission energy dataset (39 data points). It was found that the performance of the TL models was not significantly impacted by the choice of source domain or the inclusion of the extended platinum(II) salphen derivatives. Table 5.10 contains the performance of the TL models which were trained on the original models using the S_1 to S_0 transition energy source domain.

The fingerprint TL models performed the best overall, with the fingerprint + ridge TL model obtaining an R^2 score of 0.83 and an MAE of 0.077 eV (figure 5.13). This TL model also achieved an impressive R^2 score of 0.97 and an MAE of 0.030 eV on the test data (plot A). The best performing Mordred TL model was with SVM, achieving a CV R^2 score of 0.73 and an MAE of 0.11 eV. This TL model performed very strongly on the test data with an R^2 score of 0.83 and an MAE of 0.08 eV. However, the performance of these TL models were strongly influenced by the CV data splits.

Zinc(II) Salphen TL Model (CV)	Mordred		Fingerprint	
	R ²	MAE (eV)	R ²	MAE (eV)
Ridge	No fit	0.360 ± 0.280	0.831 ± 0.0954	0.0769 ± 0.0216
Lasso	No fit	0.189 ± 0.0935	No fit	0.203 ± 0.0442
RF	0.613 ± 0.195	0.113 ± 0.0337	0.828 ± 0.105	0.0662 ± 0.0210
XGB	0.377 ± 0.298	0.144 ± 0.0529	0.769 ± 0.437	0.0643 ± 0.0328
SVM	0.728 ± 0.111	0.102 ± 0.0311	0.760 ± 0.0951	0.101 ± 0.0238

Table 5.10. TL ML model performances for predicting zinc(II) salphen emission energy. Models were re-trained from the platinum(II) salphen S₁ to S₀ transition energy models. The best performing model is highlighted in bold. The model parameters are the same as the original models.

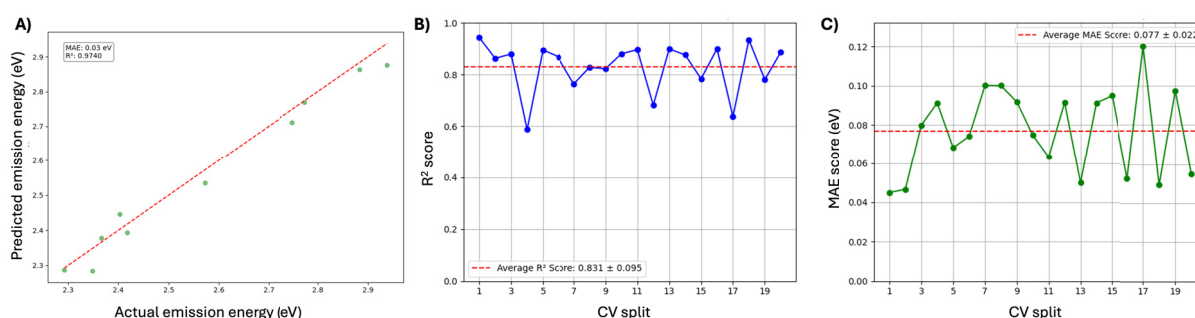


Figure 5.13. Plots showing the performance of the the fingerprint + ridge TL model for predicting zinc(II) salphen emission energies. Plots show the actual vs predicted emission energies for the zinc(II) salphen test set (A), R² score (B) and MAE (C) across CV splits.

The TL models showed a slight improvement in model performance consistency across CV folds compared to models trained only on the 49 zinc(II) salphen emission energies as seen in figure 5.14 (full results are within appendix table A.20). However, the lack of any significant performance improvement suggested that the models were not gaining additional insights from re-training the pre-trained models, and were instead just over-fitting to the training data, which is the most likely case due to the small size of the zinc(II) salphen dataset. This requires confirmation by using an external training set to validate generalisability of the TL vs experimental models.

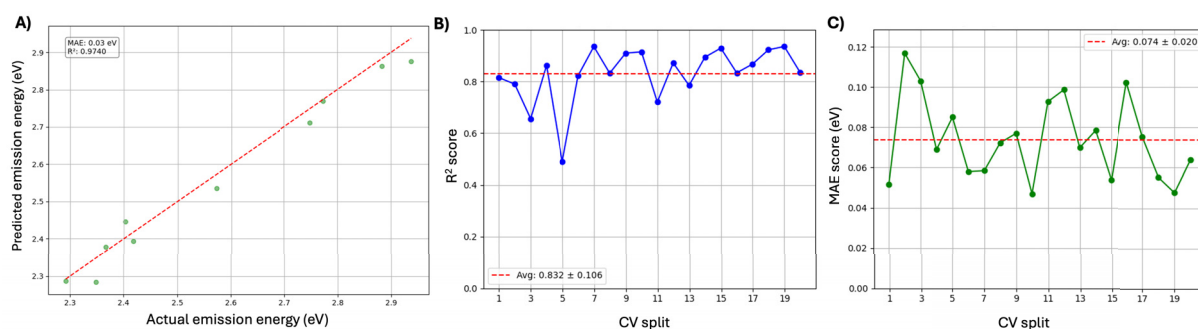


Figure 5.14. Plots showing the performance of the the fingerprint + ridge model for predicting zinc(II) salphen emission energies trained only on the zinc(II) salphen dataset. Plots show the actual vs predicted emission energies for the zinc(II) salphen test set (A), R^2 score (B) and MAE (C) across the CV data splits.

5.3.2.2 Transfer learning to predict platinum(II) NNC complex emission energy

The platinum(II) NNC emission energy TL models were re-trained on 80 % of the platinum(II) NNC emission energy dataset (32 data points). It was found that the performance of the TL models was significantly impacted by the choice of source domain, with models trained on the S_1 to S_0 transition energies performing best overall. Interestingly, the Mordred descriptors outperformed the fingerprint descriptors. This was likely a result of the Mordred descriptors being more similar across the two distinctly different metal complex structure datasets. In the case of the fingerprint descriptors, there may have been very little overlap between the representation of the platinum(II) salphen complexes and the platinum(II) NNC complexes, resulting in an inability to detect meaningful trends across the datasets by DA. Table 5.11 below shows the best TL models for both Mordred and fingerprint descriptors with each source domain. The full collection of TL models and their performances are listed within appendix table A.22.

The best performing model was Mordred + RF trained on the S_1 to S_0 transition energy source domain, with an R^2 score of 0.76 and an MAE of 0.061 eV (figure 5.15). However, this model had a poor test performance with an R^2 score of 0.18 and an MAE of 0.14 eV (plot A). The next best model was the fingerprint + ridge model, trained on both the HOMO LUMO gap and the S_1 to S_0 transition energy (both models had the same

performance) with an R^2 score of 0.63 and an MAE on 0.072 eV, as well as test R^2 score of 0.39 and an MAE of 0.14 eV. The inconsistent performance within the test data could be explained by the split sensitivity observed within these models, demonstrated by the large standard deviations in model performances across CV splits. This could have been due to the small size of the platinum(II) NNC complex dataset resulting in the model being unable to detect meaningful trends within the data. Surprisingly, models trained on the T_1 to S_0 energy transition performed the worst overall, with no R^2 scores exceeding 0.35 and all MAEs above 0.1 eV. This was not expected as the platinum(II) NNC complexes are phosphorescent, therefore their emission energy originates from the T_1 to S_0 transition, similarly to the platinum(II) salphen complexes.

Source domain	Platinum(II) NNC TL model (CV)	R^2	MAE (eV)
HOMO LUMO gap	Mordred + XGB	0.583 ± 0.299	0.0692 ± 0.0214
HOMO LUMO gap	Fingerprint + Ridge	0.632 ± 0.174	0.0723 ± 0.0210
S_1 to S_0 energy	Mordred + RF	0.760 ± 0.123	0.0610 ± 0.0173
S_1 to S_0 energy	Fingerprint + Ridge	0.632 ± 0.174	0.0723 ± 0.0210
T_1 to S_0 energy	Mordred + RF	0.257 ± 0.350	0.113 ± 0.0352
T_1 to S_0 energy	Fingerprint + SVM	0.342 ± 0.401	0.123 ± 0.0407

Table 5.11. TL ML model performances for predicting platinum(II) NNC emission energy. Models were re-trained from the original platinum(II) salphen models with the original prediction task specified within the table. The best performing model is highlighted in bold. The model parameters are the same as the original models.

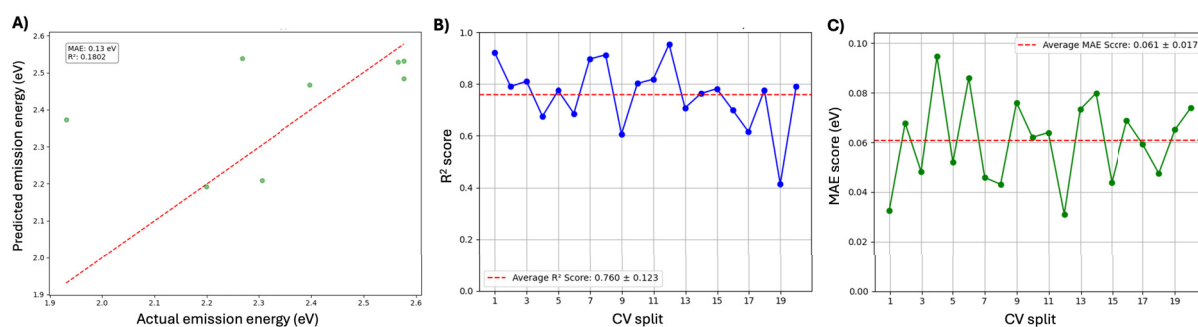


Figure 5.15. Plots showing the performance of the the Mordred + RF TL model for predicting platinum(II) NNC emission energies. Plots show the actual vs predicted emission energies for the platinum(II) NNC test set (A), R^2 score (B) and MAE (C) across the CV data splits.

When comparing the TL models to models trained on only the 40 platinum(II) NNC emission energies (figure 5.16), there was no significant improvement in performance (full results are within appendix table A.21). This again indicated that the TL model was over-fitting to the training data. This requires validation within future work by using an external training set to validate generalisability of the TL vs experimental models.

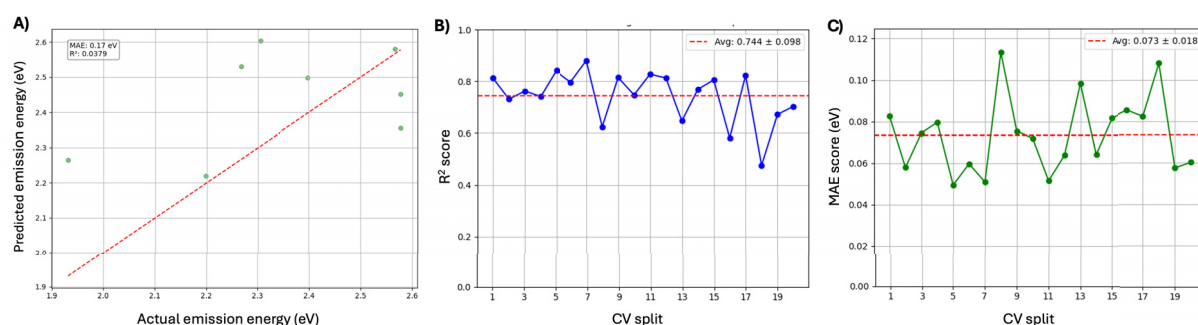


Figure 5.16. Plots showing the performance of the the Mordred + RF model for predicting platinum(II) NNC emission energies trained only on the platinum(II) NNC dataset. Plots show the actual vs predicted emission energies for the platinum(II) NNC test set (A), R^2 score (B) and MAE (C) across the CV data splits.

5.3.3 Domain adaptation for the prediction of emission energies for alternate metal complexes

As TL was unable to improve the predictive ability of platinum(II) salphen models to the new target domains, DA, an alternate TL technique was investigated. DA takes

pre-trained models and tailors them to a new target domain without retraining the model [384]. All hyperparameters and learnt underlying relationships between the input features and the target variables remain the same, but the model applies new feature pre-processing to the new targets input features. Pre-processing includes feature scaling, normalisation or other data transformations which make the new input data align with the original models' expectations. This essentially means that within DA, the models predictive hyperplane is the same and the input features are adapted to fit that plane of best fit. Whereas pure TL redefines the plane of best fit for the new data.

DA is beneficial in this scenario as the zinc(II) salphen and platinum(II) NNC datasets are too small for the models to effectively learn the relationship between the features and the emission energies. Even when using a pre-trained model, the retraining results in a loss of predictive power. As the task of the original models and the new models are both to predict emission energy, it can be assumed that the underlying relationships between the input features and the emission energies are the same across the domains, therefore making DA suitable for this TL task. DA was performed with both Mordred and fingerprint descriptors using lasso and ridge respectively, and XGB with both.

5.3.3.1 Domain adaptation to predict zinc(II) salphen emission energy

When attempting to predict the emission energies of the zinc(II) salphen complexes, the best DA model was fingerprint + XGB trained on the S_1 to S_0 transition energy source domain, with an R^2 score of 0.86 and an MAE of 0.06 eV (figure 5.17). DA provided a slight improvement to the R^2 score than the previous TL model and the model trained on the zinc(II) salphen dataset alone, alongside a significant improvement in the stability of the prediction across CV splits (table 5.12). Additionally the DA model performed significantly better on the test set than any other model with an R^2 of 0.97 and an MAE of 0.03 eV (plot A figure 5.17). The Mordred DA models were unable to achieve an R^2 score > 0.6 and showed a similar training split sensitivity as the TL and zinc only trained models (appendix table A.23).

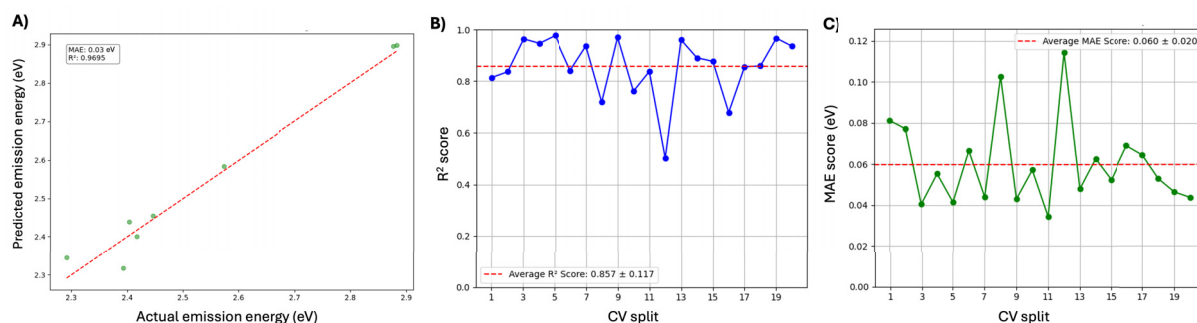


Figure 5.17. Plots showing the performance of the the fingerprint + ridge DA model for predicting zinc(II) salphen emission energies. Plots show the actual vs predicted emission energies for the zinc(II) salphen test set (A), R^2 score (B) and MAE (C) across CV splits.

S_1 to S_0 : Fingerprint + XGB	R^2	MAE (eV)
DA model	0.857 ± 0.117	0.060 ± 0.020
Pure TL model	0.769 ± 0.437	0.0643 ± 0.0328
Direct prediction from original model	No fit	0.327
Model trained on zinc(II) salphen dataset	0.825 ± 0.216	0.0706 ± 0.0429

Table 5.12. DA ML performances for the task of predicting zinc(II) salphen emission energy using the fingerprint descriptors. The table contains models trained by DA, pure TL, direct prediction from the original platinum(II) salphen model and the models trained exclusively on the zinc dataset for comparison of performances. The best performing model is highlighted in bold.

5.3.3.2 Domain adaptation to predict platinum(II) NNC complex emission energy

When attempting to the predict the emission energies of the platinum(II) NNC complexes, the best DA model was the Mordred + XGB model pre-trained on the T_1 to S_0 transition energy source domain. This DA model scored an R^2 of 0.74 and an MAE of 0.062 eV, with a low standard deviation across the different CV folds (figure 5.18). However, despite this reasonable CV performance, the DA model was still unable to accurately predict the emission energies of the platinum(II) NNC complexes within the test set (plot A figure 5.18).

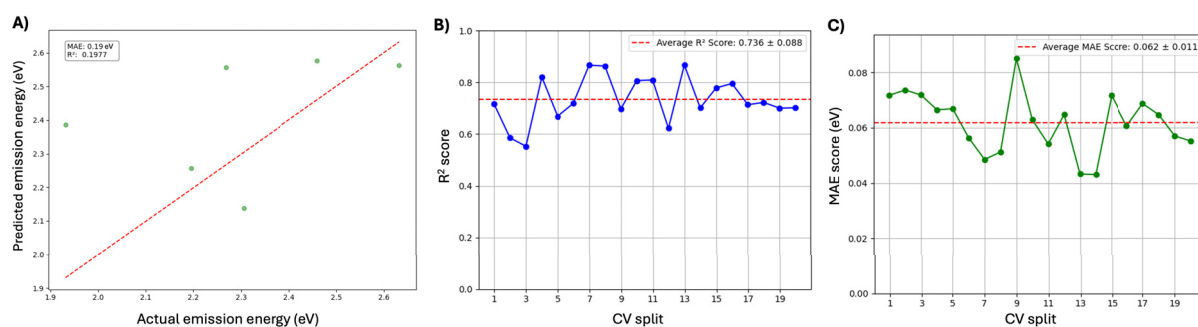


Figure 5.18. Plots showing the performance of the the fingerprint + ridge DA model for predicting platinum(II) NNC emission energies. Plots show the actual vs predicted emission energies for the platinum(II) NNC test set (A), R^2 score (B) and MAE (C) across CV splits.

All other DA models scored an R^2 of less than 0.3. The strong performance of the Mordred + XGB model mimicked the results seen when training on the platinum(II) NNC dataset alone. However, DA significantly reduced the standard deviation of R^2 and MAE performance, therefore improved the models consistency (table 5.13).

T_1 to S_0 : Mordred + XGB	R^2	MAE (eV)
DA model	0.736 ± 0.088	0.0619 ± 0.0106
TL model	No fit	0.130 ± 0.0488
Direct prediction from original model	No fit	0.2174
Model trained on platinum(II) NNC dataset	0.723 ± 0.241	0.0927 ± 0.0279

Table 5.13. DA ML performances for the task of predicting platinum(II) NNC emission energy. The table contains models trained by DA, pure TL, direct prediction from the original platinum(II) salphen model and the models trained exclusively on the zinc dataset for comparison of performances. The best performing model is highlighted in bold.

Similarly to what was seen with the platinum(II) NNC TL models, the Mordred descriptors outperformed fingerprint descriptors for predicting the platinum(II) NNC emission energies. This could be because fingerprint descriptors are solely structural, and for DA to be effective the molecular descriptors across both domains need to be similar to fit to the fixed model decision boundaries. This works well for zinc(II) salphen complexes as they share the same structural scaffolds as the platinum(II) salphen complexes. However, for the platinum(II) NNC complexes, the structures are

significantly different. Therefore, more transferable information is obtained from the Mordred descriptors which provide features related to complex properties which may be more comparable across the two ligand scaffolds. As mentioned previously, further validation is required to truly comment on the robustness of the DA models.

5.4 Determination of features important for DFT and TDDFT energies

Many of the previously explored ML models can identify which input features are significant in order to make predictions of the target property. Lasso determines feature importance by initially applying equal shrinkage to all coefficients, then as the model is optimised, features more strongly associated with the target variable maintain larger coefficients, whilst less important features shrink to zero. RF determines feature importance by building multiple decision trees which use a random selection of the input features, determining which features lead to a decrease in model accuracy when absent: features which cause the biggest decrease in model accuracy are considered important. Due to Mordred's chemical interpretability, this can be used to provide insights into which chemical properties are important for predicting target variables. The ML models explored to identify important features were lasso and RF which investigate linear and non-linear relationships respectively. Examples from the literature have shown that these models are able to interpret which molecular descriptors within chemically significant for biological and chemical based prediction tasks [386–390].

To visualise the features deemed important by each model, the top 20 ranked features were plotted within a nested pie chart. The outer pie shows which class of Mordred descriptor that the feature originates from, and the inner pie shows a number correlating to the feature name: the colours of each segment represent the molecular property represented by that feature. Each class of descriptors use different mathematical methods for calculating numerical representations of properties, including electronegativity, polarizability, charge, ionisation potential, electron distribution, connectivity, geometry, valence electrons, specific atom types, mass, surface area, hydrophobicity, and aromaticity. Using this information it is possible to visualise which molecular properties are most strongly related to the target properties. Specific names of features within each nested pie chart can be found within appendix figure A.246.

DFT and TDDFT energies were selected for feature importance determination as dataset size has a significant impact on model accuracy, and therefore the accuracy of the relationships detected between the input features and the target variable. As a result, the smaller experimental datasets were not appropriate to use.

Features determined to be important for predictions with lasso and RF for the HOMO and LUMO energies are shown within the nested pies in figure 5.19.

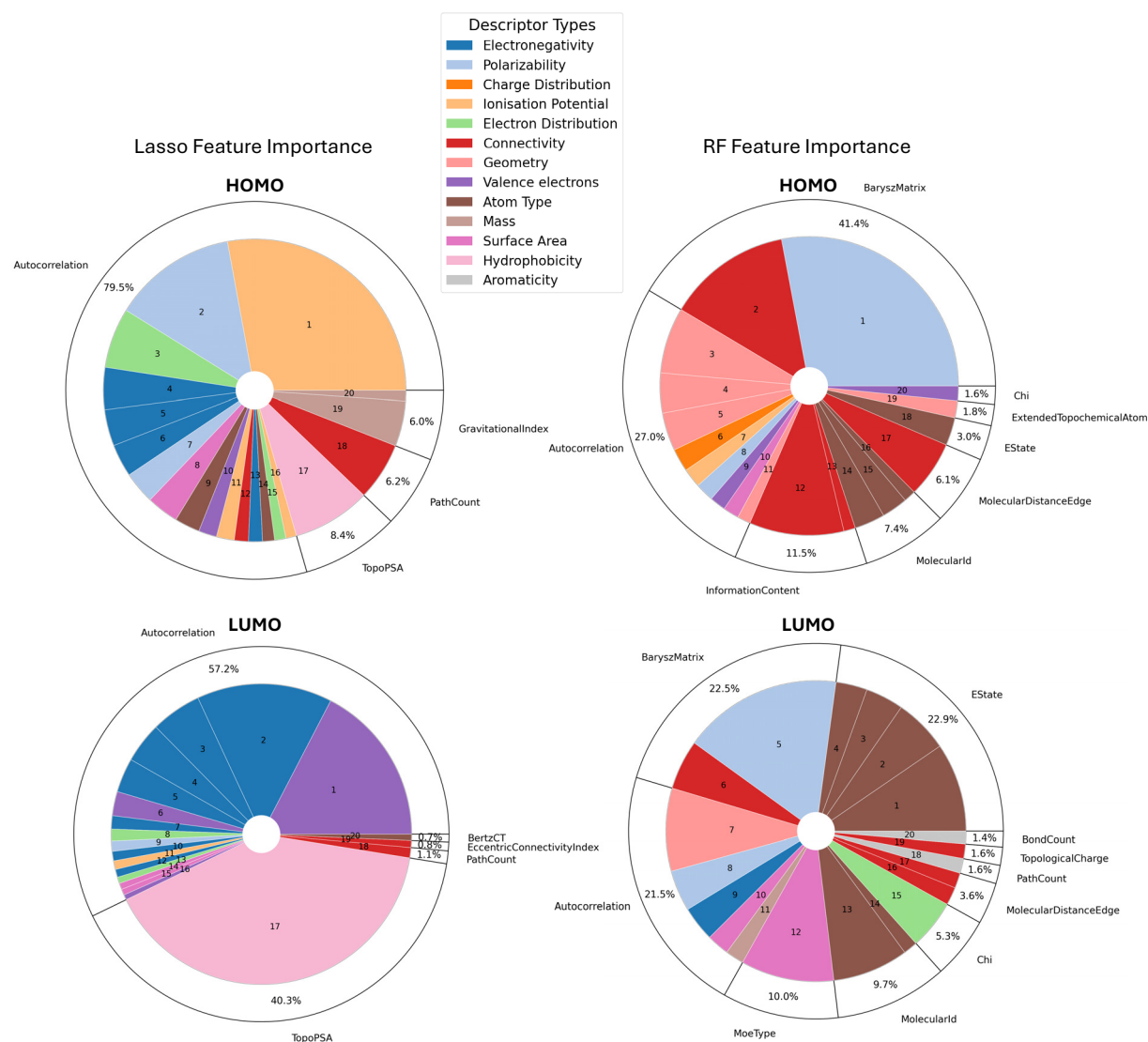


Figure 5.19. Nested pie charts showing the features determined to be important for HOMO (top) and LUMO (bottom) energies by lasso (left) and RF (right). The outer pie shows the descriptor class. The inner pie is coloured by the property the feature represents, and the number relates to the descriptor name within appendix figure A.246.

For the HOMO energy, lasso highlights a combination of features which describe electronic and geometric properties. The dominance of the autocorrelation class indicated that the distribution of these properties across the complex are important in determining the HOMO energy. When assessing feature importance by RF regression, there was a larger emphasis on the importance of geometry, connectivity, and the presence of specific atoms within the structure, indicating a more complex relationship between structural properties and the HOMO energy. RF also reiterated the importance of electronic properties within the prediction of the HOMO energy.

Lasso regression for the LUMO energy unveiled distinctively different descriptors as being significant than those for significant for the HOMO energy. The top features were dominated by a single descriptor: topological polar surface area (TPSA), which contributed to 40.3 % of the total important features. This TPSA descriptor focuses on contributions from oxygen and nitrogen atoms to the overall TPSA. The remaining features described the electronic properties of the complex. In addition to some small contributions from surface area and connectivity descriptors. Again, autocorrelation features are dominating overall, reiterating the message that the distribution of these properties across the complex are important in determining the orbital energies. The RF feature importance for the LUMO energy also highlights the importance of connectivity, geometry, and the presence of specific atoms. In this case, the specific atom descriptors dominate the important features, and a larger importance is placed on surface area and aromaticity. There also continues to be a significant contribution from electronic descriptors, making up over 25 % of the total feature importance scores.

Descriptors determined to be important by lasso and RF for the HOMO and LUMO energies align with factors which are known to influence these energies including conjugation (represented by connectivity), substitutions (represented by specific atom descriptors), and electronegativity. These effects were demonstrated for the platinum(II) salphen complexes within section 4.2, where it was seen that the position and nature of substituents on the aromatic system of the complex directly influenced the HOMO and LUMO shapes, electron density and resulting energy.

There were similarities in the features which were determined to be important for each transition energy value: HOMO LUMO gap, S_1 to S_0 , and T_1 to S_0 (figure 5.20).

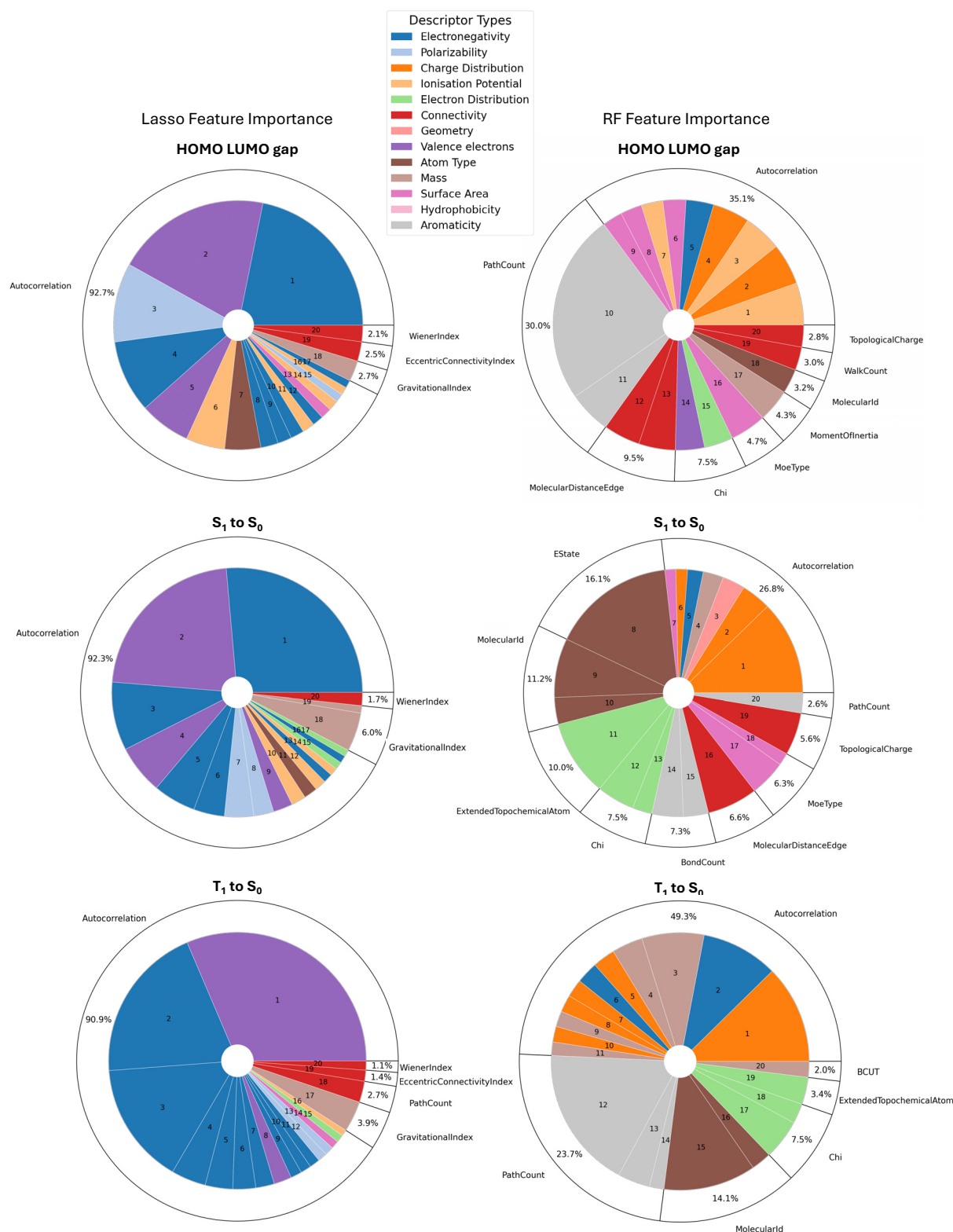


Figure 5.20. Nested pie charts showing the features determined to be important for HOMO LUMO gap energy (top), the S_1 to S_0 transition energy (middle) and the T_1 to S_0 transition energy (bottom) by lasso (left) and RF (right). The outer pie shows the descriptor class. The inner pie is coloured by the property the feature represents, and the number relates to the descriptor name within appendix figure A.246.

For the lasso feature importance, all energy transition properties were dominated by autocorrelation features representing electronic structure, with small contributions from features representing connectivity and mass. Within the RF feature importance there was a much smaller contribution from electronic properties with much more emphasis on geometry, connectivity, and the presence of aromatic groups and specific atoms.

These important features align well with what is known about the molecular properties which influence electronic transitions. Firstly, for a complex to be emissive it must contain aromatic functionalities, hence the presence of features representing conjugation and aromaticity for all transition properties. Additionally, geometry is significant for enabling energy transitions through its influence of orbital overlap and electron delocalisation. The added significance of mass descriptors in the T_1 to S_0 transition energy prediction aligns with the relationship between the mass of the platinum metal centre and the complexes ability to undergo an intersystem crossing to access the T_1 energy state. Similarly to the HOMO and LUMO energies, section 4.2 demonstrates the strong influence of substituents on the T_1 to S_0 transition energy of complexes.

5.5 Computational conclusions

Within this chapter, a semi-automated HT computational pipeline was developed to generate an *in silico* library of 263 platinum(II) salphen complexes from the SMILES strings of their building blocks. DFT and TDDFT calculations were then performed on the *in silico* library. From these calculations the T_1 to S_0 transition energy was used to represent the experimental peak emission wavelength for the development of ML predictive models. This was justified by the known relationship between the T_1 to S_0 transition energy and peak emission wavelength from the Jablonksi diagram (figure 4.1), Planck's equation (4.9), and the PC analysis of DFT and TDDFT energies to the platinum(II) salphen complexes photophysical properties.

As part of the model development, two molecular descriptors were investigated: Mordred descriptors and fingerprint descriptors. When exploring how these descriptors represented complexes using unsupervised learning, it was found that the Mordred descriptors were not able to significantly differentiate between the **CBA** and **DAB**

components of the platinum(II) salphen complexes. Upon the application of PCA dimensionality reduction to the Mordred descriptors, more structure could be seen within the t-SNE plot of the data, with **CBA** dominating the groupings. This indicated that the Mordred descriptors contained many redundant features which were leading to dilution of the important features for distinguishing between the complexes.

For the fingerprint descriptors, representation appeared to be dominated by the nature of the DAB linker group. After applying PCA, the grouping became much more strongly dominated by the **CBA** groups, similarly to Mordred PCA, except that the groups within the centre of the plot for fingerprint PCA were more defined. Overall fingerprints appeared to provide a better distinction between the difference complexes within the platinum(II) salphen dataset.

Despite the fingerprint descriptors appearing to provide the best representation of complexes in section 5.2, both the Mordred and fingerprint descriptors resulted in the best performing models across different prediction tasks. Mordred descriptors provided the best models when predicting the HOMO or LUMO energies. In comparison, fingerprint descriptors resulted in the best models when predicting energy transitions such as HOMO LUMO gap, S_1 to S_0 , and T_1 to S_0 transition energies. The development of these models has enabled the accurate, and rapid prediction of energy values for novel platinum(II) salphen complexes without the need for further DFT and TDDFT calculations.

The direct application of these models to predict the emission energies of zinc(II) salphen and platinum(II) NNC complexes resulted in poor predictions. To attempt to overcome this TL was implemented, however due to the small size of the datasets, retraining the pre-trained models resulted in poor performances. The optimal approach was found to be DA which used the original models in a frozen form and instead adapted the input features to fit the predictive hyperplane. This overcame the issues of the small dataset as it did not change the original models. As a result, a good predictive model for zinc(II) salphen emission energy was developed (fingerprint + XGB: R^2 0.87, MAE of 0.06 eV). Less success was seen for the prediction of platinum(II) NNC emission energies, likely due to the large differences between the structures of the NNC and salphen ligands, which are responsible for the origin of the emission. To further confirm

the performance of the TL and DA emission energy models, additional data is required to validate performances on unseen data.

Finally, using lasso and RF models with Mordred, it was possible to highlight features important for predicting the DFT and TDDFT energies. The properties predicted to be important aligned with known molecular properties which influence the investigated energies. This could be interpreted as the models being able to detect trends between the input features and the prediction properties based on chemical based knowledge.

Overall, the models developed within this chapter have enabled the prediction of DFT and TDDFT energies of previously unseen complexes, negating the need for additional DFT and TDDFT calculations. As well as provided an insight into the SAR of the platinum(II) salphen complexes and their orbital and transition energies.

Chapter 6

Conclusions and future outlook

Small molecule cancer drugs face many therapeutic challenges including off target side effects and drug resistance. To date, there are only 321 regulatory approved small molecule cancer drugs, and only 16% of new cancer drugs approved between 2009 and 2020 were based on new mechanisms of action [120, 145]. The challenges associated with the development of new drugs are accentuated by the fact that it takes on average 14 years for a new drug to reach the market, with many potential drugs lost to attrition as they move through the drug development pipeline outlined in chapter 1 [46]. This has therefore led to AI and automation being implemented to accelerate the drug discovery process. Within this thesis, HT synthesis, screening, and ML techniques were applied to the discovery of platinum(II) salphen complexes as photoactive anti-cancer therapeutics.

With respect to the project objectives outlines in section 1.4: a small diverse library of 22 platinum(II) salphen complexes were obtained, however, this was not achieved in HT. Although it was found that a wide variety of **CBA** functionalities were well tolerated within the platinum(II) salphen formation reaction, challenges arose with the functionalised **DABs** due to varying reactivity dependent on the nature of the substituents. **DABs** which contained ether functionalities were found to degrade overtime, and those with amide functionalities were found to have reduced reactivity, requiring dry conditions, increased reaction times, and temperatures. When using **DABs** without solubilising groups attached, complexes were found to be insoluble and precipitated out of solution within the HT synthesis conditions. Due to the observed synthetic challenges associated with varying the **DAB**, it should be concluded that the

platinum(II) salphen system is not suitable for miniaturised HT synthesis.

HT screening protocols to measure the DNA binding affinity and ROS generating ability were successfully developed using the 22 platinum(II) salphen complexes. Using the HT DNA binding assay, it was possible to rapidly screen all 22 complexes against 6 unique DNA topologies in duplicate, in less than a week. However, the development of this protocol was a time-consuming process, requiring calibration of the OT-2 liquid handling robot, and optimisation of python scripts through extensive testing to ensure pipetting was accurate and reproducible. This time investment was deemed to be worthwhile as the protocol rapidly accelerated the measurement of DNA binding affinities without the concern of human error during pipetting. The protocol was also implemented within other research projects to explore RNA binding, other emissive platinum(II) complexes, and organic fluorescent probes across institutional and international collaborations, yielding two published papers and one manuscript in preparation. The ROS assays were not translated into OT-2 protocols as only a single plate was required for the 22 complexes to be measured in triplicate.

Complexes with a range of ROS generating strengths were taken forward to explore their ability to damage DNA via ROS generation. Using a plasmid DNA assay and ROS quenching additives, it was found that complexes which had poor ROS type 1 and type 2 generation were unable to cleave plasmid DNA. Whereas complexes which had strong ROS type 1 generation, but weak ROS type 2 generation were still able to cleave plasmid DNA. It was not explored if weak ROS type 1 and strong ROS type 2 abilities would also cause plasmid DNA cleavage, but this was expected to be true as the ROS quench assay suggested that both ROS type 1 and ROS type 2 played a role in plasmid DNA cleavage. In contrast, when exploring ROS DNA damage with G4 DNA, it was found that the guanine residues were being oxidised as opposed to DNA backbone cleavage. This was likely a result of the binding modes of the platinum(II) salphen complexes to the G4 structures being π - π stacking interactions, therefore putting the complex in close proximity to the guanine residues resulting in the guanine oxidation from the generated ROS.

Cytotoxicity and phototoxicity experiments revealed that the complexes were non-toxic until irradiated, at which point 68% became highly phototoxic, making platinum(II)

salphen complexes excellent candidates as potential photosensitisers for PDT. When exploring cellular behaviours further, HT fluorescent microscopy revealed that only 2 of the 22 complexes were not cell permeable (**Pt_PPTMA_3F** and **Pt_PPTMA_5OMe**). Confocal fluorescent microscopy then revealed that the nature of the complex **DAB** influenced the cellular localisation. It was found that complexes with a fixed positively charged **DAB** (**PPTMA**) localised within mitochondria, and complexes with a protonatable piperidine or morpholine containing **DABs** (**PEM** and **AEP**) localised within nucleoli and lysosomes. It was also noted that complex cellular localisation was time-dependent, with progressive accumulation within lysosomes over time suggesting cells were actively removing the complexes from organelles.

Two lead compounds were identified within this work: **Pt_PEM_4OMe** and **Pt_AEP_4OMe**. The lead compounds were shown to be non-toxic within cancer cells until irradiated, at which point they exhibited IC_{50} values of 0.29 μ M and 2.57 μ M respectively. **Pt_PEM_4OMe** and **Pt_AEP_4OMe** both localise within nucleoli and lysosomes, demonstrated selective G4 binding over duplex DNA, and displayed moderate ROS generating abilities. Future work for these complexes should include biological studies to assess their behaviour within healthy cells, and *in vivo* studies to assess their pharmacokinetic and pharmacodynamic properties.

Energy values determined by DFT and TDDFT (HOMO, LUMO, HOMO LUMO gap, T_1 to S_0 and S_1 to S_0 transitions) were used to rationalise the observed photophysical properties across the platinum(II) salphen complexes. Direct correlations were identified between the HOMO LUMO gap and T_1 to S_0 transition energy with the peak emission wavelength. Due to the very high correlation and known physical relationships between the T_1 to S_0 transition energy and emission wavelength, it was possible to generate a dataset of experimentally relevant properties to analyse using ML. However, when utilising molecular docking to calculate complex DNA binding affinities, it was found that docking was unable to provide distinct binding scores for complexes across the series, and so was instead used qualitatively to assess binding modes.

The ML models developed for predicting DFT and TDDFT energies utilised both Mordred descriptors and fingerprint descriptors for feature representation. Notably, the predictive performance of models with each descriptor were dependant on the target

property, with Mordred descriptors best describing orbital energies, and fingerprints best describing transition energies. It was demonstrated that ML models trained on as few as 263 data points were able to accurately predict the calculated energies of platinum(II) salphen complexes. The final models were then adapted using TL techniques to predict the emission energies of much smaller metal complex datasets including zinc(II) salphen complexes and platinum(II) NNC complexes. It was found that accurate models could be built using DA with the zinc(II) salphen complexes due to strong structural similarities to the original platinum(II) salphen complex training set, varying only by the metal centre. In contrast, it was not possible to generate an accurate model to predict the emission energy of the platinum(II) NNC complexes, likely due to the large structural divergence from the platinum(II) salphen complexes.

The original objective of the ML models were to be able to predict desirable properties of untested compounds. Whilst this has been proven to be possible in the context of emission energy, the objective of predicting DNA binding affinity and ROS generating abilities may be a more challenging task. DNA binding is dependent on a multitude of factors including shape complementarity and ligand flexibility, and ROS generation can be dependent on ligand conformation and oxygen availability. As a result, these properties are more complicated to predict than emission energy, which is a more well-defined physical property dependant on molecular structure and calculable orbital energies [64, 295, 391]. Therefore, more data and alternate molecular descriptors may be required to accurately predict DNA binding and ROS generation.

To conclude, this work has laid a proof-of-concept foundation for a HT screening and ML analysis pipeline which can be implemented following the development of a successful HT synthetic protocol, which was attempted in chapter 2. Successful synthesis with as few as three 96-well plates would yield 288 complexes. These complexes could then enter the HT screening pipeline developed in chapter 3 to generate an experimental dataset containing DNA binding and ROS generation properties. The extended experimental datasets could then be analysed using ML techniques showcased in chapter 5 to attempt to train ML models to predict the DNA binding and ROS generating abilities of untested complexes. The overall goal of the final models would be to guide synthetic efforts towards high value complexes.

Chapter 7

Experimental details

7.1 Material and methods

General procedures: ^1H and ^{19}F NMR spectra were recorded on a Bruker Avance Ultrashield 400 MHz NMR spectrometer at 24 °C. ^{13}C spectra were obtained on a Bruker Avance Ultrashield 500 MHz NMR spectrometer at 24 °C. LCMS spectra were obtained using a Waters LCT Premier (ES-Tof)/Acquity i-Class spectrometer and an Agilent LCMS spectrometer. LCMS UV detectors measured absorbance at 254 nm. HRMS analysis was performed on a LCT Premier mass spectrophotometer.

DNA preparation: G4 DNA stock solutions were prepared using Tris phosphate buffer (10 mM, pH 7.4) and strand concentrations were determined using an Agilent Cary UV 60 spectrometer. The stock solutions were then annealed at 95 °C for 5 minutes and cooled to room temperature overnight after the addition of KCl (100 mM). G4 DNA sequences (see table 7.1) were purchased from IDT and purified by standard desalting. Mass specs confirmed purity of 99 % and above. CT-DNA was purchased from Sigma-Aldrich and prepared using Tris phosphate buffer (10 mM, 100 mM KCl, pH 7.4). Base-pair concentrations were determined using an Agilent Cary UV 60 spectrometer. Plasmid DNA (pUC19) was purchased from Sigma–Aldrich and prepared as a 1 mg/mL stock solution in Tris-HCl (10 mM, pH 8.0, 1 mM EDTA).

G4	Genome location	PDB code	G4 sequence	Topology
HTG4	Human Telomere	2GKU	5'-TTGGGTTAGGGTTAGG GTTAGGGA-3'	Hybrid 3+1
<i>c-myc</i>	<i>c-MYC</i> oncogene promoter	2L7V	5'-TGGGGAGGGTGGGGAG GGTGGGGAAGG-3'	Parallel
HTelo	Human telomere terminal	5MVB	5'-TTAGGGTTAGGGTTAGG GTTAGGG-3'	Hybrid-2
<i>c-kit1</i>	<i>c-kit</i> oncogene promoter	4WO2	5'-AGGGAGGGCGCTGGGA GGAGGG-3'	Anti-parallel
LTR(III)	HIV-1 LTR major G4 structure	6H1K	5'-GGGAGGCGTGGCCTGG GCGGGACTGGGG-3'	Hybrid 3+1 with duplex stem-loop

Table 7.1. List of G4 structures with their location within the genome, PDB codes, full G4 sequences, and topologies.

Photophysical characterisation: The absorbance and emission spectra of each complex were measured using a BMG LabTech Clariostar plate reader. Absorbance profiles of complexes at 100 μ M in DMSO were measured from 300 nm to 850 nm with a resolution of 2 nm. Emission profiles of complexes at 2 μ M in DMSO and water were measured from 500 nm to 840 nm with a resolution of 2 nm, following excitation at their maximum absorbance wavelength determined in the absorbance screen.

ROS generation assays: Dihydrorhodamine-123 (DHR-123) (10 μ M) was combined with each complex (1 μ M) in PBS buffer to a total volume of 200 μ L in a clear 96-well plate. The plate was irradiated with a Lumidox 457 nm 96-LED array (8 mA, 2 mW/cm²) for 30 second intervals up to 3 minutes. The emission was measured using a BMG LabTech Clariostar plate reader: λ_{ex} = 505 nm and λ_{em} = 535 nm. The change

in emission was then fitted to a linear regression and normalised to the response of $[\text{Ru}(\text{bpy})_3]^{2+}$.

9,10-Anthracenediyl-bis(methylene)dimalonic acid (ABDA) (100 μM) was combined with each complex (32 μM) in PBS buffer to a total volume of 200 μL in a clear 96-well plate. The plate was irradiated with a Lumidox 457 nm 96-LED array (8 mA, 2 mW/cm²) for 30 second intervals up to 5 minutes. The absorbance at 380 nm was measured using a BMG LabTech Clariostar plate reader. The change in absorbance was then fitted to a linear regression and normalised to the response of $[\text{Ru}(\text{bpy})_3]^{2+}$.

Emission titrations: Emission titrations were performed with 2 μM of each complex in a Tris phosphate buffer supplemented with KCl (10 mM, 100 mM KCl, pH 7.4) with concentrations of quadruplex DNA ranging from 0 to 12 μM (concentration of strands), and CT-DNA from 0 to 120 μM (concentration of base pairs). This was performed within a 96 well plate prepared using an Opentron-2 liquid handling robot with volumes of 100 μL . Complex emissions were measured using a BMG LabTech Clariostar plate reader following excitation at 380 nm. The change in emissions were plotted against the oligonucleotide concentration for each complex and fitted using GraphPad Prism 9 One site specific binding: $Y = B_{\text{max}} * X / (K_d + X)$. K_d was then converted to K_a ($K_a = 1/K_d$) and error was measured by the relative standard error.

Docking Calculations: Platinum(II) salphen complex structures obtained from DFT geometry optimisation calculations were used for docking simulations in AutoDock4 using default settings [292]. PDB structures of DNA structures were obtained from the protein databank (see DNA preparation for exact sequences). Some DNA crystal structures contained a bound ligand which was removed prior to docking (*c-myc* (2L7V), *HTelo* (5MVB), LTR(III) (6H1K), duplex (1XRW) DNA). The sequences of these structures were otherwise unchanged. Structures including a bound ligand were not available for *c-kit1* (4WO2). Results were visualised using AutoDockTools.

Oxidative damage of DNA – gel electrophoresis studies: The DNA cleavage activities of the platinum(II) complexes in the dark and upon irradiation were determined by monitoring the conversion of supercoiled plasmid DNA (form I) to nicked circular DNA (form II), compared with enzyme cut linear DNA (form III) by using agarose-gel

electrophoresis. To investigate the effect of complex light-induced ROS generation on DNA damage, pUC19 (256 ng, 30 μ M) and selected platinum(II) complexes (30 μ M) were irradiated with a Lumidox 457 nm 96-LED array (8 mA, 2 mW/cm²) for 1 hour at a final volume of 15 μ M in Tris-HCl buffer (10 mM, pH 8.0, 1 mM EDTA). [Ru(Bpy)₃]²⁺ (30 μ M) was used as a high ROS generating positive control. Negative controls included each condition in the dark.

To determine the cleavage mechanism, pUC19 (256 ng) was incubated with the highest ROS generator of the series, **Pt_PPTMA_4OMe** (30 μ M), with the addition of radical scavengers: DMSO (10 mM, hydroxyl radical scavenger), KI (40 mM, superoxide scavenger) and NaN₃ (40 mM, singlet oxygen scavenger). Mixtures were then irradiated using a Lumidox 457 nm 96-LED array (8 mA, 2 mW/cm²) for 1 hour. Controls included pUC19 + scavengers under irradiation conditions, and pUC19 + **Pt_PPTMA_4OMe** + scavengers without irradiation.

To obtain linear DNA, pUC19 (256 ng, 5 μ L) was incubated with Eco31I (1 μ L), [17] 10x Green loading buffer (2 μ L) and water (12 μ L) at 37 °C for 1 hour. After irradiation, a loading buffer (0.25 % bromophenol blue, 0.25 % xylene cyanol and 60 % glycerol, 5 μ L) was added prior to being loaded onto a 0.8 % agarose gel stained with Sybr Safe (1.0 μ g mL⁻¹). The gel was run at 50 V for 1 h in Tris-acetate EDTA buffer (10 mM). The DNA bands were visualised using UV light in the Syngene G:Box Chemi-XRQ. Fiji ImageJ software was used to quantify DNA damage.

Circular dichroism (CD) to investigate G4 DNA damage: CD spectra were measured on a Chirascan V100 CD spectrophotometer using a quartz cell with a 1 mm path length and 200 μ L volume. The spectra were recorded from 200 to 450 nm at 20 °C with a scanning rate of 15 nm/min with a 1 nm bandwidth. A background CD spectrum was subtracted of the Tris phosphate buffer (10 mM, 100 mM KCl, pH 7.4) from each experimental measurement. The reaction mixture contained 1:3 DNA to complex (15 μ M:45 μ M). Between each measurement, the cuvette was irradiated with a Lumidox 457 nm 96-LED array (8 mA, 2 mW/cm²) for 10 minutes, up to 1 hour. **Pt_PPTMA_4OMe** was irradiated for a further 2 hours to investigate the effect of prolonged ROS generation. TMPyP4 was used as a known G4 ROS damaging positive control. CD spectra were also measured for **Pt_PPTMA_4OMe** over the course of 3 hours in the absence of

irradiation.

Cytotoxicity studies: HeLa cells were grown in high glucose Dulbecco's modified Eagle medium (DMEM) containing 10 % foetal bovine serum (FBS) at 37 °C at 5 % CO₂ in humidified air. The cells were seeded at a 5000 cell per well density in Greiner-Bio black μ Clear plates. After 24 hours the platinum(II) complexes were added, and the cells incubated for a further 24 hours at concentrations ranging from 25 μ M to 0.78 μ M. Following this, the cells were washed with PBS and the media was replaced with a mixture of media, MTS and PMS at a ratio of 4:0.95:0.05. After an additional 3 hours, the absorbance of MTS was measured at 490 nm, on a BMG LabTech Clariostar plate reader. The cell viability was calculated from the dose response curve of the MTS absorbance.

Phototoxicity studies: Hela cells were seeded at a 5000 cells per well density in Greiner-Bio black μ Clear plates. After 24 hours the platinum(II) complexes were added to the cells at concentrations ranging from 25 μ M to 78 nM. After 3 hours the cells were washed with PBS and fresh media was added before irradiating with a Lumidox 457 nm 96-LED array (8 mA, 2 mW/cm²) for 12 minutes. After a further 21 hours the media was replaced with a mixture of media, MTS and PMS at a ratio of 4:0.95:0.05. Then, after an additional 3 hours, the absorbance of MTS was measured at 490 nm, on a BMG LabTech Clariostar plate reader. The cell viability was calculated from the dose response curve of the MTS absorbance.

Cellular imaging – high-throughput screen: For the initial imaging screen, Hela cells were seeded at a 5000 cells per well density in Greiner-Bio black μ Clear plates. After 24 hours the cells were incubated with platinum(II) complexes at concentrations of 6.25 μ M, 12.5 μ M and 25 μ M. After 3, 6 and 24 hours the cells were washed with PBS and fresh media was added before imaging on an Eclipse-Ti2 Widefield system (Nikon), equipped with a Prime 95B sCMOS camera (Teledyne), pE-400 LED lightsource (CoolLED), a 20x or 40x objective, and emission filters for green (515/30 nm), orange (600/50 nm), red (641/75 nm) and far-red (705/72 nm). The brightfield channel was used to autofocus each well, after which the complexes were excited using the 405 nm laser and the red emission channel was collected. The contrast of the images was adjusted using a contrast adaptive equalisation formula applied to each individual image.

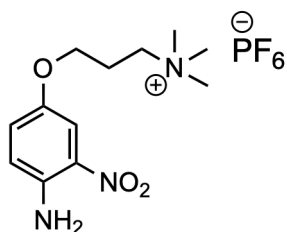
Cellular imaging – confocal microscopy: HeLa cells were seeded at 20000 cells per well density in 8-well chambered microscope slides (Lab-Tek Nunc II, part number 155411, ThermoFisher Scientific, Germany). After overnight seeding, the platinum(II) complexes at concentrations of 10 μ M were added to cellular media and imaged at different time delays. The imaging experiments were carried out using a confocal laser-scanning microscope Leica TCS SP5 II (Leica Microsystem Ltd, Germany) at room temperature, with a 100x (N.A. 1.4) HCX PL APO CS oil immersion objective lens with a correction collar (11506279, Leica Microsystem Ltd, Germany). The microscope is equipped with prism/slit system, which is flexible and provides the selection of required spectral range for each dye. Samples were excited at 514 nm with the internal continuous-wave Argon laser and the emission images were recorded in the range 540-680 nm with adjustments corresponding to the emission spectrum for each dye. The colocalization studies were performed with MitoTracker Deep Red (MTDR) or LysoTracker Deep Red (LTDR) from ThermoFisher (catalogue numbers M22426 and L12492, respectively). Cells were washed twice with PBS before and after the 5 min-staining with 50 nM of MTDR or LTDR dissolved in FluoroBrite DMEM media from ThermoFisher (A1896701). Cells in FluoroBrite DMEM were imaged in two spectral regions corresponding to the emission from platinum(II) complexes (excitation at 514 nm and detection in the range of 540-630 nm) and MTDR/LTDR (excitation at 633 nm and detection in the range of 680-740 nm). All images were processed in the Leica Application Suite X (LAS X) software package.

DFT and TDDFT: DFT and TDDFT were performed on Gaussian 16 software [392]. Complexes were generated using the STK toolkit described in chapter 5, and underwent semi-empirical optimisation before being converted to DFT input files. Complex geometries were optimised to find the lowest energy conformation using DFT B3LYP with basis sets LANL2DZ for platinum atoms, and 6–31G(d) for all other atoms and compared with crystallographic literature values. Vibrational frequency calculations were not performed, therefore geometry was assumed to reach the minima upon satisfaction of all Gaussian geometry convergence criteria. No symmetry constraints were applied to the calculations. All calculations were performed in DMSO solvent in the cPCM model. TDDFT was performed on the optimized ground state geometries in DMSO to find the vertical excitation energies to both singlet and triplet states.

7.2 Synthetic methods

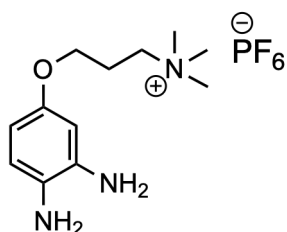
7.2.1 Intermediates

Synthesis of compound 3:



4-amino-3-nitrophenol (2.00 g, 13.0 mmol), (3-bromopropyl)trimethylammonium bromide (3.38 g, 13.0 mmol), and potassium carbonate (1.80 g, 13.0 mmol) were dissolved in DMF (30 mL) and stirred at 40 °C for 19 hours. The reaction mixture was then concentrated under reduced pressure to form a black oil. An aqueous solution of NaPF₆ (0.03 M, 200 mL) was added, resulting in the formation of a brown precipitate. The precipitate was collected by vacuum filtration and dried under high vacuum for 24 hours, to yield the product as a brown powder (2.39 g, 5.99 mmol, 46 %); ¹H NMR (400 MHz, DMSO-d₆) δ 7.43 (1H, d, *J* 2.7, Ar-OCCHCNO₂), 7.30 (2H, s, Ar-NH₂), 7.18 (1H, d, *J* 9, Ar-NH₂CCHCHCO), 7.02 (1H, d, *J* 9, Ar-NH₂CCHCHCO), 4.03 (2H, t, *J* 5.9, NMe₃CH₂), 3.09 (9H, s, N(CH₃)₂), 2.74 (2H, m, O-CH₂), 2.16 (2H, m, CH₂CH₂CH₂)

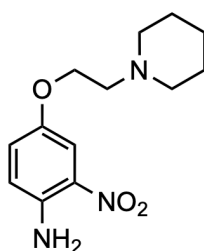
Synthesis of PPTMA:



Compound 3 (1.00 g, 2.50 mmol) was suspended in acetonitrile (50 mL) and palladium on carbon (20 wt%, 200 mg, 0.5 mmol) added to form a black suspension. The reaction vessel was first flushed with argon, and then with hydrogen. The reaction mixture was stirred under a hydrogen environment for 24 hours upon which time the solution changed from orange to colourless, with a black solid in suspension. The reaction

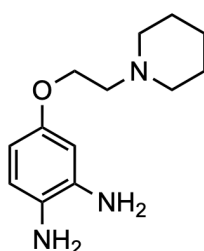
mixture was filtered through celite and concentrated under reduced pressure to give an off-white solid (796 mg, 2.16 mmol, 86 %). The reaction was determined to have gone to 100 % conversion by ^1H NMR spectroscopy and reacted without further purification. ^1H NMR (400 MHz, DMSO- d_6) δ 6.49 (1H, d, J 8.4, Ar-OCCHCHCNH $_2$), 6.20 (1H, d, J 2.6, Ar-OCCHCNH $_2$), 6.04 (1H, dd, J 8.3, 2.8, Ar-NH $_2$ CCHCHCO), 3.87 (2H, t, J 5.9, NMe $_3$ CH $_2$), 3.49 – 3.40 (2H, m, O-CH $_2$), 3.09 (9H, s, N(CH $_3$) $_3$), 2.14-2.07 (2H, m, CH $_2$)

Synthesis of compound 5:



4-amino-3-nitrophenol (500 mg, 3.25 mmol), 1-(2-chloroethyl)piperidine hydrochloride (718 mg, 3.90 mmol), and potassium carbonate (630 mg, 4.55 mmol) were dissolved in AcN (10 mL) and stirred at 80 °C for 21 hours. The reaction mixture was then concentrated under reduced pressure to form a dark purple oil. The crude mixture was purified by column chromatography on silica gel using a primary eluent of 10 % MeOH in DCM, followed by 10 % MeOH in DCM with 2 % NEt $_3$. The resulting pure product was a brown oil (269 mg, 1.01 mmol, 31 %); ^1H NMR (400 MHz, DMSO- d_6) δ 7.40 (1H, d, J 2.9, Ar-OCCHCNO $_2$), 7.27 (2H, s, Ar-NH $_2$), 7.18 (1H, dd, J 9.2, 3.0, Ar-H), 6.99 (1H, d, J 9.2, Ar-H), 4.02 (2H, t, J 5.8, Et-OCH $_2$), 2.64 (2H, m, Et-NCH $_2$), 2.43 (4H, m, pip-N(CH $_2$) $_2$), 1.50 (4H, q, J 5.6, pip-CH $_2$ (CH $_2$) $_2$), 1.39 (2H, q, J 5.8, pip-CH $_2$).

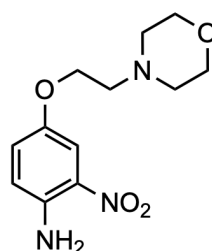
Synthesis of PEP:



Compound 5 (189 mg, 0.713 mmol) was suspended in acetonitrile (10 mL). To this

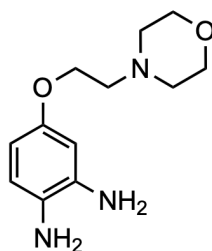
solution, palladium on carbon (20 wt%, 37.8 mg, 0.355 mmol) was added to form a black suspension. The reaction vessel was first flushed with argon, and then with hydrogen. The reaction mixture was stirred under a hydrogen environment for 24 hours and the reaction mixture changed from orange to colourless with a black suspension. The reaction mixture was filtered through celite and concentrated under reduced pressure to give an off-white oil (121.5 mg, 0.516 mmol, 72 %). The reaction was determined to have gone to 100 % conversion by NMR and reacted without further purification. ^1H NMR (400 MHz, DMSO) δ 6.39 (1H, d, J 8.3, Ar-H), 6.16 (1H, d, J 2.8, Ar-OCCHCNH₂), 5.97 (1H, dd, J 8.3, 2.8, Ar-H), 3.85 (2H, t, J 6.1, Et-OCH₂), 2.57 (2H, t, J 6.0, Et-NCH₂), 2.41 (4H, q, J 6.0, 5.5, pip-N(CH₂)₂), 1.49 (4H, p, J 5.5, pip-CH₂(CH₂)₂), 1.37 (2H, m, pip-CH₂).

Synthesis of compound 7:



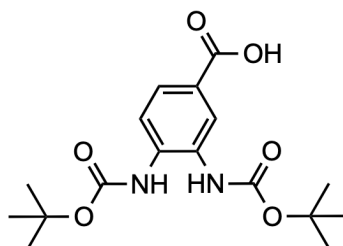
4-amino-3-nitrophenol (500 mg, 3.25 mmol), 4-(2-bromoethyl)morpholine hydrobromide (1.07 g, 3.90 mmol), and potassium carbonate (630 mg, 4.55 mmol) were dissolved in acetonitrile (10 mL) and stirred at 80 °C for 21 hours. The reaction mixture was then concentrated under reduced pressure to form a dark purple oil. The oil was then dissolved in a CH₂Cl₂ and 2 % EtOH solvent mixture (50 mL). The solution was then washed with a Na₂CO₃ aqueous solution (100 mL, 2 M), H₂O (100 mL), and brine (100 mL). The organic phase was dried with MgSO₄ and concentrated under reduced pressure to give a brown oil (220 mg, 0.82 mmol, 25 %); ^1H NMR (400 MHz, DMSO-*d*₆) δ 7.40 (1H, d, J 3.0, Ar-OCCHCNO₂), 7.27 (2H, s, Ar-NH₂), 7.16 (1H, dd, J 9.2, 3.0, Ar-H), 6.99 (1H, d, J 9.3, Ar-H), 4.02 (2H, t, J 5.7, Et-OCH₂), 3.57 (4H, t, J 4.6, morph-O(CH₂)₂), 2.65 (2H, t, J 5.6, Et-NCH₂), 2.45 (4H, t, J 4.7, morph-N(CH₂)₂).

Synthesis of PEM:



Compound 7 (755 mg, 2.82 mmol) was suspended in acetonitrile (50 mL) and palladium on carbon (20 wt%, 151 mg, 1.42 mmol) added to form a black suspension. The reaction vessel was first flushed with argon, and then with hydrogen. The reaction mixture was stirred under a hydrogen environment for 24 hours upon which time the solution changed from orange to colourless with a black solid in suspension. The reaction mixture was filtered through celite and the filtrate concentrated under reduced pressure to give an off-white oil (641.1 mg, 2.67 mmol, 96 %). The reaction was determined to have gone to 100 % conversion by ^1H NMR spectroscopy and reacted without further purification. ^1H NMR (400 MHz, DMSO) δ 6.40 (1H, d, J 8.3, Ar-H), 6.16 (1H, d, J 2.8, Ar-OCCHCNH₂), 5.98 (1H, dd, J 8.3, 2.7, Ar-H), 4.48 (2H, s, NH₂), 4.01 (2H, s, NH₂), 3.88 (2H, t, J 5.9, Et-OCH₂), 3.57 (4H, m, morph-O(CH₂)₂), 2.60 (2H, t, J 5.8, Et-NCH₂), 2.44 (4H, m, morph-N(CH₂)₂).

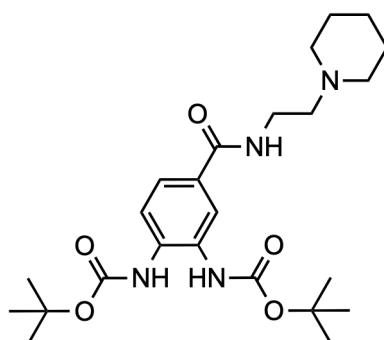
Synthesis of compound 12:



3,4-Diaminobenzoic acid (1.00 g, 6.58 mmol), di-tert-butyl-dicarbonate (8.6 g, 39.4 mmol) and guanidine hydrochloride (208 mg, 2.18 mmol) were dissolved in 1:1 ethanol:THF (60 mL) and heated at 45 °C for 18 hours. The solvent was removed under reduced pressure and the resulting product mixture was suspended in dichloromethane (60 mL). The solid was then collected by filtration and washed with dichloromethane (3 x 20 mL) and distilled water (3 x 20 mL). The solid was then dissolved in ethyl acetate (60 mL) and concentrated under reduced pressure to give a beige powder (1.39 g, 4.27

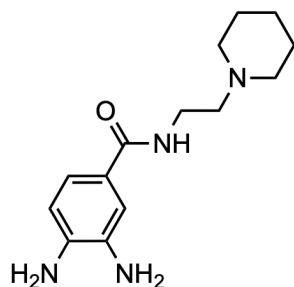
mmol, 60 %); ^1H NMR (400 MHz, DMSO-d_6) δ 12.79 (1H, s, COOH), 8.76 (1H, s, Ar-H NHCHCCOOH), 8.71 (1H, s, BocNH), 8.09 (1H, s, BocNH), 7.72 (1H, d, J 8.6 Hz, Ar-H (BocNHCHCHCOOH)), 7.64 (1H, dd, J 8.5, 2.0 Hz, Ar-H (BocNHCHCHCOOH)), 1.49 (18H, m, 2 x $\text{C}(\text{CH}_3)_3$).

Synthesis of BOC-AEP:



Compound 9 (1.00 g, 2.84 mmol), hydroxybenzotriazole· H_2O (576 mg, 4.26 mmol), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrogen chloride (817 mg, 4.26 mmol), and triethylamine (863 mg, 8.52 mmol) were dissolved in DMF (40 mL) and stirred for 30 minutes at 40 °C. 1-(2-aminoethyl)piperidine (640 mg, 4.97 mmol) was then added, and the reaction mixture was stirred at 40 °C for a further 18 hours. The reaction mixture was then quenched with water (80 mL) and extracted with ethyl acetate (3 x 20 mL). The extract was dried with MgSO_4 and concentrated under reduced pressure to afford a brown-orange crude oil (4.25 g). The crude oil was then purified by column chromatography with an eluent of ethyl acetate with 1 % triethylamine, with an increasing methanol gradient (5-20 %). The fractions containing product were then combined and concentrated under reduced pressure to yield a white solid (436.7 mg, 0.944 mmol, 33 %); ^1H NMR (400 MHz, DMSO-d_6) δ 8.65 (2H, s, 2 x NHBoc), 8.28 (1H, t, J 5.7, Ar-NHCHCC(O)NH), 7.95 (1H, m, $\text{C}(=\text{O})\text{NHCH}_2$), 7.63 (1H, d, J 8.5, Ar-CCHCHC(=O)), 7.54 (1H, dd, J 8.5, 2.1, Ar-CCHCHC(=O)), 2.40 (8H, m, $\text{C}(\text{O})\text{NHCH}_2\text{CH}_2\text{N}(\text{CH}_2\text{CH}_2)_2\text{CH}_2$), 1.48 (22H, m, 6 x BocCH_3 + Pip- $\text{N}(\text{CH}_2\text{CH}_2)_2\text{CH}_2$), 1.37 (2H, m, , Pip- $\text{N}(\text{CH}_2\text{CH}_2)_2\text{CH}_2$).

Synthesis of AEP:



Compound BOC-AEP (426 mg, 0.922 mmol) was dissolved in dichloromethane (30 mL) and trifluoroacetic acid (1.8 mL, 23.0 mmol) was added dropwise. The reaction mixture was then stirred under nitrogen at room temperature for 18 hours. Following which diethyl ether (200 mL) was added to the reaction mixture producing a pale-yellow precipitate. The precipitate was collected by vacuum filtration, washed with diethyl ether (4 x 20 mL), ran through a K₂CO₃ plug and dried under vacuum to afford a yellow solid (172 mg, 0.236 mmol, 25 %); ¹H NMR (400 MHz, DMSO-d₆) δ 7.82 (1H, t, *J* 5.7, C(=O)NHCH₂), 7.03 (1H, d, *J* 2.0, Ar-NH₃CHCC(O)NH), 6.94 (1H, dd, *J* 8.1, 2.0, Ar-CCHCHC(=O)), 6.47 (1H, d, *J* 8.1, Ar-CCHCHC(=O)), 4.95 (2H, s, NH₂), 4.56 (2H, s, NH₂), 3.32 (4H, m, C(O)NHCH₂CH₂-Pip), 2.47 (4H, m, Pip-N(CH₂CH₂)₂CH₂), 1.53 (4H, m, Pip-N(CH₂CH₂)₂CH₂), 1.39 (2H, m, Pip-N(CH₂CH₂)₂CH₂).

7.2.2 Synthesis of platinum(II) salphen complexes

All platinum complexes were synthesised following the same general synthetic protocol described below. Quantities of reagents, reaction times and yields are summarised in table S1. This is followed by characterisation data for each of the platinum(II) complexes prepared.

Synthetic protocol: 1 equivalent of the corresponding diamine and 2 equivalents of the corresponding aldehyde were dissolved in DMSO. 1.1 equivalents of Zn(OAc)₂ were added to the solution, and the reaction mixture was heated at 100 °C for 24 hours under nitrogen. Following this, 1.1 equivalents of PtCl₂ were added and the reaction mixture was heated at 100 °C to perform the transmetalation reaction; depending on the specific reagents, the reaction took different times. Once complete, as determined by ¹H NMR spectroscopy, the product was precipitated out by the addition of a NaPF₆ aqueous solution (0.03 M, 100 mL) and the resulting solid was collected by vacuum filtration.

The solid was redissolved in acetone and precipitated out with diethyl ether, collected by vacuum filtration, and dried under high vacuum for 24 hours.

Compounds for which ^1H NMR spectroscopy detected impurities were redissolved in CH_2Cl_2 and precipitated out with diethyl ether, collected by vacuum filtration, and dried under high vacuum for 24 hours.

Complex	Aldehyde identity	Diamine (mg, mmol)	Aldehyde (mg, mmol)	TM time (hours)	Yield (mg, %)
Diamine: PPTMA					
Pt_PPTMA_Ph	Salicylaldehyde	696, 1.88	0.458, 3.76	72	1.31, 92
Pt_PPTMA_Naph	2-hydroxy-1-naphthaldehyde	50.0, 0.135	0.047, 0.271	72	46.5, 39
Pt_PPTMA_5OMe	2-hydroxy-5-methoxybenzaldehyde	50.0, 0.135	0.042, 0.271	120	32.7, 29
Pt_PPTMA_4OMe	2-hydroxy-4-methoxybenzaldehyde	50.0, 0.135	0.042, 0.271	72	29.5, 26
Pt_PPTMA_5F	5-fluoro-2-hydroxybenzaldehyde	50.0, 0.135	0.038, 0.271	120	38.6, 35
Pt_PPTMA_4F	4-fluoro-2-hydroxybenzaldehyde	50.0, 0.135	0.038, 0.271	72	51.8, 48
Pt_PPTMA_5Cl	5-chloro-2-hydroxybenzaldehyde	50.0, 0.135	42.4, 0.271	72	21.7, 19
Pt_PPTMA_4Cl	4-chloro-2-hydroxybenzaldehyde	50.0, 0.135	42.4, 0.271	96	76.4, 68
Pt_PPTMA_5Br	5-bromo-2-hydroxybenzaldehyde	50.0, 0.135	54.4, 0.271	72	41.5, 33
Pt_PPTMA_4Br	4-bromo-2-hydroxybenzaldehyde	50.0, 0.135	54.4, 0.271	96	37.6, 30
Pt_PPTMA_4CF ₃	4-trifluoromethyl-2-hydroxybenzaldehyde	50.0, 0.135	51.5, 0.271	72	4.3, 4.0
Pt_PPTMA_5OCF ₃	5-trifluoromethoxy-2-hydroxybenzaldehyde	50.0, 0.135	55.8, 0.271	96	40.6, 32
Pt_PPTMA_3F	3-fluoro-2-hydroxybenzaldehyde	50.0, 0.135	37.9, 0.271	96	10.9, 10
Pt_PPTMA_6F	6-fluoro-2-hydroxybenzaldehyde	50.0, 0.135	37.9, 0.271	96	32.5, 30
Pt_PPTMA_4NEt ₂	4-(diethylamino)-2-hydroxybenzaldehyde	21.0, 0.057	22.0, 0.114	72	20.2, 16

Table 7.2. Reaction conditions for the synthesis of the **PPTMA** solubilised platinum(II) salphen complexes.

Complex	Aldehyde identity	Diamine (mg, mmol)	Aldehyde (mg, mmol)	TM time (hours)	Yield (mg, %)
Diamine: PEM					
Pt_PEM_Ph	Salicylaldehyde	103, 0.436	107, 0.872	72	222, 65
Pt_PEM_Naph	2-hydroxy-1-naphthaldehyde	50.0, 0.211	72.6, 0.422	120	117, 98
Pt_PEM_4OMe	2-hydroxy-4-methoxybenzaldehyde	50.0, 0.211	64.2, 0.422	120	112, 98
Pt_PEM_5OCF ₃	5-trifluoromethoxy-2-hydroxybenzaldehyde	50.0, 0.211	87.0, 0.422	120	103, 80
Pt_PEM_5Br	5-bromo-2-hydroxybenzaldehyde	50.0, 0.211	84.8, 0.422	120	63.2, 50

Table 7.3. Reaction conditions for the synthesis of the **PEM** solubilised platinum(II) salphen complexes.

The zinc(II) intermediate complexes with the salphen ligands containing the amide linker in the backbone (i.e. from diamine **AEP**), were isolated following the general synthetic protocol described above. The zinc(II) intermediates were isolated by filtration of the reaction mixture followed by precipitation of the complex by the addition of a NaPF₆ aqueous solution (0.03 M, 50 mL).

The zinc(II) intermediate was then redissolved in 2 mL of DMSO and reacted further. Then 1.1 equivalents of PtCl₂ was added. The reaction mixture was heated at 100 °C for 120 hours under nitrogen. The product was precipitated out by the addition of a NaPF₆ aqueous solution (0.03 M, 50 mL) and the resulting solid was collected by vacuum filtration and dried under high vacuum for 48 hours.

Complex	Aldehyde identity	Diamine (mg, mmol)	Aldehyde (mg, mmol)	TM time (hours)	Yield (mg, %)
Diamine: 4-Bromo-1,2-diaminobenzene					
Pt_PhBr_Ph	Salicylaldehyde	25.0, 0.134	32.6, 0.267	72	41.0, 52

Table 7.4. Reaction conditions for the synthesis of the **PhBr DAB** platinum(II) salphen complexes.

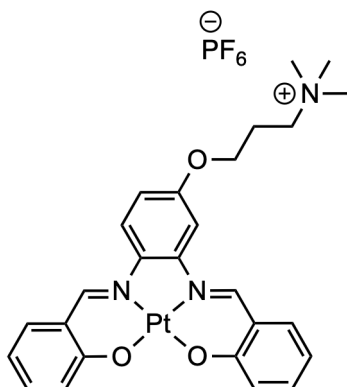
Complex	Diamine (mg, mmol)	Aldehyde (mg, mmol)	Zn(OAc) ₂ (mg, mmol)	Yield (mg, mmol, %)
Diamine: AEP				
Zn_AEP_4OMe	25.0, 0.0666	20.3, 0.133	10.2, 0.733	16.9, 34

Table 7.5. Reaction conditions for the synthesis of the **AEP** solubilised zinc(II) salphen complexes.

Complex	Diamine (mg, mmol)	Aldehyde (mg, mmol)	Zn(OAc) ₂ (mg, mmol)	Yield (mg, mmol, %)
Diamine: AEP				
Zn_AEP_4OMe	25.0, 0.0666	20.3, 0.133	10.2, 0.733	16.9, 34

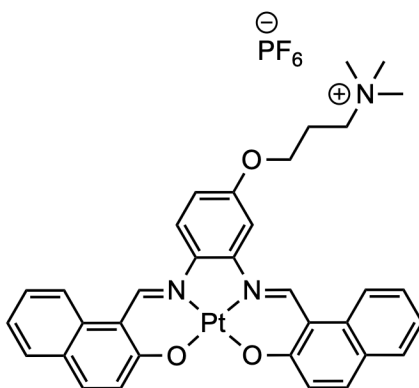
Table 7.6. Reaction conditions for the synthesis of the **AEP** solubilised platinum(II) salphen complexes.

Pt_PPTMA_Ph [hexafluorophosphate]:



Orange solid (1.31 g, 1.70 mmol, 92 %). ^1H NMR (400 MHz, DMSO-d_6) δ 9.51 (1H, s, Ar-NCH), 9.38 (1H, s, Ar-NCH), 8.38 (1H, d, J 9.1, Ar-OCCHCH), 7.98 (1H, s, Ar-OCCHCN), 7.85 (2H, m, 2 x Ar-H), 7.58 (2H, m, 2 x Ar-H), 7.13 (3H, m, 3 x Ar-H), 6.80 (2H, m, 2 x Ar-H), 4.26 (2H, m, NMe_3CH_2), 3.53 (2H, m, OCH_2CH_2), 3.14 (9H, s, $\text{N}(\text{CH}_3)_3$), 2.26 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO-d_6) δ 165, 164, 159, 152, 150, 146, 139, 136, 136, 136, 123, 123, 122, 121, 117, 117, 117, 115, 102, 66.1, 63.4, 52.8 (t), 22.9; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{28}\text{N}_3\text{O}_3]^{195}\text{Pt}^+$ requires 625.1779, found 625.1782; LCMS TOF MS ES $^+$ 4.23 min, 100 %, m/z 625 $[\text{M}]^+$.

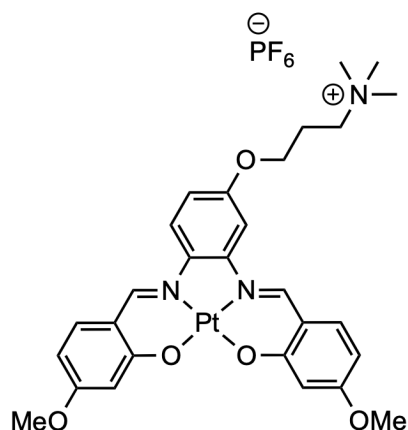
Pt_PPTMA_Naph [hexafluorophosphate]:



Brown-orange solid (35.2 mg, 0.0404 mmol, 32 %) ^1H NMR (400 MHz, DMSO-d_6) δ 9.91 (1H, s, Ar-NCH), 9.84 (1H, s, Ar-NCH), 8.65 (2H, m, 2 x Ar-H), 8.04 (3H, m, 3 x Ar-H), 7.90 – 7.84 (2H, m, 2 x Ar-H), 7.67 – 7.59 (2H, m, 2 x Ar-H), 7.41 (2H, m, 2 x Ar-H), 7.33 (2H, m, 2 x Ar-H) 7.05 (2H, m, 2 x Ar-H), 4.30 (2H, t, J 5.7, NMe_3CH_2), 3.61 – 3.53 (2H, m, OCH_2CH_2), 3.17 (9H, s, $\text{N}(\text{CH}_3)_3$), 2.28 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C

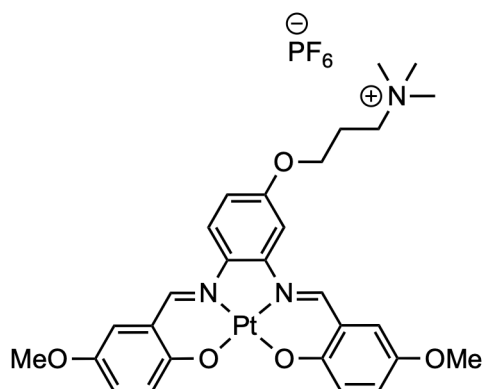
NMR (500 MHz, DMSO- d_6) δ 166, 165, 158, 147, 145, 143, 139, 137, 136, 134, 134, 129, 129, 128, 128, 128, 128, 125, 125, 124, 124, 122, 1220, 117, 114, 113, 102, 66.1, 63.6, 52.8, 23.0; HRMS m/z (ESI) $^+$ calcd. for $[C_{34}H_{32}N_3O_3^{195}Pt]^+$ requires 725.2092, found 725.2108; LCMS TOF MS ES $^+$: 5.28 min, 100 %, m/z 725 [M] $^+$.

Pt_PPTMA_4OMe [hexafluorophosphate]:



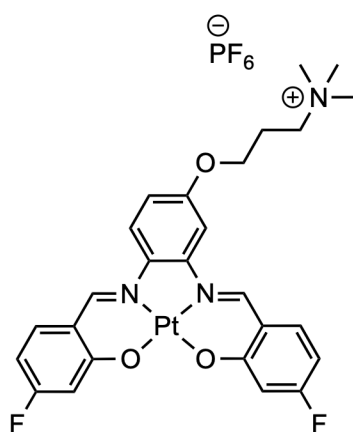
Orange-yellow powder (65.6 mg, 0.079 mmol, 59 %). 1H NMR (400 MHz, DMSO- d_6) δ 9.24 (1H, s, Ar-NCH), 9.13 (1H, s, Ar-NCH), 8.25 (1H, d, J 9.2, Ar-OCCHCH), 7.85 (1H, d, J 2.5, Ar-H), 7.71-7.66 (2H, m, 2 x Ar-H), 7.04 (1H, dd, J 9.2, 2.5, NCCHCHCO), 6.59 (2H, dd, J 6.7, 2.4, 2 x Ar-H), 6.49-6.43 (2H, m, 2 x Ar-H), 4.21 (2H, t, J 5.8, NMe $_3$ CH $_2$), 3.57 – 3.47 (2H, m, OCH $_2$ CH $_2$), 3.13 (9H, s, N(CH $_3$) $_3$), 2.55 (6H, s, 2 x O(CH $_3$)), 2.25 (2H, m, OCH $_2$ CH $_2$ CH $_2$ NMe $_3$); ^{13}C NMR (500 MHz, DMSO- d_6) δ 167, 166, 166, 166, 158, 150, 149, 146, 139, 137, 137, 117, 117, 117, 114, 108, 108, 103, 102, 102, 66.0, 63.4, 55.9, 55.8, 52.8, 23.0; HRMS m/z (ESI) $^+$ calcd. for $[C_{28}H_{32}N_3O_5^{195}Pt]^+$ requires 685.1990 found 685.2004; LCMS TOF MS ES $^+$: 4.43 min, 99 %, m/z 685 [M] $^+$.

Pt_PPTMA_5OMe [hexafluorophosphate]:



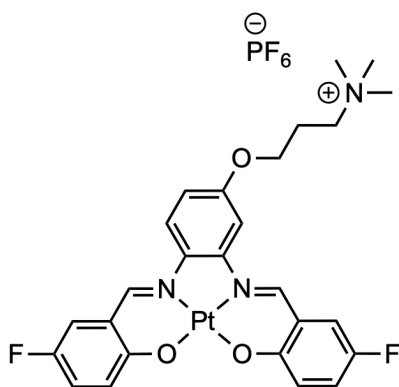
Dark purple powder (103.9 mg, 0.125 mmol, 93 %). ^1H NMR (400 MHz, DMSO- d_6) δ 9.42 (1H, s, Ar-NCH), 9.31 (1H, s, Ar-NCH), 8.30 (1H, d, J 9.3, Ar-OCCHCH), 7.90 (1H, d, J 2.5, Ar-H), 7.33 – 7.23 (4H, m, 4 x Ar-H), 7.12 – 7.01 (3H, m, 3 x Ar-H), 4.23 (2H, t, J 5.8, NMe_3CH_2), 3.58 – 3.48 (2H, m, OCH_2CH_2), 3.15 (9H, s, $\text{N}(\text{CH}_3)_3$), 2.55 (6H, s, 2 x $\text{O}(\text{CH}_3)$), 2.26 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO- d_6) δ 161, 160, 158, 151, 150, 150, 150, 146, 139, 127, 126, 123, 122, 121, 121, 117, 115, 114, 114, 102, 66.0, 63.4, 55.9, 55.8, 52.8, 23.0; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{28}\text{H}_{32}\text{N}_3\text{O}_5^{195}\text{Pt}]^+$ requires 685.1990 found 685.2004; LCMS TOF MS ES $^+$: 4.18 min, 99 %, m/z 686 $[\text{M}]^+$.

Pt_PPTMA_4F [hexafluorophosphate]:



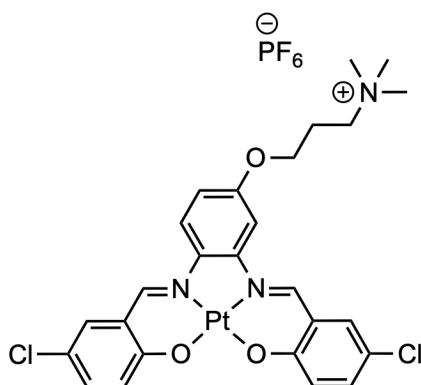
Light brown-yellow powder (108 mg, 0.134 mmol, 99 %). ^1H NMR (400 MHz, DMSO- d_6) δ 9.35 (1H, s, Ar-NCH), 9.22 (1H, s, Ar-NCH), 8.23 (1H, d, J 9.4, Ar-OCCHCH), 7.88-7.80 (3H, m, 3 x Ar-H), 7.03 (1H, dd, J 9.2, 2.4, Ar-H), 6.81 (2H, dt, J 12.1, 2.5, 2 x Ar-H), 6.70 (2H, m, 2 x Ar-H), 4.18 (2H, t, J 5.9, NMe_3CH_2), 3.56 – 3.49 (2H, m, OCH_2CH_2), 3.15 (9H, s, $\text{N}(\text{CH}_3)_3$), 2.29 – 2.20 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO- d_6) δ 168 (d, J_1 145), 166 (d, J_2 60), 166 (d, J_1 140), 166 (d, J_2 55), 159, 151, 150, 146, 139, 138 (d, J_2 55), 138 (d, J_2 50), 120, 120, 117, 115, 106, 106 (d, J_3 20), 106, 106, 102, 66, 63, 53, 23; ^{19}F NMR (400 MHz, DMSO- d_6) -69.1, -71.0; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{26}\text{N}_3\text{O}_3\text{F}_2^{195}\text{Pt}]^+$ requires 661.1590 found 661.1609; LCMS TOF MS ES $^+$: 4.60 min, 100 %, m/z 661 $[\text{M}]^+$.

Pt_PPTMA_5F [hexafluorophosphate]:



Purple powder (95.7 mg, 0.119 mmol, 88 %). ^1H NMR (400 MHz, DMSO- d_6) δ 9.30 (1H, s, Ar-NCH), 9.16 (1H, s, Ar-NCH), 8.15 (1H, d, J 9.4, Ar-OCCHCH), 7.74 (1H, d, J 2.5, Ar-H), 7.52 – 7.38 (4H, m, 4 x Ar-H), 7.04 (2H, m, 2 x Ar-H), 6.96 (1H, dd, J 9.2, 2.4, Ar-H), 4.11 (2H, t, J 6.0, NMe_3CH_2), 3.52 – 3.47 (2H, m, OCH_2CH_2), 3.15 (9H, s, $\text{N}(\text{CH}_3)_3$), 2.26 – 2.16 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO- d_6) δ 162, 161, 158.63, 154 (d, J 3 20), 152 (d, J 3 15), 151 (d, J 12), 149 (d, J 11), 146.12, 138.80, 124 (d, J 1 100), 124 (d, J 1 95), 123 (d, J 2 35), 122 (d, J 2 35), 121 (d, J 2 35), 121 (d, J 2 35), 118, 118, 117.20, 115.02, 101.84, 65.85, 63.36, 52.83, 22.93; ^{19}F NMR (400 MHz, DMSO- d_6) -69.12, -71.0; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{26}\text{N}_3\text{O}_3\text{F}_2^{195}\text{Pt}]^+$ requires 661.1590 found 661.1614; LCMS TOF MS ES $^+$: 4.32 min, 99 %, m/z 661 $[\text{M}]^+$.

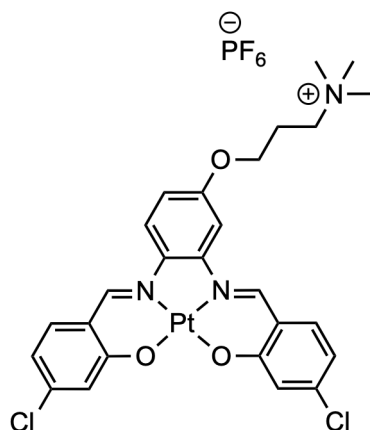
Pt_PPTMA_5Cl [hexafluorophosphate]:



Red solid (88.4 mg, 0.105 mmol, 78 %); ^1H NMR (400 MHz, DMSO- d_6) δ 9.34 (1H, s, Ar-NCH), 9.17 (1H, s, Ar-NCH), 8.15 (1H, d, J 9.3, Ar-H), 7.80 – 7.72 (3H, m, 3 x Ar-H), 7.51 (2H, m, 2 x Ar-H), 7.07 – 6.94 (3H, m, 3 x Ar-H), 4.10 (2H, t, J 6.0, NMe_3CH_2), 3.52-3.49 (3H, m, OCH_2CH_2), 3.15 (s, 9H, $\text{N}(\text{CH}_3)_3$), 2.23 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$);

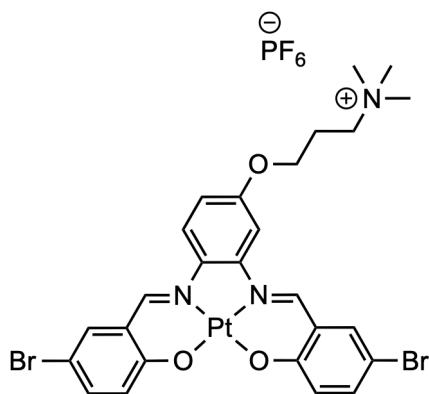
^{13}C NMR (500 MHz, DMSO-d_6) δ 163.21, 162.51, 158.55, 150.54, 149.03, 146.03, 138.69, 134.80, 134.38, 133.63, 133.51, 123, 123.16, 122.92, 122.86, 119.43, 119.22, 117.09, 115.18, 101.38, 65.64, 63.34, 52.85, 22.96; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{26}\text{Cl}_2\text{N}_3\text{O}_3^{195}\text{Pt}]^+$ requires 694.1001, found 694.1019; LCMS TOF MS ES $^+$: 2.19 min, 100 %, m/z 694 $[\text{M}]^+$.

Pt_PPTMA_4Cl [hexafluorophosphate]:



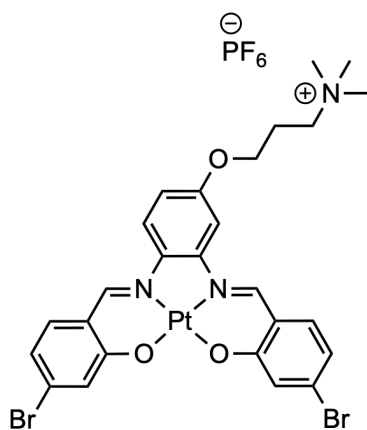
Dark red solid (76.4 mg, 0.091 mmol, 67 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.41 (1H, s, Ar-NCH), 9.27 (1H, s, Ar-NCH), 8.25 (1H, d, J 9.3, Ar-H), 7.86 – 7.76 (3H, m, 3 x Ar-H), 7.07 (3H, m, 3 x Ar-H), 6.84 (2H, td, J 8.2, 2.0, 2 x Ar-H), 4.19 (2H, t, J 5.9, NMe_3CH_2), 3.55-3.50 (3H, m, OCH_2CH_2), 3.14 (s, 9H, $\text{N}(\text{CH}_3)_3$), 2.25 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO-d_6) δ 163.98, 163.34, 158.13, 150.57, 149.01, 145.60, 139.67, 139.07, 138.29, 136.67, 136.45, 121.20, 120.97, 120.08, 119.80, 117, 116.65, 116.43, 114.42, 101.20, 65.37, 62.96, 52.37, 22.54; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{26}\text{Cl}_2\text{N}_3\text{O}_3^{195}\text{Pt}]^+$ requires 694.1001, found 694.1003; LCMS TOF MS ES $^+$: 2.33 min, 100 %, m/z 694 $[\text{M}]^+$.

Pt_PPTMA_5Br [hexafluorophosphate]:



Red solid (112 mg, 0.121 mmol, 90 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.31 (1H, s, Ar-NCH), 9.14 (1H, s, Ar-NCH), 8.12 (1H, d, J 9.3, Ar-H), 7.89 (2H, m, 2 x Ar-H), 7.73 (1H, m, Ar-H), 7.61-7.57 (2H, m, 2 x Ar-H), 6.97-6.92 (3H, m, 3 x Ar-H), 4.08 (2H, m, NMe_3CH_2), 3.50 (2H, m, OCH_2CH_2), 3.15 (9H, s, $\text{N}(\text{CH}_3)_3$), 2.22 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO-d_6) δ 163.45, 162.76, 158.52, 150.47, 148.94, 146.02, 138.68, 137.29, 136.90, 136.75, 136.63, 124.05, 123.76, 123.54, 123.31, 117.07, 115.24, 106.70, 106.47, 101.29, 65.62, 63.35, 52.87, 22.99; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{26}\text{Br}_2\text{N}_3\text{O}_3^{195}\text{Pt}]^+$ requires 782.9968, found 782.9938; LCMS TOF MS ES $^+$: 2.29 min, 100 %, m/z 783 $[\text{M}]^+$.

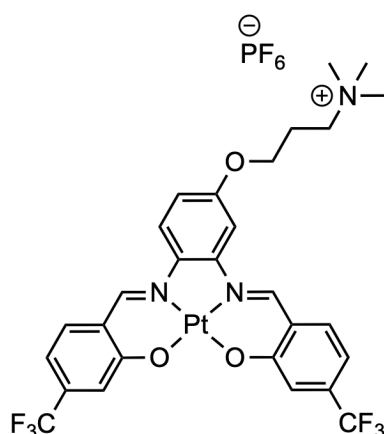
Pt_PPTMA_4Br [hexafluorophosphate]:



Red solid (116 mg, 0.125 mmol, 92 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.35 (1H, s, Ar-NCH), 9.19 (1H, s, Ar-NCH), 8.19 (1H, d, J 9.3, Ar-H), 7.79 (1H, d, J 2.5, Ar-H), 7.67 (2H, m, 2 x Ar-H), 7.17 (2H, t, J 2.0, 2 x Ar-H), 7.00 (1H, dd, J 9.2, 2.4, Ar-H), 6.94-6.91 (2H, m, 2 x Ar-H), 4.16 (2H, t, J 5.8, NMe_3CH_2), 3.57 - 3.48 (2H, m, OCH_2CH_2), 3.15

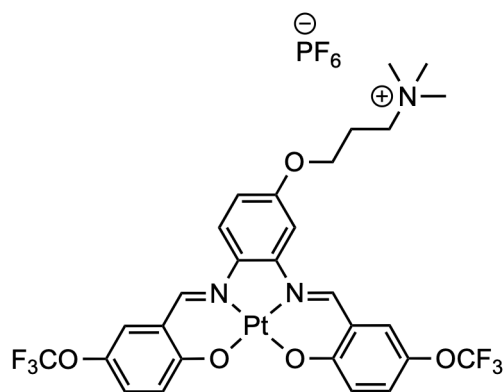
(s, 9H, N(CH₃)₃), 2.25 (2H, m, OCH₂CH₂CH₂NMe₃); ¹³C NMR (500 MHz, DMSO-d₆) δ 164.27, 163.66, 158.60, 151.15, 149.56, 146.09, 138.78, 137.01, 136.83, 129.54, 128.86, 123.66, 123.37, 121.86, 121.62, 119.80, 119.57, 117.09, 114.92, 101.63, 65.83, 63.44, 52.85 (TMA), 23.03; HRMS m/z (ESI)⁺ calcd. for [C₂₆H₂₆Br₂N₃O₃196Pt]⁺ requires 783.9970, found 784.0002; LCMS TOF MS ES⁺: 2.36 min, 100 %, m/z 783 [M]⁺.

Pt_PPTMA_4CF₃ [hexafluorophosphate]:



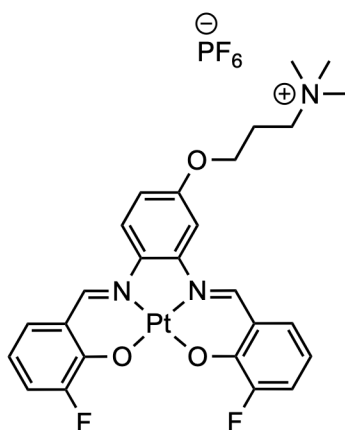
Red solid (60.1 mg, 0.066 mmol, 49 %); ¹H NMR (400 MHz, DMSO-d₆) δ 9.53 (1H, s, Ar-NCH), 9.36 (1H, s, Ar-NCH), 8.24 (1H, d, *J* 9.4, Ar-H), 7.97 (2H, dd, *J* 12.7, 8.4, 2 x Ar-H), 7.86 (1H, d, *J* 2.5, Ar-H), 7.23 (2H, d, *J* 6.9, 2 x Ar-H), 7.02 (3H, dt, *J* 6.3, 3.4, 3 x Ar-H), 4.17 (2H, t, *J* 6.0, NMe₃CH₂), 3.53 - 3.49 (2H, m, OCH₂CH₂), 3.14 (s, 9H, N(CH₃)₃), 2.24 (2H, m, OCH₂CH₂CH₂NMe₃); ¹³C NMR (500 MHz, DMSO-d₆) δ 163.5, 162.9, 158.8, 151.4, 149.7, 146.2, 138.8, 137.2, 136.9, 134.7-133.4 (2C, m, 2 x Ar-CCF₃), 125.8 (d, *J* 3 20), 124.9, 124.6, 123.1 (d, *J* 3 20), 118.2, 117.8, 117.4, 115.4, 111.5, 101.7, 65.8, 63.4, 52.8, 22.9; ¹⁹F NMR (376 MHz, DMSO-d₆) δ -62.30, -62.46; HRMS m/z (ESI)⁺ calcd. for [C₂₈H₂₆N₃O₃F₆¹⁹⁵Pt]⁺ requires 761.1526, found 761.1559; LCMS TOF MS ES⁺: 2.43 min, 100 %, m/z 761 [M]⁺.

Pt_PPTMA_5OCF₃ [hexafluorophosphate]:



Red solid (106 mg, 0.113 mmol, 83 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.50 (1H, s, Ar-NCH), 9.36 (1H, s, Ar-NCH), 8.25 (1H, d, J 9.3, Ar-H), 7.89 - 7.79 (3H, m, 3 x Ar-H), 7.58 - 7.52 (2H, m, 2 x Ar-H), 7.14 - 7.08 (3H, m, 3 x Ar-H), 4.20 (2H, t, J 5.8, NMe_3CH_2), 3.54 - 3.40 (2H, m, OCH_2CH_2), 3.14 (s, 9H, $\text{N}(\text{CH}_3)_3$), 2.26 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO-d_6) δ 163.3, 162.6, 158.8, 151.3, 149.7, 146.1, 138.74, 138.3, 138.3, 129.0, 128.5, 126.5, 123.0, 126.6, 122.6, 122.2, 121.8, 121.5, 119.6, 117.3, 115.4, 101.8, 65.8, 63.4, 52.8, 22.9; ^{19}F NMR (376 MHz, DMSO-d_6) δ -57.4, -57.4; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{28}\text{H}_{26}\text{N}_3\text{O}_5\text{F}_6]^{195}\text{Pt}^+$ requires 793.1425, found 793.1440; LCMS TOF MS ES $^+$: 2.44 min, 100 %, m/z 793 $[\text{M}]^+$.

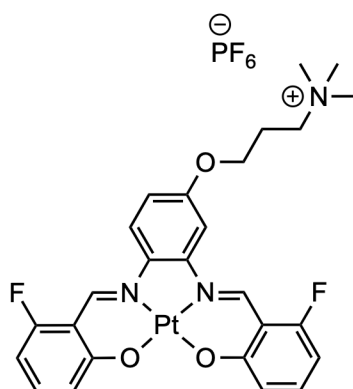
Pt_PPTMA_3F [hexafluorophosphate]:



Red-orange solid (114 mg, 0.141 mmol, 98 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.49 (1H, s, Ar-NCH), 9.35 (1H, s, Ar-NCH), 8.33 (1H, d, J 9.3, Ar-H), 7.93 (1H, m, Ar-H), 7.65 (2H, t, J 8.8, 2 x Ar-H), 7.50 (2H, q, J 10.8, 2 x Ar-H), 7.11-7.05 (1H, m, Ar-H), 6.73 (2H, m, 2 x Ar-H), 4.20 (2H, t, J 5.8, NMe_3CH_2), 3.52 (2H, m, OCH_2CH_2), 3.14 (s, 9H, $\text{N}(\text{CH}_3)_3$).

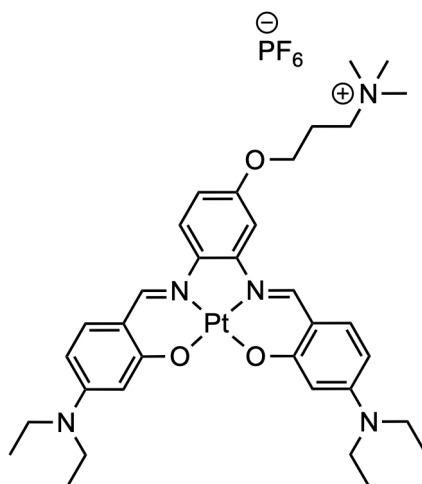
$\text{N}(\text{CH}_3)_3$), 2.26 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO-d_6) δ 158.80, 155.4 (d, J 3 28), 153.5 (d, J 2 50), 153.0 (d, J 3 28), 152.8 (d, J 2 50), 151.5, 150.0, 146.2, 138.8, 131.2, 131.0, 124.8, 124.5, 119.9, 119.7, 119.5, 119.4, 117.4, 115.3, 102.0, 65.9, 63.4, 52.8, 22.9; ^{19}F NMR (400 MHz, DMSO-d_6) -132.4, -132.6; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{26}\text{N}_3\text{O}_3\text{F}_2^{195}\text{Pt}]^+$ requires 661.1590, found 661.1581; LCMS TOF MS ES $^+$: 1.88 min, 100 %, m/z 661 $[\text{M}]^+$.

Pt_PPTMA_6F [hexafluorophosphate]:



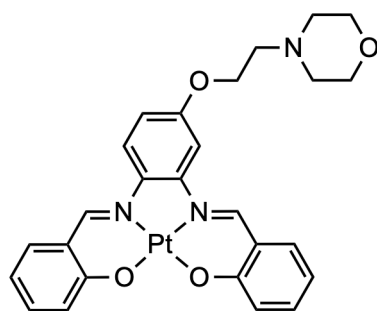
Dark red solid (112 mg, 139 μmol , 100 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.18 (1H, s, Ar-NCH), 9.04 (1H, s, Ar-NCH), 8.18 (1H, d, J 9.3, Ar-H), 7.63 (1H, d, J 2.5 Ar-H), 7.47 (3H, p, J 7.8, Ar-H), 6.93 – 6.90 (1H, dd, J 9.2, 2.4, Ar-H), 6.85 (2H, d, J 8.8, Ar-H), 6.51 (2H, m, 2 x Ar-H), 4.12 (2H, t, J 5.9, NMe_3CH_2), 3.53-3.48 (2H, m, OCH_2CH_2), 3.15 (s, 9H, $\text{N}(\text{CH}_3)_3$), 2.21 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$); ^{13}C NMR (500 MHz, DMSO-d_6) δ 165.2, 164.5, 163.9, 161.3, 158.6, 146.1, 143.3, 143.2, 141.6, 141.4, 138.8, 134.6 (d, J 2 52), 134.1 (d, J 2 52), 118.0, 117.7, 117.1, 114.9, 112.4 (t, J 32), 101.6 (m), 101.5 (m), 65.6, 63.4, 52.8, 22.9; ^{19}F NMR (400 MHz, DMSO-d_6) -113.4 -114.0; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{26}\text{N}_3\text{O}_3\text{F}_2^{195}\text{Pt}]^+$ requires 661.1590, found 661.1594; LCMS TOF MS ES $^+$: 1.97 min, 100 %, m/z 661 $[\text{M}]^+$.

Pt_PPTMA_4NEt₂ [hexafluorophosphate]:



Black solid (20.2, mmol, %); ^1H NMR (400 MHz, DMSO-d_6) δ 8.94 (1H, s, Ar-H), 8.84 (1H, s, Ar-H), 8.15 (1H, d, J 9.2, Ar-H), 7.77 (1H, s, Ar-H), 7.55 - 7.47 (2H, m, Ar-H), 6.95 (1H, d, J 8.8, Ar-H), 6.43 – 6.28 (3H, m, Ar-H), 6.17 (1H, m, Ar-H), 4.21 (2H, J 4.0, NMe_3CH_2), 3.53 (2H, m, OCH_2CH_2), 3.42 (8H, t, J 6.9, 4 x NCH_2CH_3), 3.12 (9H, s, $\text{N}(\text{CH}_3)_3$), 2.24 (2H, m, $\text{OCH}_2\text{CH}_2\text{CH}_2\text{NMe}_3$), 1.17 (12H, m, 4 x NCH_2CH_3); ^{13}C NMR (500 MHz, DMSO-d_6) δ 166.0, 165.2, 156.8, 153.1, 152.7, 147.0, 146.0, 145.4, 138.5, 136.6, 136.4, 115.6, 113.5, 113.5, 112.6, 105.1, 104.5, 100.7, 98.4, 65.5, 63.0, 52.3, 44.1, 44.1, 22.5, 12.7; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{34}\text{H}_{46}\text{N}_5\text{O}_3^{195}\text{Pt}]^+$ requires 767.3243, found 767.3274; LCMS TOF MS ES $^+$: 2.78 min, 100 %, m/z 767 $[\text{M}]^+$.

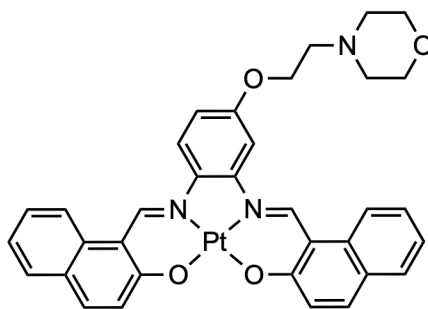
Pt_PEM_Ph:



Dark orange solid (222 mg, 0.283 mmol, 65 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.47 (1H, s, Ar imine-CH), 9.30 (1H, s, Ar imine-CH), 8.29 (1H, d, J 9.3, Ar-H), 7.94 (1H, d, J 2.5, Ar-H), 7.82 (2H, ddd, J 21.3, 8.1, 1.9, 2 x Ar-H), 7.54 (2H, dddd, J 15.3, 8.6, 6.7, 1.8, 2 x Ar-H), 7.09 (3H, m, 3 x Ar-H), 6.76 (2H, m, Ar-H), 4.23 (2H, t, J 5.7, ethyl- $\text{OCH}_2\text{CH}_2\text{N}$), 3.61 (4H, t, J 4.7, 2 x morph- $\text{NCH}_2\text{CH}_2\text{O}$), 2.74 (2H,

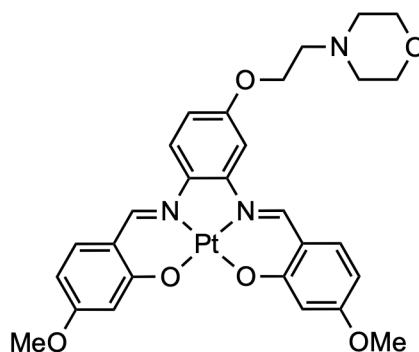
t, J 5.7, ethyl-OCH₂CH₂N); ¹³C NMR (500 MHz, DMSO-d₆) δ 164.5, 163.8, 157.4, 151.1, 149.9, 145.7, 138.8, 135.5, 135.4, 135.4, 135.0, 122.1, 121.8, 121.3, 120.9, 116.7, 116.3, 116.1, 114.7, 101.8, 63.4, 62.8, 55.1, 51.9; HRMS m/z (ESI)⁺ calcd. for [C₂₆H₂₆N₃O₄¹⁹⁵Pt]⁺ requires 639.1566, found 639.1647; LCMS TOF MS ES⁺: 2.01 min, 100 %, m/z 639 [M]⁺.

Pt_PEM_Naph:



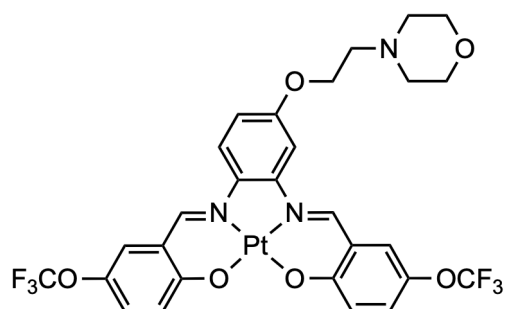
Dark red solid (117 mg, 0.132 mmol, 98 %); ¹H NMR (400 MHz, DMSO-d₆) δ 9.87 (1H, s, Ar imine-CH), 9.79 (1H, s, Ar imine-CH), 8.68 – 8.54 (3H, m, Ar-H), 8.07 (1H, d, J 2.4, Ar-H), 7.99 (2H, dd, J 13.5, 9.3, Ar-H), 7.84 (2H, dd, J 8.2, 3.3, Ar-H), 7.62 (2H, q, J 8.6, 2 x Ar-H), 7.38 (2H, q, J 7.5, Ar-H), 7.29 (2H, d, J 9.2, Ar-H), 7.03 (1H, dd, J 9.2, 2.4, Ar-H), 4.29 (2H, t, J 5.7, ethyl-OCH₂CH₂), 3.63 (4H, t, J 4.6, 2 x morph-NCH₂CH₂O), 2.75 (2H, t, J 5.8, ethyl-OCH₂CH₂N); ¹³C NMR (500 MHz, DMSO-d₆) δ 165.9, 164.8, 145.8, 143.6, 142.5, 138.9, 135.9, 135.4, 133.9, 133.8, 128.9, 128.8, 127.7, 127.3, 127.2, 126.6, 124.2, 124.1, 123.3, 123.1, 121.4, 121.3, 116.7, 113.7, 112.0, 112.0, 101.6, 63.9, 63.9, 55.7, 52.2; HRMS m/z (ESI)⁺ calcd. for [C₃₄H₃₀N₃O₄¹⁹⁵Pt]⁺ requires 739.1879, found 739.1941; LCMS TOF MS ES⁺: 2.77 min, 90 %, m/z 739 [M]⁺.

Pt_PEM_4OMe:



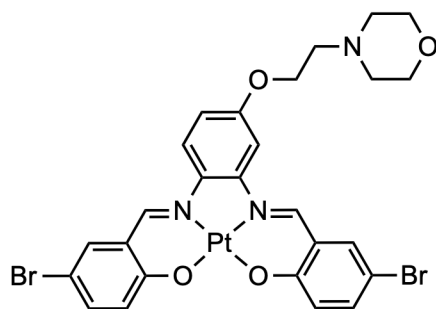
Brown solid (112 mg, 0.133 mmol, 98 %); ^1H NMR (400 MHz, DMSO- d_6) δ 9.27 (1H, s, Ar imine-CH), 9.11 (1H, s, Ar imine-CH), 8.21 (1H, d, J 9.2, Ar-H), 7.86 (1H, s, Ar-H), 7.69 (2H, dd, J 20.5, 9.0, Ar-H), 7.04 (1H, dd, J 9.1, 2.4, Ar-H), 6.58 (2H, dd, J 4.4, 2.4, Ar-H), 6.44 (2H, m, Ar-H), 4.22 (2H, t, J 5.7, ethyl-OCH₂CH₂), 3.83 (6H, d, J 4.5, 2 x OCH₃), 3.61 (4H, t, J 4.6, 2 x morph-NCH₂CH₂O), 2.75 (2H, t, J 5.7, ethyl-OCH₂CH₂N); ^{13}C NMR (500 MHz, DMSO- d_6) δ 166.6, 165.8, 165.5, 165.1, 157.1, 149.5, 148.3, 145.5, 138.6, 136.6, 136.4, 116.6, 116.5, 116.3, 113.8, 107.9, 107.3, 102.1, 102.0, 101.3, 63.9, 63.1, 55.6, 55.3, 52.3; HRMS m/z (ESI)⁺ calcd. for [C₂₈H₃₀N₃O₆¹⁹⁵Pt]⁺ requires 699.1777, found 699.1789; LCMS TOF MS ES⁺: 2.15 min, 100 %, m/z 699 [M]⁺.

Pt_PEM_5OCF₃:



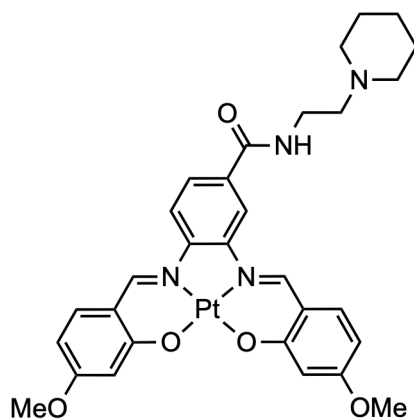
Red purple solid (103 mg, 0.108 mmol, 80 %); ^1H NMR (400 MHz, DMSO- d_6) δ 9.36 (1H, s, Ar imine-CH), 9.19 (1H, s, Ar imine-CH), 8.08 (1H, d, J 9.2, Ar-H), 7.76 – 7.69 (3H, m, Ar-H), 7.45 (2H, m, Ar-H), 7.02 (2H, dd, J 9.3, 1.5, Ar-H), 6.93 (1H, dd, J 9.2, 2.3, Ar-H), 4.11 (2H, t, J 5.9, ethyl-OCH₂CH₂), 3.61 (4H, t, J 4.7, 2 x morph-NCH₂CH₂O), 2.70 (2H, t, J 5.9, ethyl-OCH₂CH₂N); ^{13}C NMR (500 MHz, DMSO- d_6) δ 162.8, 162.2, 157.7, 150.8, 149.4, 145.6, 138.5, 137.8, 137.7, 128.6, 128.1, 126.0, 123.5, 122.6, 122.2, 121.4, 121.3, 121.0, 119.4, 117.4, 116.8, 115.1, 101.6, 63.8, 63.3, 55.3, 52.2; ^{19}F NMR (500 MHz, DMSO- d_6) -69.4, -70.9; HRMS m/z (ESI)⁺ calcd. for [C₂₈H₂₄F₆N₃O₆¹⁹⁵Pt]⁺ requires 807.1212, found 807.1215; LCMS TOF MS ES⁺: 2.83 min, 100 %, m/z 807 [M]⁺.

Pt_PEM_5Br:



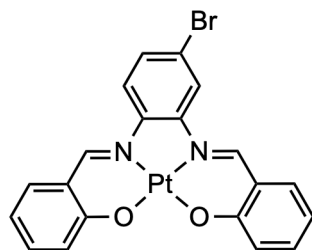
Red solid (63.2 mg, 0.0671 mmol, 50 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.46 (1H, s, Ar imine-CH), 9.28 (1H, s, Ar imine-CH), 8.21 (1H, d, J 9.2, Ar-H), 8.05 – 7.94 (2H, m, Ar-H), 7.86 (1H, s, Ar-H), 7.67 – 7.56 (2H, m, Ar-H), 7.11 (1H, d, J 9.6, Ar-H), 7.03 (2H, dd, J 9.2, 5.4, Ar-H), 4.20 (2H, m, ethyl- OCH_2CH_2), 3.62 (4H, t, J 4.7, 2 x morph- $\text{NCH}_2\text{CH}_2\text{O}$), 2.76 (2H, t, J 5.7, ethyl- $\text{OCH}_2\text{CH}_2\text{N}$); ^{13}C NMR (500 MHz, DMSO-d_6) δ 163.2, 162.5, 157.9, 150.7, 149.3, 145.7, 138.6, 137.3, 136.9, 136.3, 136.3, 123.6, 123.4, 123.1, 117.0, 115.2, 106.4, 106.2, 101.7, 64.5, 64.0, 55.6, 52.4; HRMS m/z (ESI) $^+$ calcd. for $[\text{C}_{26}\text{H}_{24}\text{Br}_2\text{N}_3\text{O}_4^{195}\text{Pt}]^+$ requires 796.9761, found 796.9800 $[\text{M}+2\text{H}]$; LCMS TOF MS ES $^+$: 4.70 min, 99 %, m/z 798 $[\text{M}]^+$.

Pt_AEP_4OMe:



Red solid (2 mg, 0.00230 mmol, 10 %); ^1H NMR (400 MHz, DMSO-d_6) δ 9.36 (2H, s, Ar imine-CH), 8.79 (1H, m, Ar-H), 8.45 (1H, d, J 8.8, Ar-H), 7.83 (2H, m, Ar-H), 7.75 (1H, d, J 8, Ar-H), 7.09 (1H, s, (O)CNH), 6.63 (2H, m, Ar-H), 6.5 (2H, m, Ar-H), 3.65 (2H, m, (O)CNH CH_2CH_2 -pip), 3.07 (4H, m, pip- $\text{N}(\text{CH}_2)_2(\text{CH}_2)_2\text{CH}_2$), 1.71 (4H, m, pip- $\text{N}(\text{CH}_2\text{CH}_2)_2\text{CH}_2$), 1.53 (2H, m, (O)CNH CH_2CH_2 -pip), 1.24 (2H, m, pip- $\text{N}(\text{CH}_2\text{CH}_2)_2\text{CH}_2$); LCMS TOF MS ES $^+$: 4.23 min, 95 %, m/z 724.3 $[\text{M}+\text{H}]^+$.

Pt_PhBr_Ph:



Dark red solid (41.0 mg, 0.0696 mmol, 52 %); ^1H NMR (400 MHz, DMSO- d_6) δ 9.51 (2H, d, J 20, Ar-imine-CH), 8.71 (1H, s, Ar-CBrCHN), 8.38 (1H, d, J 9, Ar-H), 7.84 (2H, t, J 7, Ar-H), 7.65 (1H, d, J 8.8, Ar-H), 7.59 (1H, d, J 10, Ar-H), 7.10 (2H, d, J 8, Ar-H), 6.81 (2H, m, Ar-H), LCMS TOF MS ES^+ : 6.29 min, 95 %, m/z 589 $[\text{M}+\text{H}]^+$.

References

- [1] Christine Tesseromatis. Drugs in greek mythology and medicine. *Journal of Medicinal Plants Studies*, 8:38–43, 2020.
- [2] Nikolaos A Kostakopoulos, Themistoklis C Bellos, Stamatios Katsimperi, and Lazaros Tzelves. Hippocrates of Kos (460-377 BC): The Founder and Pioneer of Clinical Medicine. *Cureus*, 16:e70602, 10 2024.
- [3] Alexander Fleming. The Discovery of Penicillin. *British Medical Bulletin*, 2:4–5, 1 1944.
- [4] Fulong Liao. Discovery of Artemisinin (Qinghaosu). *Molecules*, 14:5362, 12 2009.
- [5] Daniel L. Klayman. Qinghaosu (artemisinin): An antimalarial drug from China. *Science*, 228:1049–1055, 1985.
- [6] Mansukhlal C. Wani, Harold Lawrence Taylor, Monroe E. Wall, Philip Coggon, and Andrew T. McPhail. Plant Antitumor Agents .VI. The Isolation and Structure of Taxol, a Novel Antileukemic and Antitumor Agent from *Taxus brevifolia*. *Journal of the American Chemical Society*, 93:2325–2327, 5 1971.
- [7] Alan W. Jones. Early drug discovery and the rise of pharmaceutical chemistry. *Drug Testing and Analysis*, 3:337–344, 6 2011.
- [8] Walter Sneader. The discovery of aspirin: a reappraisal. *British Medical Journal*, 321: 1591, 12 2000.
- [9] Emil Fischer. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der Deutschen Chemischen Gesellschaft*, 27:2985–2993, 10 1894.
- [10] Maria Batool, Bilal Ahmad, and Sangdun Choi. A Structure-Based Drug Discovery Paradigm. *International Journal of Molecular Sciences*, 20:2783, 6 2019.

- [11] Edgar E. Meyer. The first years of the Protein Data Bank. *Protein Science*, 6:1591–1597, 1997.
- [12] Athina Geronikaki, Gurudutt Dubey, Anthi Petrou, and Sivapriya Kirubakaran. Computer-aided drug design: An overview. *Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development*, pages 39–68, 1 2023.
- [13] Daniel E. Koshland. The Key–Lock Theory and the Induced Fit Theory. *Angewandte Chemie International Edition*, 33:2375–2378, 1 1995.
- [14] Chandrabose Selvaraj, Ondipilliraja Rudhra, Abdulaziz S. Alothaim, Mustfa Alkhanani, and Sanjeev Kumar Singh. Structure and chemistry of enzymatic active sites that play a role in the switch and conformation mechanism. *Advances in Protein Chemistry and Structural Biology*, 130:59–83, 1 2022.
- [15] Huiwen Wang. Prediction of protein-ligand binding affinity via deep learning models. *Briefings in Bioinformatics*, 25, 3 2024.
- [16] Marta Bon, Alan Bilsland, Justin Bower, and Kirsten McAulay. Fragment based drug discovery the importance of high quality molecule libraries. *Molecular Oncology*, 16:3761, 11 2022.
- [17] Jean Louis Reymond. The Chemical Space Project. *Accounts of Chemical Research*, 48:722–730, 3 2015.
- [18] Asher Mullard. The drug-maker’s guide to the galaxy. *Nature*, 549:445–447, 9 2017.
- [19] Ravinderpal Kour Sodhi and Satya Paul. Metal Complexes in Medicine: An Overview and Update from Drug Design Perspective. *Cancer Therapy and Oncology International Journal*, 14:1–8, 6 2019.
- [20] Umar Ndagi, Ndumiso Mhlono, and Mahmoud E. Soliman. Metal complexes in cancer therapy – an update from drug design perspective. *Drug Design, Development and Therapy*, 11:599, 3 2017.
- [21] Chung H. Leung, Sheng Lin, Hai J. Zhong, and Dik L. Ma. Metal complexes as potential modulators of inflammatory and autoimmune responses. *Chemical Science*, 6:871, 2 2014.

- [22] Yi Liu, Jiaying Ma, Qianling Zhang, Yi Wang, and Qi Sun. Mechanism of Metal Complexes in Alzheimer's Disease. *International Journal of Molecular Sciences*, 25, 11 2024.
- [23] Angelo Frei. Metal Complexes, an Untapped Source of Antibiotic Potential? *Antibiotics*, 9:90, 2 2020.
- [24] Thomas W. Rees, Po Yu Ho, and Jeannine Hess. Recent Advances in Metal Complexes for Antimicrobial Photodynamic Therapy. *ChemBioChem*, 24:e202200796, 8 2023.
- [25] Susan Monro, Katsuya L. Colón, Huimin Yin, John Roque, Prathyusha Konda, Shashi Gujar, Randolph P. Thummel, Lothar Lilge, Colin G. Cameron, and Sherri A. McFarland. Transition Metal Complexes and Photodynamic Therapy from a Tumor-Centered Approach: Challenges, Opportunities, and Highlights from the Development of TLD1433. *Chemical Reviews*, 119:797, 1 2018.
- [26] Yanping Wu, Shumeng Li, Yuncong Chen, Weijiang He, and Zijian Guo. Recent advances in noble metal complex based photodynamic therapy. *Chemical Science*, 13:5085, 4 2022.
- [27] Zhongxian Fan, Juyang Huang, Huaiyi Huang, and Samya Banerjee. Metal-Based Catalytic Drug Development for Next-Generation Cancer Therapy. *ChemMedChem*, 16: 2480–2486, 8 2021.
- [28] Nora Graf and Stephen J. Lippard. Redox activation of metal-based prodrugs as a strategy for drug delivery. *Advanced Drug Delivery Reviews*, 64:993, 8 2012.
- [29] Stephane Bandeira, Jorge Gonzalez-Garcia, Evangelina Pensa, Ramon Vilar,] S Bandeira, J Gonzalez-Garcia, E Pensa, T Albrecht, and R Vilar. A Redox-Activated G-Quadruplex DNA Binder Based on a Platinum(IV)–Salphen Complex. *Angewandte Chemie International Edition*, 57:310–313, 1 2018.
- [30] Brondwyn S. McGhie and Janice R. Aldrich-Wright. Photoactive and Luminescent Transition Metal Complexes as Anticancer Agents: A Guiding Light in the Search for New and Improved Cancer Treatments. *Biomedicines*, 10:578, 3 2022.
- [31] Tudor I. Oprea, Cristian G. Bologa, Søren Brunak, Allen Campbell, Gregory N. Gan, Anna Gaulton, Shawn M. Gomez, Rajarshi Guha, Anne Hersey, Jayme Holmes, Ajit Jadhav, Lars Juhl Jensen, Gary L. Johnson, Anneli Karlson, Andrew R. Leach, Avi Ma'ayan, Anna

- Malovannaya, Subramani Mani, Stephen L. Mathias, Michael T. McManus, Terrence F. Meehan, Christian Von Mering, Daniel Muthas, Dac Trung Nguyen, John P. Overington, George Papadatos, Jun Qin, Christian Reich, Bryan L. Roth, Stephan C. Schürer, Anton Simeonov, Larry A. Sklar, Noel Southall, Susumu Tomita, Ilinca Tudose, Oleg Ursu, Dušica Vidović, Anna Waller, David Westergaard, Jeremy J. Yang, and Gergely Zahoránszky-Köhalmi. Unexplored therapeutic opportunities in the human genome. *Nature Reviews. Drug Discovery*, 17:317–332, 4 2018.
- [32] G. Sitta Sittampalam, Dobrila D. Rudnicki, Danilo A. Tagle, Anton Simeonov, and Christopher P. Austin. Mapping biologically active chemical space to accelerate drug discovery. *Nature Reviews. Drug Discovery*, 18:83, 2 2019.
- [33] Jayesh A. Kulkarni, Dominik Witzigmann, Sarah B. Thomson, Sam Chen, Blair R. Leavitt, Pieter R. Cullis, and Roy van der Meel. The current landscape of nucleic acid therapeutics. *Nature Nanotechnology* 2021 16:6, 16:630–643, 5 2021.
- [34] Peng Wu, Dik Lung Ma, Chung Hang Leung, Siu Cheong Yan, Nianyong Zhu, R. Abagyan, and Chi Ming Che. Stabilization of G-Quadruplex DNA with Platinum(II) Schiff Base Complexes: Luminescent Probe and Down-Regulation of c-myc Oncogene Expression. *Chemistry – A European Journal*, 15:13008–13021, 12 2009.
- [35] Jasmine S. Smith, Qijun Chen, Liliya A. Yatsunyk, John M. Nicoludis, Mark S. Garcia, Ramon Kranaster, Shankar Balasubramanian, David Monchaud, Marie Paule Teulade-Fichou, Lara Abramowitz, David C. Schultz, and F. Brad Johnson. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nature Structural and Molecular Biology*, 18:478–485, 3 2011.
- [36] Merve Becit, Sevtap A. Dilsiz, and Nurşen Başaran. Interaction of curcumin on cisplatin cytotoxicity in HeLa and HepG2 carcinoma cells. *Istanbul Journal of Pharmacy*, 50: 202–210, 2020.
- [37] James D. Watson and Francis H.C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- [38] Serguei V. Kuznetsov, Cha Chi Ren, Sarah A. Woodson, and Anjum Ansari. Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Research*, 36:1098, 3 2007.

- [39] Jixin Chen, Nitesh K. Poddar, Lawrence J. Tauzin, David Cooper, Anatoly B. Kolomeisky, and Christy F. Landes. Single-Molecule FRET Studies of HIV TAR–DNA Hairpin Unfolding Dynamics. *The Journal of Physical Chemistry B*, 118:12130, 10 2014.
- [40] Jin H. Song, Hyun J. Kang, Libia A. Luevano, Vijay Gokhale, Kui Wu, Ritu Pandey, H. H. Sherry Chow, Laurence H. Hurley, and Andrew S. Kraft. Small-molecule targeting hairpin loop of hTERT promoter G-quadruplex induces cancer cell death. *Cell Chemical Biology*, 26:1110, 8 2019.
- [41] Julian L. Huppert and Shankar Balasubramanian. Prevalence of quadruplexes in the human genome. *Nucleic Acids Research*, 33:2908–2916, 5 2005.
- [42] Susie L. Brown and Samantha Kendrick. The i-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure. *Pharmaceuticals*, 14:96, 2 2021.
- [43] Elaine Sanij, Katherine M. Hannan, Jiachen Xuan, Shunfei Yan, Jessica E. Ahern, Anna S. Trigos, Natalie Brajanovski, Jinbae Son, Keefe T. Chan, Olga Kondrashova, Elizabeth Lieschke, Matthew J. Wakefield, Daniel Frank, Sarah Ellis, Carleen Cullinane, Jian Kang, Gretchen Poortinga, Purba Nag, Andrew J. Deans, Kum Kum Khanna, Linda Mileschkin, Grant A. McArthur, John Soong, Els M.J.J. Berns, Ross D. Hannan, Clare L. Scott, Karen E. Sheppard, and Richard B. Pearson. CX-5461 activates the DNA damage response and demonstrates therapeutic efficacy in high-grade serous ovarian cancer. *Nature Communications*, 11:1–18, 5 2020.
- [44] Amit Khot, Natalie Brajanovski, Donald P. Cameron, Nadine Hein, Kylee H. Maclachlan, Elaine Sanij, John Lim, John Soong, Emma Link, Piers Blombery, Ella R. Thompson, Andrew Fellowes, Karen E. Sheppard, Grant A. McArthur, Richard B. Pearson, Ross D. Hannan, Gretchen Poortinga, and Simon J. Harrison. First-in-human RNA polymerase I transcription inhibitor CX-5461 in patients with advanced hematologic cancers: Results of a phase I dose-escalation study. *Cancer Discovery*, 9:1036–1049, 8 2019.
- [45] Ahmed A. Ahmed, William Greenhalf, Daniel H. Palmer, Nicole Williams, Jenny Worthington, Tariq Arshad, Shozeb Haider, Effrosyni Alexandrou, Dilek Guneri, Zoe A.E. Waller, and Stephen Neidle. The Potent G-Quadruplex-Binding Compound QN-302 Downregulates S100P Gene Expression in Cells and in an In Vivo Model of Pancreatic Cancer. *Molecules*, 28:2452, 3 2023.

- [46] Duxin Sun, Wei Gao, Hongxiang Hu, and Simon Zhou. Why 90 percent of clinical drug development fails and how to improve it. *Acta Pharmaceutica Sinica B*, 12:3049–3062, 7 2022.
- [47] Denise B. Catacutan, Jeremie Alexander, Autumn Arnold, and Jonathan M. Stokes. Machine learning in preclinical drug discovery. *Nature Chemical Biology*, 20:960–973, 7 2024.
- [48] Kang Zhang, Xin Yang, Yifei Wang, Yunfang Yu, Niu Huang, Gen Li, Xiaokun Li, Joseph C. Wu, and Shengyong Yang. Artificial intelligence in drug development. *Nature Medicine*, 31:45–59, 1 2025.
- [49] Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, and Shanrong Zhao. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18:463–477, 4 2019.
- [50] Friederike Maite Siemers, Christian Feldmann, and Jürgen Bajorath. Minimal data requirements for accurate compound activity prediction using machine learning methods of different complexity. *Cell Reports Physical Science*, 3:101113, 11 2022. ISSN 2666-3864.
- [51] Gisbert Schneider. Automating drug discovery. *Nature Reviews Drug Discovery*, 17: 97–113, 12 2017.
- [52] Steven M. Mennen, Carolina Alhambra, C. Liana Allen, Mario Barberis, Simon Berritt, Thomas A. Brandt, Andrew D. Campbell, Jesús Castañón, Alan H. Cherney, Melodie Christensen, David B. Damon, J. Eugenio De Diego, Susana García-Cerrada, Pablo García-Losada, Rubén Haro, Jacob Janey, David C. Leitch, Ling Li, Fangfang Liu, Paul C. Lobben, David W.C. Macmillan, Javier Magano, Emma McInturff, Sebastien Monfette, Ronald J. Post, Danielle Schultz, Barbara J. Sitter, Jason M. Stevens, Iulia I. Strambeanu, Jack Twilton, Ke Wang, and Matthew A. Zajac. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Organic Process Research and Development*, 23:1213–1242, 6 2019.
- [53] Joseph Zaia. The 2022 Nobel Prize in Chemistry for the development of click chemistry and bioorthogonal chemistry. *Analytical and Bioanalytical Chemistry*, 415:527–532, 2

2023.

- [54] Melissa J. Buskes and Maria J. Blanco. Impact of Cross-Coupling Reactions in Drug Discovery and Development. *Molecules*, 25:3493, 8 2020.
- [55] Sujiet Puthenveetil. Utilizing Solid-Phase to Enable High-Throughput, Site-Specific Conjugation and Dual-Labeled Antibody and Fab Conjugates. *Methods in Molecular Biology*, 2078:99–112, 2020.
- [56] Fanny Guzmán, Adriana Gauna, Tanya Roman, Omar Luna, Claudio Álvarez, Claudia Pareja-Barrueto, Luis Mercado, Fernando Albericio, and Constanza Cárdenas. Tea Bags for Fmoc Solid-Phase Peptide Synthesis: An Example of Circular Economy. *Molecules*, 26:5035, 8 2021.
- [57] Robert B. Merrifield. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society*, 85:2149–2154, 7 1963.
- [58] Rajavel Srinivasan, Lay Pheng Tan, Hao Wu, Peng Yu Yang, Karunakaran A. Kalesh, and Shao Q. Yao. High-throughput synthesis of azide libraries suitable for direct “click” chemistry and in situ screening. *Organic and Biomolecular Chemistry*, 7:1821–1828, 5 2009.
- [59] Kavya Ganabady, Nicola Contessi Negrini, Jacob C. Scherba, Brandon M. Nitschke, Morgan R. Alexander, Kyle H. Vining, Melissa A. Grunlan, David J. Mooney, and Adam D. Celiz. High-Throughput Screening of Thiol-ene Click Chemistries for Bone Adhesive Polymers. *ACS Applied Materials and Interfaces*, 15:50908–50915, 11 2023.
- [60] María Hechavarría Fonseca and Benjamin List. Combinatorial chemistry and high-throughput screening for the discovery of organocatalysts. *Current Opinion in Chemical Biology*, 8:319–326, 6 2004.
- [61] Zinia Jaman, Ahmed Mufti, Samyukta Sah, Larisa Avramova, and David H. Thompson. High Throughput Experimentation and Continuous Flow Validation of Suzuki–Miyaura Cross-Coupling Reactions. *Chemistry – A European Journal*, 24:9546–9554, 7 2018.
- [62] Bailee Wathen, Emily Lanehart, Lena A. Woodis, and Anthony J. Rojas. Buchwald-Hartwig Amination, High-Throughput Experimentation, and Process Chemistry: An Introduction via Undergraduate Laboratory Experimentation. *Journal of*

Chemical Education, 98:996–1000, 3 2021.

- [63] Athi Welsh, David Husbands, and Angelo Frei. High-Throughput Combinatorial Metal Complex Synthesis. *Angewandte Chemie International Edition*, 64:e202420204, 2 2025.
- [64] Timothy Kench, Arielle Rahardjo, Gianmarco G. Terrones, Adinarayana Bellamkonda, Thomas E. Maher, Marko Storch, Heather J. Kulik, and Ramon Vilar. A Semi-Automated, High-Throughput Approach for the Synthesis and Identification of Highly Photo-Cytotoxic Iridium Complexes. *Angewandte Chemie International Edition*, 63:e202401808, 4 2024.
- [65] Christopher J. Ziegler, Adam P. Silverman, and Stephen J. Lippard. High-throughput synthesis and screening of platinum drug candidates. *Journal of Biological Inorganic Chemistry*, 5:774–783, 2 2000.
- [66] Vincent Blay, Bhairavi Tolani, Sunita P. Ho, and Michelle R. Arkin. High-Throughput Screening: today's biochemical and cell-based approaches. *Drug Discovery Today*, 25: 1807–1821, 10 2020.
- [67] Xiaoni Fang, Yongzan Zheng, Yaokai Duan, Yang Liu, and Wenwan Zhong. Recent Advances in Design of Fluorescence-based Assays for High-throughput Screening. *Analytical Chemistry*, 91:482, 1 2018.
- [68] Johan Fenneteau, Dany Chauvin, Andrew D. Griffiths, Clément Nizak, and Janine Cossy. Synthesis of new hydrophilic rhodamine based enzymatic substrates compatible with droplet-based microfluidic assays. *Chemical Communications*, 53:5437–5440, 5 2017.
- [69] Yu Jing Lu, Qiang Deng, Jin Qiang Hou, Dong Ping Hu, Zheng Ya Wang, Kun Zhang, Leonard G. Luyt, Wing Leung Wong, and Cheuk Fai Chow. Molecular Engineering of Thiazole Orange Dye: Change of Fluorescent Signaling from Universal to Specific upon Binding with Nucleic Acids in Bioassay. *ACS Chemical Biology*, 11:1019–1029, 4 2016.
- [70] Michal Avital-Shmilovici, Xiaohe Liu, Thomas Shaler, Andrew Lowenthal, Pauline Bourbon, Janey Snider, Arlyn Tambo-Ong, Claire Repellin, Kenya Yniguez, Lidia Sambucetti, Peter B. Madrid, and Nathan Collins. Mega-High-Throughput Screening Platform for the Discovery of Biologically Relevant Sequence-Defined Non-Natural Polymers. *ACS Central Science*, 8:86–101, 1 2022.
- [71] Sarah L. Wicks and Amanda E. Hargrove. Fluorescent indicator displacement assays to

- identify and characterize small molecule interactions with RNA. *Methods*, 167:3–14, 9 2019.
- [72] Tomonori Shibata, Yasumasa Matsumoto, Akiko Iihara, Kazunori Yamada, Hiroshi Ochiai, Ryo Saito, Shinichi Kusaka, and Toshiyuki Kume. Fluorescent indicator displacement assay for the discovery of UGGAA repeat-targeted small molecules. *Chemical Communications*, 59:5071–5074, 4 2023.
- [73] Adrien Boes, Samir Olatunji, Tamimount Mohammadi, Eefjan Breukink, and Mohammed Terrak. Fluorescence anisotropy assays for high throughput screening of compounds binding to lipid II, PBP1b, FtsW and MurJ. *Scientific Reports*, 10:1–8, 4 2020.
- [74] Sven T. Sowa, Carlos Vela-Rodríguez, Albert Galera-Prat, Mariana Cázares-Olivera, Renata Prunskaitė-Hyyryläinen, Alexander Ignatev, and Lari Lehtiö. A FRET-based high-throughput screening platform for the discovery of chemical probes targeting the scaffolding functions of human tankyrases. *Scientific Reports*, 10:1–12, 7 2020.
- [75] Nikolett Szalóki, Quang Minh Doan-Xuan, János Szöllosi, Katalin Tóth, György Vámosi, and Zsolt Bacsó. High throughput FRET analysis of protein-protein interactions by slide-based imaging laser scanning cytometry. *Cytometry Part A*, 83:818–829, 9 2013.
- [76] Daniel R. Stroik, Samantha L. Yuen, Kevyn A. Janicek, Tory M. Schaaf, Ji Li, Delaine K. Ceholski, Roger J. Hajjar, Razvan L. Cornea, and David D. Thomas. Targeting protein-protein interactions for therapeutic discovery via FRET-based high-throughput screening in living cells. *Scientific Reports*, 8:1–13, 8 2018.
- [77] Kaibo Wang, Daniel P. Flaherty, Lan Chen, and Danzhou Yang. High-Throughput Screening of G-Quadruplex Ligands by FRET Assay. *Methods in Molecular Biology*, 2035: 323, 2019.
- [78] Maria Emilia Dueñas, Rachel E Peltier-Heap, Melanie Leveridge, Roland S Annan, Frank H Büttner, and Matthias Trost. Advances in high throughput mass spectrometry in drug discovery. *EMBO Molecular Medicine*, 15:e14850, 1 2022.
- [79] David G. McLaren, Vinit Shah, Thomas Wisniewski, Lucien Ghislain, Chang Liu, Hui Zhang, and S. Adrian Saldanha. High-Throughput Mass Spectrometry for Hit Identification: Current Landscape and Future Perspectives. *SLAS Discovery*, 26: 168–191, 2 2021.

- [80] Jim Nonomiya, Ke Sherry Li, Brett M. Babin, and Melinda M. Mulvihill. Covalent Library Screening by Targeted Mass Spectrometry for Rapid Binding Site Identification. *Analytical Chemistry*, 95:3779–3788, 2 2023.
- [81] Bainan Wu, Elisa Barile, Surya K De, Jun Wei, Angela Purves, and Maurizio Pellecchia. High-throughput screening by Nuclear Magnetic Resonance (HTS by NMR) for the identification of PPIs antagonists. *Current Topics in Medicinal Chemistry*, 15:2032, 7 2015.
- [82] Gregory Nierode, Paul S. Kwon, Jonathan S. Dordick, and Seok Joon Kwon. Cell-Based Assay Design for High-Content Screening of Drug Candidates. *Journal of Microbiology and Biotechnology*, 26:213, 10 2016.
- [83] Peter Larsson and Toshima Z. Parris. Optimization of Cell Viability Assays for Drug Sensitivity Screens. *Methods in Molecular Biology*, 2644:287–302, 2023.
- [84] Ru Zang, Ding Li, I-Ching Tang, Jufang Wang, and Shang-Tian Yang. Cell-Based Assays in High-Throughput Screening for Drug Discovery. *International Journal of Biotechnology for Wellness Industries*, 1:31–51, 2012.
- [85] Sonia Chadha and Sharad P. Kale. Simple fluorescence based high throughput cell viability assay for filamentous fungi. *Letters in Applied Microbiology*, 61:238–244, 9 2015.
- [86] Daniel F. Gilbert and Michael Boutros. A protocol for a high-throughput multiplex cell viability assay. *Methods in Molecular Biology*, 1470:75–84, 2016.
- [87] Paul W. Hastwell, Li Leng Chai, Kevin J. Roberts, Thomas W. Webster, James S. Harvey, Robert W. Rees, and Richard M. Walmsley. High-specificity and high-sensitivity genotoxicity assessment in a human cell line: validation of the GreenScreen HC GADD45a-GFP genotoxicity assay. *Mutation Research*, 607:160–175, 9 2006.
- [88] Andrew W. Knight, Stephen Little, Keith Houck, David Dix, Richard Judson, Ann Richard, Nancy McCarroll, Gregory Akerman, Chihae Yang, Louise Birrell, and Richard M. Walmsley. Evaluation of high-throughput genotoxicity assays used in profiling the US EPA ToxCast chemicals. *Regulatory Toxicology and Pharmacology*, 55:188–199, 11 2009.
- [89] Irena Bronstein, John Fortin, Philip E. Stanley, Gordon S.A.B. Stewart, and Larry J. Kricka. Chemiluminescent and Bioluminescent Reporter Gene Assays. *Analytical Biochemistry*,

219:169–181, 6 1994.

- [90] Thomas J. Crisman, Christian N. Parker, Jeremy L. Jenkins, Josef Scheiber, Mathis Thoma, Zhao Bin Kang, Richard Kim, Andreas Bender, James H. Nettles, John W. Davies, and Meir Glick. Understanding false positives in reporter gene assays: In silico chemogenomics approaches to prioritize cell-based HTS data. *Journal of Chemical Information and Modeling*, 47:1319–1327, 2007.
- [91] Mark C. Fishman and Jeffery A. Porter. A new grammar for drug discovery. *Nature*, 437: 491–493, 9 2005.
- [92] Michael J. Delves, Celia Miguel-Blanco, Holly Matthews, Irene Molina, Andrea Ruecker, Sabrina Yahiya, Ursula Straschil, Matthew Abraham, María Luisa León, Oliver J. Fischer, Ainoa Rueda-Zubiaurre, Jochen R. Brandt, Álvaro Cortés, Anna Barnard, Matthew J. Fuchter, Félix Calderón, Elizabeth A. Winzeler, Robert E. Sinden, Esperanza Herreros, Francisco J. Gamo, and Jake Baum. A high throughput screen for next-generation leads targeting malaria parasite transmission. *Nature Communications*, 9:1–13, 9 2018.
- [93] Vidhya Prasad, Denis Semwogerere, and Eric R. Weeks. Confocal microscopy of colloids. *Journal of Physics: Condensed Matter*, 19:113102, 2 2007.
- [94] Celia Miguel-Blanco, Joël Lelièvre, Michael J. Delves, Ana I. Bardera, Jesús L. Presa, María José López-Barragán, Andrea Ruecker, Sara Marques, Robert E. Sinden, and Esperanza Herreros. Imaging-based high-throughput screening assay to identify new molecules with transmission-blocking potential against *Plasmodium falciparum* female gamete formation. *Antimicrobial Agents and Chemotherapy*, 59:3298–3305, 6 2015.
- [95] Beverly L. Falcon, Julie Stewart, Scharri Ezell, Jeff Hanson, John Wijsman, Xiang Ye, Eric Westin, Greg Donoho, Kelly Credille, and Mark T. Uhlik. High-content multiplexed tissue imaging and quantification for cancer drug discovery. *Drug Discovery Today*, 18: 510–522, 6 2013.
- [96] Mai Yamagishi, Kaede Miyata, Takashi Kamatani, Hiroki Kabata, Rie Baba, Yumiko Tanaka, Nobutake Suzuki, Masako Matsusaka, Yasutaka Motomura, Tsuyoshi Kiniwa, Satoshi Koga, Keisuke Goda, Osamu Ohara, Takashi Funatsu, Koichi Fukunaga, Kazuyo Moro, Sotaro Uemura, and Yoshitaka Shirasaki. Quantitative live-cell imaging of secretion activity reveals dynamic immune responses. *iScience*, 27:109840, 6 2024.

- [97] Ayodele J. Oyejide, Adeyemi A. Adekunle, Oluwatosin D. Abodunrin, and Ebenezer O. Atoyebe. Artificial intelligence, computational tools and robotics for drug discovery, development, and delivery. *Intelligent Pharmacy*, 1 2025.
- [98] Mariusz Butkiewicz, Yanli Wang, Stephen H Bryant, Jr Edward W Lowe, David C Weaver, and Jens Meiler. High-Throughput Screening Assay Datasets from the PubChem Database. *Chemical Informatics*, 3:1, 2017.
- [99] Gabriel H.S. Dreiman, Magda Bictash, Paul V. Fish, Lewis Griffin, and Fredrik Svensson. Changing the HTS Paradigm: AI-Driven Iterative Screening for Hit Finding. *Slas Discovery*, 26:257, 2 2020.
- [100] Eduard Porta-Pardo, Victoria Ruiz-Serra, Samuel Valentini, and Alfonso Valencia. The structural coverage of the human proteome before and after AlphaFold. *PLoS Computational Biology*, 18:e1009818, 1 2022.
- [101] Dev Desai, Shiv V Kantliwala, Jyothi Vybhavi, Renju Ravi, Harshkumar Patel, and Jitendra Patel. Review of AlphaFold 3: Transformative Advances in Drug Design and Therapeutics. *Cureus*, 16:e63646, 7 2024.
- [102] Sally R. Ellingson, Brian Davis, and Jonathan Allen. Machine learning and ligand binding predictions: A review of data, methods, and obstacles. *Biochimica et biophysica acta. General subjects*, 1864, 6 2020.
- [103] Francesco Gentile, Jean Charle Yaacoub, James Gleave, Michael Fernandez, Anh Tien Ton, Fuqiang Ban, Abraham Stern, and Artem Cherkasov. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nature Protocols*, 17: 672–697, 3 2022.
- [104] Thomas C. Terwilliger, Dorothee Liebschner, Tristan I. Croll, Christopher J. Williams, Airlie J. McCoy, Billy K. Poon, Pavel V. Afonine, Robert D. Oeffner, Jane S. Richardson, Randy J. Read, and Paul D. Adams. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nature Methods*, 21: 110–116, 11 2023.
- [105] Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15:3130–3139, 2 2024.

- [106] Varnavas D. Mouchlis, Antreas Afantitis, Angela Serra, Michele Fratello, Anastasios G. Papadiamantis, Vassilis Aidinis, Iseult Lynch, Dario Greco, and Georgia Melagraki. Advances in De Novo Drug Design: From Conventional to Machine Learning Methods. *International Journal of Molecular Sciences*, 22:1676, 2 2021.
- [107] Feng Ren, Alex Aliper, Jian Chen, Heng Zhao, Sujata Rao, Christoph Kuppe, Ivan V. Ozerov, Man Zhang, Klaus Witte, Chris Kruse, Vladimir Aladinskiy, Yan Ivanenkov, Daniil Polykovskiy, Yanyun Fu, Eugene Babin, Junwen Qiao, Xing Liang, Zhenzhen Mou, Hui Wang, Frank W. Pun, Pedro Torres Ayuso, Alexander Veviorskiy, Dandan Song, Sang Liu, Bei Zhang, Vladimir Naumov, Xiaoqiang Ding, Andrey Kukhareenko, Evgeny Izumchenko, and Alex Zhavoronkov. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nature Biotechnology*, 43, 1 2025.
- [108] Petrina Kamya, Ivan V. Ozerov, Frank W. Pun, Kyle Tretina, Tatyana Fokina, Shan Chen, Vladimir Naumov, Xi Long, Sha Lin, Mikhail Korzinkin, Daniil Polykovskiy, Alex Aliper, Feng Ren, and Alex Zhavoronkov. PandaOmics: An AI-Driven Platform for Therapeutic Target and Biomarker Discovery. *Journal of Chemical Information and Modeling*, 64: 3961–3969, 5 2024.
- [109] Yan A. Ivanenkov, Daniil Polykovskiy, Dmitry Bezrukov, Bogdan Zagribelnyy, Vladimir Aladinskiy, Petrina Kamya, Alex Aliper, Feng Ren, and Alex Zhavoronkov. Chemistry42: An AI-Driven Platform for Molecular Design and Optimization. *Journal of Chemical Information and Modeling*, 63:695–701, 2 2023.
- [110] Natesh Singh, Philippe Vayer, Shivalika Tanwar, Jean-Luc Poyet, Katya Tsaïoun, Bruno O Villoutreix, Vasanthanathan Poongavanam, Duc Duy Vo, and Priyanka Banerjee. Drug discovery and development: introduction to the general public and patient groups. *Frontiers in Drug Discovery*, 3:1201419, 5 2023.
- [111] Frank W. Pun, Ivan V. Ozerov, and Alex Zhavoronkov. AI-powered therapeutic target discovery. *Trends in Pharmacological Sciences*, 44:561–572, 9 2023.
- [112] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O. Anatole Von Lilienfeld, Klaus Robert Müller, and Alexandre Tkatchenko. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *Journal of Physical Chemistry Letters*, 6:2326–2331, 6 2015.

- [113] Bara A Badwan, Gerry Liaropoulos, Efthymios Kyrodimos, Dimitrios Skaltsas, Aristotelis Tsirigos, and Vassilis G Gorgoulis. Machine learning approaches to predict drug efficacy and toxicity in oncology. *Cell Reports Methods*, 3:100413, 2023.
- [114] Vishal B. Siramshetty, Xin Xu, and Pranav Shah. Artificial Intelligence in ADME Property Prediction. *Methods in Molecular Biology*, 2714:307–327, 2024.
- [115] Claudio N. Cavasotto and Valeria Scardino. Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point. *ACS Omega*, 7:47536–47546, 12 2022.
- [116] Ha Y. Jang, Jihyeon Song, Jae H. Kim, Howard Lee, In W. Kim, Bongki Moon, and Jung M. Oh. Machine learning-based quantitative prediction of drug exposure in drug-drug interactions using drug label information. *Digital Medicine*, 5:1–11, 7 2022.
- [117] Zhi Z. Wang, Xing X. Shi, Guang Y/ Huang, Ge F. Hao, and Guang F. Yang. Fragment-based drug design facilitates selective kinase inhibitor discovery. *Trends in Pharmacological Sciences*, 42:551–565, 7 2021.
- [118] Feixiong Cheng, Weiqiang Lu, Chuang Liu, Jiansong Fang, Yuan Hou, Diane E. Handy, Ruisheng Wang, Yuzheng Zhao, Yi Yang, Jin Huang, David E. Hill, Marc Vidal, Charis Eng, and Joseph Loscalzo. A genome-wide positioning systems network algorithm for in silico drug repurposing. *Nature Communications*, 10:1–14, 8 2019.
- [119] Arseniy E Yuzhalin. Redefining cancer research for therapeutic breakthroughs. *British Journal of Cancer*, 130:1078, 4 2024.
- [120] Pan Pantziarka, Rica Capistrano I, Arno De Potter, Liese Vandeborne, and Gauthier Bouche. An Open Access Database of Licensed Cancer Drugs. *Frontiers in Pharmacology*, 12, 3 2021.
- [121] Sohini Kulavi, Chandreyi Ghosh, Moumita Saha, and Sirshendu Chatterjee. One Size Does Not Fit All: An Overview of Personalized Treatment in Cancer. *Journal of Pharmaceutical Research International*, 33:87–103, 5 2021.
- [122] Rob Jones. Cytotoxic chemotherapy: Clinical aspects. *Medicine*, 44:25–29, 1 2016.
- [123] Alex D. Waldman, Jill M. Fritz, and Michael J. Lenardo. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology*, 20:651–668,

5 2020.

- [124] Kyoji Tsuchikama, Yasuaki Anami, Summer Y.Y. Ha, and Chisato M. Yamazaki. Exploring the next generation of antibody–drug conjugates. *Nature Reviews Clinical Oncology*, 21: 203–223, 1 2024.
- [125] Jacinta Abraham and John Staffurth. Hormonal therapy for cancer. *Medicine*, 44:30–33, 1 2016.
- [126] Tafere Mulaw Belete. The Current Status of Gene Therapy for the Treatment of Cancer. *Biologics: Targets and Therapy*, 15:67, 2021.
- [127] Sri Murugan Poongkavithai Vadevoo, Smriti Gurung, Hyun Su Lee, Gowri Rangaswamy Gunassekaran, Seok Min Lee, Jae Won Yoon, Yun Ki Lee, and Byungheon Lee. Peptides as multifunctional players in cancer therapy. *Experimental and Molecular Medicine*, 55: 1099–1109, 6 2023.
- [128] D. Arango, A. J. Wilson, Q. Shi, G. A. Corner, M. J. Arañes, C. Nicholas, M. Lesser, J. M. Mariadason, and L. H. Augenlicht. Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. *British Journal of Cancer*, 91:1931–1946, 11 2004.
- [129] Patricia A. Sinoway and Jeffrey P. Callen. Chlorambucil. *Arthritis and Rheumatism*, 36: 319–324, 3 1993.
- [130] Kelly Johnson-Arbor and Ramin Dubey. Doxorubicin. *The Comprehensive Pharmacology Reference*, pages 1–5, 10 2017.
- [131] Patrick J. Loehrer and Lawrence H. Einhorn. Cisplatin. *Annals of Internal Medicine*, 100: 704–713, 12 2008.
- [132] E. Raymond, S. G. Chaney, A. Taamma, and E. Cvitkovic. Oxaliplatin: A review of preclinical and clinical studies. *Annals of Oncology*, 9:1053–1071, 10 1998.
- [133] Renzo Canetta, Marcel Rozenzweig, and Stephen K. Carter. Carboplatin: the clinical spectrum to date. *Cancer Treatment Reviews*, 12:125–136, 9 1985.
- [134] Guoqiang Sun, Dawei Rong, Zhouxiao Li, Guangshun Sun, Fan Wu, Xiao Li, Hongyong Cao, Ye Cheng, Weiwei Tang, and Yangbai Sun. Role of Small Molecule Targeted

- Compounds in Cancer: Progress, Opportunities, and Challenges. *Frontiers in Cell and Developmental Biology*, 9:694363, 9 2021.
- [135] Uttpal Anand, Abhijit Dey, Arvind K.Singh Chandel, Rupa Sanyal, Amarnath Mishra, Devendra Kumar Pandey, Valentina De Falco, Arun Upadhyay, Ramesh Kandimalla, Anupama Chaudhary, Jaspreet Kaur Dhanjal, Saikat Dewanjee, Jayalakshmi Vallamkondu, and José M. Pérez de la Lastra. Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes and Diseases*, 10:1367–1401, 7 2023.
- [136] Ankit Kumar Singh, Adarsh Kumar, Harshwardhan Singh, Pankaj Sonawane, Prateek Pathak, Maria Grishina, Jagat Pal Yadav, Amita Verma, and Pradeep Kumar. Metal Complexes in Cancer Treatment: Journey So Far. *Chemistry and Biodiversity*, 20:e202300061, 4 2023.
- [137] Ron R. Allison and Keyvan Moghissi. Photodynamic Therapy (PDT): PDT Mechanisms. *Clinical Endoscopy*, 46:24, 2013.
- [138] Chaw-Ning Lee, Rosie Hsu, Hsuan Chen, Tak-Wah Wong, M Amparo, F Faustino, Carlos J P Monteiro, and Catarina I V Ramos. Daylight Photodynamic Therapy: An Update. *Molecules*, 25:5195, 11 2020.
- [139] Prabal Singh Maharjan and Hitesh Kumar Bhattarai. Singlet Oxygen, Photodynamic Therapy, and Mechanisms of Cancer Cell Death. *Journal of Oncology*, 2022:7211485, 2022.
- [140] Girish S. Kulkarni, Lothar Lilge, Michael Nesbitt, Roger J. Dumoulin-White, Arkady Mandel, and Michael A.S. Jewett. A Phase 1b Clinical Study of Intravesical Photodynamic Therapy in Patients with Bacillus Calmette-Guérin-unresponsive Non-muscle-invasive Bladder Cancer. *European Urology Open Science*, 41:105, 7 2022.
- [141] José H. Correia, José A. Rodrigues, Sara Pimenta, Tao Dong, and Zhaochu Yang. Photodynamic Therapy Review: Principles, Photosensitizers, Applications, and Future Directions. *Pharmaceutics*, 13:1332, 9 2021.
- [142] Gurcan Gunaydin, M. Emre Gedik, and Seylan Ayan. Photodynamic Therapy—Current Limitations and Novel Approaches. *Frontiers in Chemistry*, 9:691697, 6 2021.

- [143] Martin Ochsner. Photophysical and photobiological processes in the photodynamic therapy of tumours. *Journal of Photochemistry and Photobiology B: Biology*, 39:1–18, 1997.
- [144] Genevieve Housman, Shannon Byler, Sarah Heerboth, Karolina Lapinska, McKenna Longacre, Nicole Snyder, and Sibaji Sarkar. Drug Resistance in Cancer: An Overview. *Cancers*, 6:1769, 9 2014.
- [145] Timothée Olivier, Alyson Haslam, and Vinay Prasad. Anticancer Drugs Approved by the US Food and Drug Administration From 2009 to 2020 According to Their Mechanism of Action. *JAMA Network Open*, 4:e2138793–e2138793, 12 2021.
- [146] Jiajie Peng, Yuxian Wang, Jiaojiao Guan, Jingyi Li, Ruijiang Han, Jianye Hao, Zhongyu Wei, and Xuequn Shang. An end-to-end heterogeneous graph representation learning-based framework for drug–target interaction prediction. *Briefings in Bioinformatics*, 22:1–9, 9 2021.
- [147] Kanghao Shao, Yunhao Zhang, Yuqi Wen, Zhongnan Zhang, Song He, and Xiaochen Bo. DTI-HETA: prediction of drug–target interactions based on GCN and GAT on heterogeneous graph. *Briefings in Bioinformatics*, 23:1–16, 5 2022.
- [148] Ziduo Yang, Weihe Zhong, Lu Zhao, and Calvin Yu Chian Chen. ML-DTI: Mutual learning mechanism for interpretable drug-target interaction prediction. *Journal of Physical Chemistry Letters*, 12:4247–4261, 5 2021.
- [149] Rawan S. Olayan, Haitham Ashoor, and Vladimir B. Bajic. DDR: efficient computational method to predict drug–target interactions using graph mining and machine learning approaches. *Bioinformatics*, 34:1164–1173, 4 2018.
- [150] Arwa Raies, Ewa Tulodziecka, James Stainer, Lawrence Middleton, Ryan S. Dhindsa, Pamela Hill, Ola Engkvist, Andrew R. Harper, Slavé Petrovski, and Dimitrios Vitsios. DrugnomeAI is an ensemble machine-learning framework for predicting druggability of candidate drug targets. *Communications Biology*, 5:1–16, 11 2022.
- [151] Alan Huang, Levi A. Garraway, Alan Ashworth, and Barbara Weber. Synthetic lethality as an engine for cancer drug target discovery. *Nature Reviews Drug Discovery*, 19:23–38, 11 2019.

- [152] Bing Hui Li, Jun Qi Ge, Ya Li Wang, Li Jun Wang, Qi Zhang, and Cong Bian. Ligand-Based and Docking-Based Virtual Screening of MDM2 Inhibitors as Potent Anticancer Agents. *Computational and Mathematical Methods in Medicine*, 2021:3195957, 1 2021.
- [153] Michael C. Hutter. Differential Multimolecule Fingerprint for Similarity Search—Making Use of Active and Inactive Compound Sets in Virtual Screening. *Journal of Chemical Information and Modeling*, 62:2726–2736, 6 2022.
- [154] Wenqi Liu, Soumen K. Samanta, Bradley D. Smith, and Lyle Isaacs. Synthetic mimics of biotin/(strept)avidin. *Chemical Society Reviews*, 46:2391–2403, 5 2017.
- [155] Jannis Born, Matteo Manica, Ali Oskooei, Joris Cadow, Greta Markert, and María Rodríguez Martínez. PaccMannRL: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning. *iScience*, 24:102269, 4 2021.
- [156] Seung Hwan Hong, Seongok Ryu, Jaechang Lim, and Woo Youn Kim. Molecular Generative Model Based on an Adversarially Regularized Autoencoder. *Journal of Chemical Information and Modeling*, 60:29–36, 1 2020.
- [157] Bidisha Samanta, Abir De, Gourhari Jana, Pratim Kumar Chattaraj, Niloy Ganguly, and Manuel Gomez Rodriguez. NeVAE: A Deep Generative Model for Molecular Graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:1110–1117, 7 2019.
- [158] Francesca Grisoni, Michael Moret, Robin Lingwood, and Gisbert Schneider. Bidirectional Molecule Generation with Recurrent Neural Networks. *Journal of Chemical Information and Modeling*, 60:1175–1183, 3 2020.
- [159] Łukasz Maziarka, Agnieszka Pocha, Jan Kaczmarczyk, Krzysztof Rataj, Tomasz Danel, and Michał Warchoń. Mol-CycleGAN: A generative model for molecular optimization. *Journal of Cheminformatics*, 12:1–18, 1 2020.
- [160] Maryam Abbasi, Beatriz P. Santos, Tiago C. Pereira, Raul Sofia, Nelson R.C. Monteiro, Carlos J.V. Simões, Rui Brito, Bernardete Ribeiro, José L. Oliveira, and Joel P. Arrais. Designing optimized drug candidates with Generative Adversarial Network. *Journal of Cheminformatics*, 14:1–16, 12 2022.
- [161] Manan Goel, Shampa Raghunathan, Siddhartha Laghuvarapu, and U. Deva Priyakumar.

- MoleGuLAR: Molecule Generation Using Reinforcement Learning with Alternating Rewards. *Journal of Chemical Information and Modeling*, 61:5815–5826, 12 2021.
- [162] Celia Chen, Jin Liu, Jianyu Wu, Xin Cheng, Luoheng Qin, Supriya Bavadekar, Man Zhang, Sujata Rao, Feng Ren, and Alex Zhavoronkov. 873 ISM5939, a novel, potent, orally available, selective ENPP1 inhibitor enhances the anti-tumor effects of immune checkpoint inhibitors in syngeneic murine cancer models. *Journal for ImmunoTherapy of Cancer*, 11:A972–A972, 11 2023.
- [163] Aparna Bansal, Shikha Kaushik, and Shrikant Kukreti. Non-canonical DNA structures: Diversity and disease association. *Frontiers in Genetics*, 13:959258, 9 2022.
- [164] Xiaoyi Deng, Yanxun V. Yu, and Youngnam N. Jin. Non-canonical translation in cancer: significance and therapeutic potential of non-canonical ORFs, m6A-modification, and circular RNAs. *Cell Death Discovery*, 10:1–18, 9 2024.
- [165] Nils Kosiol, Stefan Juranek, Peter Brossart, Annkristin Heine, and Katrin Paeschke. G-quadruplexes: a promising target for cancer therapy. *Molecular Cancer*, 20:1–18, 2 2021.
- [166] Tilmann Richl, Jochen Kuper, and Caroline Kisker. G-quadruplex-mediated genomic instability drives SNVs in cancer. *Nucleic Acids Research*, 52:2198–2211, 3 2024.
- [167] Joana Figueiredo, Jean Louis Mergny, and Carla Cruz. G-quadruplex ligands in cancer therapy: Progress, challenges, and clinical perspectives. *Life Sciences*, 340:122481, 3 2024.
- [168] Robert Hänsel-Hertsch, Angela Simeone, Abigail Shea, Winnie W.I. Hui, Katherine G. Zyner, Giovanni Marsico, Oscar M. Rueda, Alejandra Bruna, Alistair Martin, Xiaoyun Zhang, Santosh Adhikari, David Tannahill, Carlos Caldas, and Shankar Balasubramanian. Landscape of G-quadruplex DNA structural regions in breast cancer. *Nature Genetics*, 52:878–883, 8 2020.
- [169] Chuya Nakanishi and Hiroyuki Seimiya. G-quadruplex in cancer biology and drug discovery. *Biochemical and Biophysical Research Communications*, 531:45–50, 10 2020.
- [170] Adam Siddiqui-Jain, Cory L. Grand, David J. Bearss, and Laurence H. Hurley. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule

- to repress c-MYC transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 99:11593–11598, 9 2002.
- [171] Stephen Neidle. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nature Reviews Chemistry*, 1:1–10, 5 2017.
- [172] Jochen Spiegel, Santosh Adhikari, and Shankar Balasubramanian. The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry*, 2:123–136, 2 2020.
- [173] Debmalya Bhattacharyya, Gayan Mirihana Arachchilage, and Soumitra Basu. Metal Cations in G-Quadruplex Folding and Stability. *Frontiers in Chemistry*, 4, 2016.
- [174] Sarah Burge, Gary N. Parkinson, Pascale Hazel, Alan K. Todd, and Stephen Neidle. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*, 34: 5402–5415, 11 2006.
- [175] Andrew N. Lane, J. Brad Chaires, Robert D. Gray, and John O. Trent. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Research*, 36:5482–5515, 10 2008.
- [176] Jun Zhou, Anne Bourdoncle, Frøddoric Rosu, Valørie Gabelica, Jean-Louis Mergny, J Zhou, A Bourdoncle, F Rosu, and V Gabelica. Tri-G-Quadruplex: Controlled Assembly of a G-Quadruplex Structure from Three G-Rich Strands. *Angewandte Chemie International Edition*, 51:11002–11005, 10 2012.
- [177] Gary N. Parkinson, Michael P.H. Lee, and Stephen Neidle. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 2002, 417:876–880, 5 2002.
- [178] Giulia Biffi, David Tannahill, John McCafferty, and Shankar Balasubramanian. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature Chemistry*, 5:182–186, 3 2013.
- [179] Vicki S. Chambers, Giovanni Marsico, Jonathan M. Boutell, Marco Di Antonio, Geoffrey P. Smith, and Shankar Balasubramanian. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature Biotechnology*, 33:877–881, 7 2015.
- [180] Emilia Puig Lombardi and Arturo Londoño-Vallejo. A guide to computational methods for G-quadruplex prediction. *Nucleic Acids Research*, 48:1–15, 1 2020.

- [181] Oleg Kikin, Lawrence D'Antonio, and Paramjeet S. Bagga. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Research*, 34:W676–W682, 7 2006.
- [182] Václav Brázda, Jan Kolomazník, Jiří Lýsek, Martin Bartas, Miroslav Fojta, Jiří Šťastný, and Jean Louis Mergny. G4Hunter web application: a web server for G-quadruplex prediction. *Bioinformatics*, 35:3493–3495, 9 2019.
- [183] Robert K. Moyzis, James M. Buckingham, L. Scott Cram, Mark Dani, Larry L. Deaven, Michael D. Jones, Julianne Meyne, Robert L. Ratliff, and Jung-Rung Wu. A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 85:6622–6626, 1988.
- [184] Anthony Bugaut and Shankar Balasubramanian. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Research*, 40:4727–4741, 6 2012.
- [185] Enid Yi Ni Lam, Dario Beraldi, David Tannahill, and Shankar Balasubramanian. G-quadruplex structures are stable and detectable in human genomic DNA. *Nature Communications*, 4:1–8, 4 2013.
- [186] Stephen Neidle. Quadruplex Nucleic Acids as Novel Therapeutic Targets. *Journal of Medicinal Chemistry*, 59:5987–6011, 7 2016.
- [187] Giulia Biffi, David Tannahill, Jodi Miller, William J. Howat, and Shankar Balasubramanian. Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues. *PLOS ONE*, 9:e102711, 7 2014.
- [188] Calvin B. Harley, A. Bruce Futcher, and Carol W. Greider. Telomeres shorten during ageing of human fibroblasts. *Nature*, 345:458–460, 1990.
- [189] Nam W. Kim, Mieczyslaw A. Piatyszek, Karen R. Prowse, Calvin B. Harley, Michael D. West, Peter L.C. Ho, Gina M. Coviello, Woodring E. Wright, Scott L. Weinrich, and Jerry W. Shay. Specific association of human telomerase activity with immortal cells and cancer. *Science*, 266:2011–2015, 1994.
- [190] Mieczyslaw A. Piatyszek, Nam W. Kim, Scott L. Weinrich, Keiko Hiyama, Eiso Hiyama, Woodring E. Wright, and Jerry W. Shay. Detection of telomerase activity in human cells

- and tumors by a telomeric repeat amplification protocol (TRAP). *Methods in Cell Science*, 17:1–15, 3 1995.
- [191] Anne De Cian, Gael Cristofari, Patrick Reichenbach, Elsa De Lemos, David Monchaud, Marie Paule Teulade-Fichou, Kazuo Shin-ya, Laurent Lacroix, Joachim Lingner, and Jean Louis Mergny. Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proceedings of the National Academy of Sciences of the United States of America*, 104:17347–17352, 10 2007.
- [192] Aaron L. Moye, Karina C. Porter, Scott B. Cohen, Tram Phan, Katherine G. Zyner, Natsuki Sasaki, George O. Lovrecz, Jennifer L. Beck, and Tracy M. Bryan. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nature Communications*, 6:1–12, 7 2015.
- [193] Kim Ngoc Luu, Anh Tuân Phan, Vitaly Kuryavyi, Laurent Lacroix, and Dinshaw J. Patel. Structure of the human telomere in K⁺ solution: An intramolecular (3 + 1) G-quadruplex scaffold. *Journal of the American Chemical Society*, 128:9963–9970, 8 2006.
- [194] Jixun Dai, Megan Carver, Chandanamali Punchihewa, Roger A. Jones, and Danzhou Yang. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Research*, 35:4927–4940, 8 2007.
- [195] Attila Ambrus, Ding Chen, Jixun Dai, Tiffanie Bialis, Roger A. Jones, and Danzhou Yang. Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Research*, 34:2723–2735, 5 2006.
- [196] Donald M. Miller, Shelia D. Thomas, Ashraful Islam, David Muench, and Kara Sedoris. c-Myc and Cancer Metabolism. *Clinical Cancer Research*, 18:5546, 10 2012.
- [197] Daekyu Sun, Kexiao Guo, Jadrian J. Rusche, and Laurence H. Hurley. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Research*, 33:6070–6080, 10 2005.
- [198] Thomas S. Dexheimer, Daekyu Sun, and Laurence H. Hurley. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream

- of the bcl-2 P1 promoter. *Journal of the American Chemical Society*, 128:5404–5415, 4 2006.
- [199] Susanna Cogoi and Luigi E. Xodo. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Research*, 34: 2536–2549, 5 2006.
- [200] Sarah Rankin, Anthony P. Reszka, Julian Huppert, Mire Zloh, Gary N. Parkinson, Alan K. Todd, Sylvain Ladame, Shankar Balasubramanian, and Stephen Neidle. Putative DNA quadruplex formation within the human c-kit oncogene. *Journal of the American Chemical Society*, 127:10584–10589, 8 2005.
- [201] Leo Kretzner, Anna Scuto, Pamela M. Dino, Claudia M. Kowolik, Jun Wu, Patrick Ventura, Richard Jove, Stephen J. Forman, Yun Yen, and Mark H. Kirschbaum. Combining histone deacetylase inhibitor vorinostat with aurora kinase inhibitors enhances lymphoma cell killing with repression of c-Myc, hTERT, and microRNA levels. *Cancer Research*, 71: 3912–3920, 6 2011.
- [202] Shang Te Danny Hsu, Peter Varnai, Anthony Bugaut, Anthony P. Reszka, Stephen Neidle, and Shankar Balasubramanian. A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. *Journal of the American Chemical Society*, 131:13399–13409, 9 2009.
- [203] Wen Lan Wang, Mary Ellen Healy, Martin Sattler, Shalini Verma, Jeffrey Lin, Gautam Maulik, Charles D. Stiles, James D. Griffin, Bruce E. Johnson, and Ravi Salgia. Growth inhibition and modulation of kinase pathways of small cell lung cancer cell lines by the novel tyrosine kinase inhibitor STI 571. *Oncogene*, 19:3521–3528, 7 2000.
- [204] Seiichi Hirota, Koji Isozaki, Yasuhiro Moriyama, Koji Hashimoto, Toshirou Nishida, Shingo Ishiguro, Kiyoshi Kawano, Masato Hanada, Akihiko Kurata, Masashi Takeda, Ghulam Muhammad Tunio, Yuji Matsuzawa, Yuzuru Kanakura, Yasuhisa Shinomura, and Yukihiko Kitamura. Gain-of-Function Mutations of c-kit in Human Gastrointestinal Stromal Tumors. *Science*, 279:577–580, 1 1998.
- [205] First Generative AI Drug Begins Phase II Trials with Patients | Insilico Medicine. URL https://insilico.com/blog/first_phase2.
- [206] Mallesham Bejugam, Mekala Gunaratnam, Sebastian Müller, Deborah A. Sanders,

- Sven Sewitz, Jonathan A. Fletcher, Stephen Neidle, and Shankar Balasubramanian. Targeting the c-Kit Promoter G-quadruplexes with 6-Substituted Indenoisoquinolines. *ACS Medicinal Chemistry Letters*, 1:306, 10 2010.
- [207] Stefania Mazzini, Raimundo Gargallo, Loana Musso, Francesca De Santis, Anna Aviñó, Leonardo Scaglioni, Ramon Eritja, Massimo Di Nicola, Franco Zunino, Annabella Amatulli, and Sabrina Dallavalle. Stabilization of c-KIT G-Quadruplex DNA Structures by the RNA Polymerase I Inhibitors BMH-21 and BA-41. *International Journal of Molecular Sciences*, 20:4927, 10 2019.
- [208] David Monchaud and Marie Paule Teulade-Fichou. A hitchhiker's guide to G-quadruplex ligands. *Organic and Biomolecular Chemistry*, 6:627–636, 2 2008.
- [209] Tiago Santos, Gilmar F. Salgado, Eurico J. Cabrita, and Carla Cruz. G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals*, 14:769, 8 2021.
- [210] Yu Xia Yao, Bao Hong Xu, and Yan Zhang. CX-3543 Promotes Cell Apoptosis through Downregulation of CCAT1 in Colon Cancer Cells. *BioMed Research International*, 2018: 9701957, 1 2018.
- [211] Denis Drygin, Adam Siddiqui-Jain, Sean O'Brien, Michael Schwaebe, Amy Lin, Josh Bliesath, Caroline B. Ho, Chris Proffitt, Katy Trent, Jeffrey P. Whitten, John K.C. Lim, Daniel Von Hoff, Kenna Anderes, and William G. Rice. Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Research*, 69:7653–7661, 10 2009.
- [212] Yu Huan Wang, Qian Fan Yang, Xiao Lin, Die Chen, Zhi Yin Wang, Bin Chen, Hua Yi Han, Hao Di Chen, Kai Cong Cai, Qian Li, Shu Yang, Ya Lin Tang, and Feng Li. G4LDB 2.2: a database for discovering and studying G-quadruplex and i-Motif ligands. *Nucleic Acids Research*, 50:D150–D160, 1 2022.
- [213] Sara Iachettini, Annamaria Biroccio, and Pasquale Zizza. Therapeutic Use of G4-Ligands in Cancer: State-of-the-Art and Future Perspectives. *Pharmaceuticals*, 17:771, 6 2024.
- [214] Angelika M. Burger, Fangping Dai, Christoph M. Schultes, Anthony P. Reszka, Michael J. Moore, John A. Double, and Stephen Neidle. The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Research*, 65:1489–1496, 2 2005.

- [215] Peisi Li, Dawang Zhou, Yumo Xie, Ze Yuan, Mingzhe Huang, Gaopo Xu, Junfeng Huang, Zhuokai Zhuang, Yanxin Luo, Huichuan Yu, and Xiaolin Wang. Targeting G-quadruplex by TMPyP4 for inhibition of colorectal cancer through cell cycle arrest and boosting anti-tumor immunity. *Cell Death and Disease*, 15:1–10, 11 2024.
- [216] Liu Yi Liu, Tian Zhu Ma, You Liang Zeng, Wenting Liu, and Zong Wan Mao. Structural Basis of Pyridostatin and Its Derivatives Specifically Binding to G-Quadruplexes. *Journal of the American Chemical Society*, 144:11878–11887, 7 2022.
- [217] Anirban Ghosh, Marko Trajkovski, Marie Paule Teulade-Fichou, Valérie Gabelica, and Janez Plavec. Phen-DC3 Induces Refolding of Human Telomeric DNA into a Chair-Type Antiparallel G-Quadruplex through Ligand Intercalation. *Angewandte Chemie*, 61, 10 2022.
- [218] Hikaru Fukuda, Tingting Zou, Satoshi Fujii, Shinobu Sato, Daiki Wakahara, Sen Higashi, Ting Yuan Tseng, Ta Chau Chang, Naomi Yada, Kou Matsuo, Manabu Habu, Kazuhiro Tominaga, Hiroshi Takeuchi, and Shigeori Takenaka. Cyclic anthraquinone derivatives, unique G-quadruplex binders, selectively induce cancer cell apoptosis and inhibit tumor growth. *PNAS Nexus*, 2:211, 7 2023.
- [219] Jun Fujimori, Toshihiro Matsuo, Shoji Shimose, Tadahiko Kubo, Masakazu Ishikawa, Yuji Yasunaga, and Mitsuo Ochi. Antitumor effects of telomerase inhibitor TMPyP4 in osteosarcoma cell lines. *Journal of Orthopaedic Research*, 29:1707–1711, 11 2011.
- [220] Giulia Miglietta, Marco Russo, Renée C. Duardo, and Giovanni Capranico. G-quadruplex binders as cytostatic modulators of innate immune genes in cancer cells. *Nucleic Acids Research*, 49:6673–6686, 7 2021.
- [221] Savvas N. Georgiades, Nurul H. Abd Karim, Kogularamanan Suntharalingam, and Ramon Vilar. Interaction of Metal Complexes with G-Quadruplex DNA. *Angewandte Chemie International Edition*, 49:4020–4034, 6 2010.
- [222] Elisa Palma, Josué Carvalho, Carla Cruz, and António Paulo. Metal-Based G-Quadruplex Binders for Cancer Theranostics. *Pharmaceuticals*, 14:605, 7 2021.
- [223] Tim Kench and Ramon Vilar. Metal complexes as G-quadruplex binders. *Annual Reports in Medicinal Chemistry*, 54:485–515, 1 2020.

- [224] Frank Lovering, Jack Bikker, and Christine Humblet. Escape from flatland: Increasing saturation as an approach to improving clinical success. *Journal of Medicinal Chemistry*, 52:6752–6756, 11 2009.
- [225] Elzbieta Izbicka, Richard T. Wheelhouse, Eric Raymond, Karen K. Davidson, Richard A. Lawrence, Daekyu Sun, Bradford E. Windle, Laurence H. Hurley, and Daniel D. Von Hoff. Effects of Cationic Porphyrins as G-Quadruplex Interactive Agents in Human Tumor Cells. *Cancer Research*, 59:639–644, 2 1999.
- [226] Dong-Fang Shi, Richard T. Wheelhouse, Daekyu Sun, and Laurence H. Hurley. Quadruplex-interactive agents as telomerase inhibitors: Synthesis of porphyrins and structure-activity relationship for the inhibition of telomerase. *Journal of Medicinal Chemistry*, 44:4509–4523, 12 2001.
- [227] Subhrangsu S. Mandal, Umesh Varshney, and Santanu Bhattacharya. Role of the Central Metal Ion and Ligand Charge in the DNA Binding and Modification by Metallosalen Complexes. *Bioconjugate Chemistry*, 8:798–812, 1997.
- [228] Julie E. Reed, Anna Arola Arnal, Stephen Neidle, and Ramón Vilar. Stabilization of G-quadruplex DNA and inhibition of telomerase activity by square-planar nickel(II) complexes. *Journal of the American Chemical Society*, 128:5992–5993, 5 2006.
- [229] Asfa Ali, Mohini Kamra, Soma Roy, K. Muniyappa, and Santanu Bhattacharya. Novel Oligopyrrole Carboxamide based Nickel(II) and Palladium(II) Salens, Their Targeting of Human G-Quadruplex DNA, and Selective Cancer Cell Toxicity. *Chemistry – An Asian Journal*, 11:2542–2554, 9 2016.
- [230] Luisa D’Anna, Simona Rubino, Candida Pipitone, Graziella Serio, Carla Gentile, Antonio Palumbo Piccionello, Francesco Giannici, Giampaolo Barone, and Alessio Terenzi. Salphen metal complexes as potential anticancer agents: interaction profile and selectivity studies toward the three G-quadruplex units in the KIT promoter. *Dalton Transactions*, 52:2966–2975, 3 2023.
- [231] Julie E. Reed, Stephen Neidle, and Ramon Vilar. Stabilisation of human telomeric quadruplex DNA and inhibition of telomerase by a platinum–phenanthroline complex. *Chemical Communications*, pages 4366–4368, 10 2007.
- [232] Vivian Wing Wah Yam and Angela Sin Yee Law. Luminescent d8 metal complexes of

- platinum(II) and gold(III): From photophysics to photofunctional materials and probes. *Coordination Chemistry Reviews*, 414:213298, 7 2020.
- [233] Nurul H. Abd Karim, Oscar Mendoza, Arun Shivalingam, Alexander J. Thompson, Sushobhan Ghosh, Marina K. Kuimova, and Ramon Vilar. Salphen metal complexes as tunable G-quadruplex binders and optical probes. *RSC Advances*, 4:3355–3363, 12 2013.
- [234] Shahratul Ain Mohd Sukri, Lee Yook Heng, and Nurul Huda Abd Karim. Synthesis, Characterization and DNA-Binding Studies of Hydroxyl Functionalized Platinum(II) Salphen Complexes. *Journal of Fluorescence*, 27:1009–1023, 5 2017.
- [235] Samya Banerjee, Miles S. Capper, Guy J. Clarkson, Huaiyi Huang, and Peter J. Sadler. Dual-action platinum(II) Schiff base complexes: Photocytotoxicity and cellular imaging. *Polyhedron*, 172:157–166, 11 2019.
- [236] Keiko Kawauchi, Ryoto Urano, Natsuki Kinoshita, Shin Kuwamoto, Takeru Torii, Yoshiki Hashimoto, Shinya Taniguchi, Mitsuki Tsuruta, and Daisuke Miyoshi. Photosensitizers Based on G-Quadruplex Ligand for Cancer Photodynamic Therapy. *Genes*, 11:1340, 11 2020.
- [237] Marco Deiana, José M. A. Castán, Pierre Josse, Abraha Kahsay, Darío P. Sánchez, Korentin Morice, Natacha Gillet, Ranjitha Ravindranath, Ankit K. Patel, Pallabi Sengupta, Ikenna Obi, Eva Rodriguez-Marquez, Lhoussain Khrouz, Elise Dumont, Laura A. Galán, Magali Allain, Bright Walker, Hyun S. Ahn, Olivier Maury, Philippe Blanchard, Tangui Le Bahers, Daniel Öhlund, Jonas von Hofsten, Cyrille Monnereau, Clément Cabanetos, and Nasim Sabouri. A new G-quadruplex-specific photosensitizer inducing genome instability in cancer cells by triggering oxidative DNA damage and impeding replication fork progression. *Nucleic Acids Research*, 51:6264–6285, 7 2023.
- [238] Marco Caterino, Federica D’Aria, Andrey V. Kustov, Dmitrii V. Belykh, Irina S. Khudyaeva, Olga M. Starseva, Dmitriy B. Berezin, Yana I. Pylina, Tatiana Usacheva, Jussara Amato, and Concetta Giancola. Selective binding of a bioactive porphyrin-based photosensitizer to the G-quadruplex from the KRAS oncogene promoter. *International Journal of Biological Macromolecules*, 145:244–251, 2 2020.
- [239] Enrico Cadoni, Alex Manicardi, Mathieu Fossépré, Kaat Heirwegh, Mathieu Surin,

- and Annemieke Madder. Teaching photosensitizers a new trick: red light-triggered G-quadruplex alkylation by ligand co-localization. *Chemical Communications*, 57: 1010–1013, 2 2021.
- [240] Junya Muramoto and Takashi Sakamoto. Tripodal Quinone-Cyanine G-Quadruplex Ligands as Novel Photosensitizers on Photoinduced Cancer Cell Death. *Molecules*, 29: 5094, 11 2024.
- [241] Ariadna Gil-Martínez, Adrián Hernández, Cristina Galiana-Roselló, Sònia López-Molina, Javier Ortiz, Ángela Sastre-Santos, Enrique García-España, and Jorge González-García. Development and application of metallo-phthalocyanines as potent G-quadruplex DNA binders and photosensitizers. *Journal of Biological Inorganic Chemistry*, 28:495–507, 8 2023.
- [242] Molly Bartlett, Joseph Burke, Petr Sherin, Marina K. Kuimova, Mauricio Barahona, and Ramon Vilar. Platinum(II)-Salphen Complexes as DNA Binders and Photosensitisers. *Chemistry – A European Journal*, 30:e202402465, 11 2024.
- [243] Carmen L. Ruehl, Aaron H.M. Lim, Timothy Kench, David J. Mann, and Ramon Vilar. An Octahedral Cobalt(III) Complex with Axial NH₃ Ligands that Templates and Selectively Stabilises G-quadruplex DNA. *Chemistry – A European Journal*, 25:9691–9700, 7 2019.
- [244] Chatna Rajput, Ramune Rutkaite, Linda Swanson, Ihtshamul Haq, and Jim A. Thomas. Dinuclear Monointercalating RuII Complexes That Display High Affinity Binding to Duplex and Quadruplex DNA. *Chemistry – A European Journal*, 12:4611–4619, 6 2006.
- [245] Justin Weynand, Aurélie Diman, Michaël Abraham, Lionel Marcélis, Hélène Jamet, Anabelle Decottignies, Jérôme Dejeu, Eric Defrancq, and Benjamin Elias. Towards the Development of Photo-Reactive Ruthenium(II) Complexes Targeting Telomeric G-Quadruplex DNA. *Chemistry – A European Journal*, 24:19216–19227, 12 2018.
- [246] Tom Wilson, Paulo J. Costa, Vítor Félix, Mike P. Williamson, and Jim A. Thomas. Structural studies on dinuclear ruthenium(II) complexes that bind diastereoselectively to an antiparallel folded human telomere sequence. *Journal of Medicinal Chemistry*, 56: 8674–8683, 11 2013.
- [247] Zhen Feng Chen, Qi Pin Qin, Jiao Lan Qin, Jie Zhou, Yu Lan Li, Nan Li, Yan Cheng Liu, and Hong Liang. Water-Soluble Ruthenium(II) Complexes with Chiral

- 4-(2,3-Dihydroxypropyl)-formamide Oxoaporphine (FOA): In Vitro and in Vivo Anticancer Activity by Stabilization of G-Quadruplex DNA, Inhibition of Telomerase Activity, and Induction of Tumor Cell Apoptosis. *Journal of Medicinal Chemistry*, 58:4771–4789, 6 2015.
- [248] Qiong Wu, Kangdi Zheng, Siyan Liao, Yang Ding, Yangqiu Li, and Wenjie Mei. Arene Ruthenium(II) Complexes as Low-Toxicity Inhibitor against the Proliferation, Migration, and Invasion of MDA-MB-231 Cells through Binding and Stabilizing c-myc G-Quadruplex DNA. *Organometallics*, 35:317–326, 1 2016.
- [249] Kane T. McQuaid, Shuntaro Takahashi, Lena Baumgaertner, David J. Cardin, Neil G. Paterson, James P. Hall, Naoki Sugimoto, and Christine J. Cardin. Ruthenium Polypyridyl Complex Bound to a Unimolecular Chair-Form G-Quadruplex. *Journal of the American Chemical Society*, 144:5956–5964, 4 2022.
- [250] Lige Ren, Anming Zhang, Jing Huang, Ping Wang, Xiaocheng Weng, Lixia Zhang, Feng Liang, Zheng Tan, and Xiang Zhou. Quaternary Ammonium Zinc Phthalocyanine: Inhibiting Telomerase by Stabilizing G quadruplexes and Inducing G-Quadruplex Structure Transition and Formation. *ChemBioChem*, 8:775–780, 5 2007.
- [251] Lixia Zhang, Jing Huang, Lige Ren, Minghui Bai, Lin Wu, Baoping Zhai, and Xiang Zhou. Synthesis and evaluation of cationic phthalocyanine derivatives as potential inhibitors of telomerase. *Bioorganic and Medicinal Chemistry*, 16:303–312, 1 2008.
- [252] Hélène Bertrand, David Monchaud, Anne De Cian, Régis Guillot, Jean Louis Mergny, and Marie Paule Teulade-Fichou. The importance of metal geometry in the recognition of G-quadruplex-DNA by metal-terpyridine complexes. *Organic and Biomolecular Chemistry*, 5:2555–2559, 8 2007.
- [253] Elodie Morel, Claire Beauvineau, Delphine Naud-Martin, Corinne Landras-Guetta, Daniela Verga, Deepanjan Ghosh, Sylvain Achelle, Florence Mahuteau-Betzer, Sophie Bombard, and Marie Paule Teulade-Fichou. Selectivity of Terpyridine Platinum Anticancer Drugs for G-quadruplex DNA. *Molecules*, 24:404, 1 2019.
- [254] Julie E. Reed, Andrew J.P. White, Stephen Neidle, and Ramón Vilar. Effect of metal coordination on the interaction of substituted phenanthroline and pyridine ligands with quadruplex DNA. *Dalton Transactions*, pages 2558–2568, 3 2009.

- [255] Roxanne Kieltyka, Johans Fakhoury, Nicolas Moitessier, and Hanadi F. Sleiman. Platinum Phenanthroimidazole Complexes as G-Quadruplex DNA Selective Binders. *Chemistry – A European Journal*, 14:1145–1154, 1 2008.
- [256] Ma Dik-Lung, Chi Ming Che, and Siu Cheong Yan. Platinum(II) complexes with dipyridophenazine ligands as human telomerase inhibitors and luminescent probes for G-quadruplex DNA. *Journal of the American Chemical Society*, 131:1835–1846, 2 2009.
- [257] Aurane Froux, Luisa D’Anna, Aurianne Rainot, Camille Neybecker, Angelo Spinello, Riccardo Bonsignore, Raphaël Rouget, Guillaume Harlé, Alessio Terenzi, Antonio Monari, Stéphanie Grandemange, and Giampaolo Barone. Metal centers and aromatic moieties in Schiff base complexes: impact on G-quadruplex stabilization and oncogene downregulation. *Inorganic Chemistry Frontiers*, 11:5725–5740, 8 2024.
- [258] Keli Kuang, Chunyan Li, Fatlinda Maksut, Deepanjan Ghosh, Robin Vinck, Maolin Wang, Joël Poupon, Run Xiang, Wen Li, Fei Li, Zhu Wang, Junrong Du, Marie Paule Teulade-Fichou, Gilles Gasser, Sophie Bombard, and Tao Jia. A G-quadruplex-binding platinum complex induces cancer mitochondrial dysfunction through dual-targeting mitochondrial and nuclear G4 enriched genome. *Journal of Biomedical Science*, 31:1–21, 12 2024.
- [259] Donn Liew, Zi Way Lim, and Ee Hou Yong. Machine learning-based prediction of DNA G-quadruplex folding topology with G4ShapePredictor. *Scientific Reports*, 14:1–13, 10 2024.
- [260] H. Busra Cagirici, Hikmet Budak, and Taner Z. Sen. G4Boost: a machine learning-based tool for quadruplex identification and stability prediction. *BMC Bioinformatics*, 23:1–18, 12 2022.
- [261] Mira Barshai, Barak Engel, Idan Haim, and Yaron Orenstein. G4mismatch: Deep neural networks to predict G-quadruplex propensity based on G4-seq data. *PLOS Computational Biology*, 19:e1010948, 3 2023.
- [262] Aleksandr B. Sahakyan, Vicki S. Chambers, Giovanni Marsico, Tobias Santner, Marco Di Antonio, and Shankar Balasubramanian. Machine learning model for sequence-driven DNA G-quadruplex formation. *Scientific Reports*, 7:1–11, 11 2017.
- [263] Anna Korsakova and Anh Tuân Phan. Prediction of G4 formation in live cells with

- epigenetic data: a deep learning approach. *NAR Genomics and Bioinformatics*, 5, 7 2023.
- [264] Vincent Rocher, Matthieu Genais, Elissar Nassereddine, and Raphael Mourad. DeepG4: A deep learning approach to predict cell-type specific active G-quadruplex regions. *PLOS Computational Biology*, 17:e1009308, 8 2021.
- [265] Luisa D', Ugo Perricone, Giada De Simone, Antonio Monari, Giampaolo Barone, and Alessio Terenzi. AGAPE (computAtional G-quadruplex Affinitiy PrEdiction): The first AI In-silico workflow for G-quadruplex binding affinity prediction. *bioRxiv*, 11 2024.
- [266] Piermaria Pinter, Johannes Soellner, Thomas Strassner, P Pinter, J Soellner, and T Strassner. Metallophilic Interactions in Bimetallic Cyclometalated Platinum(II) N-Heterocyclic Carbene Complexes. *European Journal of Inorganic Chemistry*, 2021: 3104–3107, 8 2021.
- [267] Valentina Iannace, Ferran Sabaté, Molly Bartlett, Jessica Berrones Reyes, Ariadna Lázaro, Alessia Fantoni, Ramon Vilar, Laura Rodríguez, and Antonella Dalla Cort. Syntheses, Characterization, and DNA Binding Studies of a Series of Copper(II), Nickel(II), Platinum(II) and Zinc(II) Complexes Derived from Schiff Base Ligands. *European Journal of Inorganic Chemistry*, 26:e202300144, 10 2023.
- [268] Chandan L. Barhate, Andrew F. Donnell, Merrill Davies, Ling Li, Yong Zhang, Fukang Yang, Regina Black, Greg Zipp, Yingru Zhang, Cullen L. Cavallaro, E. Scott Priestley, and Harold N. Weller. Microscale purification in support of high-throughput medicinal chemistry. *Chemical Communications*, 57:11037–11040, 10 2021.
- [269] Mathis Andreas Muth, Miguel Carrasco-Orozco, and Mukundan Thelakkat. Liquid-Crystalline Perylene Diester Polymers with Tunable Charge-Carrier Mobility. *Advanced Functional Materials*, 21:4510–4518, 12 2011.
- [270] Stephen C. Mitchell, Paul Carmichael, and Rosemary Waring. Aminophenols. *Kirk-Othmer Encyclopedia of Chemical Technology*, 6 2003.
- [271] Yan Tan, Jia Jia Lang, Li Mei Chen, Mei Lun Tang, Zhi Xing Yu, Peng Bing Mi, and Xing Zheng. The Sonogashira Carbonylation Reaction between Aromatic Bromide and Alkyne with N-Formylsaccharin as a CO Source. *ChemistrySelect*, 7:e202201063, 6 2022.

- [272] Alemany Ribes M, García Díaz M, Acedo P, Agut M, and Nonell S. Why Not Introducing the Third Dimension in Photodynamic Therapy Research? *Journal of Analytical and Bioanalytical Techniques*, 2013.
- [273] Ellen F. Reid, Vernon C. Cook, David J.D. Wilson, and Conor F. Hogan. Facile Tuning of Luminescent Platinum(II) Schiff Base Complexes from Yellow to Near-Infrared: Photophysics, Electrochemistry, Electrochemiluminescence and Theoretical Calculations. *Chemistry – A European Journal*, 19:15907–15917, 11 2013.
- [274] Chi Ming Che, Chi Chung Kwok, Siu Wai Lai, Andreas F. Rausch, Walter J. Finkenzeller, Nianyong Zhu, and Hartmut Yersin. Photophysical Properties and OLED Applications of Phosphorescent Platinum(II) Schiff Base Complexes. *Chemistry – A European Journal*, 16:233–247, 1 2010.
- [275] Chi Ming Che, Siu Chung Chan, Hai Feng Xiang, Michael C.W. Chan, Yu Liu, and Yue Wang. Tetradentate Schiff base platinum(II) complexes as new class of phosphorescent materials for high-efficiency and white-light electroluminescent devices. *Chemical Communications*, 4:1484–1485, 6 2004.
- [276] Till Jahnke, Hendrik Sann, Tilo Havermeier, Katharina Kreidi, Christian Stuck, Moritz Meckel, Markus Schöffler, Nadine Neumann, Robert Wallauer, Stefan Voss, Achim Czasch, Ottmar Jagutzki, Abdollah Malakzadeh, Feras Afaneh, Thorsten Weber, Horst Schmidt-Böcking, and Reinhard Dörner. Ultrafast energy transfer between water molecules. *Nature Physics*, 6:139–142, 1 2010.
- [277] Darl H. McDaniel and Herbert C. Brown. An Extended Table of Hammett Substituent Constants Based on the Ionization of Substituted Benzoic Acids. *Journal of Organic Chemistry*, 23:420–427, 3 1958.
- [278] Guilian Luchini and Robert S. Paton. Bottom-Up Atomistic Descriptions of Top-Down Macroscopic Measurements: Computational Benchmarks for Hammett Electronic Parameters. *ACS Physical Chemistry Au*, 4:259–267, 5 2024.
- [279] Chandrabhan Verma, Akram Alfantazi, M. A. Quraishi, and Kyong Yop Rhee. Significance of Hammett and Taft substituent constants on bonding potential of organic corrosion inhibitors: Tailoring of reactivity and performance. *Coordination Chemistry Reviews*, 495: 215385, 11 2023.

- [280] Eun Joung Choi and Seung Bum Park. Unique photophysical properties of 9-styryl-1,2-dihydropyrrolo[3,4- β]indolizin-3-one and its efficient synthesis via direct C–H activation. *Organic and Biomolecular Chemistry*, 13:5202–5208, 4 2015.
- [281] Manjusha Verma, Aneese F. Chaudhry, and Christoph J. Fahrni. Predicting the photoinduced electron transfer thermodynamics in polyfluorinated 1,3,5-triarylpyrazolines based on multiple linear free energy relationships. *Organic and Biomolecular Chemistry*, 7:1536–1546, 2009.
- [282] Nina Sadlej-Sosnowska and Michal Kijak. Excited state substituent constants: To Hammett or not? *Structural Chemistry*, 23:359–365, 4 2012.
- [283] Tiago Entradas, Sarah Waldron, and Martin Volk. The detection sensitivity of commonly used singlet oxygen probes in aqueous environments. *Journal of Photochemistry and Photobiology B: Biology*, 204:111787, 3 2020.
- [284] Lydia M. Henderson and J. Brian Chappell. Dihydrorhodamine 123: a fluorescent probe for superoxide generation? *European Journal of Biochemistry*, 217:973–980, 1993.
- [285] Ayman A. Abdel-Shafi, Paul D. Beer, Roger J. Mortimer, and Francis Wilkinson. Photosensitized Generation of Singlet Oxygen from Vinyl Linked Benzo-Crown-Ether-Bipyridyl Ruthenium(II) Complexes. *Journal of Physical Chemistry A*, 104:192 – 202, 1 2000.
- [286] Yujie Sun, Sibrina N. Collins, Lauren E. Joyce, and Claudia Turro. Unusual photophysical properties of a ruthenium(II) complex related to [Ru(bpy)₂(dppz)]²⁺. *Inorganic Chemistry*, 49:4257–4262, 5 2010.
- [287] Pavel Müller and Klaus Brettel. [Ru(bpy)₃]²⁺ as a reference in transient absorption spectroscopy: differential absorption coefficients for formation of the long-lived 3MLCT excited state. *Photochemical and Photobiological Sciences*, 11:632–636, 3 2012.
- [288] Chao Liang, Jiaen Xie, Shuangling Luo, Can Huang, Qianling Zhang, Huaiyi Huang, and Pingyu Zhang. A highly potent ruthenium(II)-sonosensitizer and sonocatalyst for in vivo sonotherapy. *Nature Communications*, 12:5001, 12 2021.
- [289] Jessica Berrones Reyes, Peter S. Sherin, Amrita Sarkar, Marina K. Kuimova, and Ramon Vilar. Platinum(II)-Based Optical Probes for Imaging Quadruplex DNA Structures via

- Phosphorescence Lifetime Imaging Microscopy. *Angewandte Chemie*, 135:e202310402, 10 2023.
- [290] Elena Butovskaya, Brahim Heddi, Blaž Bakalar, Sara N. Richter, and Anh Tuấn Phan. Major G-Quadruplex Form of HIV-1 LTR Reveals a (3 + 1) Folding Topology Containing a Stem-Loop. *Journal of the American Chemical Society*, 140:13654–13662, 10 2018.
- [291] Eric Largy, Florian Hamon, and Marie Paule Teulade-Fichou. Development of a high-throughput G4-FID assay for screening and evaluation of small molecules binding quadruplex nucleic acid structures. *Analytical and Bioanalytical Chemistry*, 400: 3419–3427, 7 2011.
- [292] Garrett M. Morris, Huey Ruth, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30: 2785–2791, 12 2009.
- [293] Clarisse G. Ricci and Paulo A. Netz. Docking studies on DNA-ligand interactions: building and application of a protocol to identify the binding mode. *Journal of Chemical Information and Modeling*, 49:1925–1935, 8 2009.
- [294] Xing Du, Yi Li, Yuan Ling Xia, Shi Meng Ai, Jing Liang, Peng Sang, Xing Lai Ji, and Shu Qun Liu. Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods. *International Journal of Molecular Sciences*, 17:144, 1 2016.
- [295] Tatu Pääsjarvi and Antti Poso. Binding Affinity via Docking: Fact and Fiction. *Molecules*, 23:1899, 2018.
- [296] Sérgio F. Sousa, António J. M. Ribeiro, João T. S. Coimbra, Rui P. P. Neves, Sílvia Martins, Hari N. Moorthy, Pedro A. Fernandes, and Maria J. Ramos. Protein-Ligand Docking in the New Millennium – A Retrospective of 10 Years in the Field. *Current Medicinal Chemistry*, 20:2296–2314, 5 2013.
- [297] Jixun Dai, Megan Carver, Laurence H. Hurley, and Danzhou Yang. Solution structure of a 2:1 quindoline-c-MYC G-quadruplex: Insights into G-quadruplex-interactive small molecule drug design. *Journal of the American Chemical Society*, 133:17673–17680, 11 2011.

- [298] Oleg Trott and Arthur J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31:455–461, 1 2010.
- [299] William A. LaMarr, Lu Yu, Kyriacos C. Nicolaou, and Peter C. Dedon. Supercoiling affects the accessibility of glutathione to DNA-bound molecules: Positive supercoiling inhibits calicheamicin-induced DNA damage. *Proceedings of the National Academy of Sciences of the United States of America*, 95:102–107, 1 1998.
- [300] Kogularamanan Suntharalingam, Douglas J. Hunt, Alexandra A. Duarte, Andrew J.P. White, David J. Mann, and Ramon Vilar. A tri-copper(II) complex displaying DNA-cleaving properties and antiproliferative activity against cancer cells. *Chemistry*, 18:15133–15141, 11 2012.
- [301] Josué Carvalho, João A. Queiroz, and Carla Cruz. Circular dichroism of G-Quadruplex: A laboratory experiment for the study of topology and ligand binding. *Journal of Chemical Education*, 94:1547–1551, 10 2017.
- [302] Raveendra I. Mathad, Emmanuel Hatzakis, Jixun Dai, and Danzhou Yang. c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Research*, 39:9023, 11 2011.
- [303] Valeria Libera, Francesca Ripanti, Caterina Petrillo, Francesco Sacchetti, Javier Ramos-Soriano, Maria Carmen Galan, Giorgio Schirò, Alessandro Paciaroni, and Lucia Comez. Stability of Human Telomeric G-Quadruplexes Complexed with Photosensitive Ligands and Irradiated with Visible Light. *International Journal of Molecular Sciences*, 24: 9090, 5 2023.
- [304] Terry L Riss, Richard A Moravec, Andrew L Niles, Sarah Duellman, Hélène A Benink, Tracy J Worzella, and Lisa Minor. Cell Viability Assays. *Assay Guidance Manual*, 7 2016.
- [305] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, Jean Yves Tinevez, Daniel James White, Volker Hartenstein, Kevin Eliceiri, Pavel Tomancak, and Albert Cardona. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9:676–682, 6 2012.

- [306] Olivier Gavet and Jonathon Pines. Activation of cyclin B1-Cdk1 synchronizes events in the nucleus and the cytoplasm at mitosis. *Journal of Cell Biology*, 189:247–259, 4 2010.
- [307] Jacek Zielonka, Joy Joseph, Adam Sikora, Micael Hardy, Olivier Ouari, Jeannette Vasquez-Vivar, Gang Cheng, Marcos Lopez, and Balaraman Kalyanaraman. Mitochondria-Targeted Triphenylphosphonium-Based Compounds: Syntheses, Mechanisms of Action, and Therapeutic and Diagnostic Applications. *Chemical Reviews*, 117:10043, 8 2017.
- [308] Hannah Crawford, Maria Dimitriadi, Jatinder Bassin, Michael T. Cook, Thais Fedatto Abelha, and Jesus Calvo-Castro. Mitochondrial Targeting and Imaging with Small Organic Conjugated Fluorophores: A Review. *Chemistry – A European Journal*, 28:e202202366, 12 2022.
- [309] Sadhan Jana, Jialin Shang, Jun Young Hong, Michael K. Fenwick, Rishi Puri, Xuan Lu, Ari M. Melnick, Meng Li, and Hening Lin. A Mitochondria-Targeting SIRT3 Inhibitor with Activity against Diffuse Large B Cell Lymphoma. *Journal of Medicinal Chemistry*, 67: 15428–15437, 9 2024.
- [310] Gantumur Battogtokh, Yeon Su Choi, Dong Seop Kang, Sang Jun Park, Min Suk Shim, Kang Moo Huh, Yong Yeon Cho, Joo Young Lee, Hye Suk Lee, and Han Chang Kang. Mitochondria-targeting drug conjugates for cytotoxic, anti-oxidizing and sensing purposes: current strategies and future perspectives. *Acta Pharmaceutica Sinica B*, 8: 862–880, 10 2018.
- [311] Michael D. Feit, Joseph A. Fleck, and Arno Steiger. Solution of the Schrödinger equation by a spectral method. *Journal of Computational Physics*, 47:412–433, 9 1982.
- [312] Wayne Kohn, Axel D. Becke, and Robert G. Parr. Density Functional Theory of Electronic Structure. *Journal of Physical Chemistry*, 100:12974–12980, 8 1996.
- [313] Wayne Kohn and Lu J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133–A1138, 11 1965.
- [314] Yan Peng, Hui Zhong, Zhen Feng Chen, Yan Cheng Liu, Guo Hai Zhang, Qi Pin Qin, and Hong Liang. A planar Schiff base platinum(II) complex: crystal structure, cytotoxicity and interaction with DNA. *Chemical and Pharmaceutical Bulletin*, 62:221–228, 2014.

- [315] Mahboube Eslami Moghadam, Maryam Hasanzadeh Esfahani, Mahdi Behzad, Samaneh Zolghadri, Nadali Ramezani, and Yasaman Azadi. New platinum (II) complexes based on schiff bases: synthesis, specification, X-ray structure, ADMET, DFT, molecular docking, and anticancer activity against breast cancer. *Journal of Biological Inorganic Chemistry*, 28:519–529, 8 2023.
- [316] Erich Runge and E. K.U. Gross. Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters*, 52:997, 3 1984.
- [317] Miguel A.L. Marques and Eberhard K.U. Gross. Time-dependent density functional theory. *Annual Review of Physical Chemistry*, 55:427–455, 2004.
- [318] Danuta Frackowiak. The Jablonski diagram. *Journal of Photochemistry and Photobiology B: Biology*, 2:399, 11 1988.
- [319] Patrick Schober and Lothar A. Schwarte. Correlation coefficients: Appropriate use and interpretation. *Anesthesia and Analgesia*, 126:1763–1768, 5 2018.
- [320] Frédéric Dumur, Emmanuel Contal, Guillaume Wantz, and Didier Gigmes. Photoluminescence of zinc complexes: Easily tunable optical properties by variation of the bridge between the Imido groups of schiff base ligands. *European Journal of Inorganic Chemistry*, 2014:4186–4198, 9 2014.
- [321] Ku Hsien Chang, Chiung Cheng Huang, Yi Hung Liu, Ya Hui Hu, Pi Tai Chou, and Ying Chih Lin. Synthesis of photo-luminescent Zn(II) Schiff base complexes and its derivative containing Pd(II) moiety. *Dalton Transactions*, pages 1731–1738, 5 2004.
- [322] Siu Wai Lai, Michael Chi Wang Chan, Tsz Chun Cheung, Shie Ming Peng, and Chi Ming Che. Probing d8-d8 interactions in luminescent mono- and binuclear cyclometalated platinum(II) complexes of 6-phenyl-2,2'-bipyridines. *Inorganic Chemistry*, 38:4046–4055, 1999.
- [323] Rebecca J. Salthouse, Amit Sil, Louise F. Gildea, Dmitry S. Yufit, and J. A. Gareth Williams. Platinum(II) Complexes of Nonsymmetrical NCN-Coordinating Ligands: Unimolecular and Excimeric Luminescence Properties and Comparison with Symmetrical Analogues. *Inorganic Chemistry*, 62:12356–12371, 8 2023.
- [324] Baozhu Yang, Shuang Huang, Jing Zhong, and Hongxing Zhang. A comparison of excited

- state properties between two different N-heterocyclic platinum(II) complexes. *Journal of Molecular Structure*, 1097:23–28, 10 2015. ISSN 0022-2860.
- [325] Zixing Wang, Eric Turner, Vanessa Mahoney, Sijesh Madakuni, Thomas Groy, and Jian Li. Facile synthesis and characterization of phosphorescent Pt(NCN)X complexes. *Inorganic Chemistry*, 49:11276–11286, 12 2010.
- [326] Toshiaki Kayano, Satoshi Takayasu, Kohei Sato, and Kenji Shinozaki. Luminescence Color Tuning of PtII Complexes and a Kinetic Study of Trimer Formation in the Photoexcited State. *Chemistry - A European Journal*, 20:16583–16589, 12 2014.
- [327] Lukas Turcani, Enrico Berardo, and Kim E. Jelfs. stk: A python toolkit for supramolecular assembly. *Journal of Computational Chemistry*, 39:1931–1942, 9 2018.
- [328] Lukas Turcani, Andrew Tarzia, Filip T. Szczypiński, and Kim E. Jelfs. stk: An extendable Python framework for automated molecular and supramolecular structure assembly and discovery. *Journal of Chemical Physics*, 154:214102, 6 2021.
- [329] David Weininger. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *Journal of Chemical Information and Computer Sciences*, 28:31–36, 2 1988.
- [330] Noel M. O’Boyle, Michael Banck, Craig A. James, Chris Morley, Tim Vandermeersch, and Geoffrey R. Hutchison. Open Babel: An Open chemical toolbox. *Journal of Cheminformatics*, 3:1–14, 10 2011.
- [331] Julian D. Gale and Andrew L. Rohl. The General Utility Lattice Program (GULP). *Molecular Simulation*, 29:291–341, 1 2003.
- [332] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation*, 15:1652–1671, 3 2019.
- [333] Justus Johann Lange, Andrea Anelli, Jochem Alsenz, Martin Kuentz, Patrick J. O’Dwyer, Wiebke Saal, Nicole Wyttenbach, and Brendan T. Griffin. Comparative Analysis of Chemical Descriptors by Machine Learning Reveals Atomistic Insights into Solute-Lipid Interactions. *Molecular Pharmaceutics*, 21:3343–3355, 7 2024.

- [334] Ana E. Comesana, Tyler T. Huntington, Corinne D. Scown, Kyle E. Niemeyer, and Vi H. Rapp. A systematic method for selecting molecular descriptors as features when training models for predicting physiochemical properties. *Fuel*, 321:123836, 8 2022.
- [335] Yuankai Zhao, Roger J. Mulder, Shadi Houshyar, and Tu C. Le. A review on the application of molecular descriptors and machine learning in polymer design. *Polymer Chemistry*, 14:3325–3346, 7 2023.
- [336] Daniel S. Wigh, Jonathan M. Goodman, and Alexei A. Lapkin. A review of molecular representation in the age of machine learning. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12:e1603, 9 2022.
- [337] Hiroto Moriaki, Yu Shi Tian, Norihito Kawashita, and Tatsuya Takagi. Mordred: A molecular descriptor calculator. *Journal of Cheminformatics*, 10:1–14, 2 2018.
- [338] Alice Capecchi, Daniel Probst, and Jean Louis Reymond. One molecular fingerprint to rule them all: Drugs, biomolecules, and the metabolome. *Journal of Cheminformatics*, 12:1–15, 6 2020.
- [339] Jonathan M. Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M. Donghia, Craig R. MacNair, Shawn French, Lindsey A. Carfrae, Zohar Bloom-Ackerman, Victoria M. Tran, Anush Chiappino-Pepe, Ahmed H. Badran, Ian W. Andrews, Emma J. Chory, George M. Church, Eric D. Brown, Tommi S. Jaakkola, Regina Barzilay, and James J. Collins. A Deep Learning Approach to Antibiotic Discovery. *Cell*, 180:688–702.e13, 2 2020.
- [340] Natalia Łapińska, Adam Paclawski, Jakub Szlęk, and Aleksander Mendyk. Integrated QSAR Models for Prediction of Serotonergic Activity: Machine Learning Unveiling Activity and Selectivity Patterns of Molecular Descriptors. *Pharmaceutics*, 16:349, 3 2024.
- [341] Marcin Cieślak, Tomasz Danel, Olga Krzysztyńska-Kuleta, and Justyna Kalinowska-Tłuścik. Machine learning accelerates pharmacophore-based virtual screening of MAO inhibitors. *Scientific Reports*, 14:1–15, 4 2024.
- [342] Keerthana Jaganathan, Mobeem Ur Rehman, Hilal Tayara, and Kil To Chong. XML-CIMT: Explainable Machine Learning (XML) Model for Predicting Chemical-Induced Mitochondrial Toxicity. *International Journal of Molecular Sciences*, 23:15655, 12 2022.

- [343] Kai Leong, Goh Kai Leong, and Lu Yunpeng Asst Xia Kelin Wee Jun Jie. XGBoost, mordred and RDKit for the prediction of glass transition temperature of polymers. *Nanyang Technological University*, pages 2020–2041, 2025.
- [344] Dimitri Panagopoulos Abrahamsson, June Soo Park, Randolph R. Singh, Marina Sirota, and Tracey J. Woodruff. Applications of Machine Learning to in silico Quantification of Chemicals without Analytical Standards. *Journal of Chemical Information and Modeling*, 60:2718, 6 2020.
- [345] Muhammad Zafar Irshad Khan, Jia Nan Ren, Cheng Cao, Hong Yu Xiang Ye, Hao Wang, Ya Min Guo, Jin Rong Yang, and Jian Zhong Chen. Comprehensive hepatotoxicity prediction: ensemble model integrating machine learning and deep learning. *Frontiers in Pharmacology*, 15:1441587, 2024.
- [346] Myeonghun Lee and Kyoungmin Min. A Comparative Study of the Performance for Predicting Biodegradability Classification: The Quantitative Structure–Activity Relationship Model vs the Graph Convolutional Network. *ACS Omega*, 7:3649–3655, 2 2022.
- [347] Hiroto Moriaki, Yu Shi Tian, Norihito Kawashita, and Tatsuya Takagi. Mordred documentation, 2018. mordred-descriptor.github.io/documentation/master/descriptors.html.
- [348] Hongyi Zhou and Jeffrey Skolnick. Utility of the Morgan Fingerprint in Structure-Based Virtual Ligand Screening. *The Journal of Physical Chemistry. B*, 128:5363, 6 2024.
- [349] Fabio Urbina, Ana C. Puhl, and Sean Ekins. Recent Advances in Drug Repurposing Using Machine Learning. *Current Opinion in Chemical Biology*, 65:74, 12 2021.
- [350] Bulat Zagidullin, Ziyang Wang, Yuanfang Guan, Esa Pitkänen, and Jing Tang. Comparative analysis of molecular fingerprints in prediction of drug combination effects. *Briefings in Bioinformatics*, 22:1–15, 11 2021.
- [351] Fernanda I. Saldívar-González, Victor D. Aldas-Bulos, Jose L. Medina-Franco, and Fabien Plisson. Natural product drug discovery in the artificial intelligence era. *Chemical Science*, 13:1526–1546, 2 2022.
- [352] Abbas Kabir and Aaron Muth. Polypharmacology: The science of multi-targeting

- molecules. *Pharmacological Research*, 176:106055, 2 2022.
- [353] Roshan A. Patel and Michael A. Webb. Data-Driven Design of Polymer-Based Biomaterials: High-throughput Simulation, Experimentation, and Machine Learning. *ACS Applied Bio Materials*, 7:510–527, 2 2024.
- [354] Philippe Schwaller, Daniel Probst, Alain C. Vaucher, Vishnu H. Nair, David Kreutter, Teodoro Laino, and Jean Louis Reymond. Mapping the space of chemical reactions using attention-based neural networks. *Nature Machine Intelligence*, 3:144–152, 1 2021.
- [355] José L. Medina-Franco, Ana L. Chávez-Hernández, Edgar López-López, and Fernanda I. Saldivar-González. Chemical Multiverse: An Expanded View of Chemical Space. *Molecular Informatics*, 41:2200116, 11 2022.
- [356] Weikuan Jia, Meili Sun, Jian Lian, and Sujuan Hou. Feature dimensionality reduction: a review. *Complex and Intelligent Systems*, 8:2663–2693, 1 2022.
- [357] Laurens Van Der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- [358] T. Tony Cai and Rong Ma. Theoretical Foundations of t-SNE for Visualizing High-Dimensional Clustered Data. *Journal of Machine Learning Research*, 23:1–54, 5 2021.
- [359] Kristina P. Sinaga and Miin Shen Yang. Unsupervised K-means clustering algorithm. *IEEE Access*, 8:80716–80727, 2020.
- [360] Abiodun M. Ikotun, Absalom E. Ezugwu, Laith Abualigah, Belal Abuhaija, and Jia Heming. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622:178–210, 4 2023.
- [361] Andrzej Maćkiewicz and Waldemar Ratajczak. Principal components analysis (PCA). *Computers and Geosciences*, 19:303–342, 3 1993.
- [362] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 374:20150202, 4 2016.
- [363] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. UMAP: Uniform

- Manifold Approximation and Projection. *Journal of Open Source Software*, 3:861, 9 2018.
- [364] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, David Cournapeau, Matthieu Brucher, and Matthieu Perrot. Scikit-learn: Machine Learning in Python. Pedregosa, Varoquaux, Gramfort et al. *Journal of Machine Learning Research*, 12: 2825–2830, 2011.
- [365] Dastan Maulud, Dastan Maulud, and Adnan M. Abdulazeez. A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1:140–147, 12 2020.
- [366] Trevor Hastie. Ridge Regularization: An Essential Concept in Data Science. *Technometrics*, 62:426, 10 2020.
- [367] R. Muthukrishnan and Rohini Rajagopalan. LASSO: A feature selection technique in predictive modeling for machine learning. *IEEE International Conference on Advances in Computer Applications*, pages 18–20, 3 2017.
- [368] Laura Freijeiro-González, Manuel Febrero-Bande, and Wenceslao González-Manteiga. A Critical Review of LASSO and Its Derivatives for Variable Selection Under Dependence Among Covariates. *International Statistical Review*, 90:118–145, 4 2022.
- [369] Ash Bullement, A K Ehsanes Saleh, Mohammad Arashi, and B M Golam. Theory of Ridge Regression Estimation with Applications. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185:742–743, 4 2022.
- [370] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016:785–794, 8 2016.
- [371] Victor Rodriguez-Galiano, Manuel Sanchez-Castillo, Mario Chica-Olmo, and Mario Chica-Rivas. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71:804–818, 12 2015.
- [372] Mudasir A. Ganaie, Minghui Hu, Ashwani K. Malik, Mohammad Tanveer, and

- Ponnuthurai N. Suganthan. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 4 2021.
- [373] Andy Liaw and Matthew Wiener. Classification and Regression by Random Forest. *R News*, 2, 2002.
- [374] Parveen Kumari, Sagar Paruthi, Ahmad Alyaseen, Afzal Husain Khan, and Alpana Jijja. Predictive performance assessment of recycled coarse aggregate concrete using artificial intelligence: A review. *Cleaner Materials*, 13:100263, 9 2024.
- [375] Harris Drucker, Chris J C Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. In *Proceedings of the 10th International Conference on Neural Information Processing Systems*, pages 155–161. MIT Press, 1996.
- [376] Raquel Rodríguez-Pérez and Jürgen Bajorath. Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. *Journal of Computer-Aided Molecular Design*, 36:355, 5 2022.
- [377] Payal Chhabra and Sonali Goyal. A Thorough Review on Deep Learning Neural Network. *2023 International Conference on Artificial Intelligence and Smart Communication, AISC 2023*, pages 220–226, 2023.
- [378] Heba Askr, Enas Elgeldawi, Heba Aboul Ella, Yaseen A.M.M. Elshaier, Mamdouh M. Gomaa, and Aboul Ella Hassanien. Deep learning in drug discovery: an integrative review and future challenges. *Artificial Intelligence Review*, 56:5975–6037, 7 2023.
- [379] Johannes Allgaier and Rüdiger Pryss. Cross-Validation Visualized: A Narrative Guide to Advanced Methods. *Machine Learning and Knowledge Extraction*, 6:1378–1388, 6 2024.
- [380] Luke A. Yates, Zach Aandahl, Shane A. Richards, and Barry W. Brook. Cross validation for model selection: A review with examples from ecology. *Ecological Monographs*, 93: e1557, 2 2023.
- [381] Henry E Niles. Correlation, Causation and Wright's Theory of "Path Coefficients". *Genetics*, 7:258, 5 1922.
- [382] Timothy O. Hodson. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15:5481–5487, 7 2022.

- [383] Davide Chicco, Matthijs J. Warrens, and Giuseppe Jurman. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7:1–24, 2021.
- [384] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Senior Member, Hui Xiong, and Qing He. A Comprehensive Survey on Transfer Learning. *arXiv*, 2020.
- [385] Asmaul Hosna, Ethel Merry, Jigmey Gyalmo, Zulfikar Alom, Zeyar Aung, and Mohammad Abdul Azim. Transfer learning: a friendly introduction. *Journal of Big Data*, 9:1–19, 12 2022.
- [386] Iori Shimada, Mitsumasa Osada, Hiroshi Fukunaga, and Michihisa Koyama. Interpretable Machine Learning with Physics-Based Feature Engineering: Application to the Catalytic Cracking Reaction of Polycyclic Aromatic Hydrocarbons. *Industrial and Engineering Chemistry Research*, 62:21087–21099, 12 2023.
- [387] Yuming Su, Yiheng Dai, Yifan Zeng, Caiyun Wei, Yangtao Chen, Fuchun Ge, Peikun Zheng, Da Zhou, Pavlo O. Dral, and Cheng Wang. Interpretable Machine Learning of Two Photon Absorption. *Advanced Science*, 10:2204902, 3 2023.
- [388] Abbas Pak, Abdullah Kaviani Rad, Mohammad Javad Nematollahi, and Mohammadreza Mahmoudi. Application of the Lasso regularisation technique in mitigating overfitting in air quality prediction models. *Scientific Reports*, 15:1–17, 1 2025.
- [389] Shijie Jiang, Lily belle Sweet, Georgios Blougouras, Alexander Brenning, Wantong Li, Markus Reichstein, Joachim Denzler, Wei Shangguan, Guo Yu, Feini Huang, and Jakob Zscheischler. How Interpretable Machine Learning Can Benefit Process Understanding in the Geosciences. *Earth's Future*, 12:e2024EF004540, 7 2024.
- [390] Fubo Yu, Changhong Wei, Peng Deng, Ting Peng, and Xiangang Hu. Deep exploration of random forest model boosts the interpretability of machine learning studies of complicated immune responses and lung burden of nanoparticles. *Science Advances*, 7:eabf4130, 5 2021.
- [391] Cheng Wei Ju, Hanzhi Bai, Bo Li, and Rizhang Liu. Machine Learning Enables Highly Accurate Predictions of Photophysical Properties of Organic Fluorescent Materials: Emission Wavelengths and Quantum Yields. *Journal of Chemical Information and*

- [392] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision A.1. Gaussian Inc. Wallingford CT 2009.

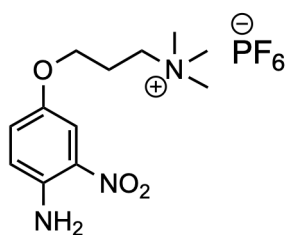
Appendix A

Appendix

A.1 Synthetic spectra

A.1.1 Intermediates

Synthesis of compound 3:



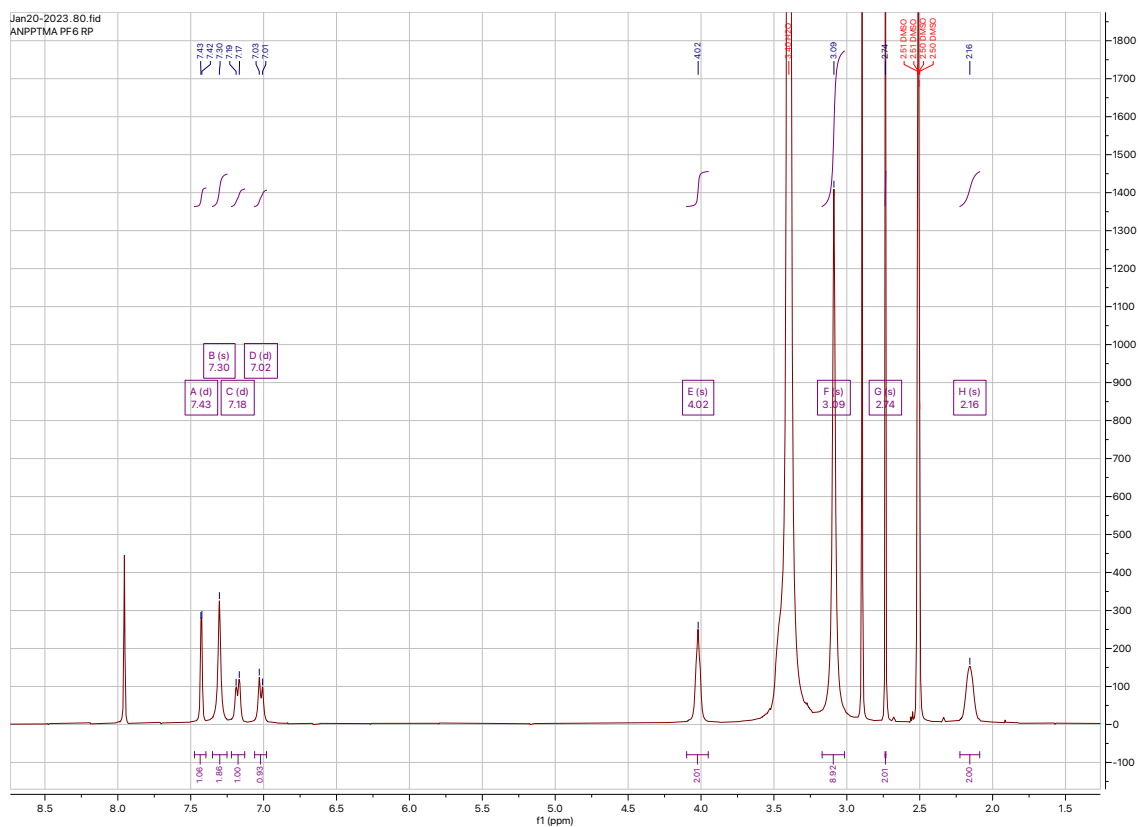
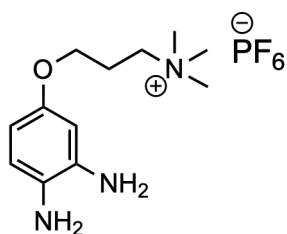


Figure A.1. ^1H NMR spectrum of compound 3: . DMF peaks can be seen at 7.95 and 2.89 ppm.

Synthesis of PPTMA:



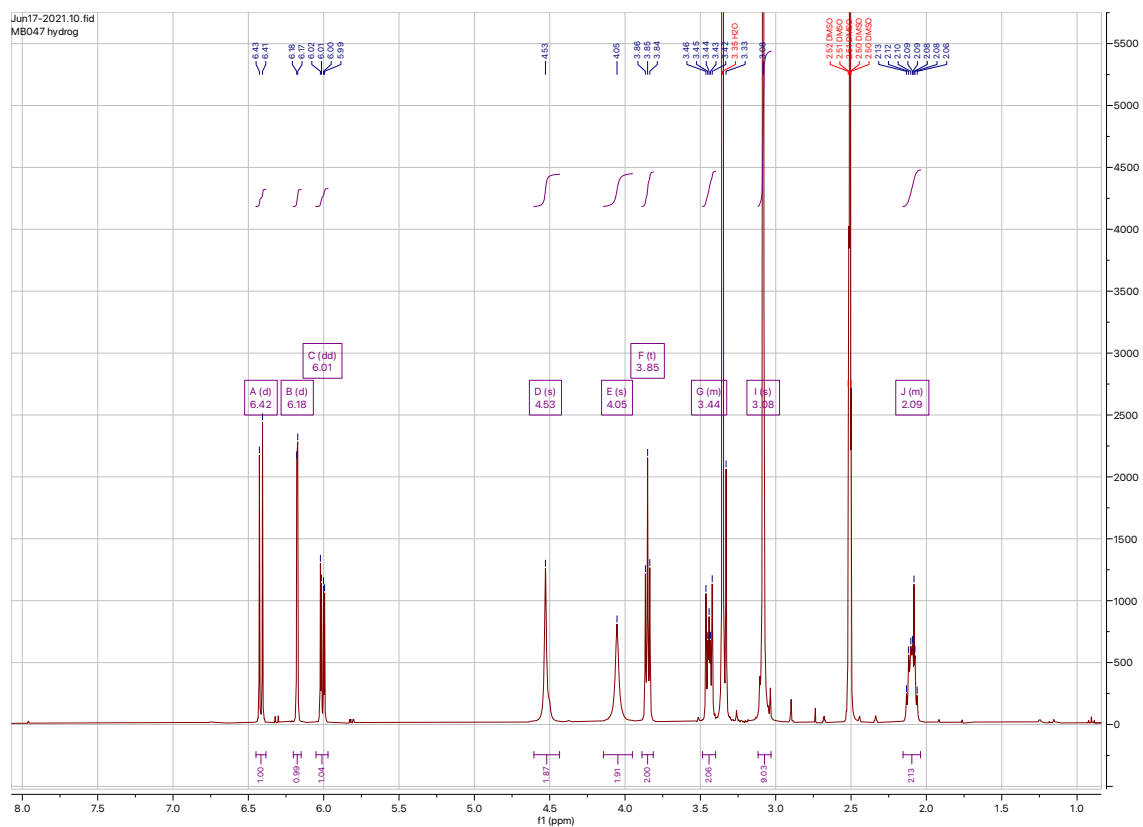
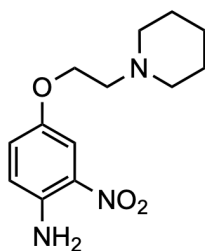


Figure A.2. ^1H NMR spectrum of compound PPTMA.

Synthesis of compound 5:



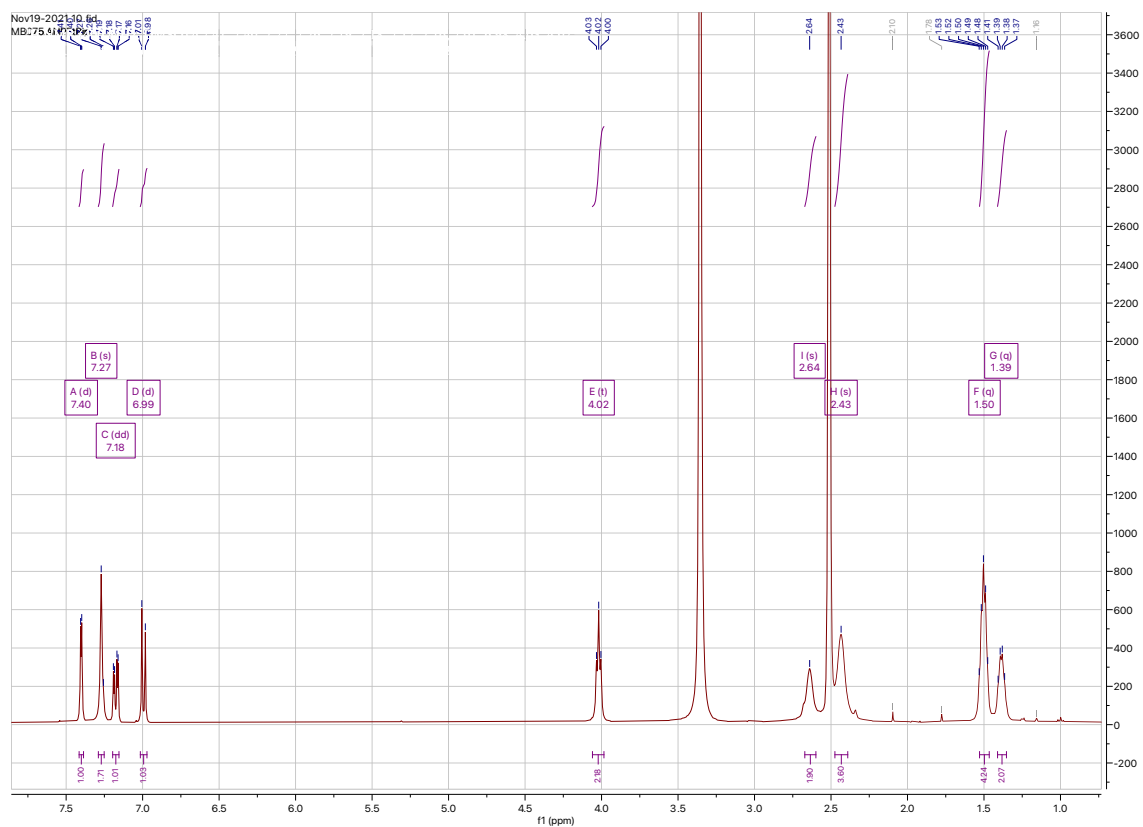
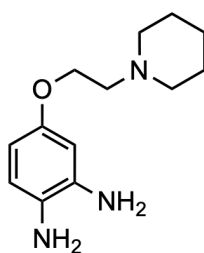


Figure A.3. ^1H NMR spectrum of compound 5.

Synthesis of PEP:



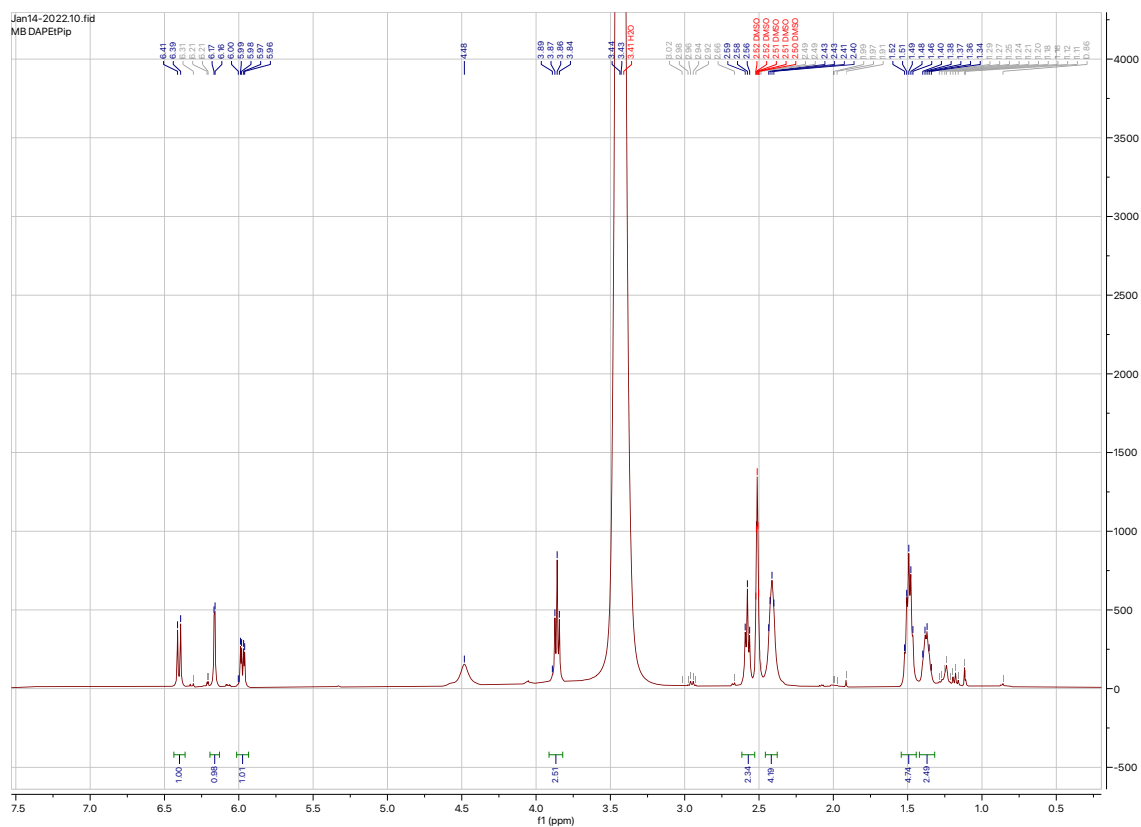
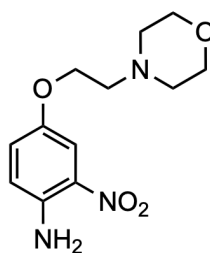
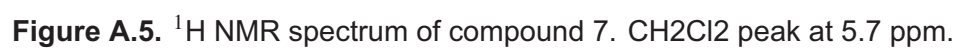


Figure A.4. ^1H NMR spectrum of PEP.

Synthesis of compound 7:



Nc1cc(N)ccc1OCCN2CCCCC2

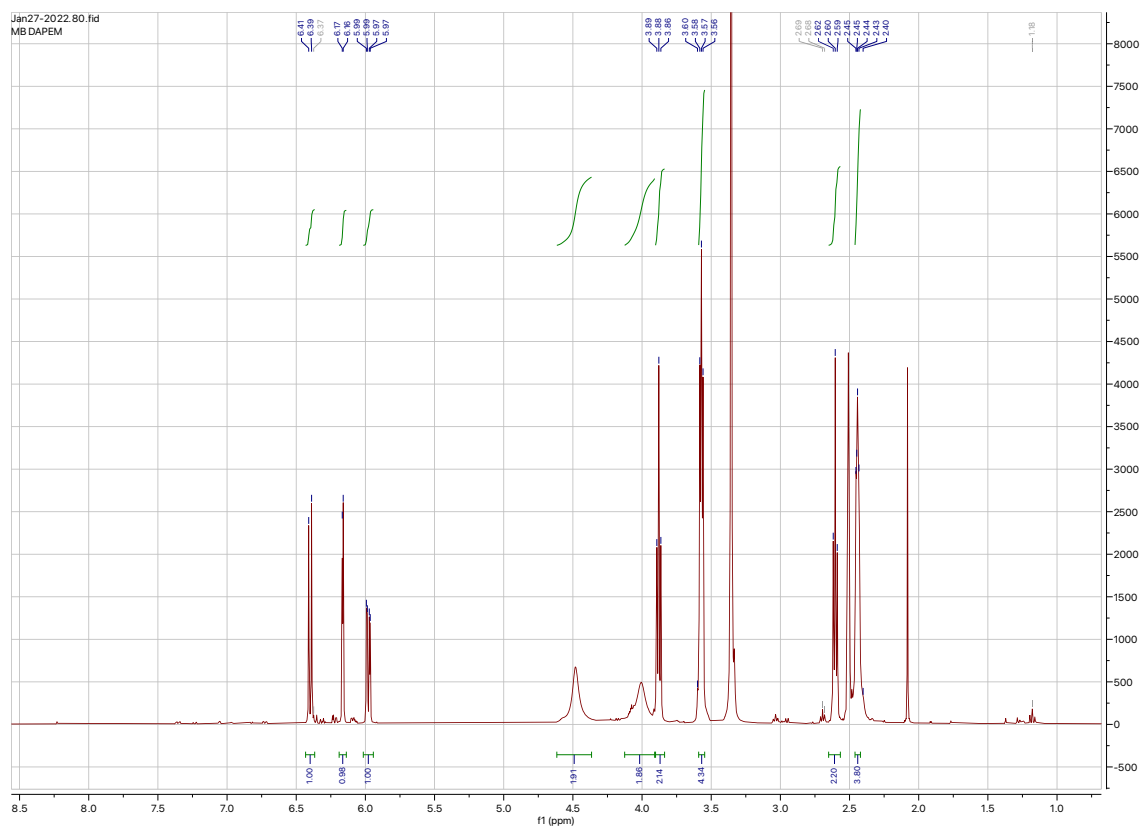
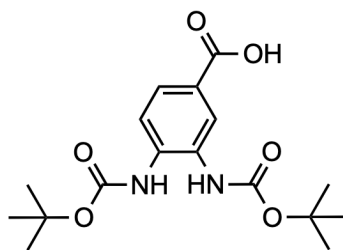


Figure A.6. ^1H NMR spectrum of PEM. Acetonitrile peak at 2.07 ppm.

Synthesis of compound 12:



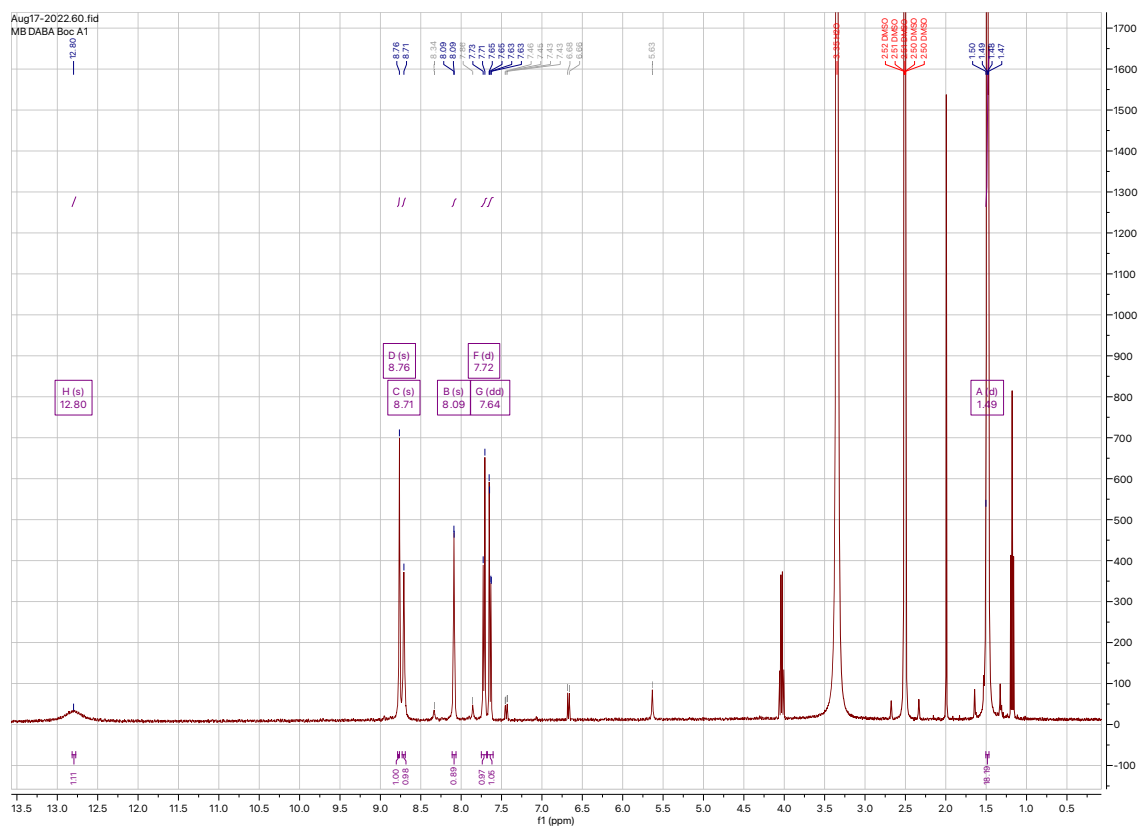
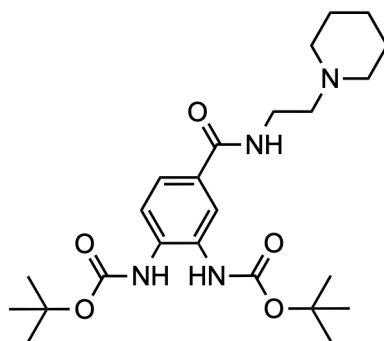


Figure A.7. ^1H NMR spectrum of compound 12. Ethyl acetate peaks at 1.17, 1.99 and 4.03 ppm.

Synthesis of BOC-AEP:



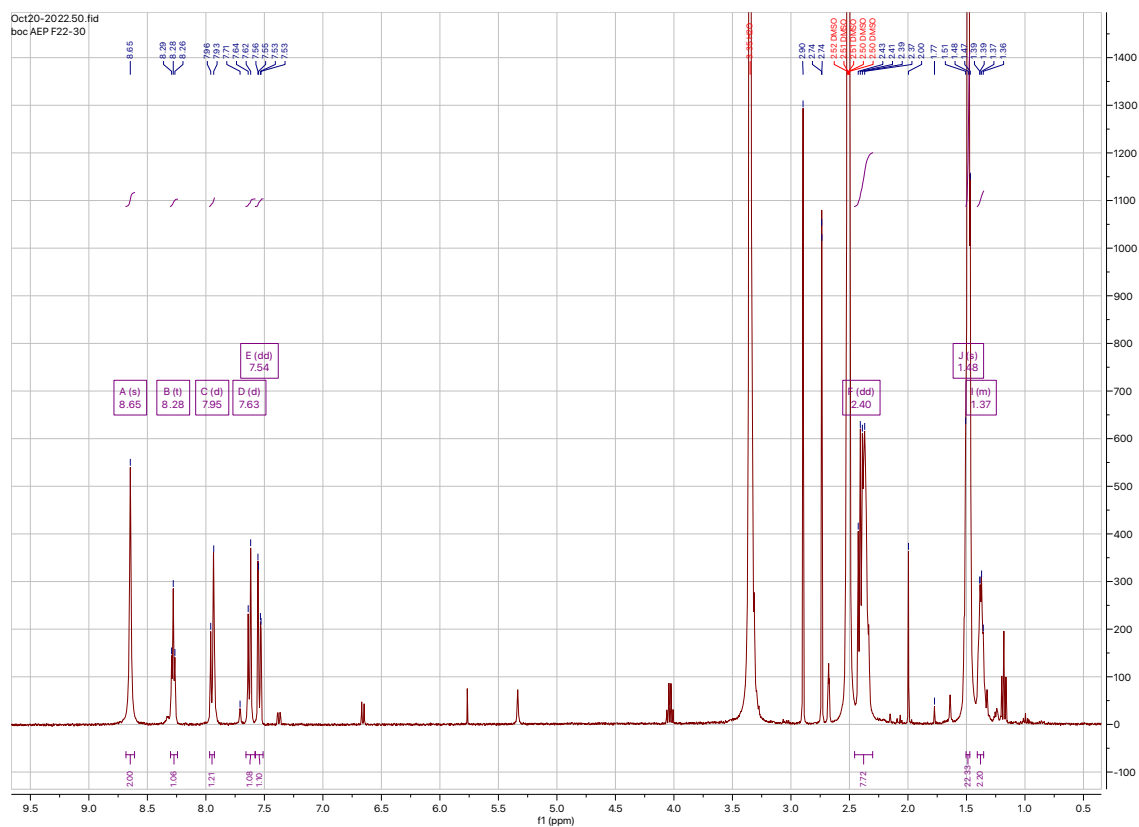
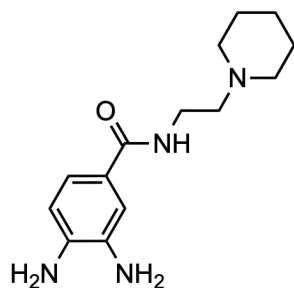


Figure A.8. ^1H NMR spectrum of compound BOC-AEP. Ethyl acetate peaks at 1.17, 1.99 and 4.03 ppm. CH_2Cl_2 peak at 5.77 ppm.

Synthesis of AEP:



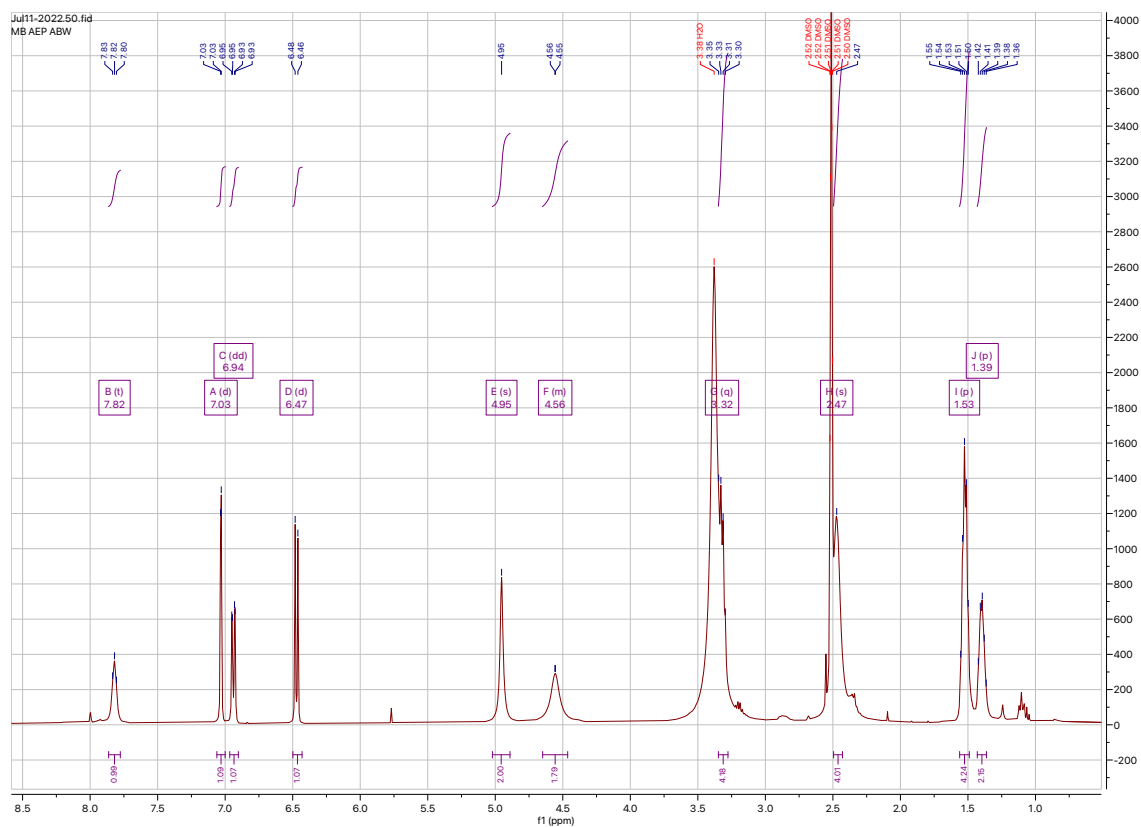
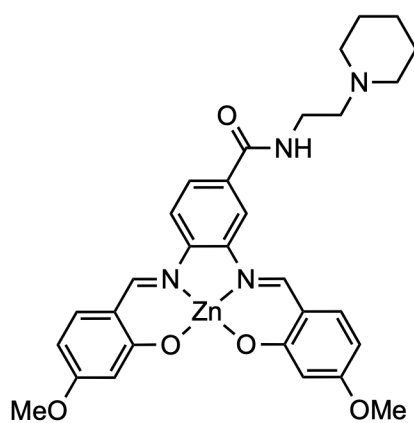


Figure A.9. ^1H NMR spectrum of compound AEP.



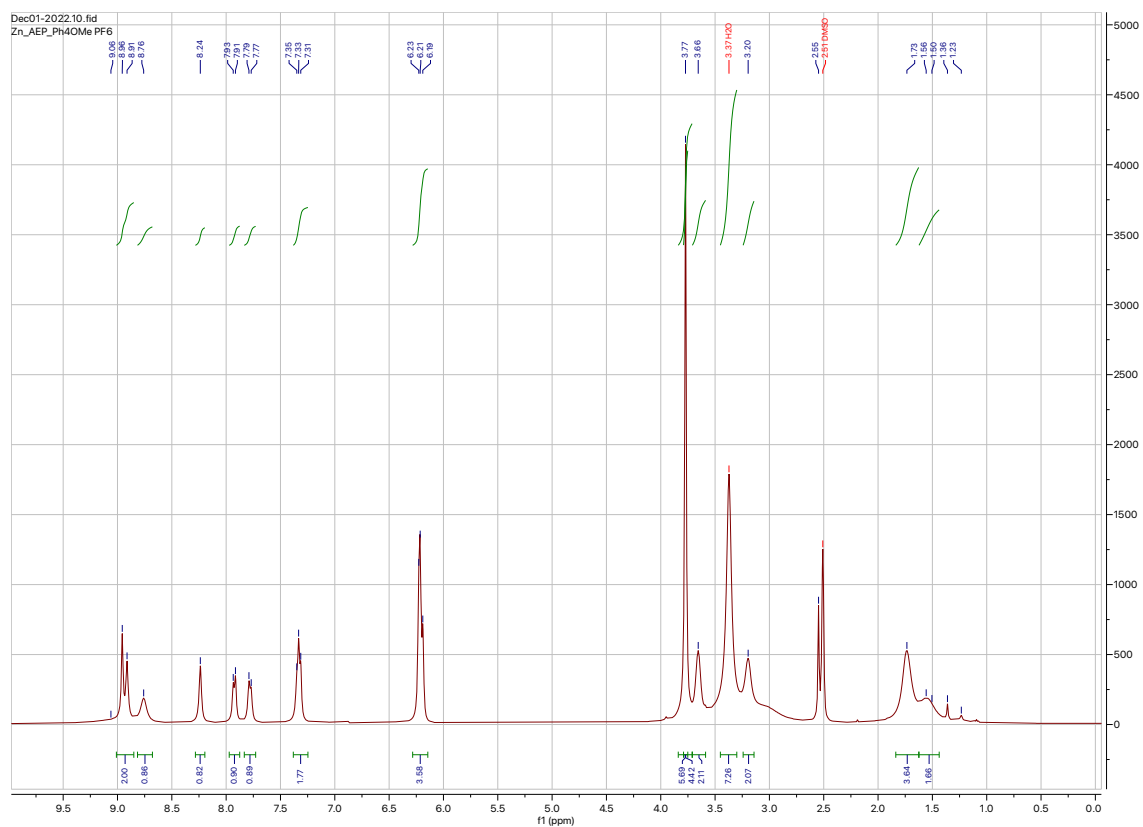


Figure A.10. ^1H NMR spectrum of Zn_AEP_4OMe.

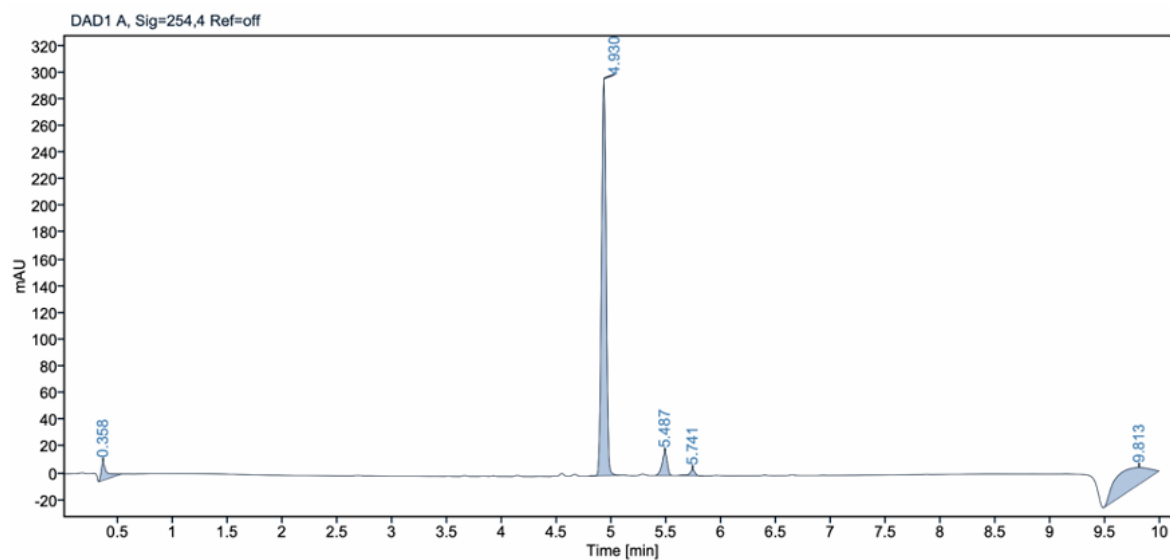


Figure A.11. LCMS spectrum of Zn_AEP_4OMe.

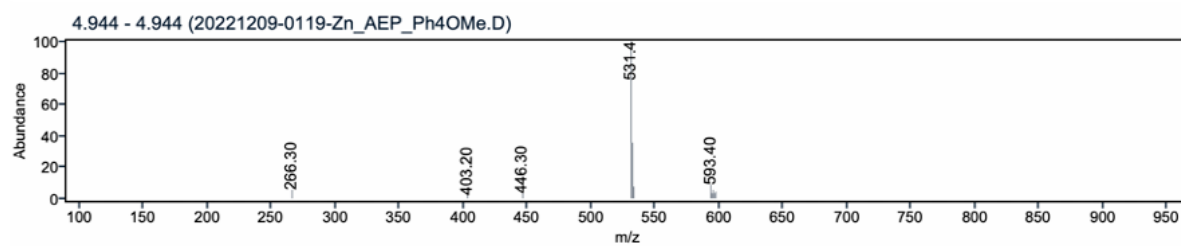
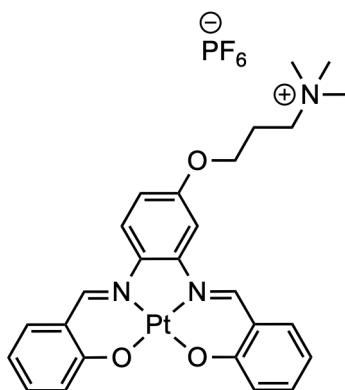


Figure A.12. Mass spec for Zn_AEP_4OMe under LCMS UV peak.

A.1.2 Platinum(II) salphen complexes

Pt_PPTMA_Ph [hexafluorophosphate]:



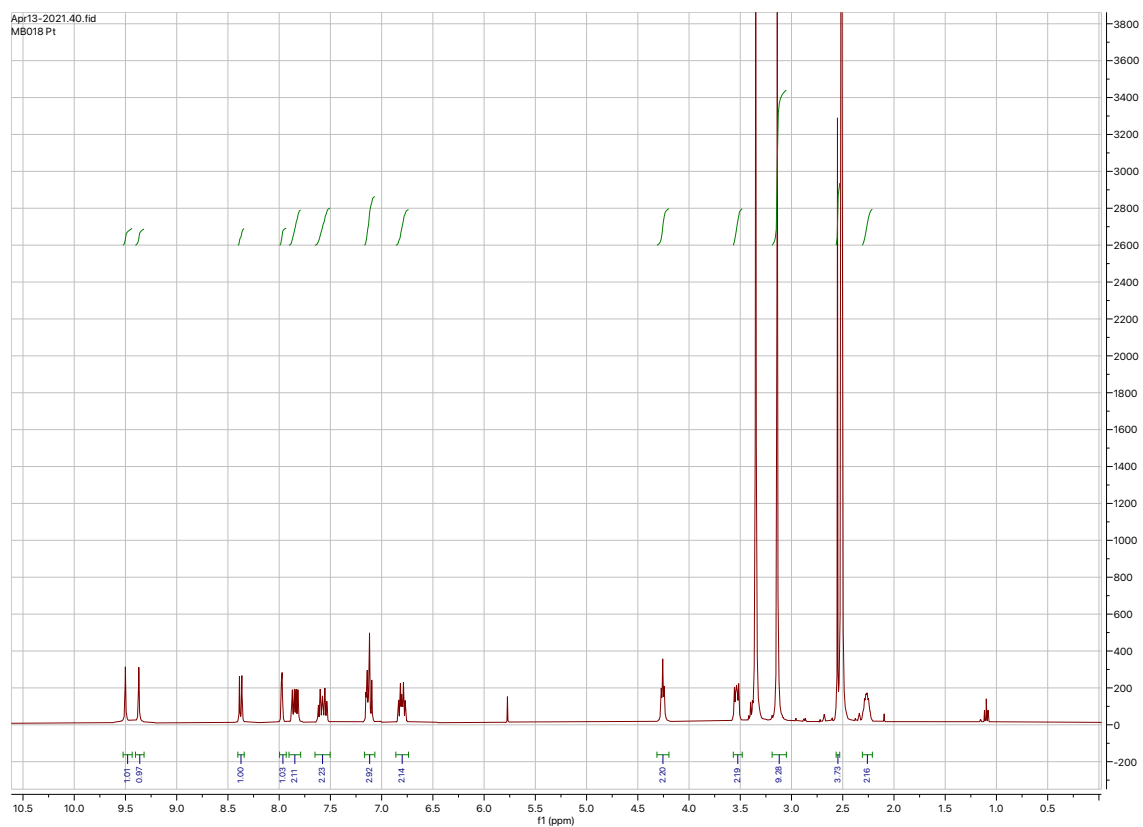


Figure A.13. Full ^1H NMR spectrum of Pt_PPTMA_Ph.

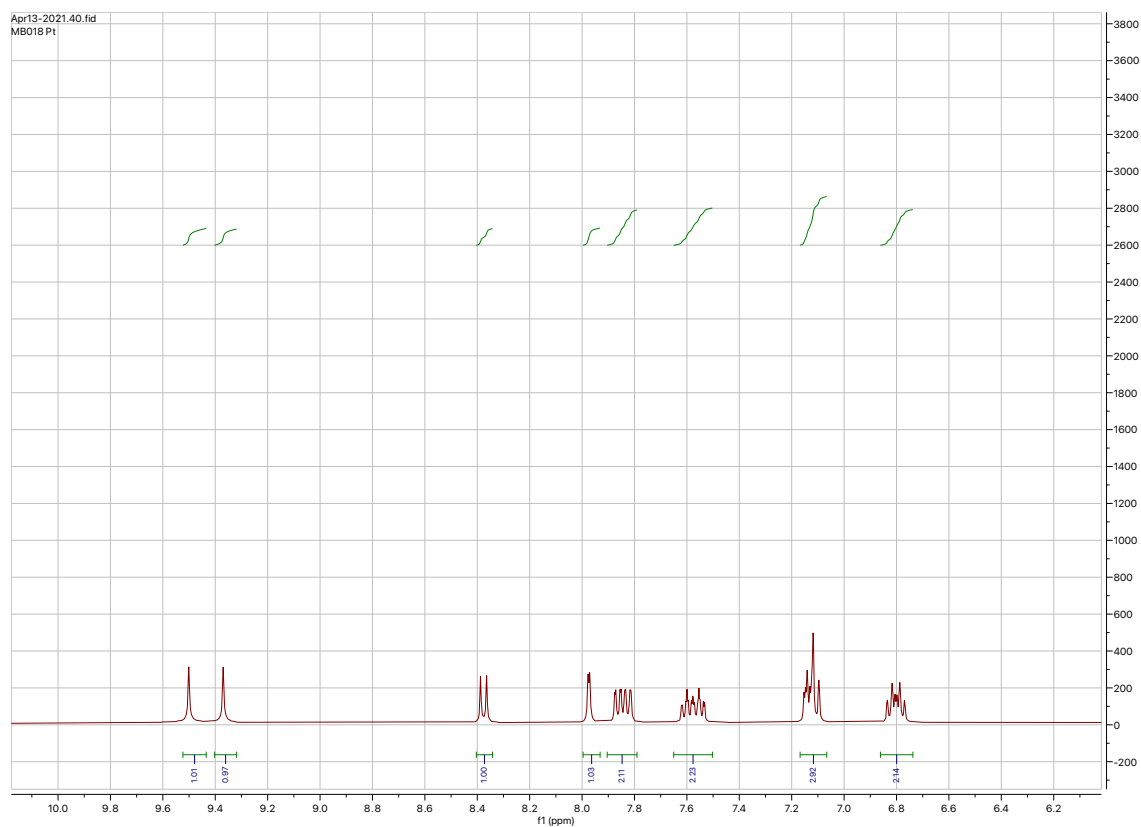


Figure A.14. Aromatic region of the ^1H NMR spectrum of Pt_PPTMA_Ph.

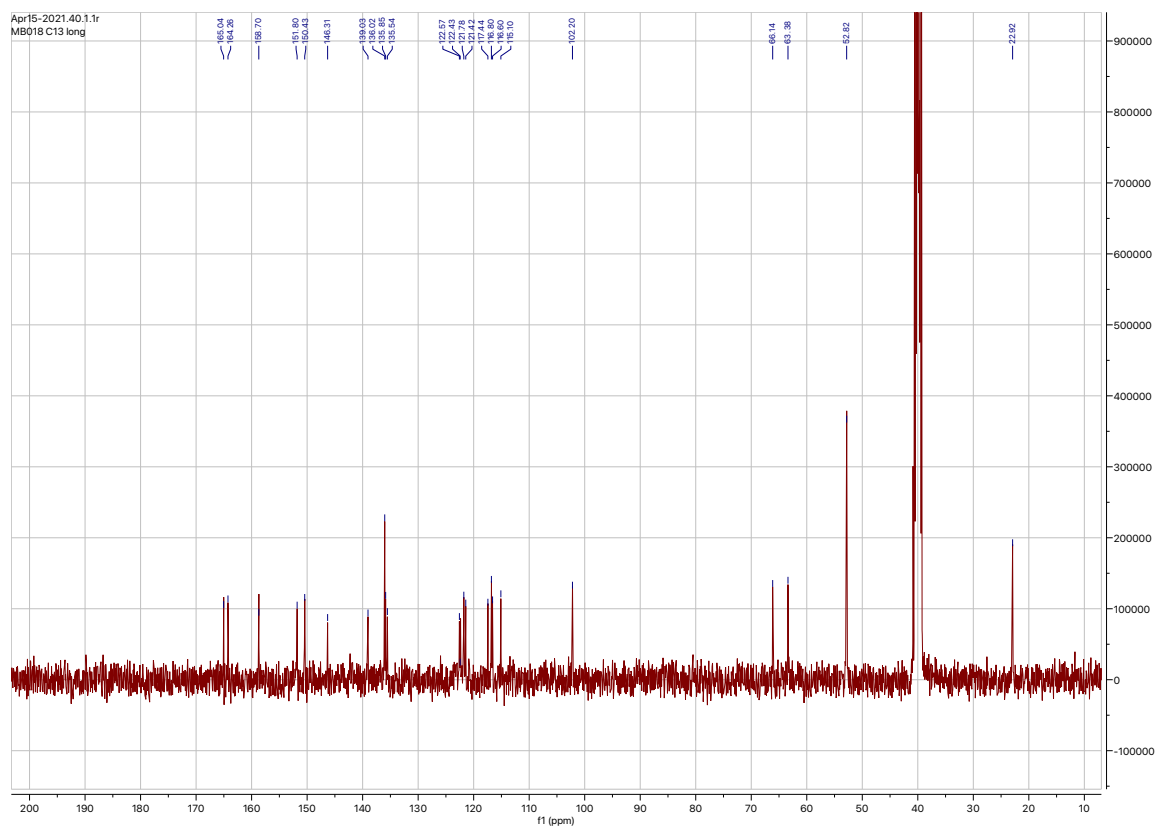


Figure A.15. ^{13}C NMR spectrum of Pt_PPTMA_Ph.

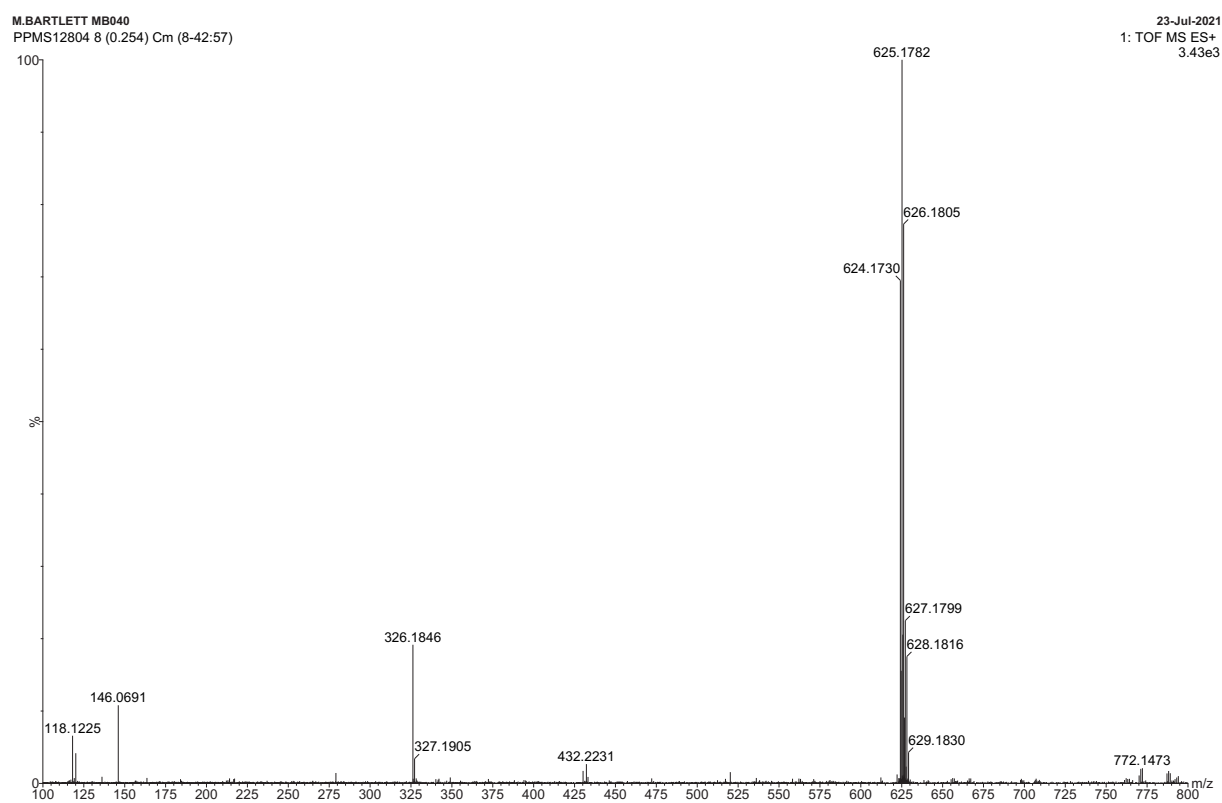


Figure A.16. HRMS of Pt_PPTMA_Ph.

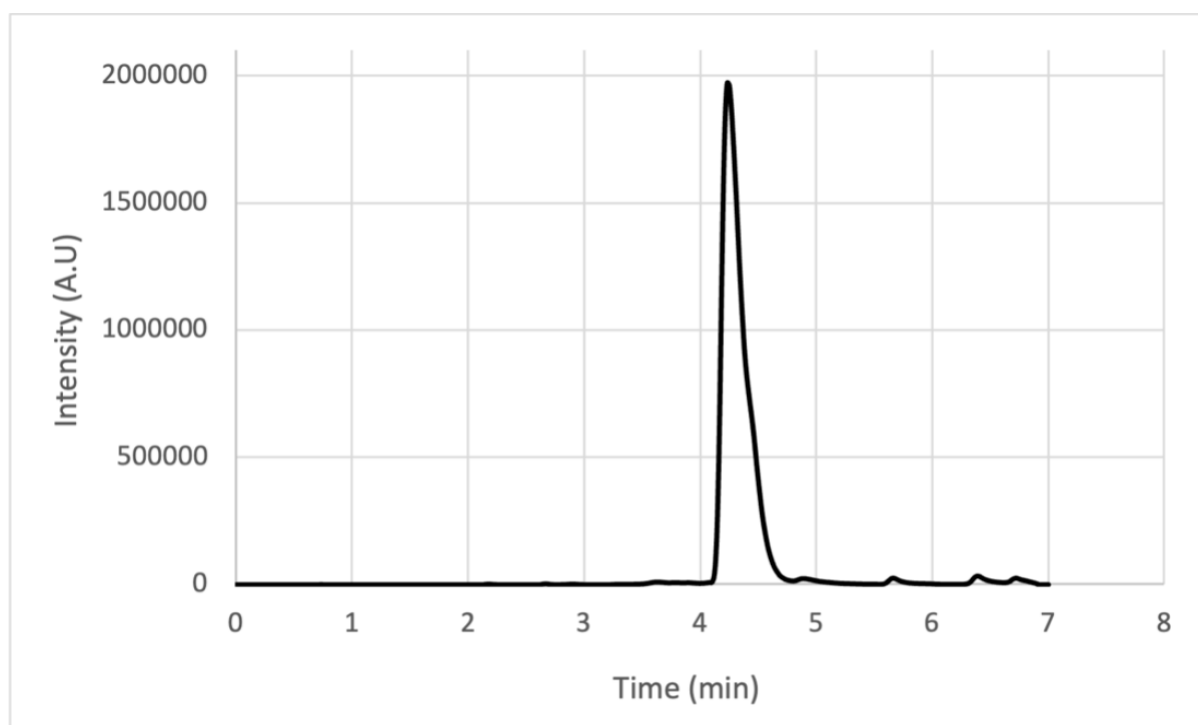


Figure A.17. LCMS UV trace for Pt_PPTMA_Ph.

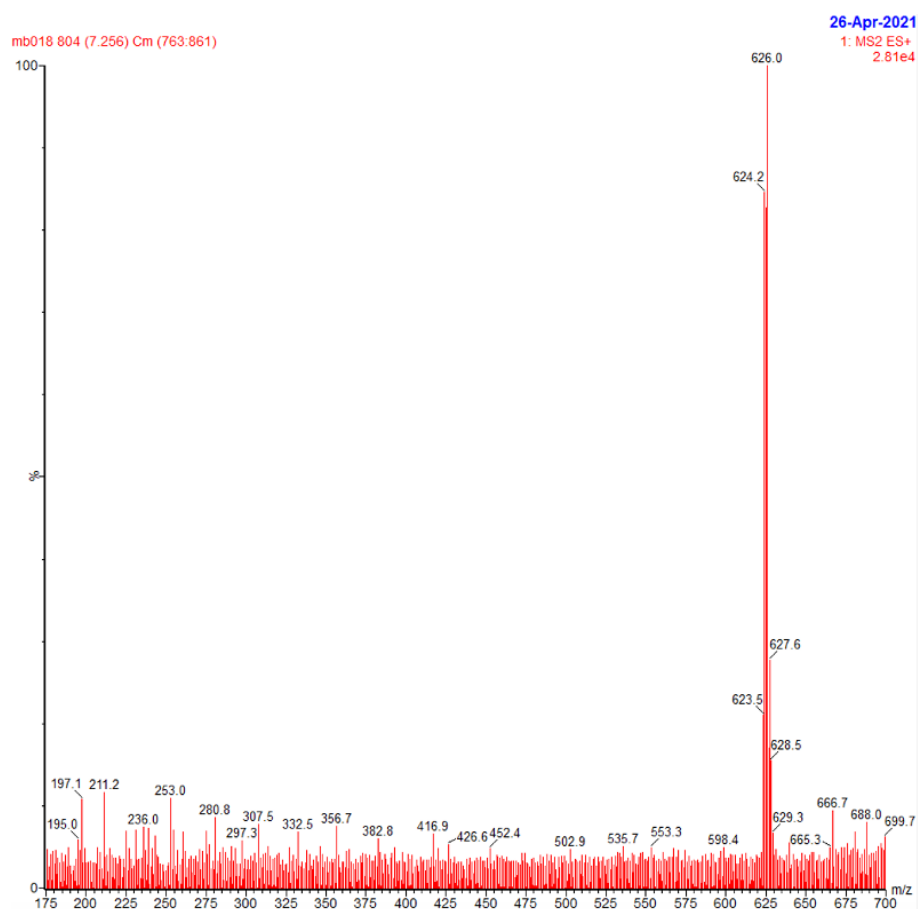


Figure A.18. Mass spec of Pt_PPTMA_Ph LCMS under UV peak at 4.23 min.

Pt_PPTMA_Naph [hexafluorophosphate]:

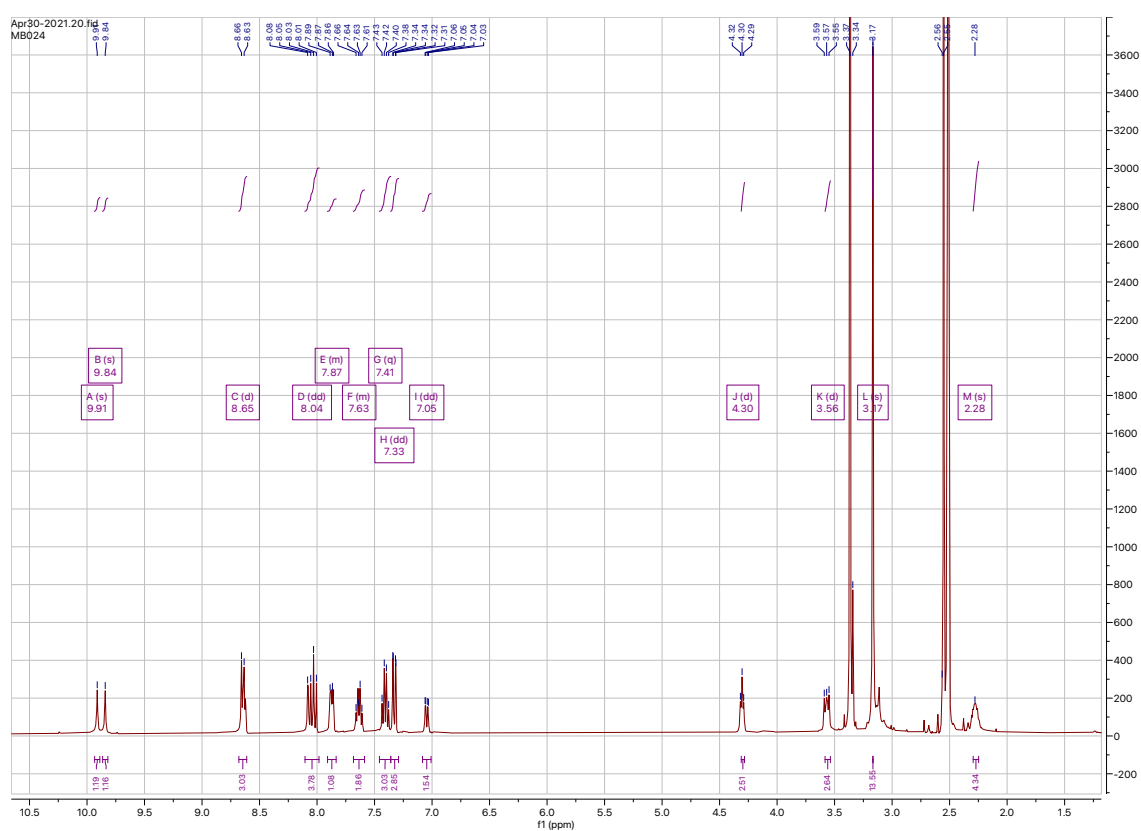
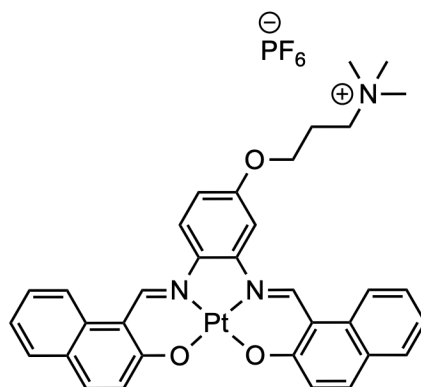


Figure A.19. Full ^1H NMR spectrum of Pt_PPTMA_Naph.

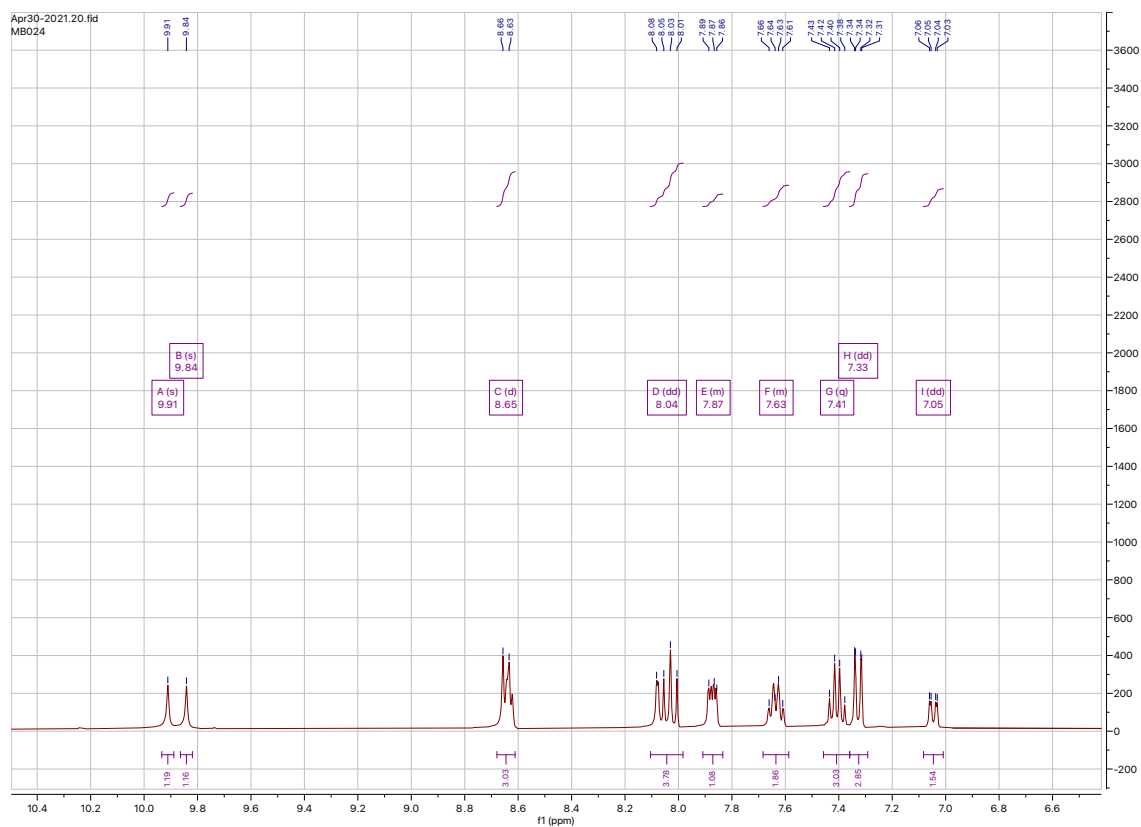


Figure A.20. Aromatic region of the ^1H NMR spectrum of Pt_PPTMA_Naph.

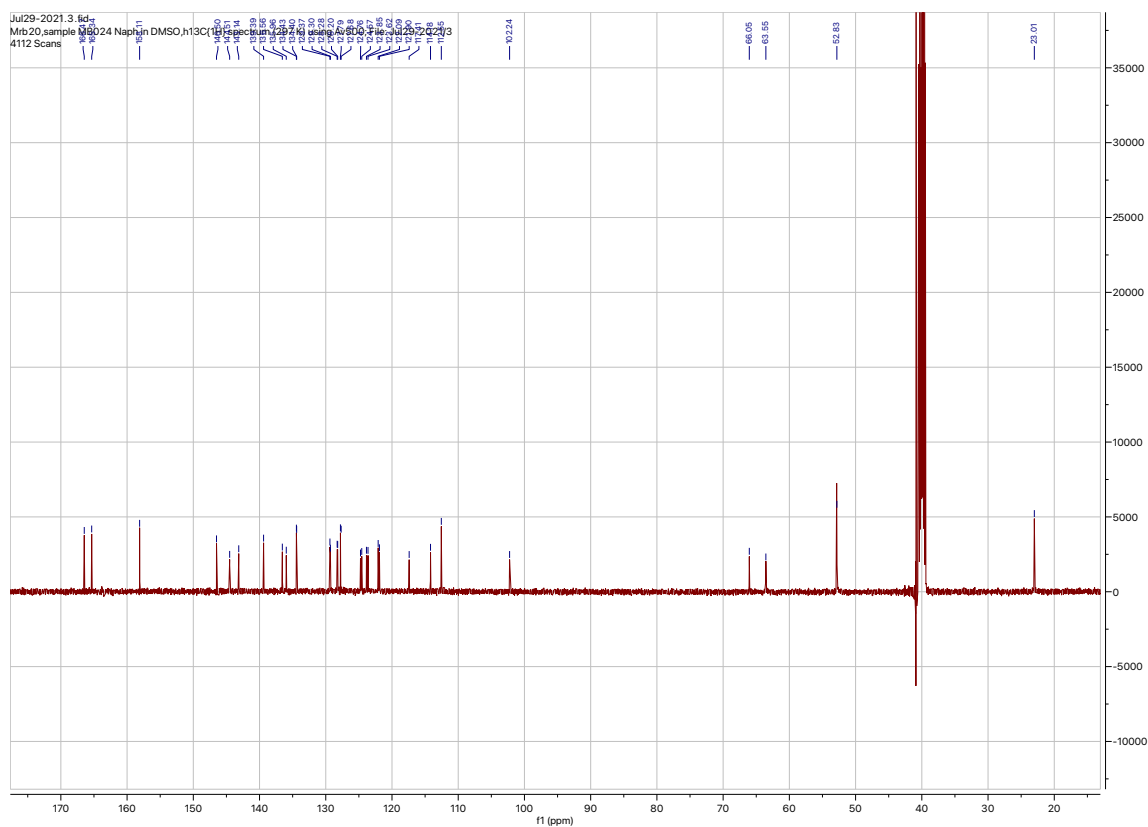


Figure A.21. ^{13}C NMR spectrum of Pt_PPTMA_Naph.

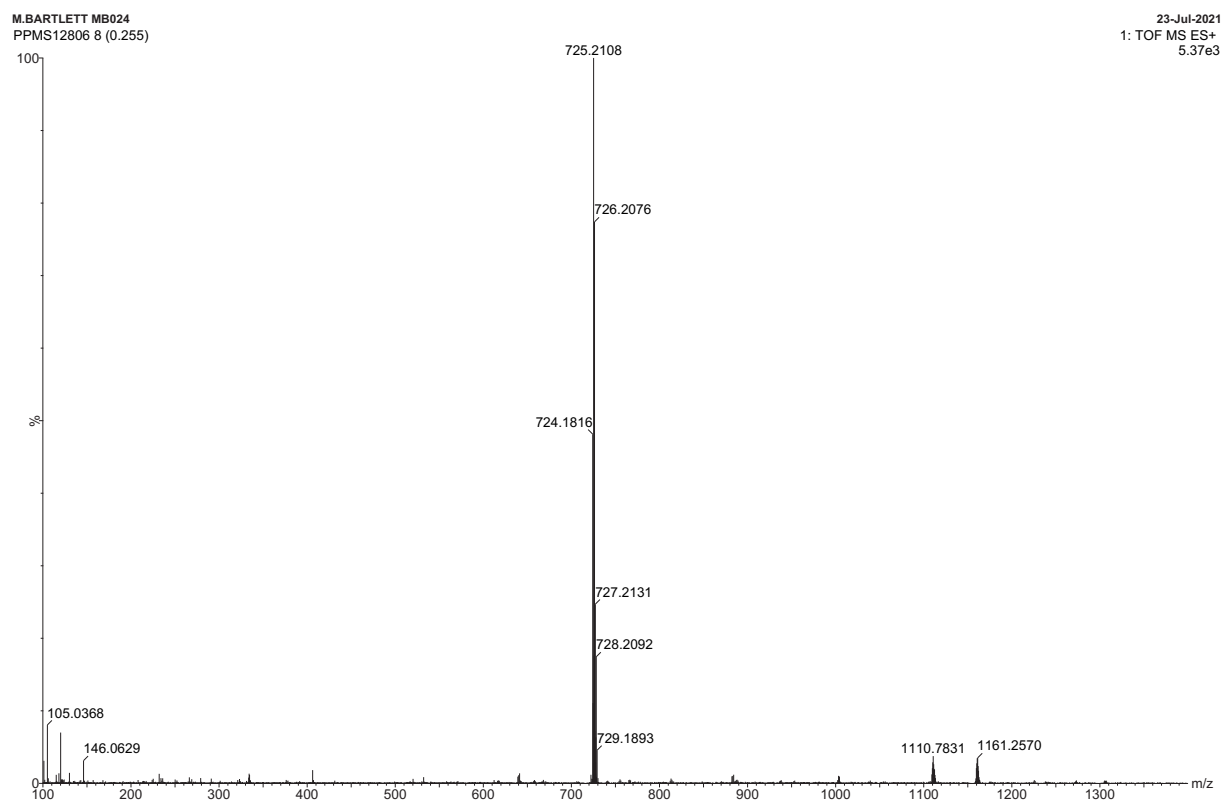


Figure A.22. HRMS of Pt_PPTMA_Naph.

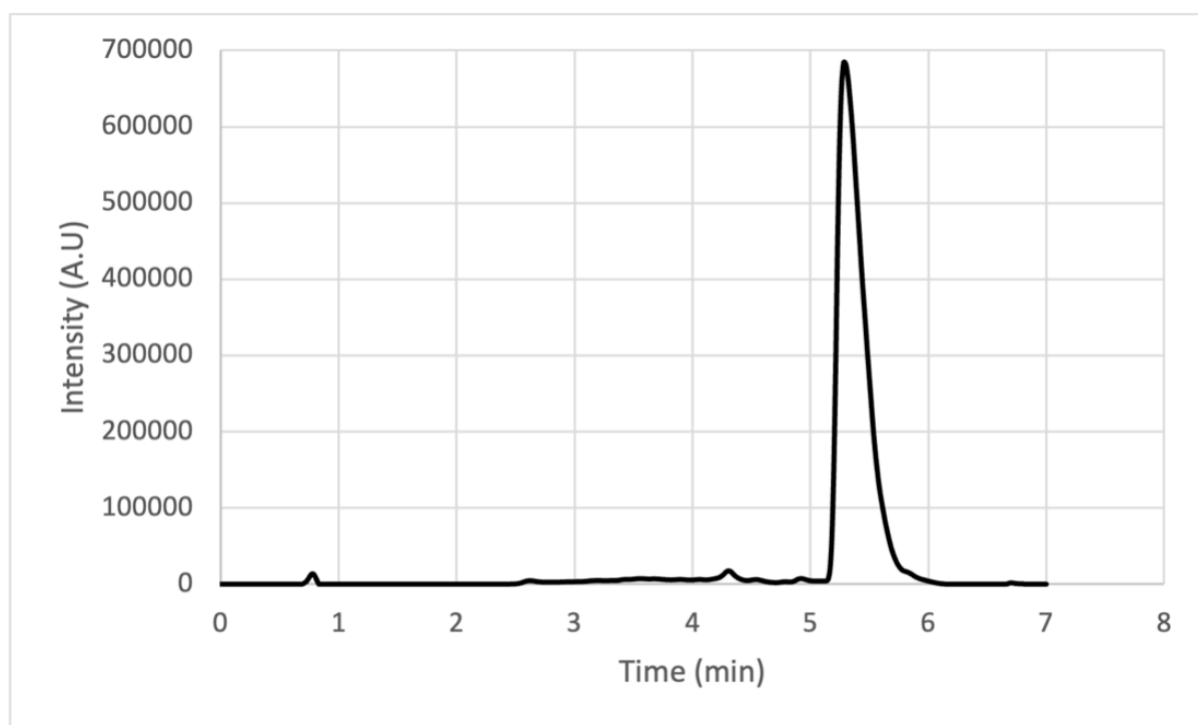


Figure A.23. LCMS UV trace for Pt_PPTMA_Naph

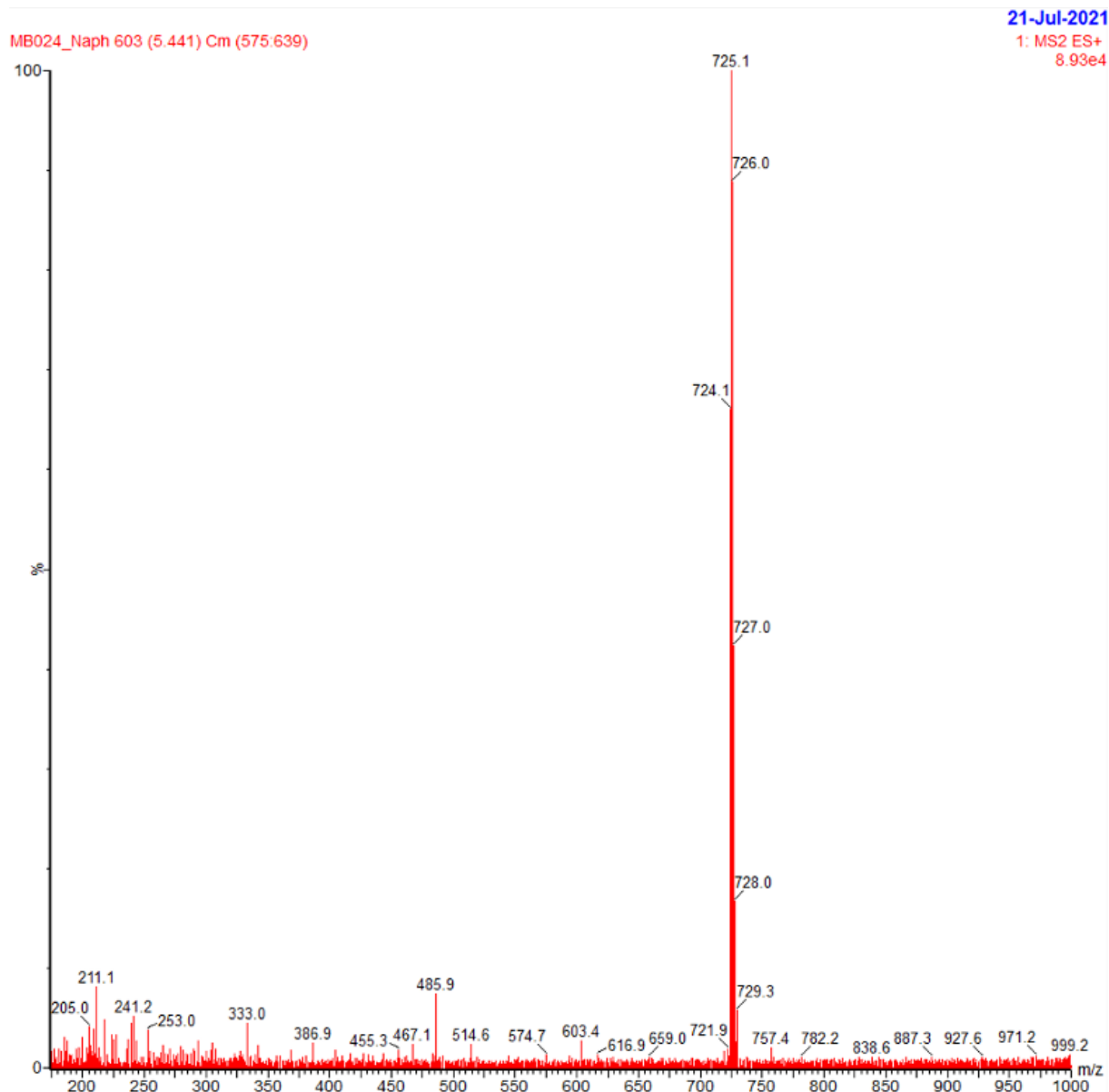
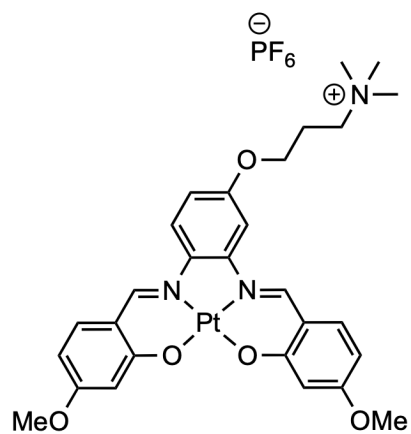


Figure A.24. Mass spec of Pt_PPTMA_Naph LCMS under UV peak at 5.28 min.

Pt_PPTMA_4OMe [hexafluorophosphate]:



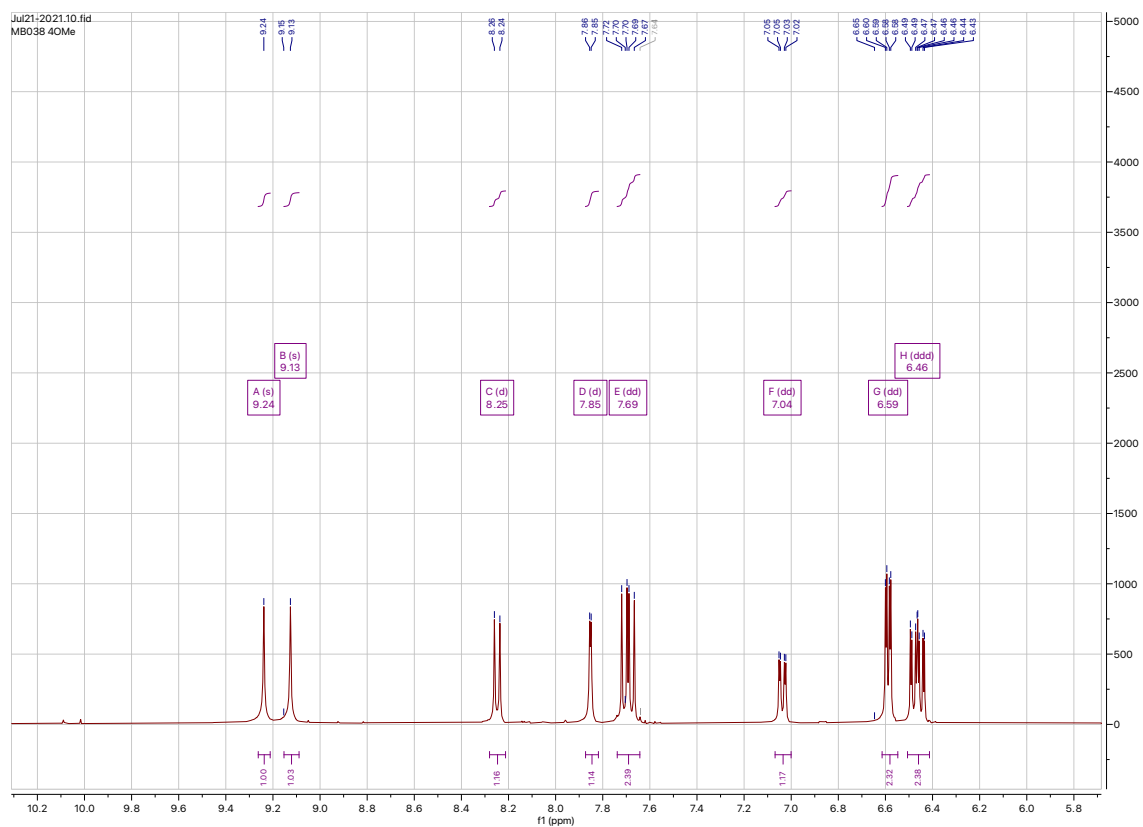


Figure A.26. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_4OMe.

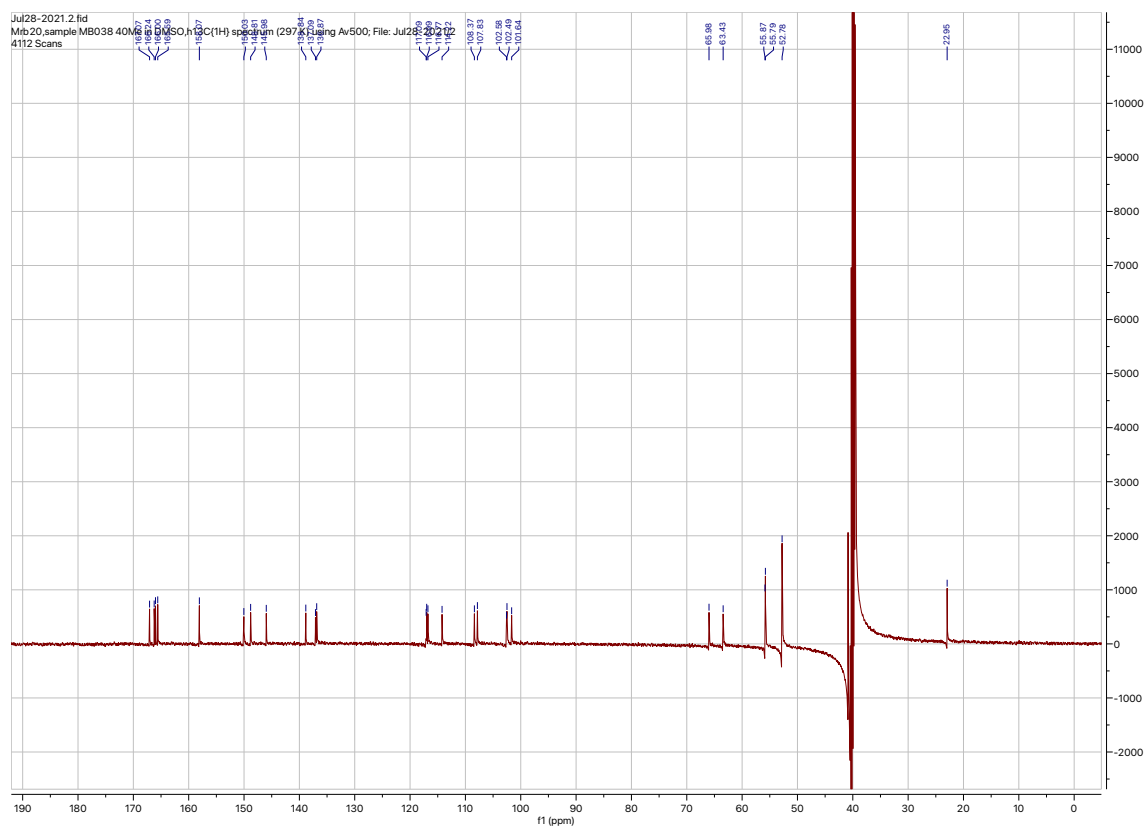


Figure A.27. ^{13}C NMR spectrum of Pt_PPTMA_4OMe.

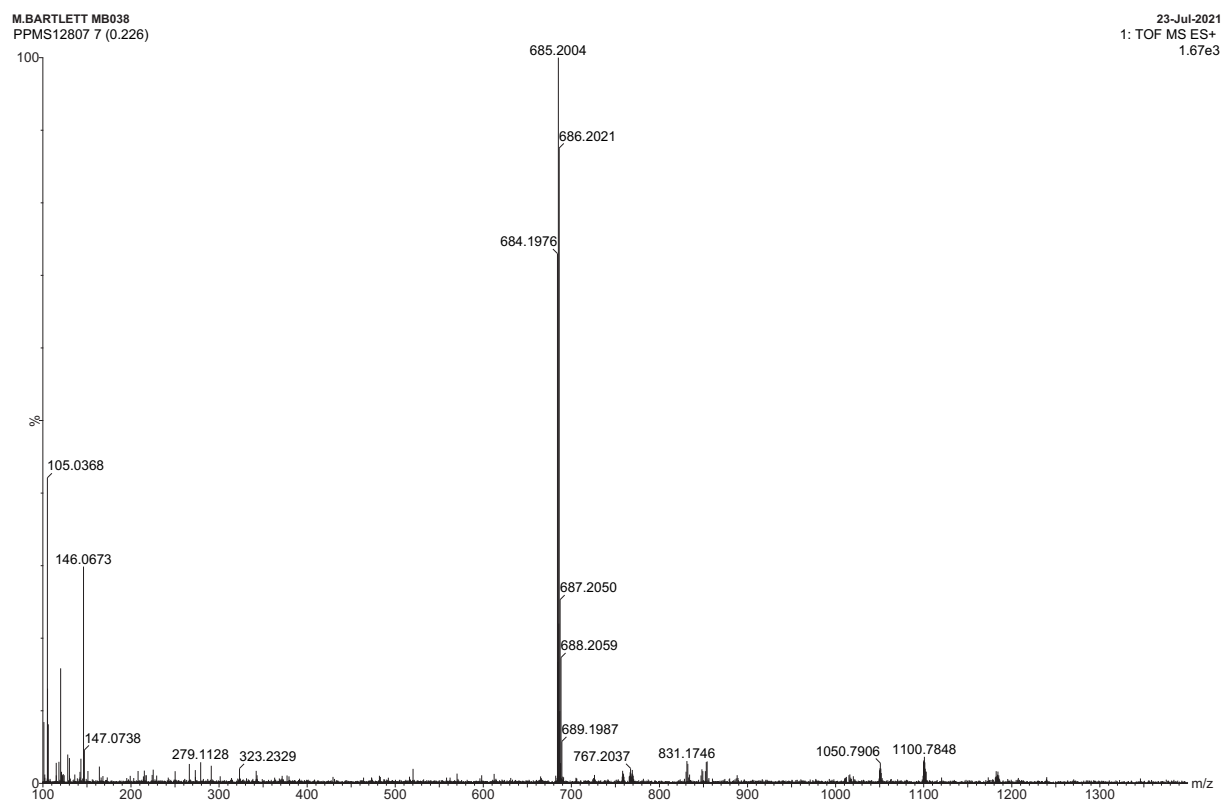


Figure A.28. HRMS of Pt_PPTMA_4OMe.

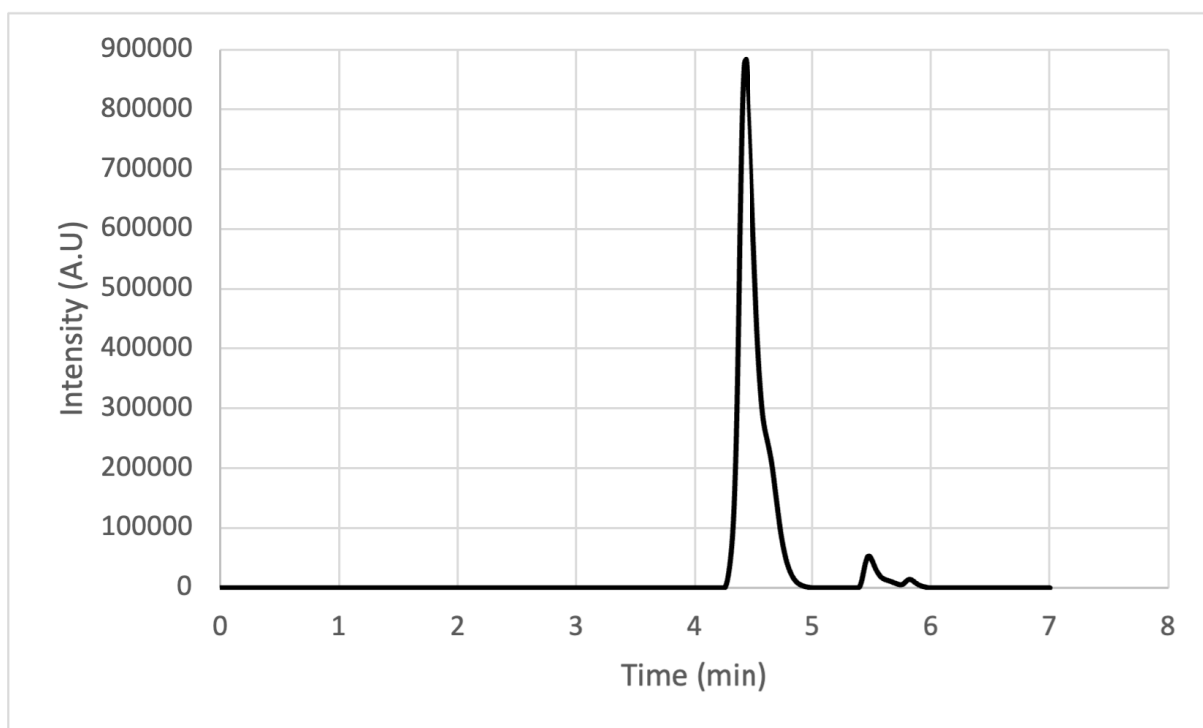


Figure A.29. LCMS UV trace of Pt_PPTMA_4OMe.

MB038_4OMePh 488 (4.402) Cm (457:536)

21-Jul-2021

1: MS2 ES+
3.63e4

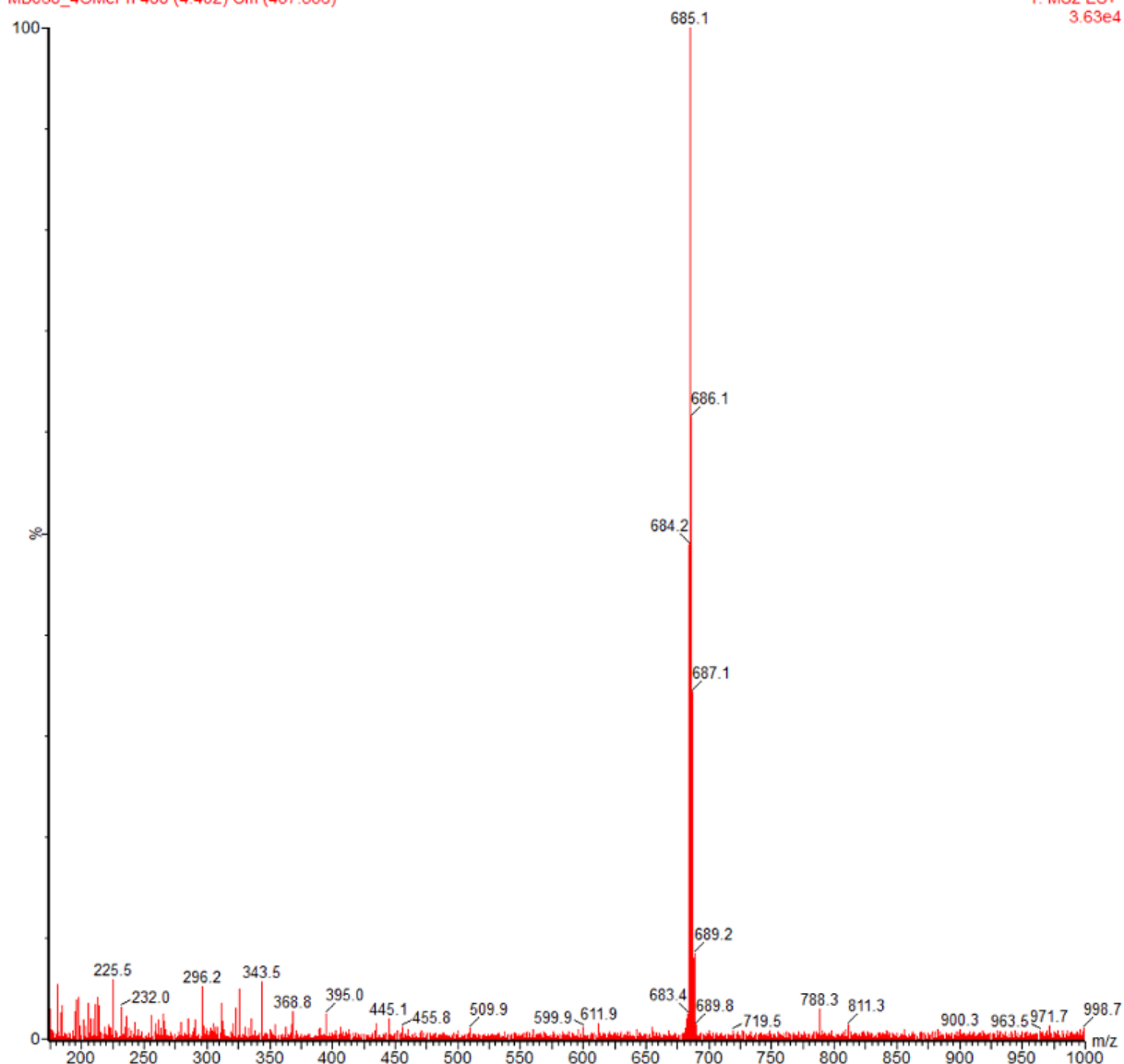
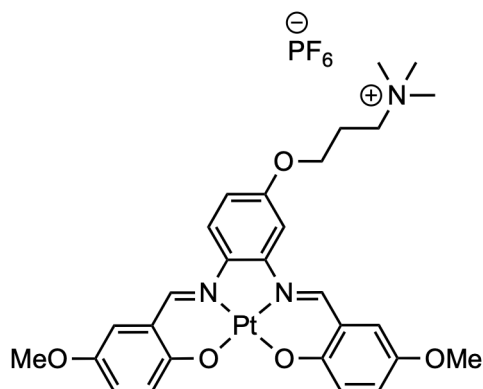


Figure A.30. Mass spec of Pt_PPTMA_4OMe LCMS under UV peak at 4.43 min

Pt_PPTMA_5OMe [hexafluorophosphate]:



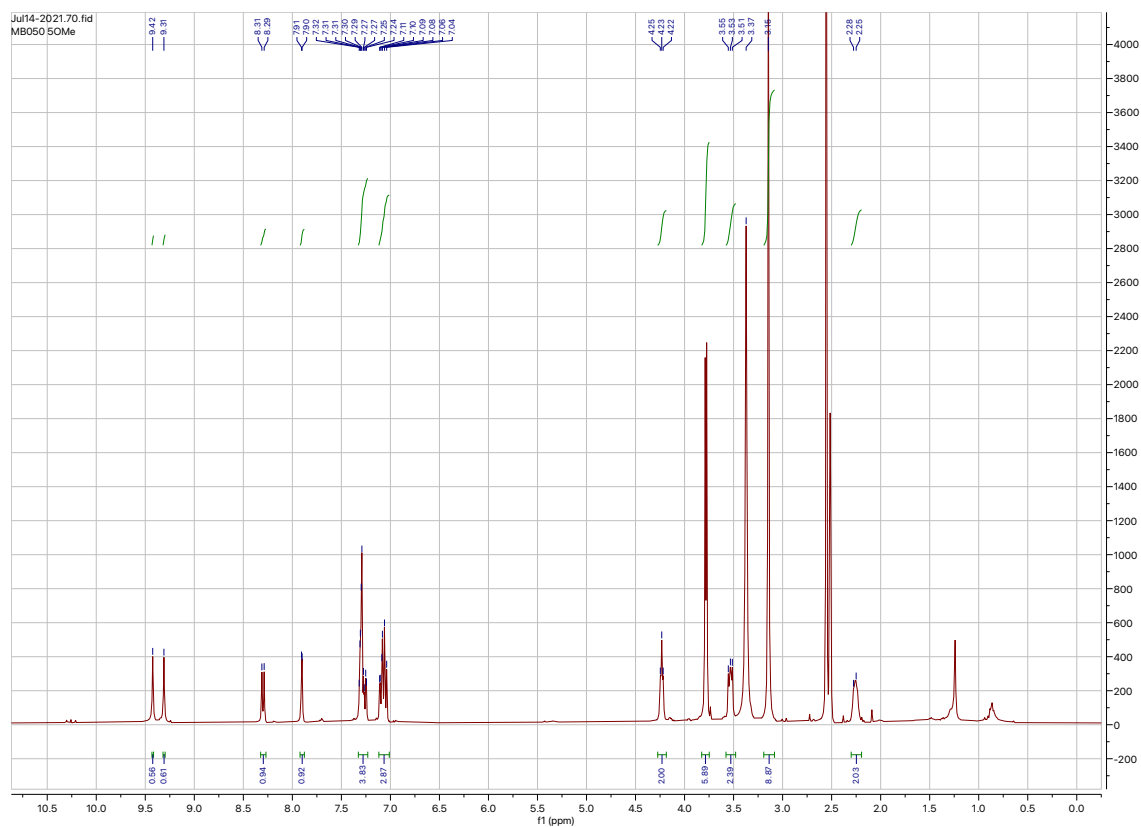


Figure A.31. ^1H NMR spectrum of Pt_PPTMA_5OMe.

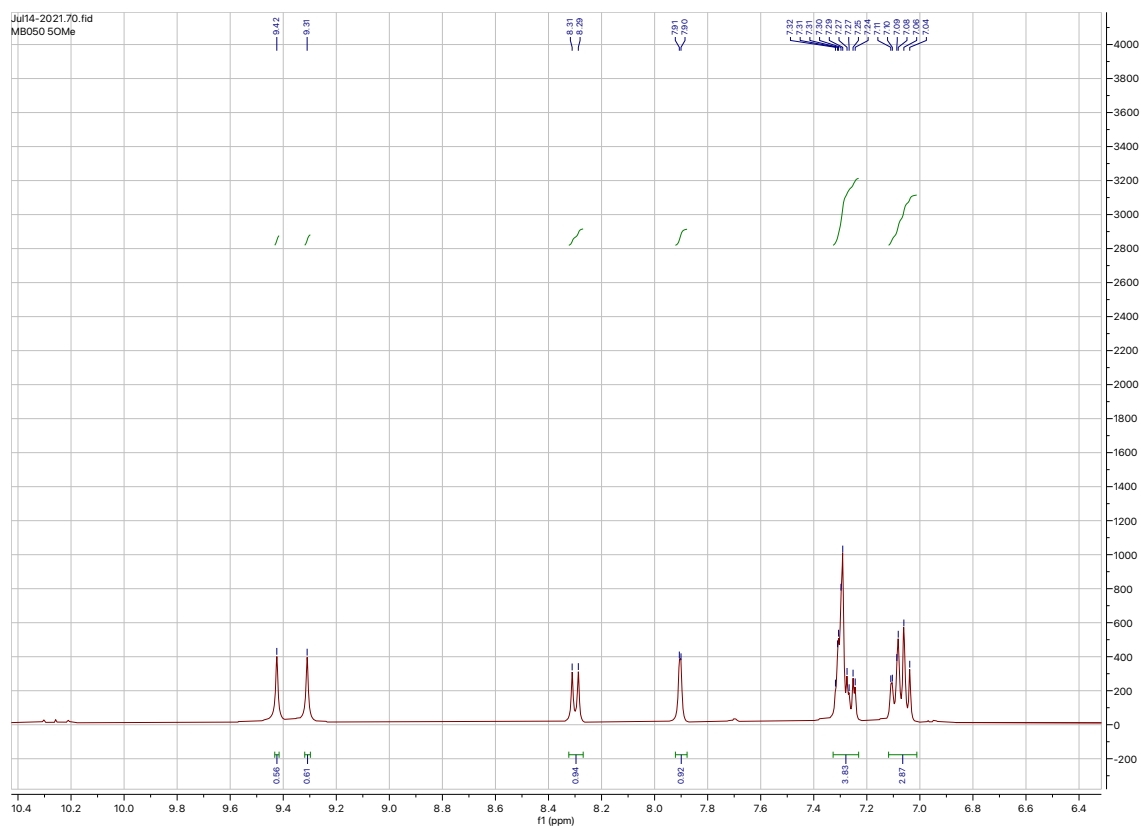


Figure A.32. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_5OMe.

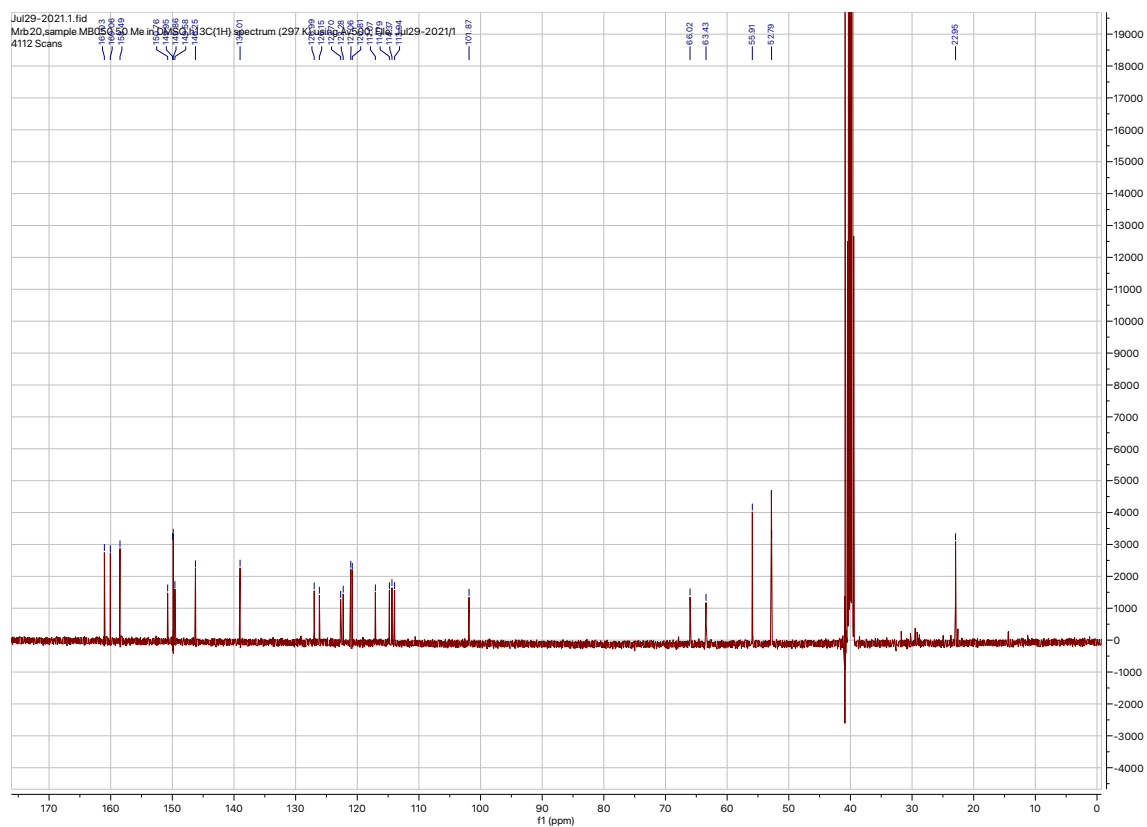


Figure A.33. ^{13}C NMR spectrum of Pt_PPTMA_5OMe.

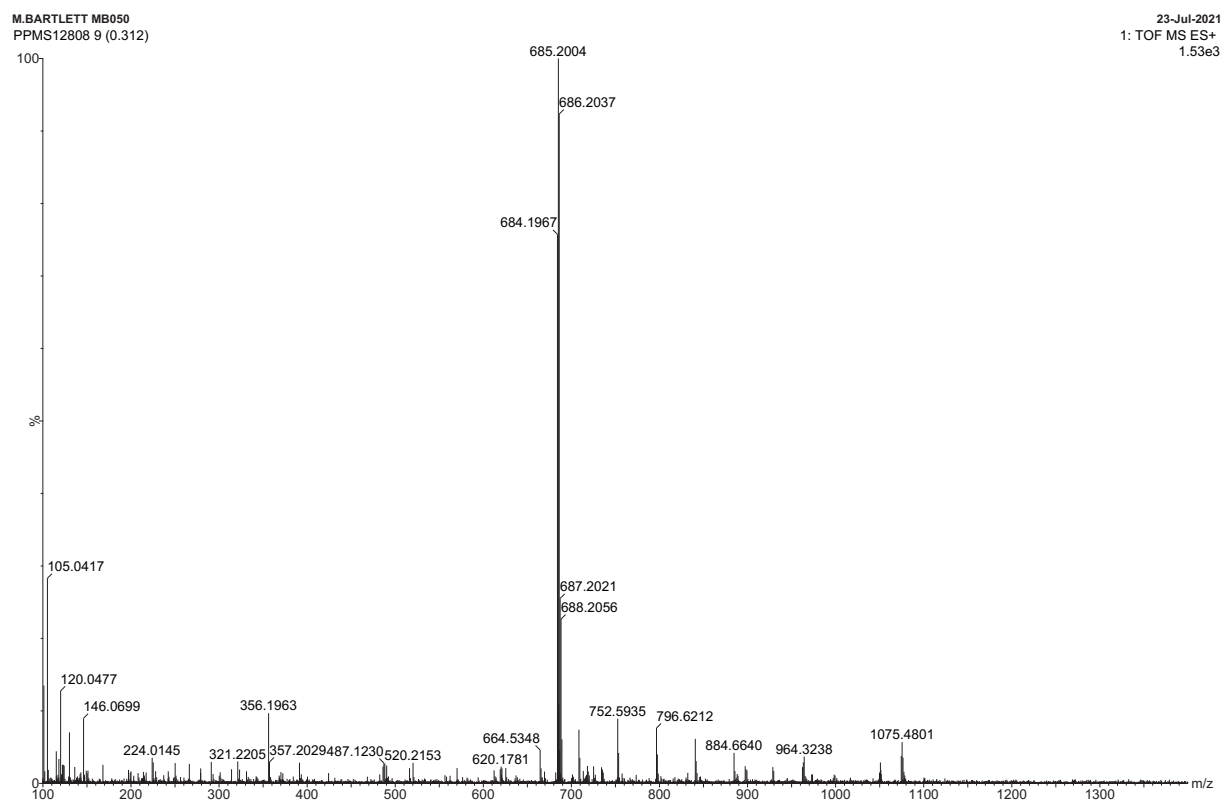


Figure A.34. HRMS of Pt_PPTMA_5OMe.

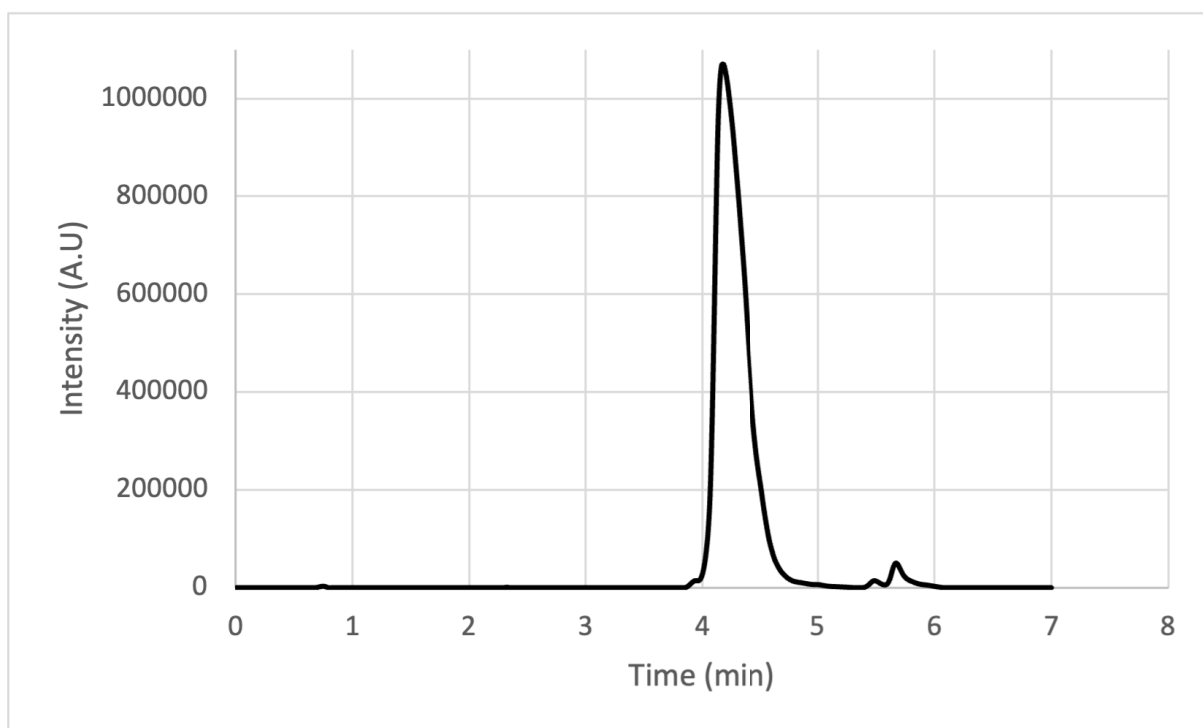


Figure A.35. LCMS UV trace of Pt_PPTMA_5OMe.

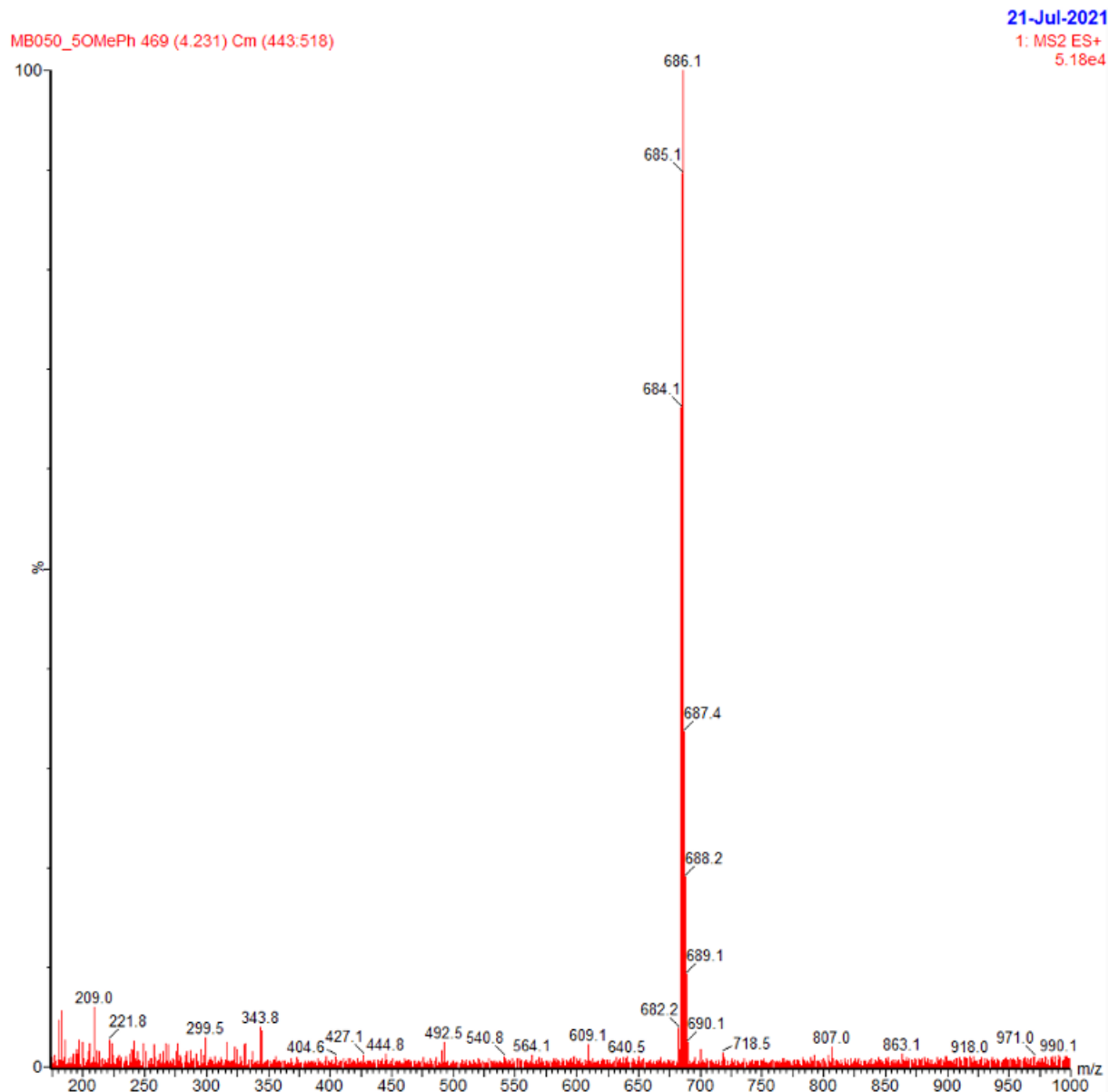
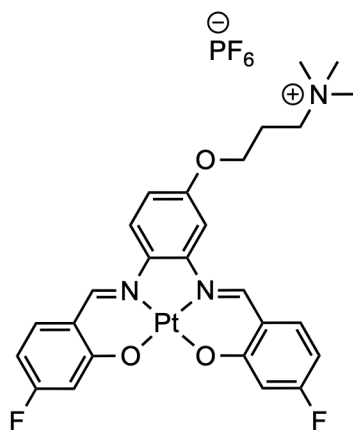


Figure A.36. Mass spec of Pt_PPTMA_5OMe LCMS under UV peak at 4.18 mins.

Pt_PPTMA_4F [hexafluorophosphate]:



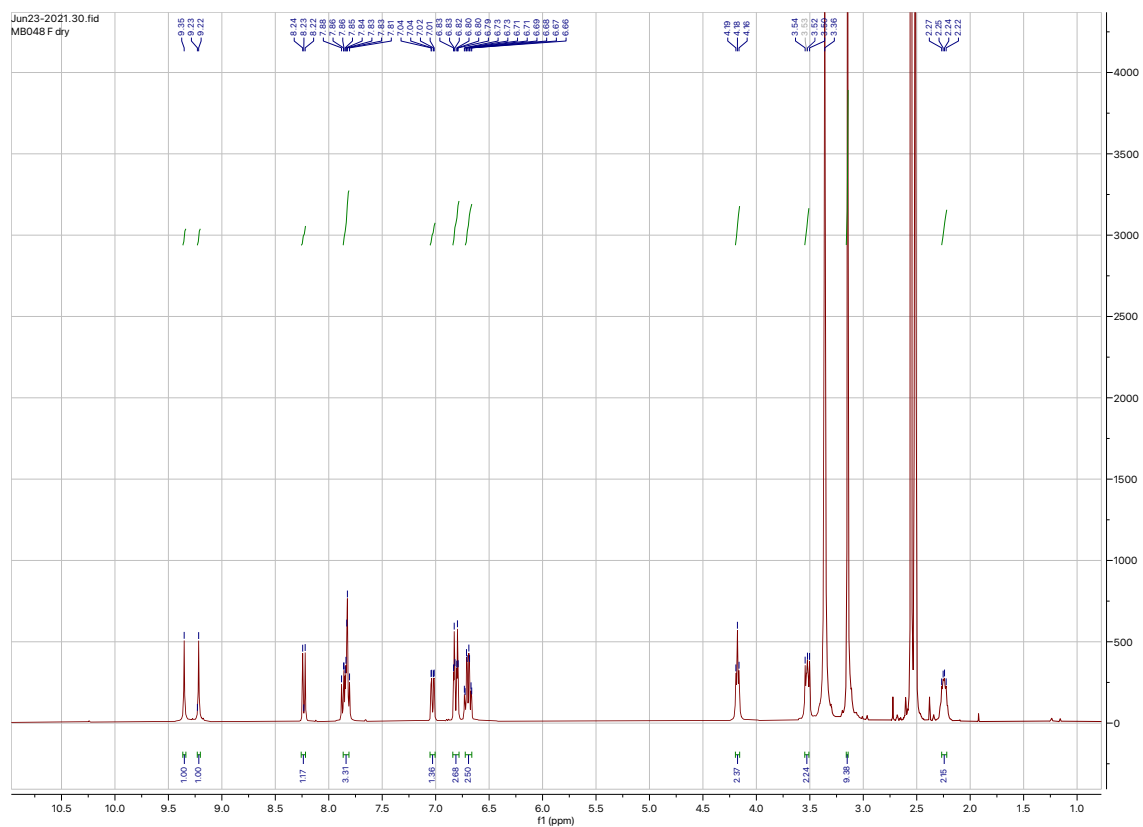


Figure A.37. Full ^1H NMR spectrum of Pt_PPTMA_4F.

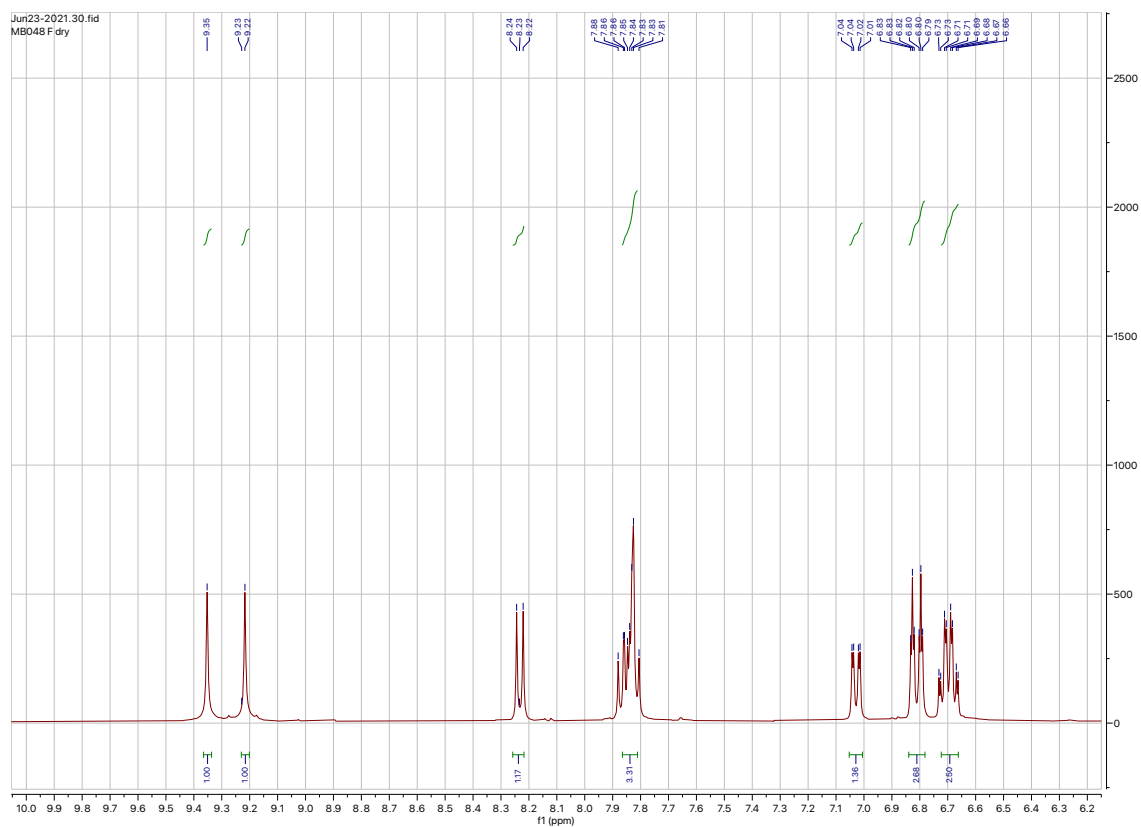
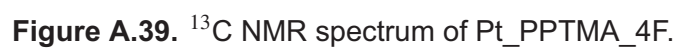


Figure A.38. Aromatic region of ^1H NMR spectrum for Pt_PPTMA_4F.



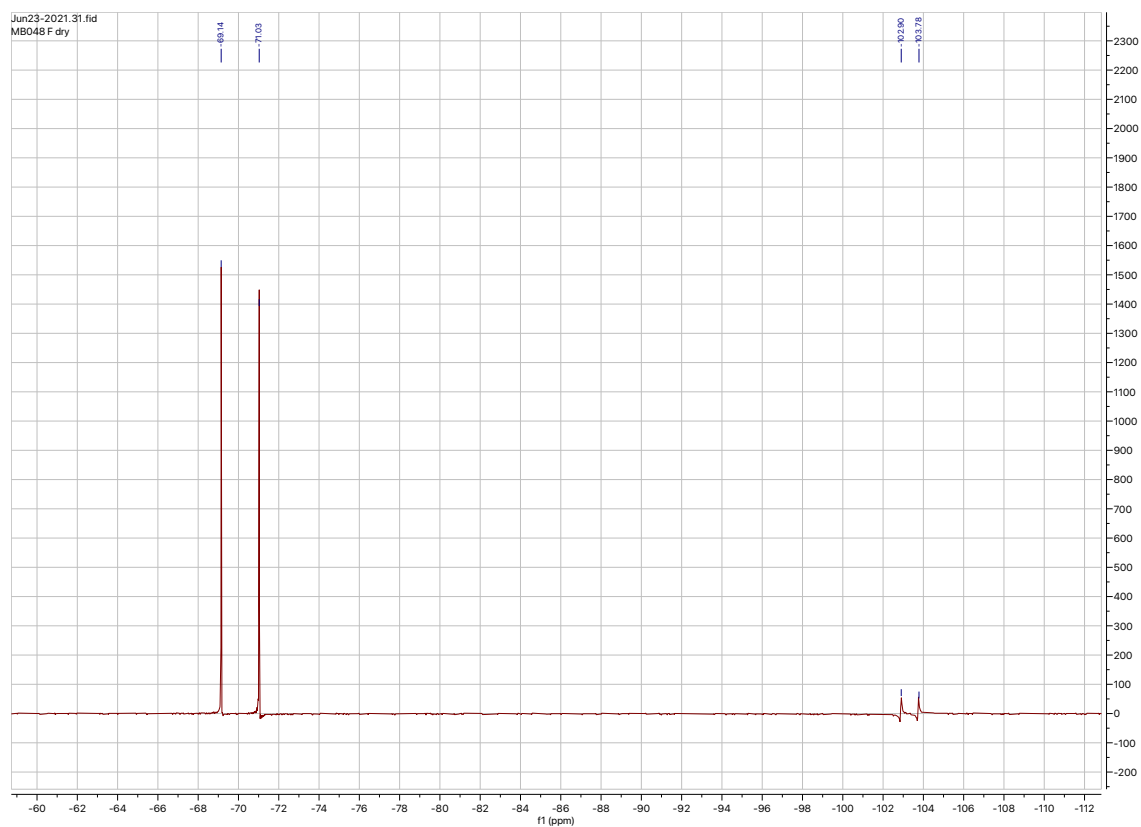


Figure A.40. ^{19}F NMR spectrum of Pt_PPTMA_4F.

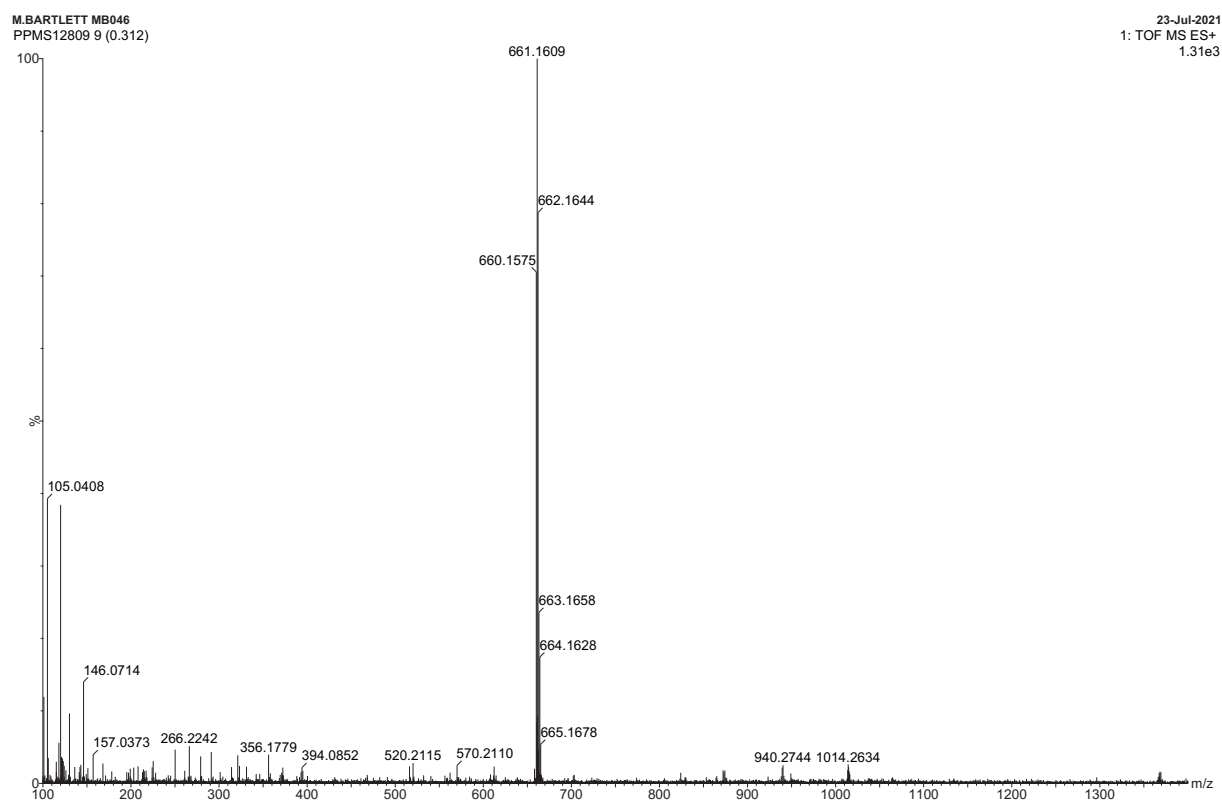


Figure A.41. HRMS of Pt_PPTMA_4F.

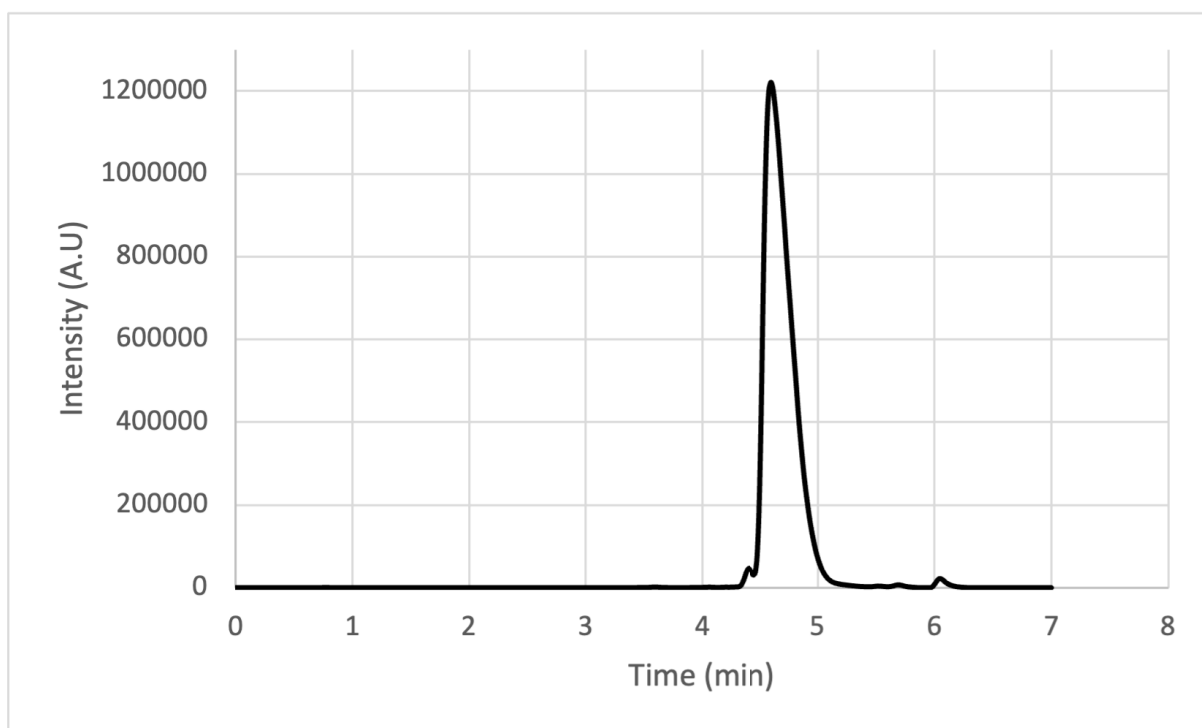


Figure A.42. LCMS UV trace of Pt_PPTMA_4F.

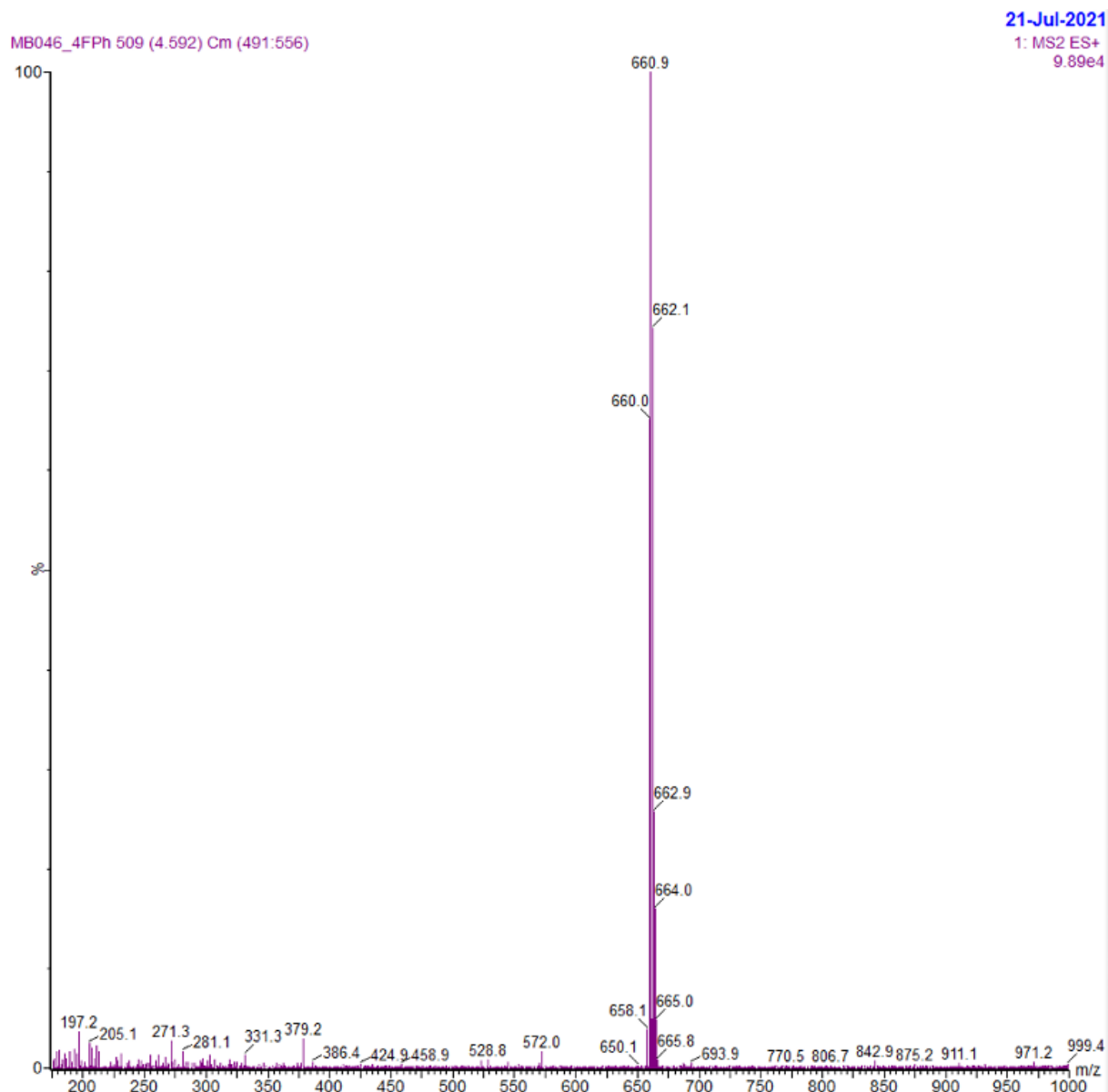
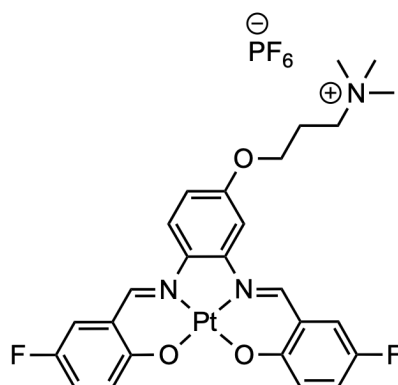


Figure A.43. Mass spec of Pt_PPTMA_4F under LCMS UV peak at 4.60 minutes.

Pt_PPTMA_5F [hexafluorophosphate]:



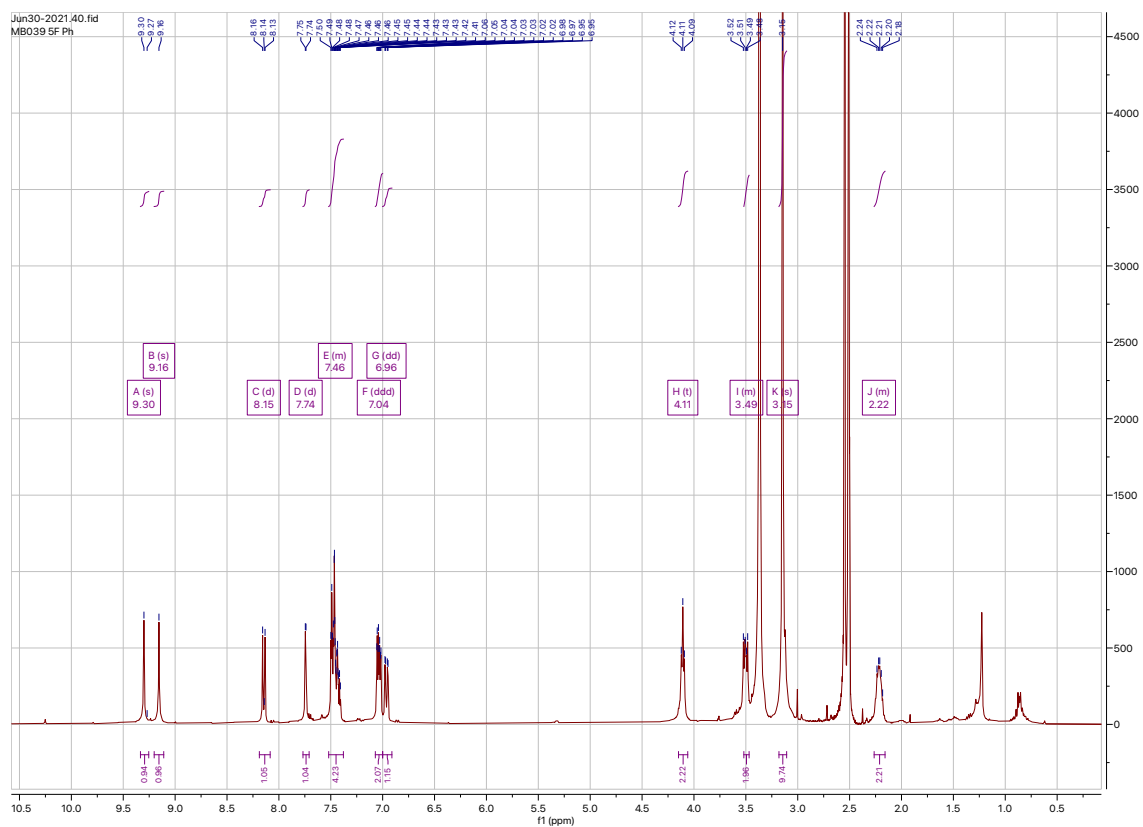


Figure A.44. Full ^1H NMR spectrum of Pt_PPTMA_5F.

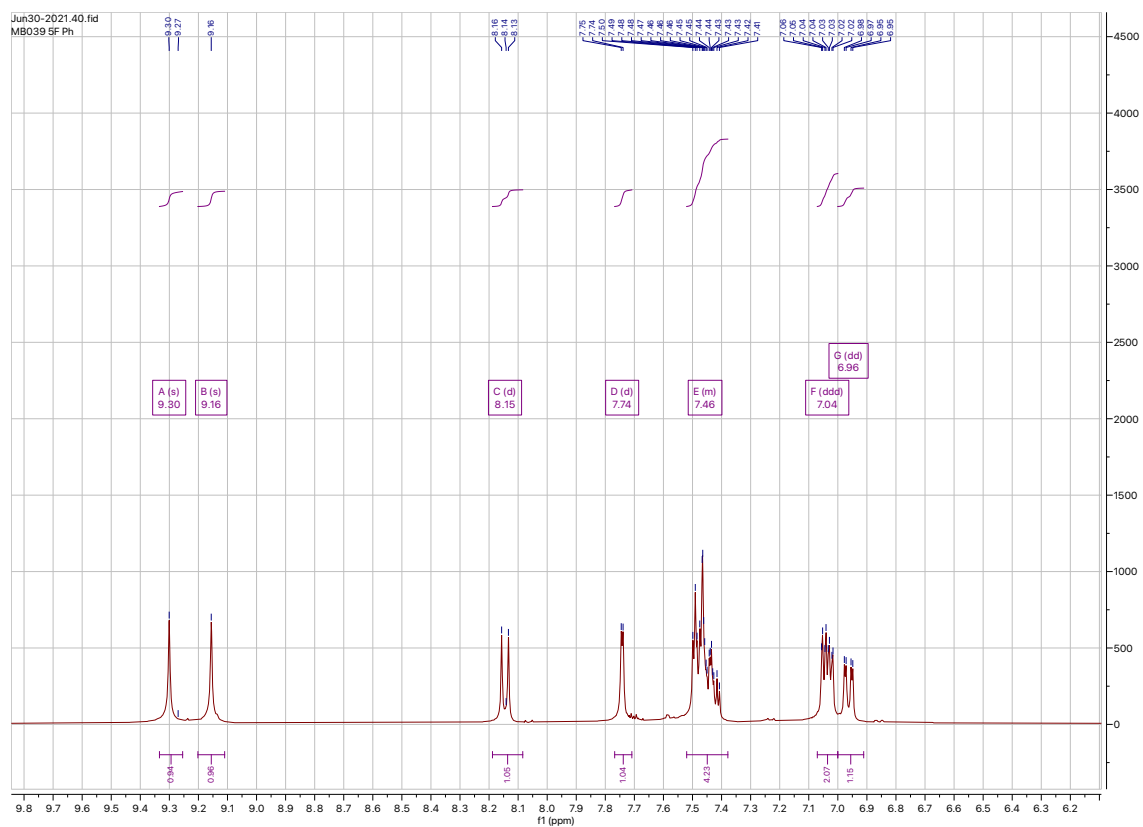


Figure A.45. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_5F.

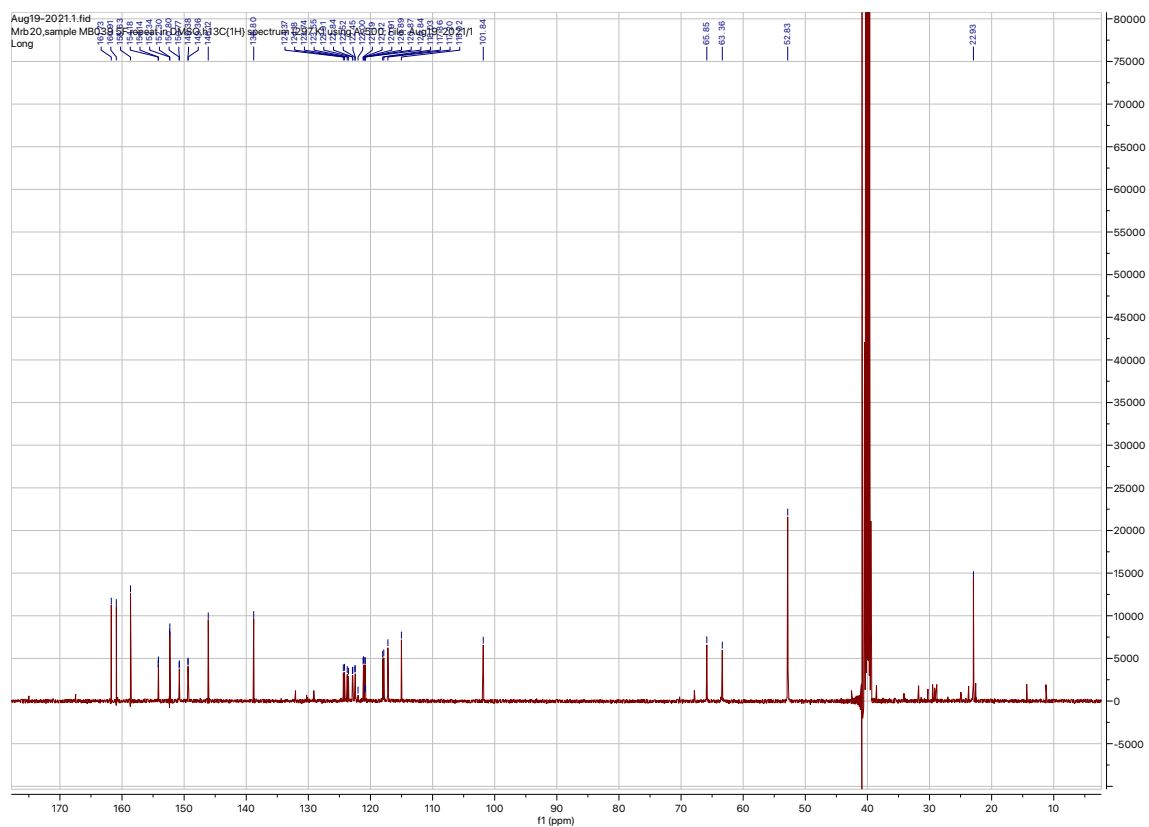


Figure A.46. ^{13}C NMR spectrum of Pt_PPTMA_5F.

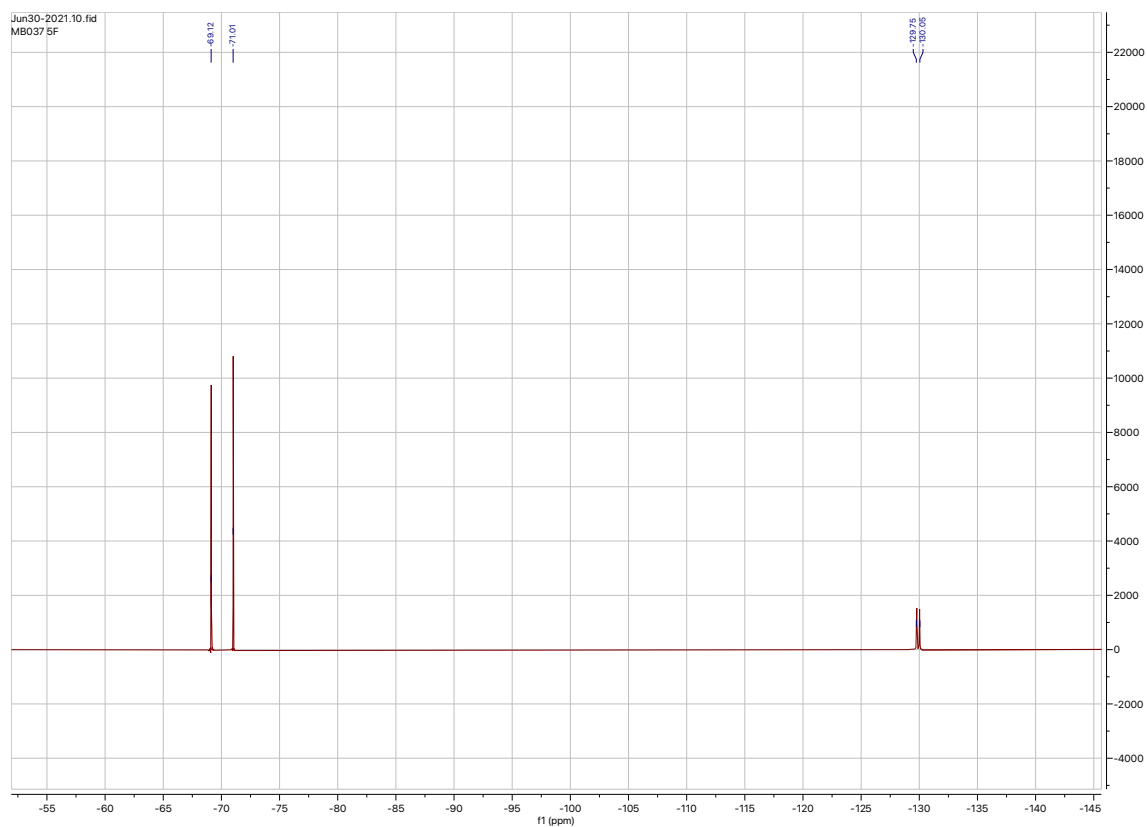


Figure A.47. ^{19}F NMR spectrum of Pt_PPTMA_5F.

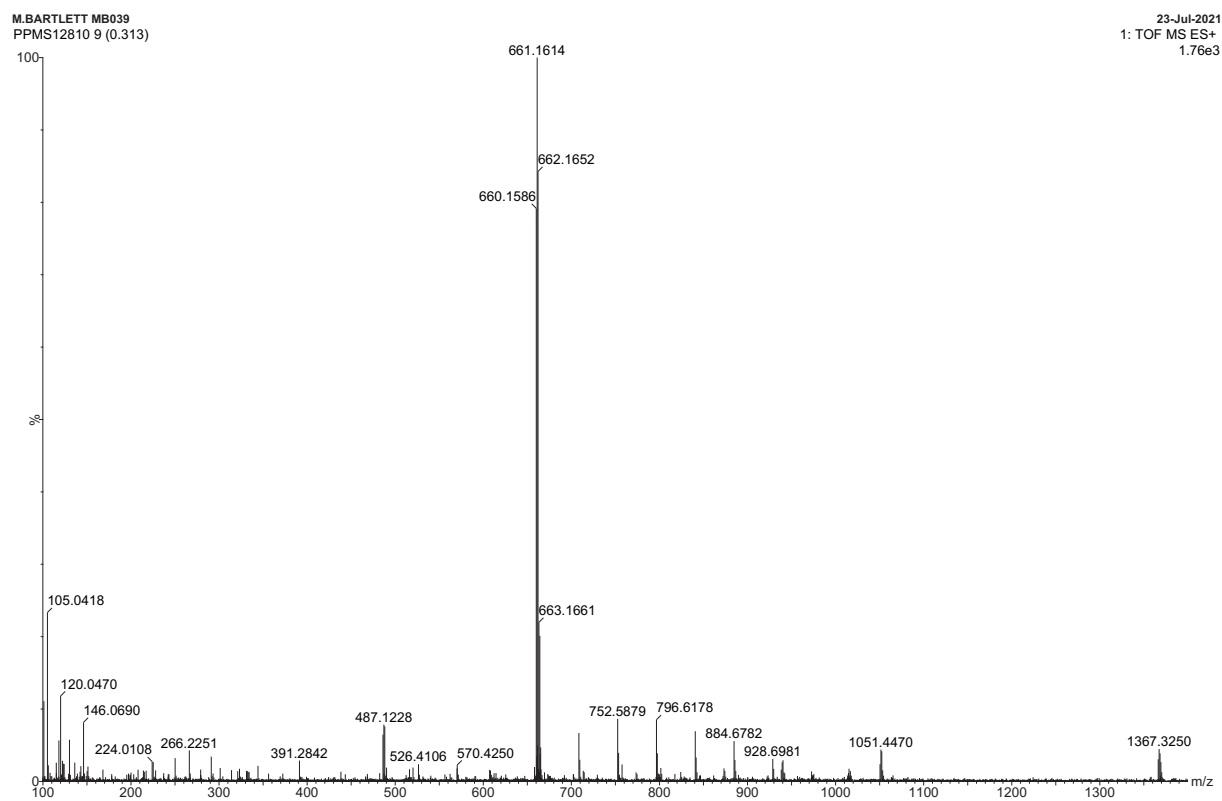


Figure A.48. HRMS of Pt_PPTMA_5F.

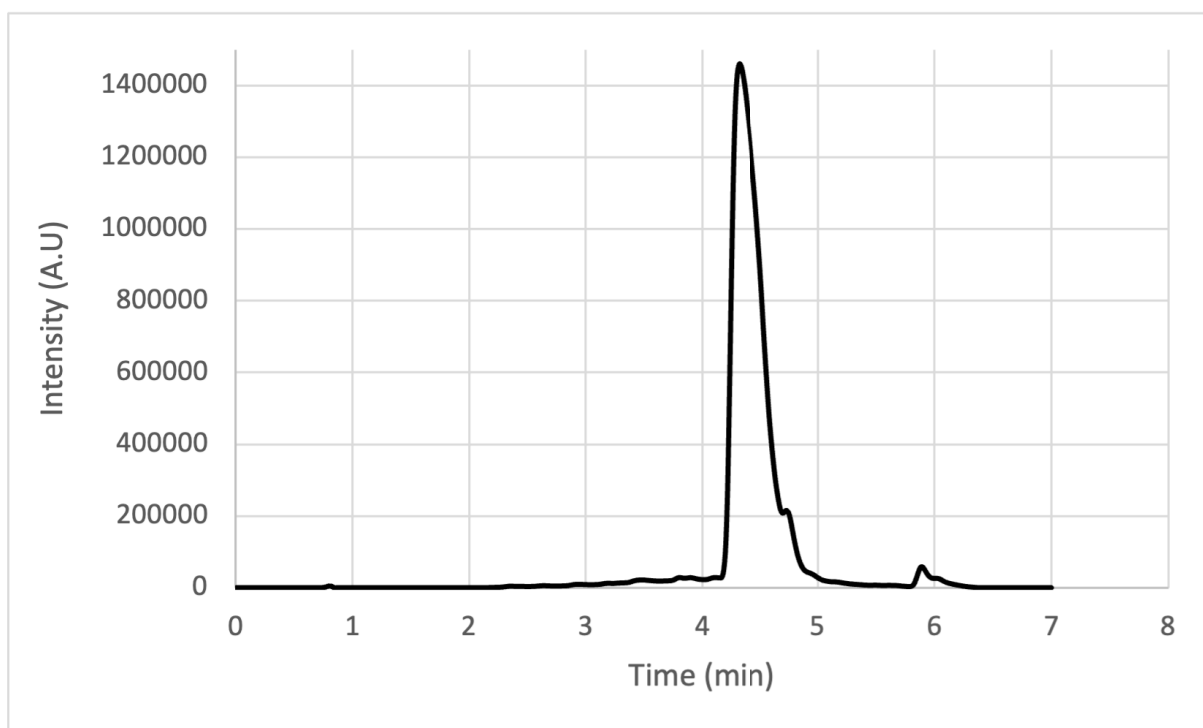


Figure A.49. LCMS UV trace of Pt_PPTMA_5F.

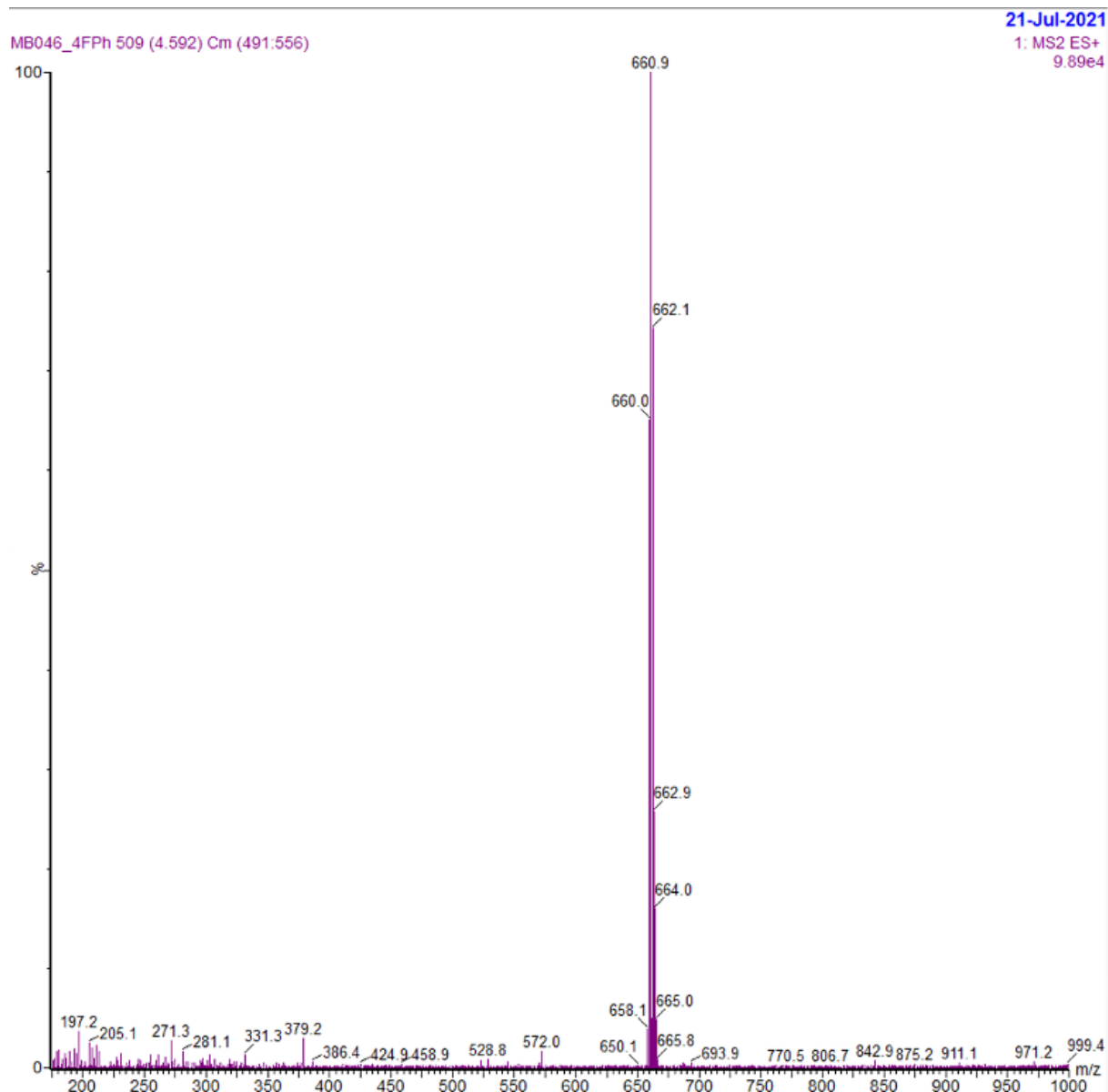
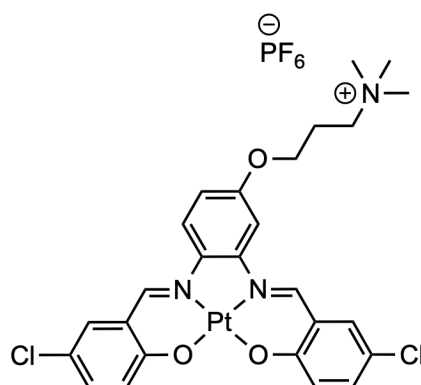


Figure A.50. Mass spec of Pt_PPTMA_5F under LCMS UV peak at 4.32 minutes.

Pt_PPTMA_5Cl [hexafluorophosphate]:



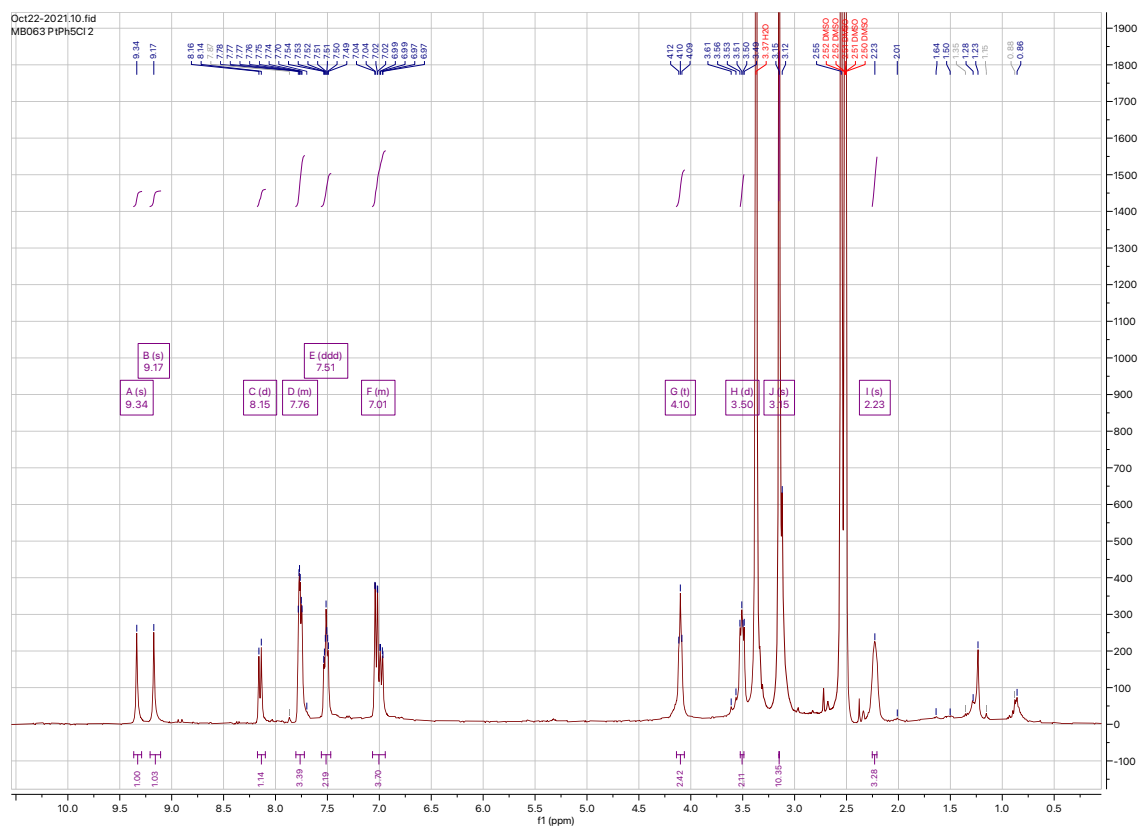


Figure A.51. Full ^1H NMR spectrum of Pt_PPTMA_5Cl.

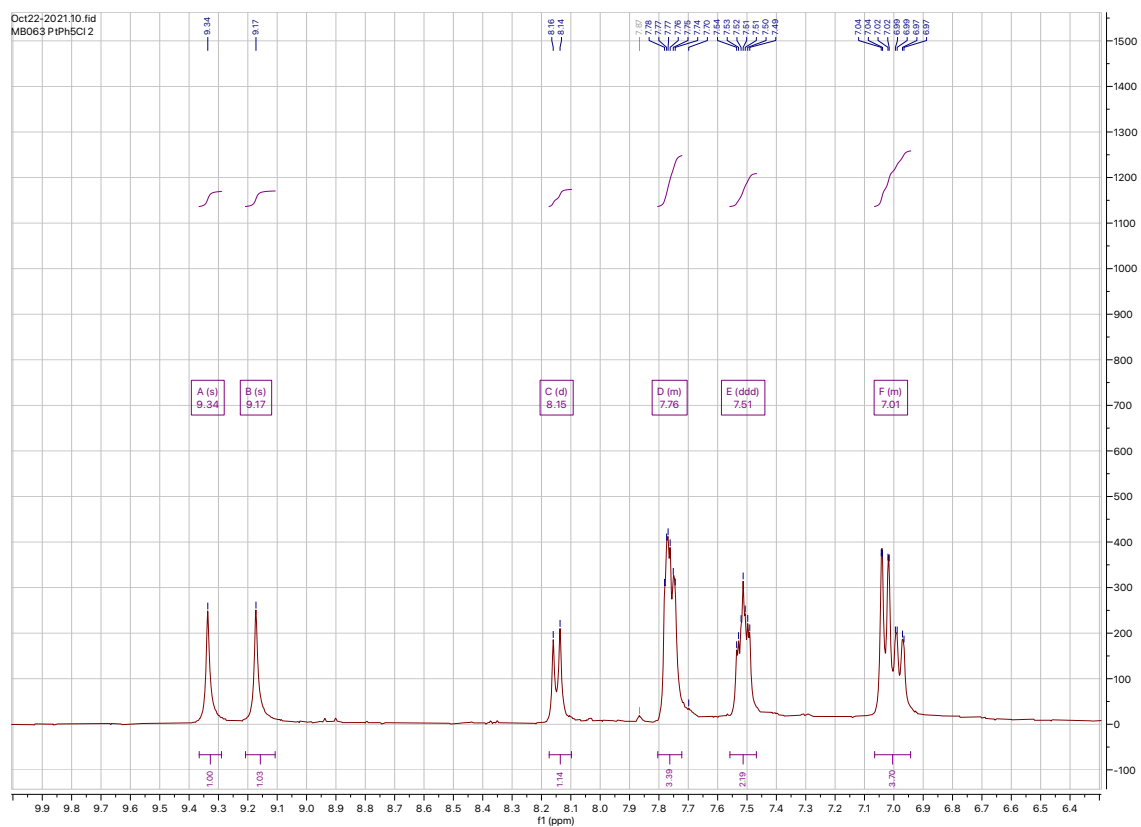


Figure A.52. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_5Cl.

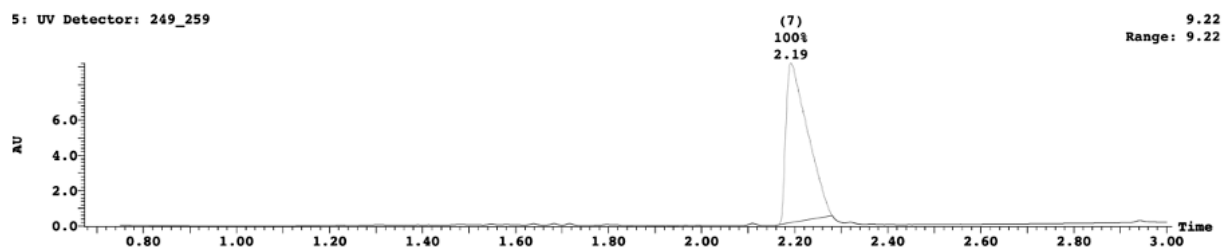


Figure A.55. LCMS UV trace of Pt_PPTMA_5Cl.

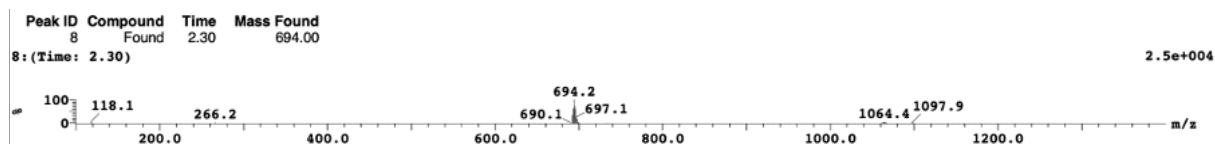
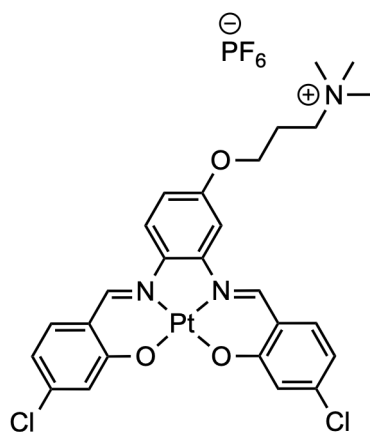


Figure A.56. Mass spec of Pt_PPTMA_5Cl under LCMS UV peak at 2.19 minutes.

Pt_PPTMA_4Cl [hexafluorophosphate]:



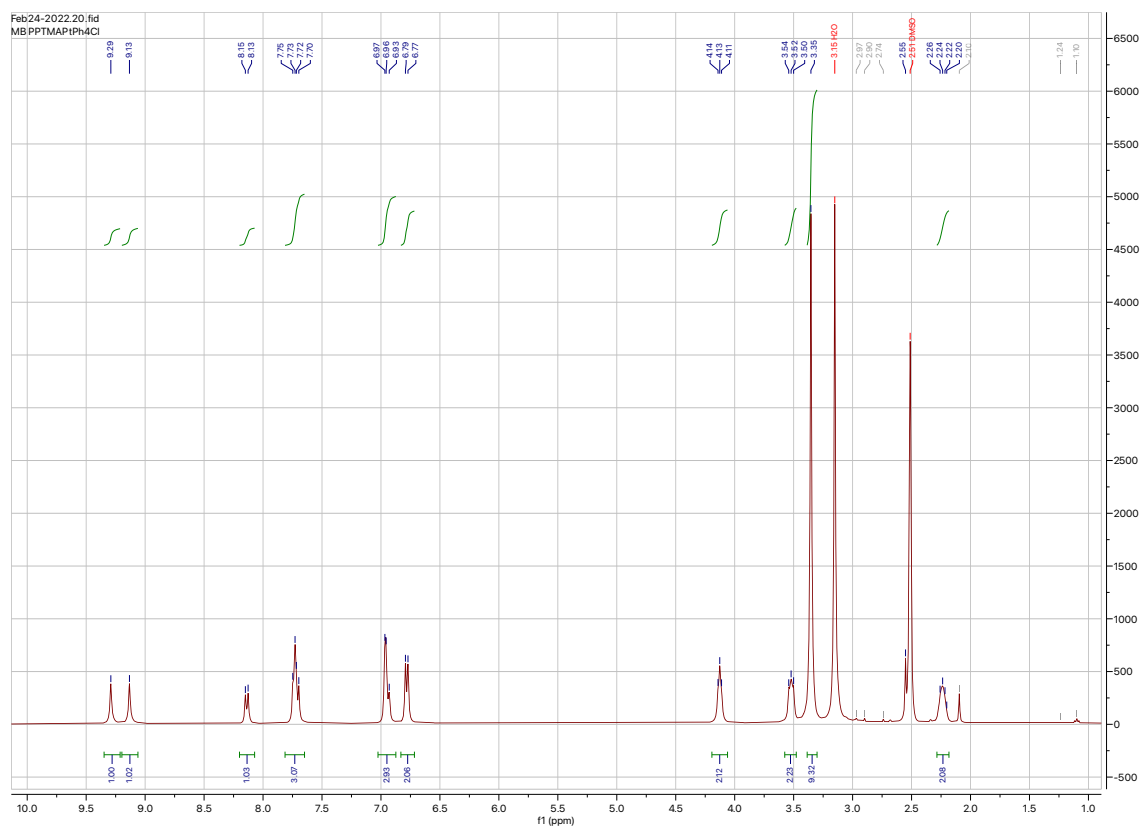


Figure A.57. Full ^1H NMR spectrum of Pt_PPTMA_4Cl.

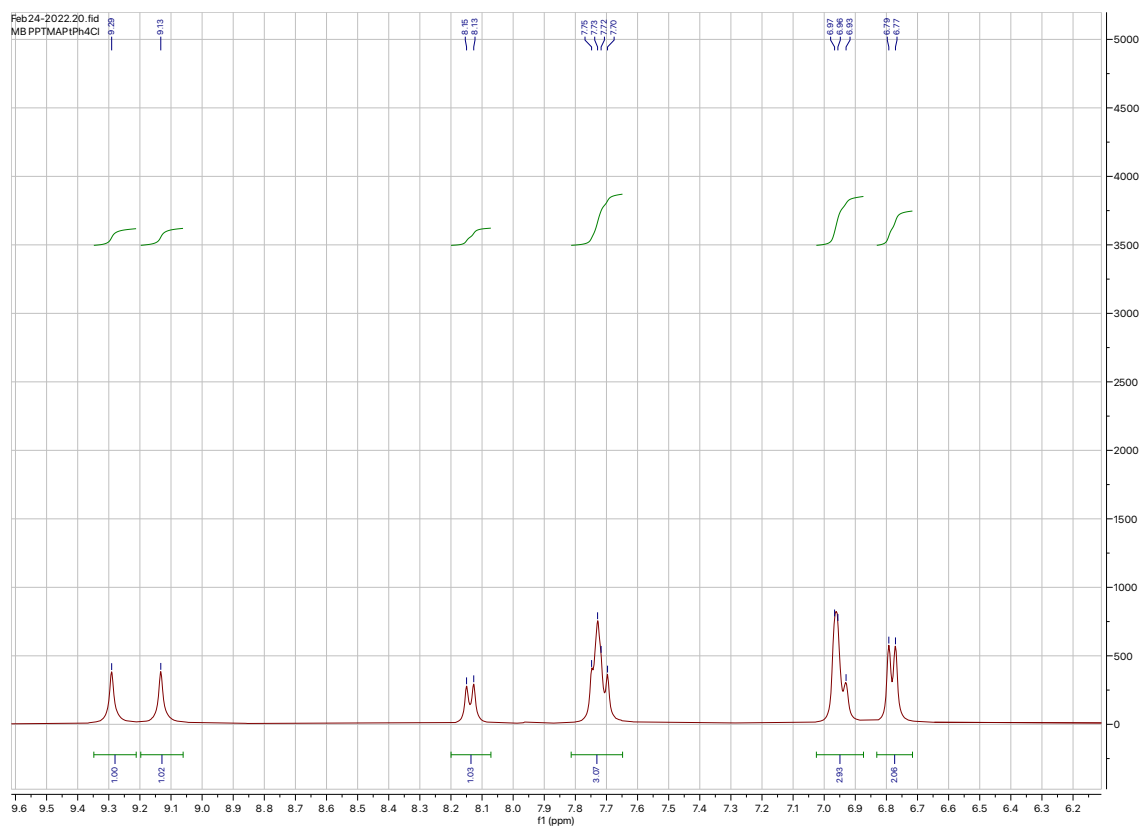


Figure A.58. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_4Cl.

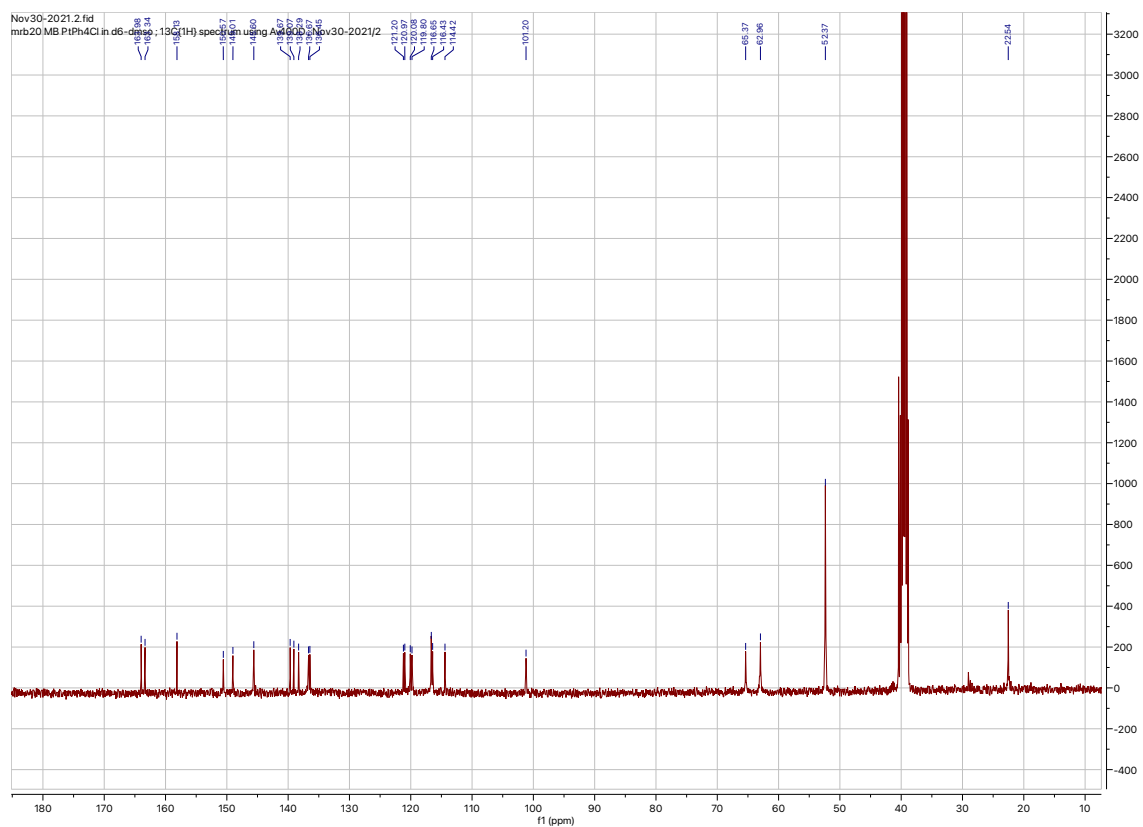


Figure A.59. ^{13}C NMR spectrum of Pt_PPTMA_4Cl.

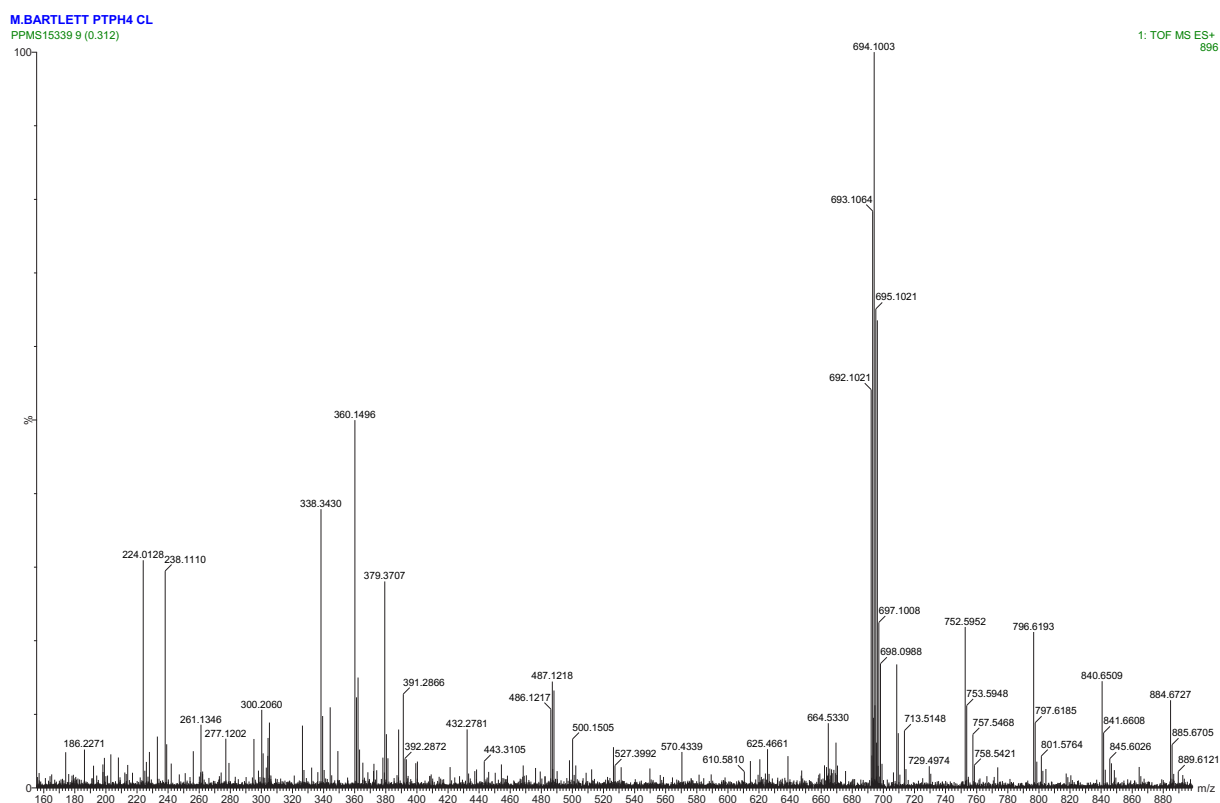


Figure A.60. HRMS of Pt_PPTMA_4Cl. PEG impurities seen from machine contamination, not sample.

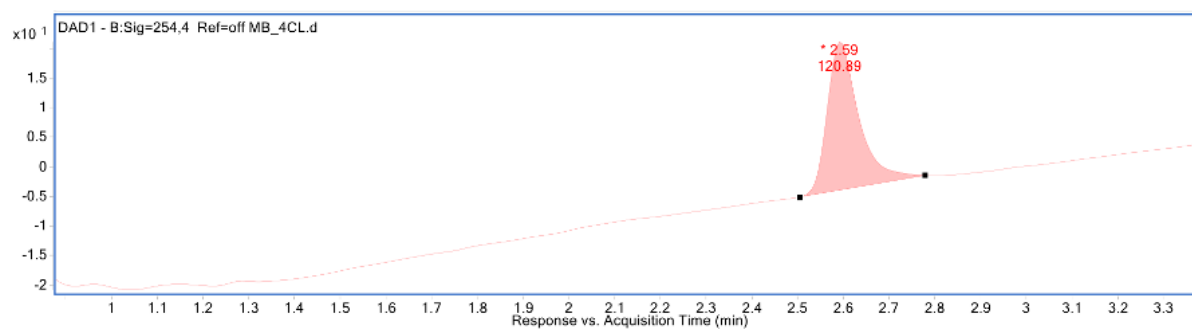


Figure A.61. LCMS UV vis trace of Pt_PPTMA_4Cl.

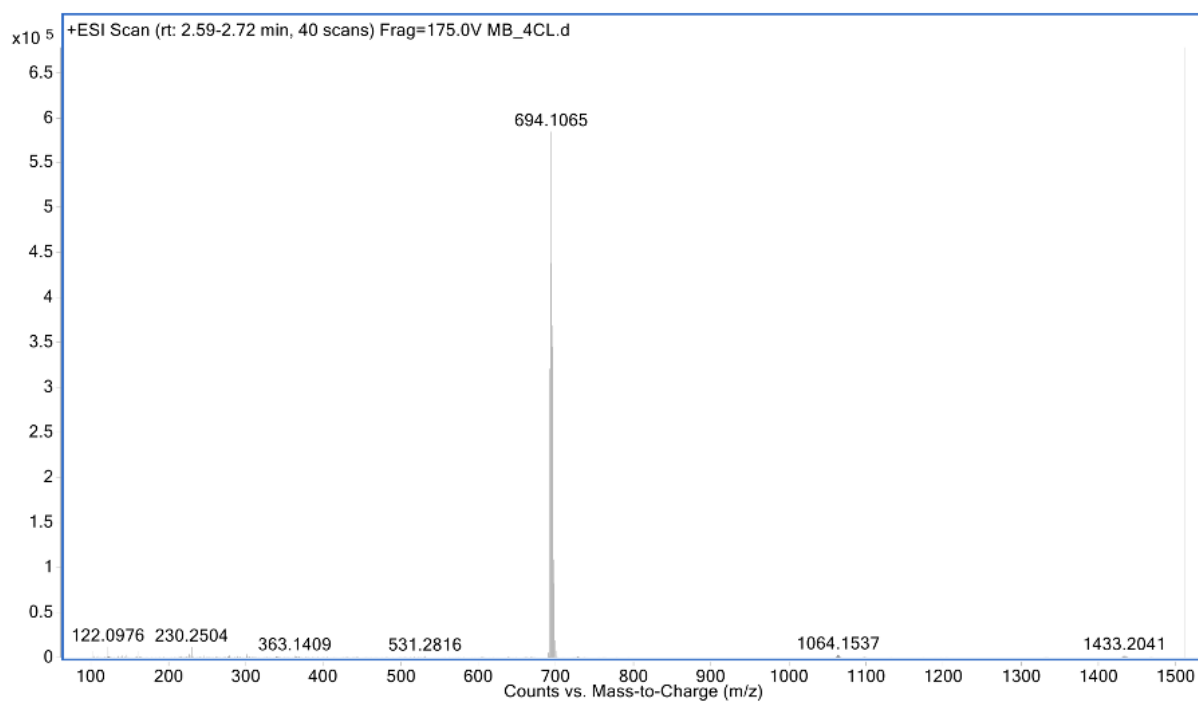
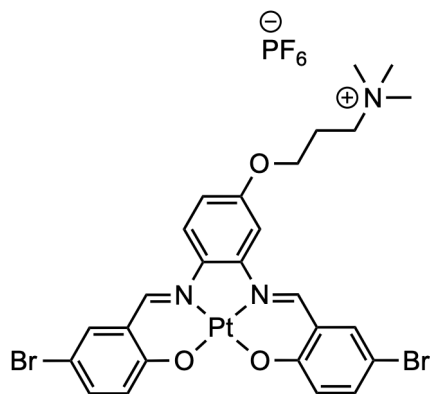


Figure A.62. Mass spec of Pt_PPTMA_4Cl under LCMS UV peak at 2.59 minutes.

Pt_PPTMA_5Br [hexafluorophosphate]:



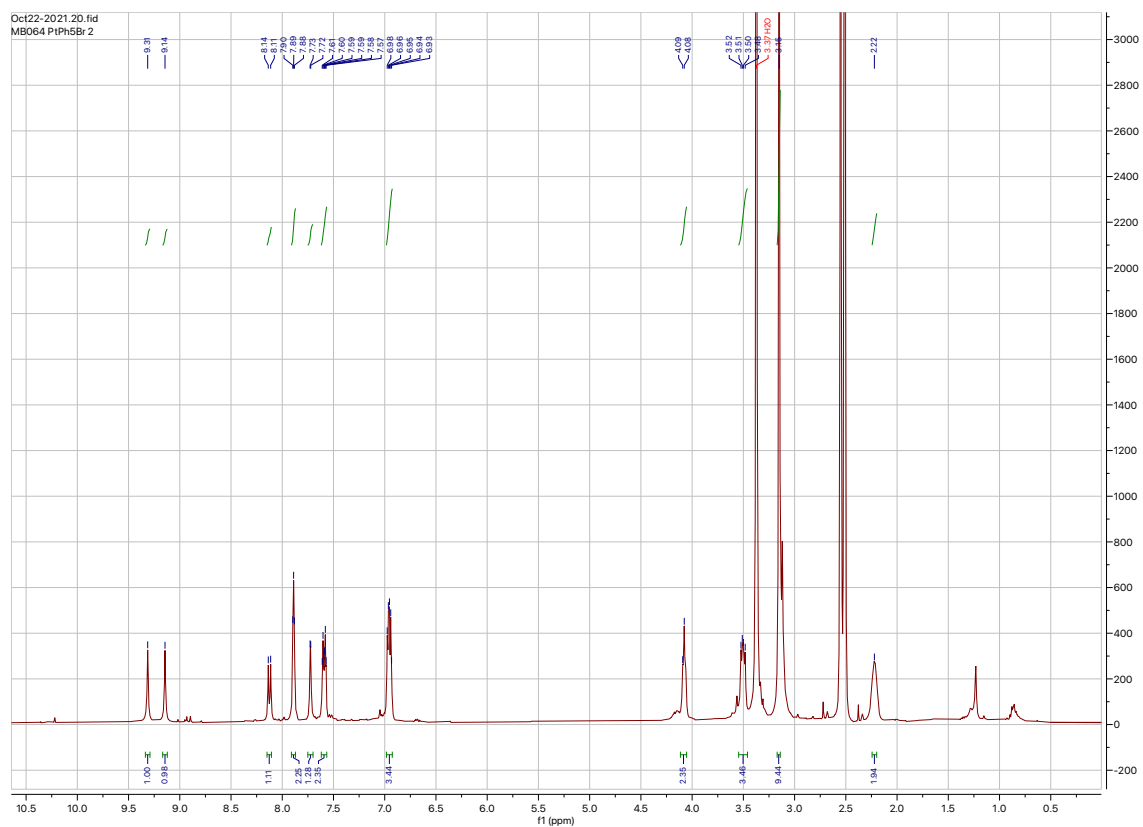


Figure A.63. Full ^1H NMR spectrum of Pt_PPTMA_5Br.

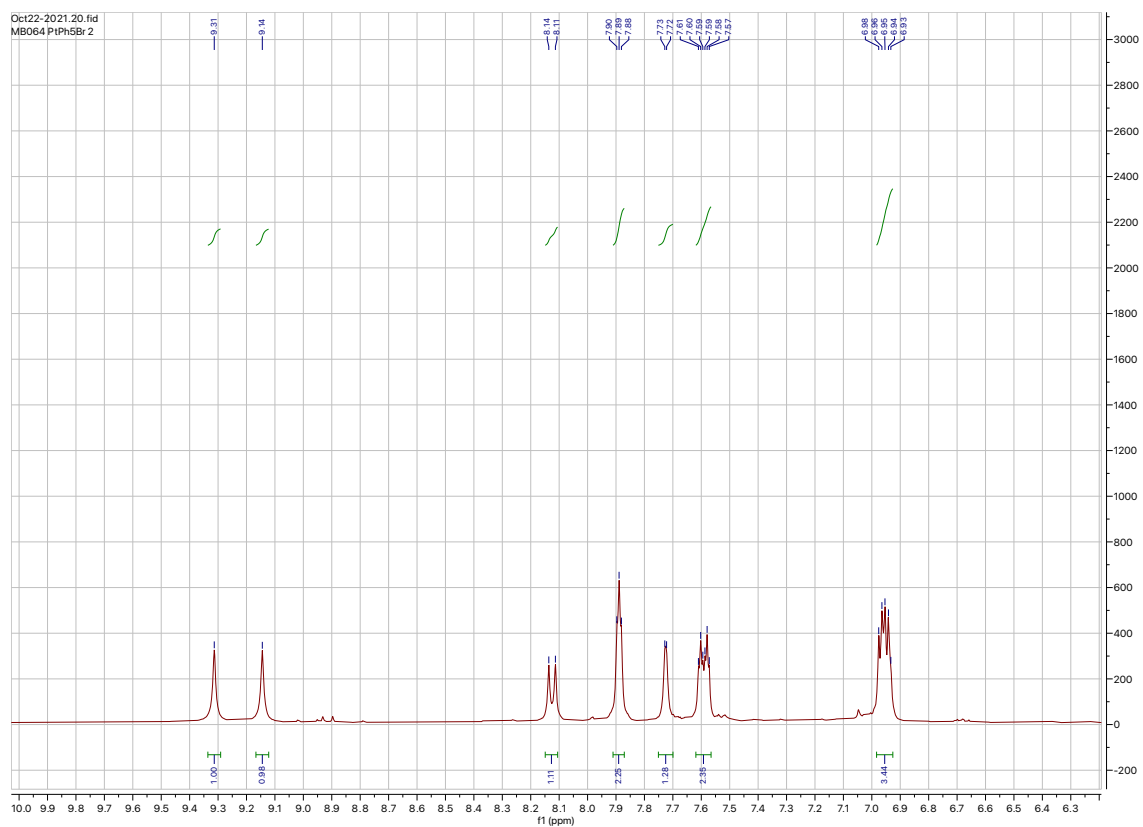


Figure A.64. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_5Br.

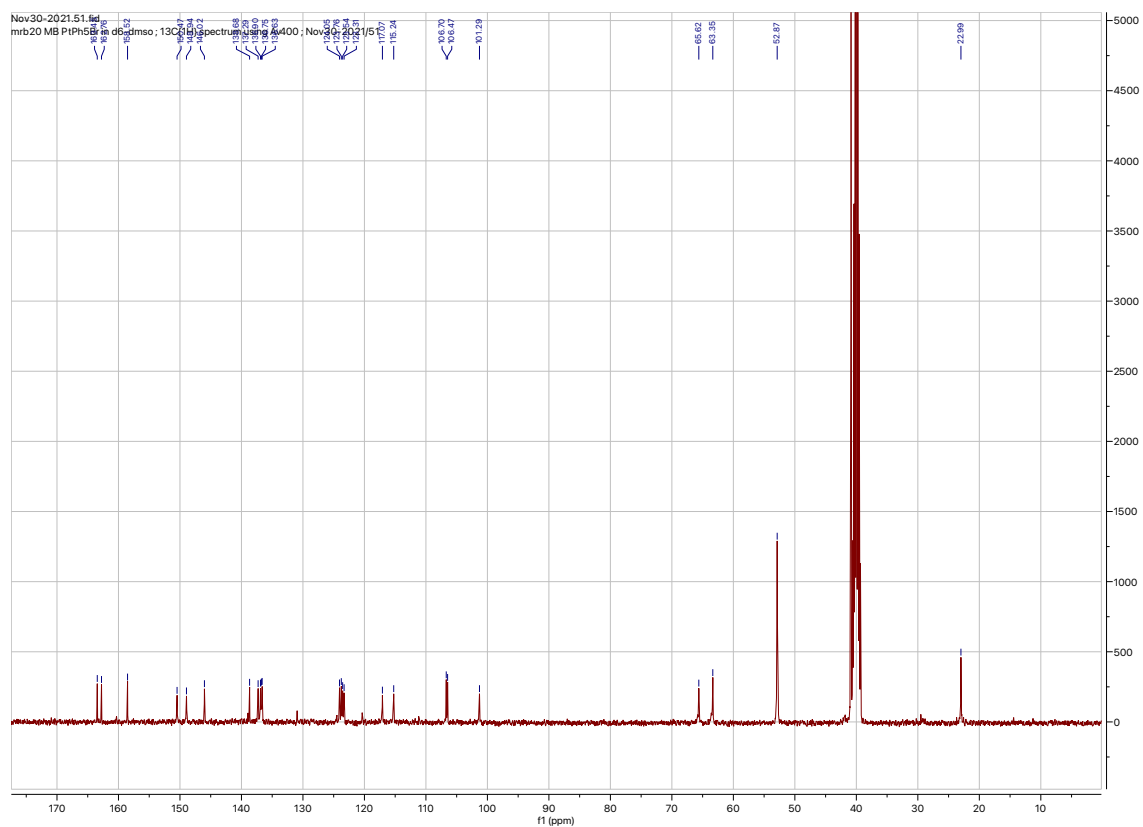


Figure A.65. ^{13}C NMR spectrum of Pt_PPTMA_5Br.

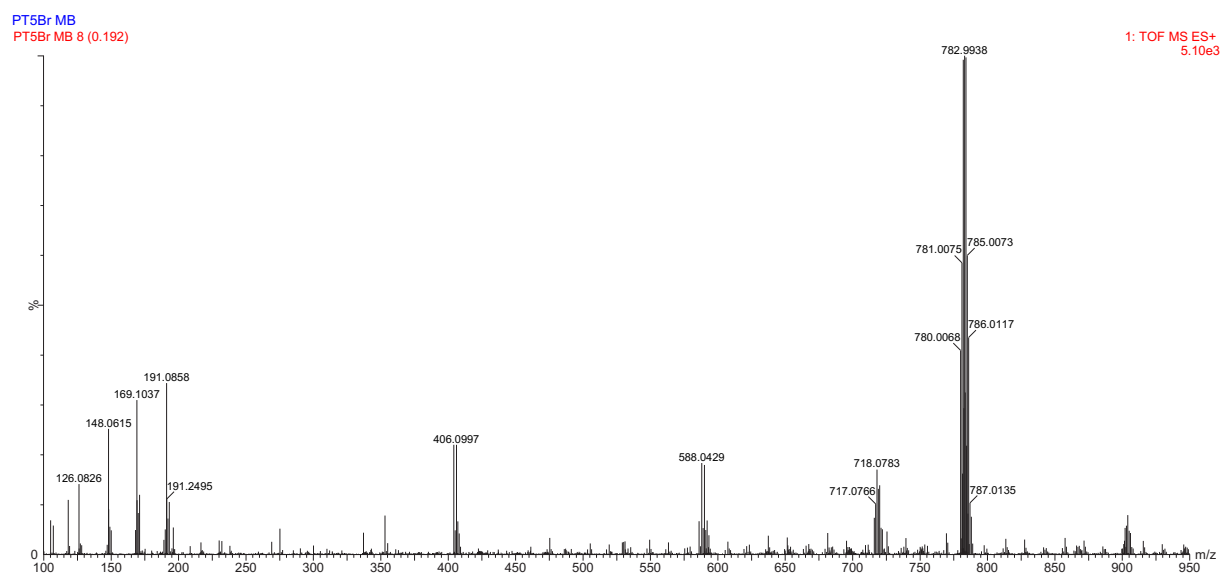


Figure A.66. HRMS of Pt_PPTMA_5Br.

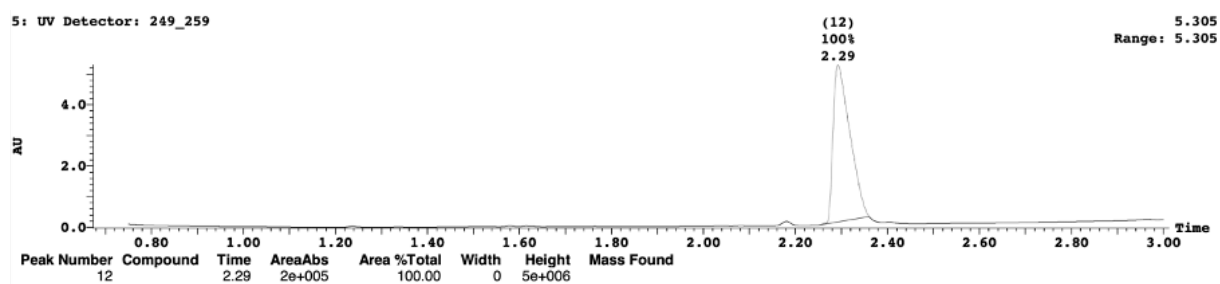


Figure A.67. LCMS UV trace of Pt_PPTMA_5Br.

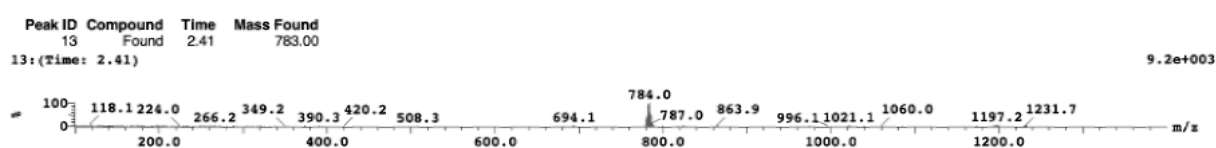
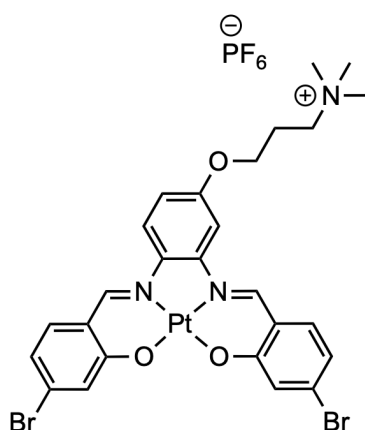


Figure A.68. Mass spec of Pt_PPTMA_5Br of LCMS UV peak at 2.29 minutes.

Pt_PPTMA_4Br [hexafluorophosphate]:



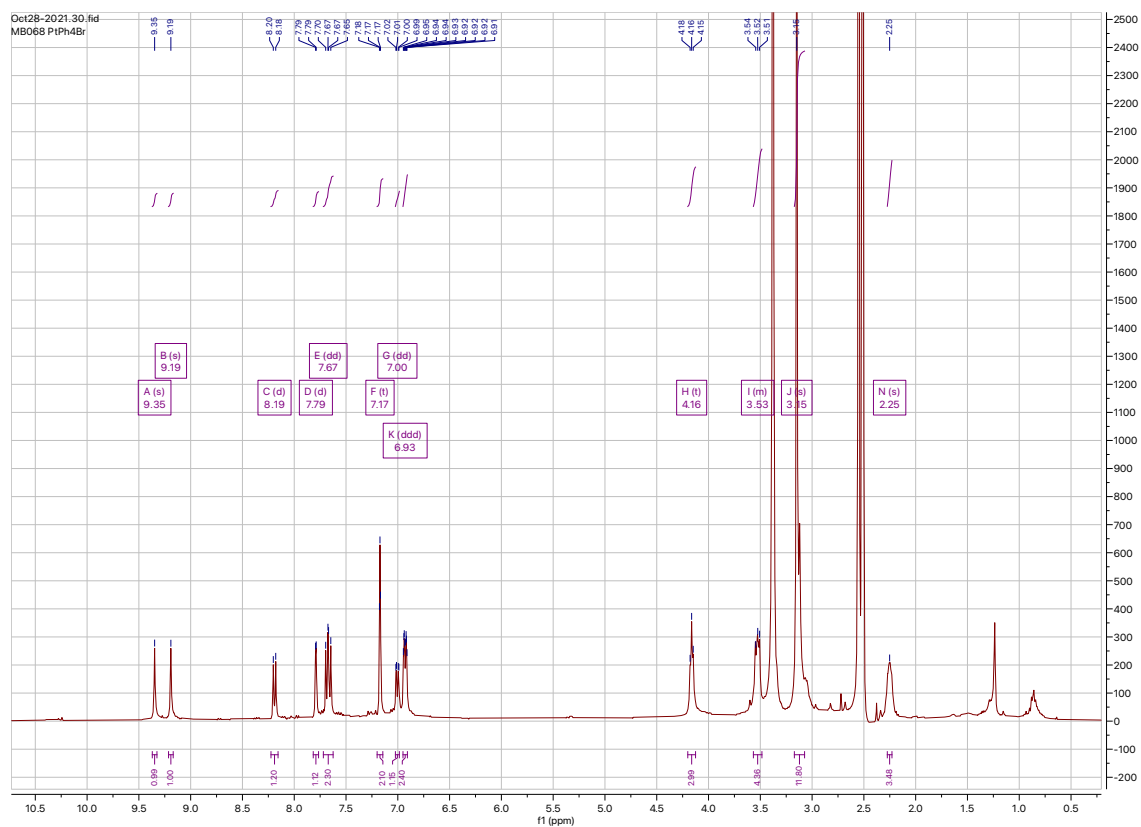


Figure A.69. Full ^1H NMR spectrum of Pt_PPTMA_4Br.

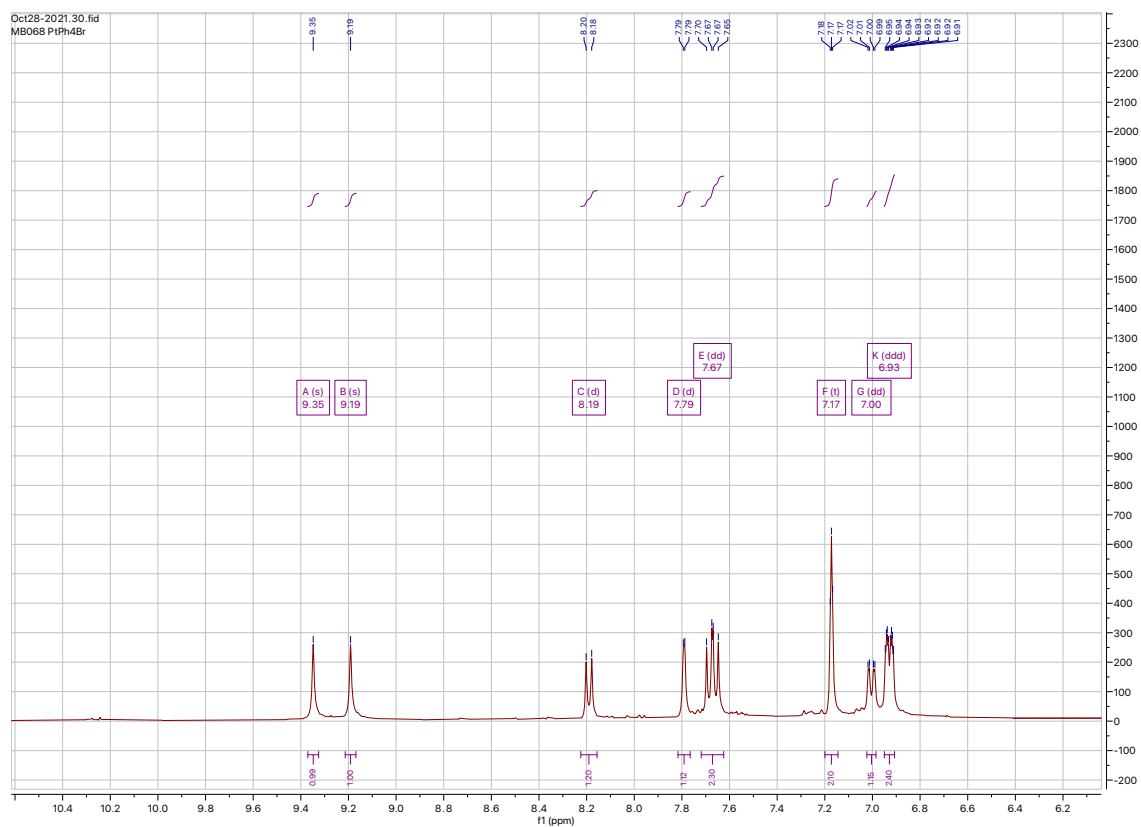


Figure A.70. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_4Br.



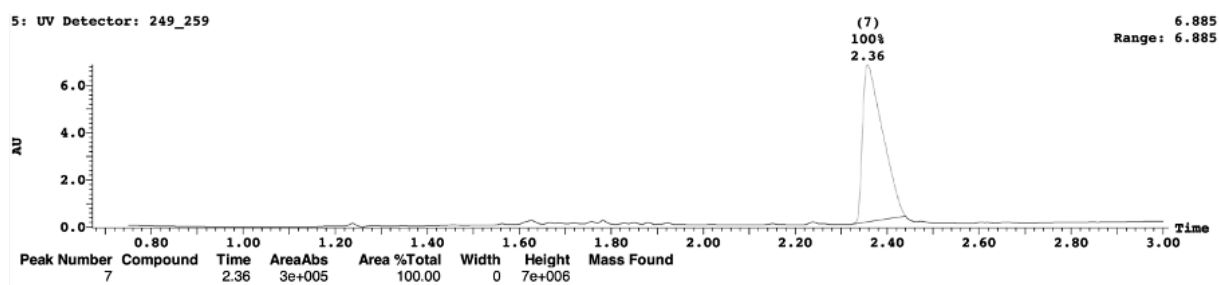


Figure A.73. LCMS UV trace of Pt_PPTMA_4Br.

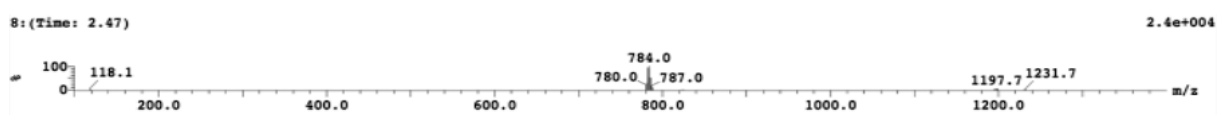
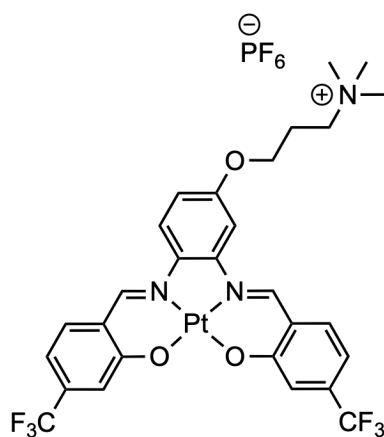


Figure A.74. Mass spec of Pt_PPTMA_4Br under LCMS UV peak at 2.36 minutes.

Pt_PPTMA_4CF₃ [hexafluorophosphate]:



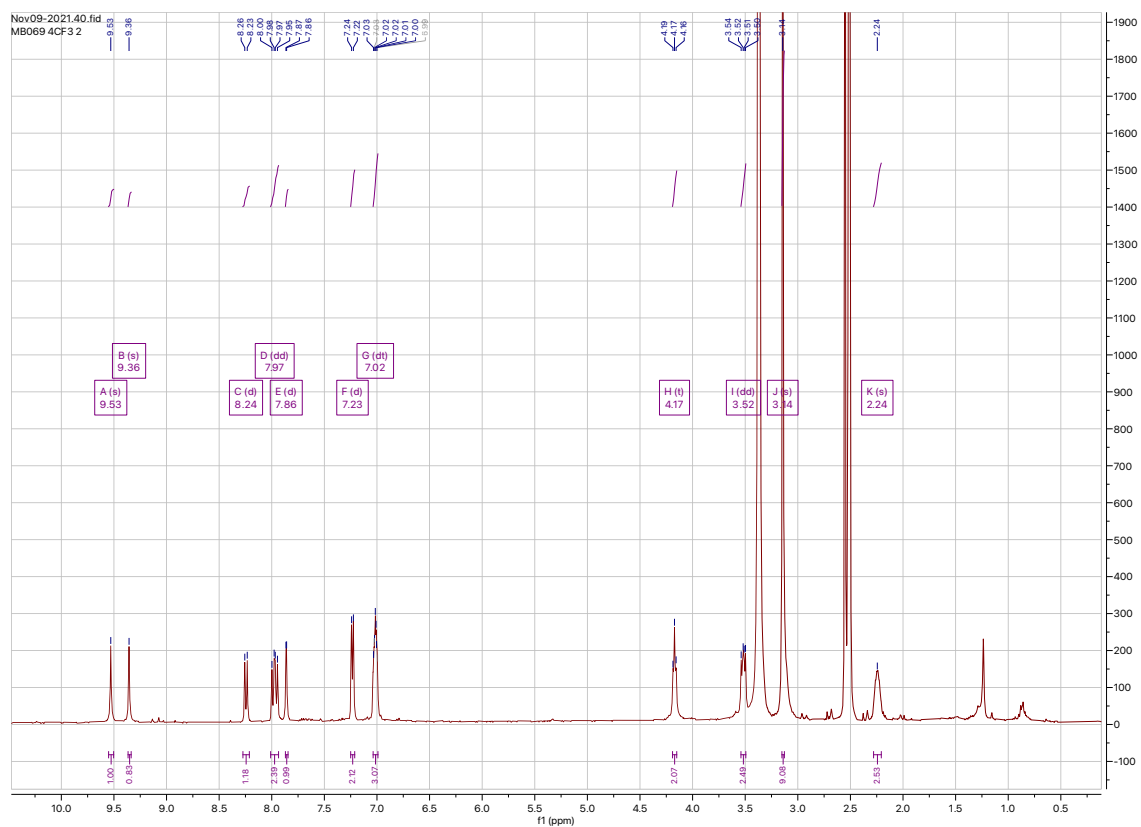


Figure A.75. Full ^1H NMR spectrum of Pt_PPTMA_4CF₃.

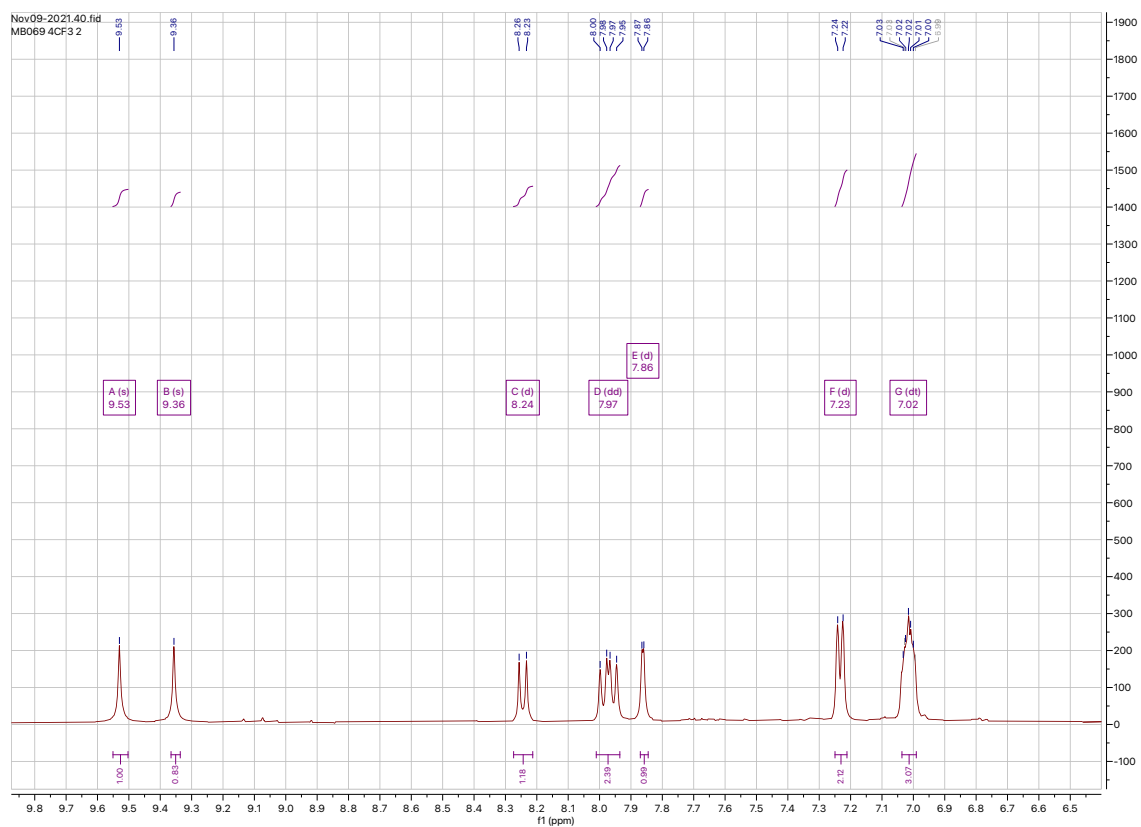


Figure A.76. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_4CF₃.

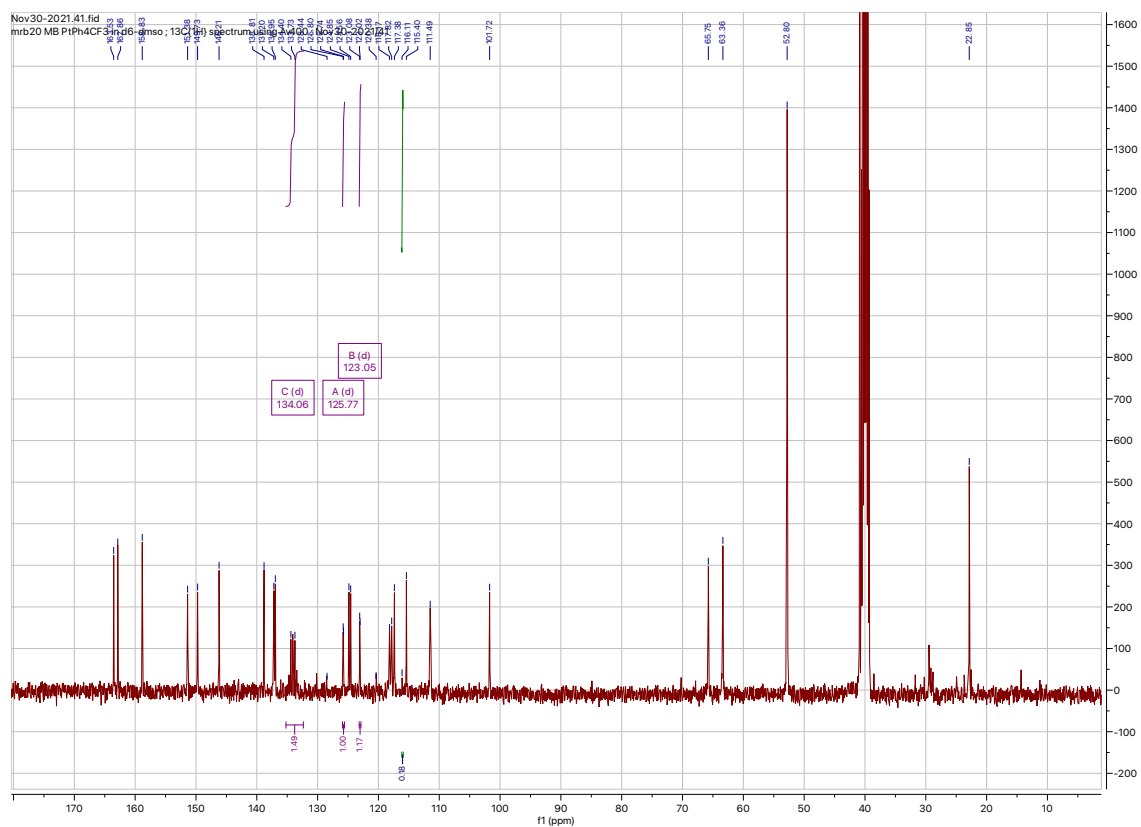


Figure A.77. ^{13}C NMR spectrum of Pt_PPTMA_4CF₃.

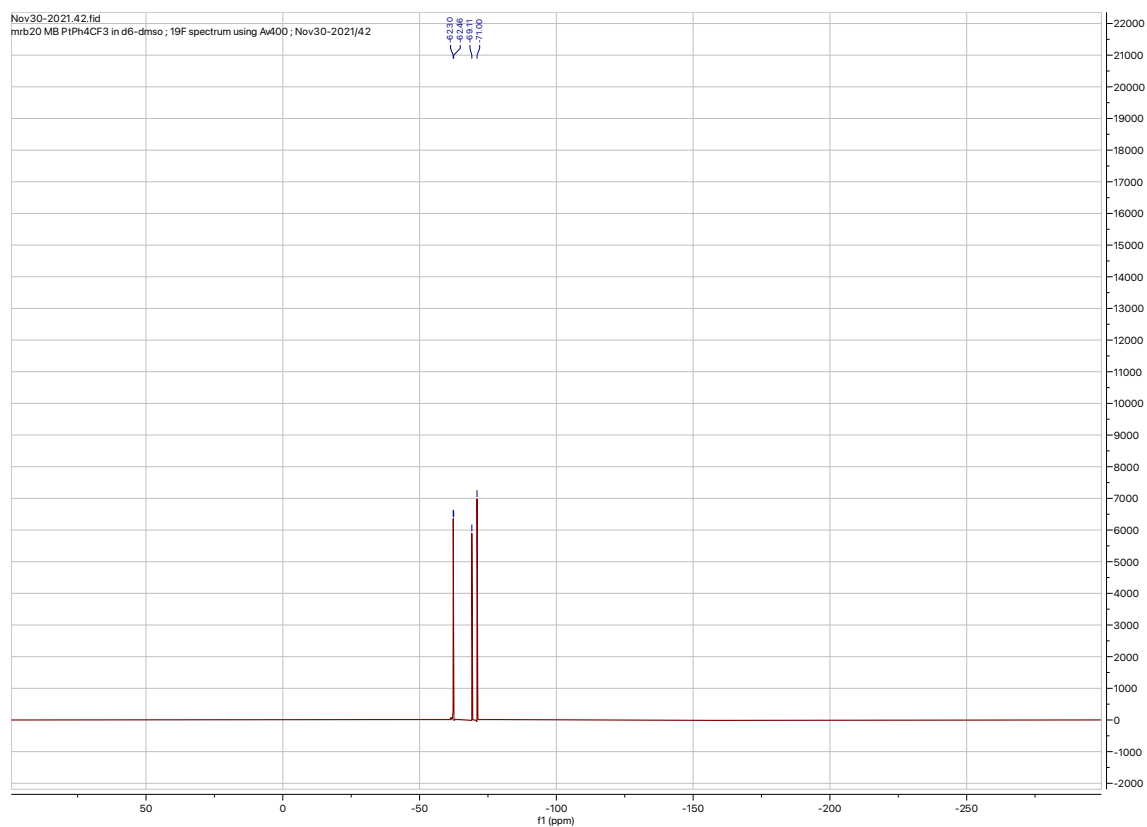


Figure A.78. ^{19}F NMR spectrum of Pt_PPTMA_4CF₃.

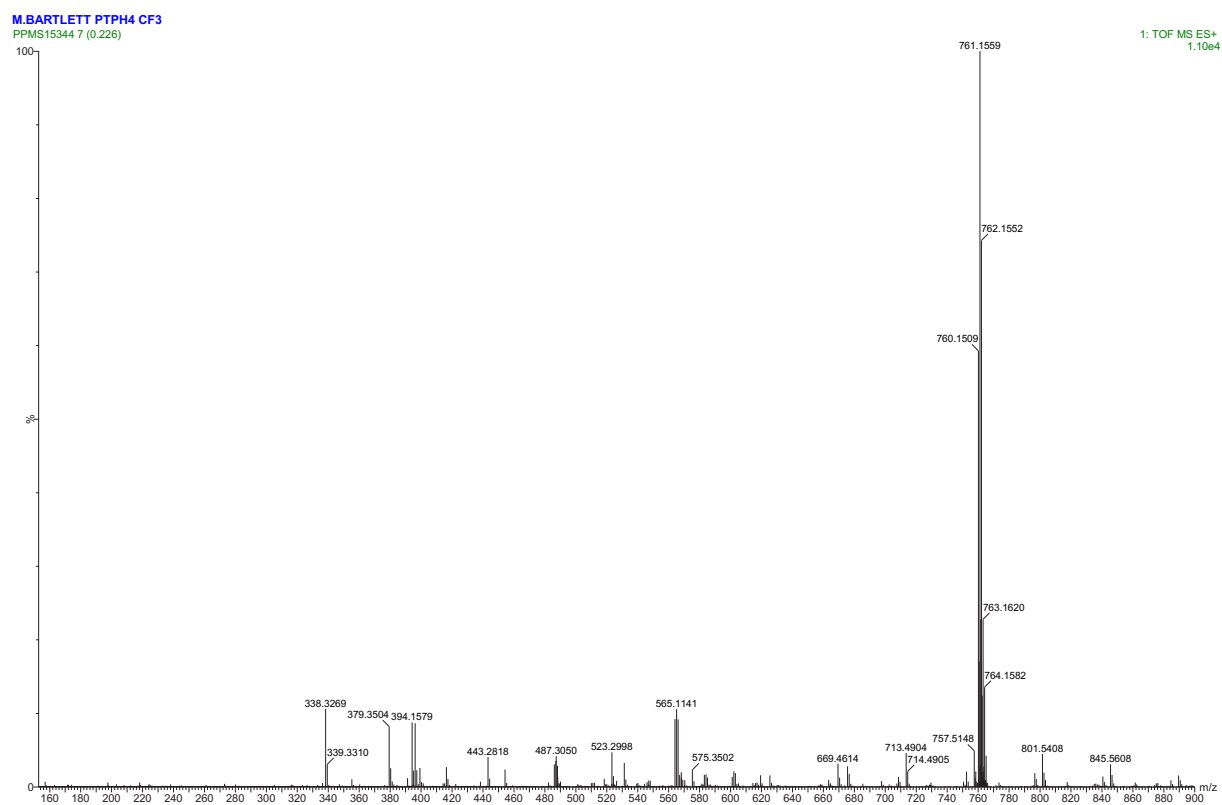


Figure A.79. HRMS of Pt_PPTMA_4CF₃.

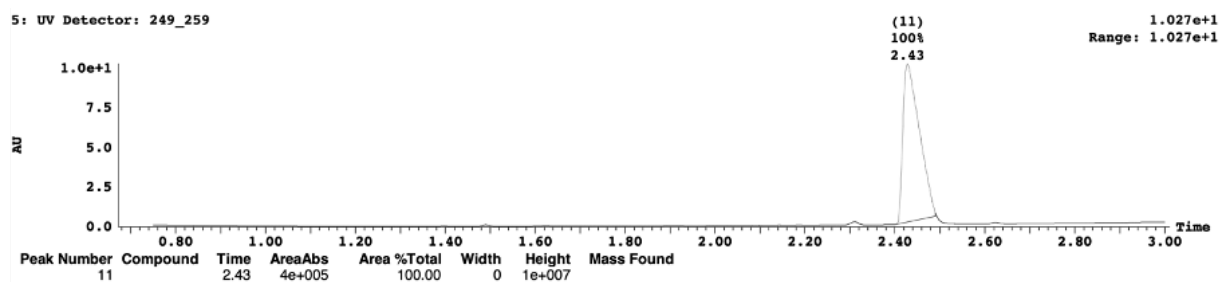


Figure A.80. LCMS UV trace of Pt_PPTMA_4CF₃.

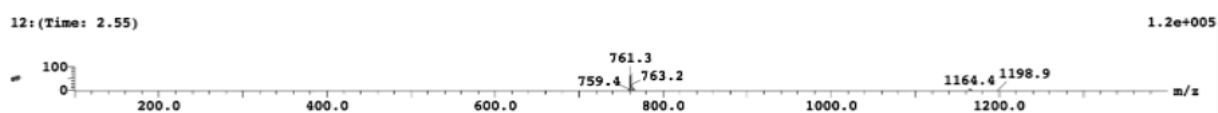
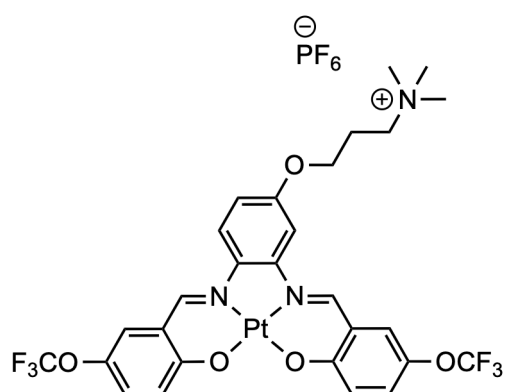


Figure A.81. Mass spec of Pt_PPTMA_4CF₃ under LCMS UV peak at 2.43 minutes.

Pt_PPTMA_5OCF₃ [hexafluorophosphate]:



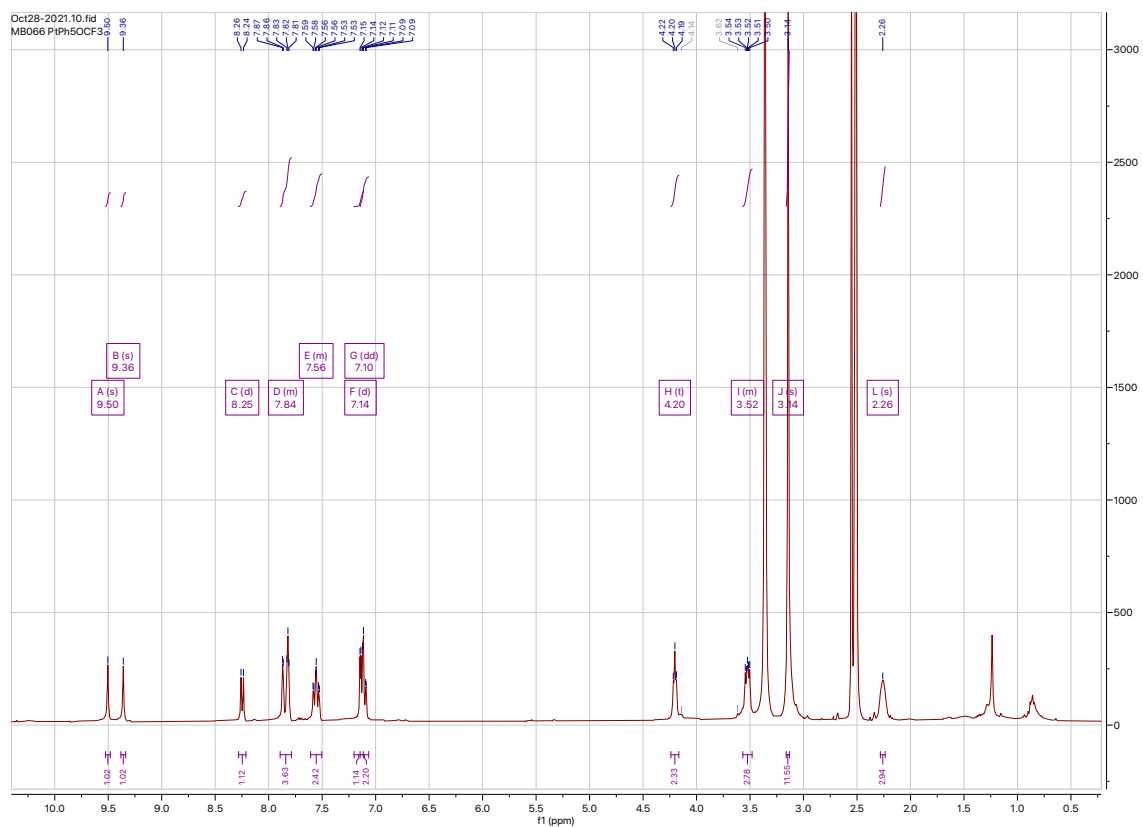


Figure A.82. Full ¹H NMR spectrum of Pt_PPTMA_5OCF₃.

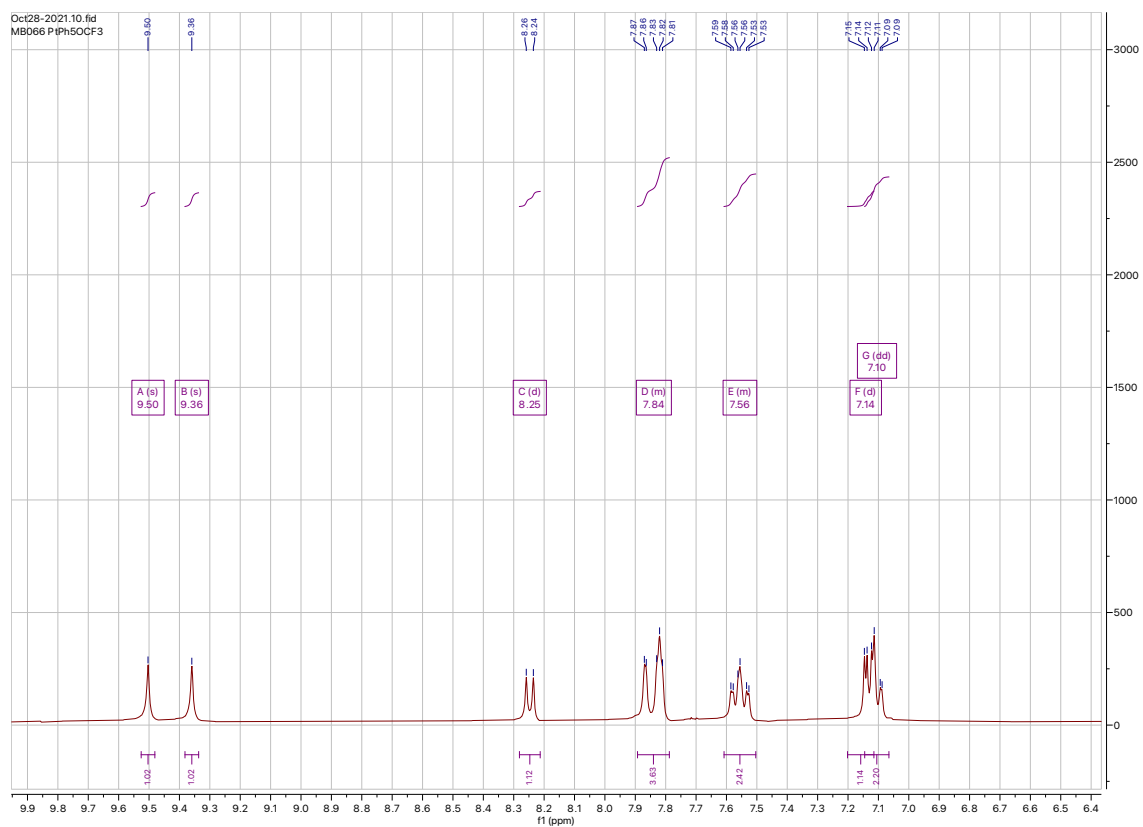


Figure A.83. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_5OCF₃.

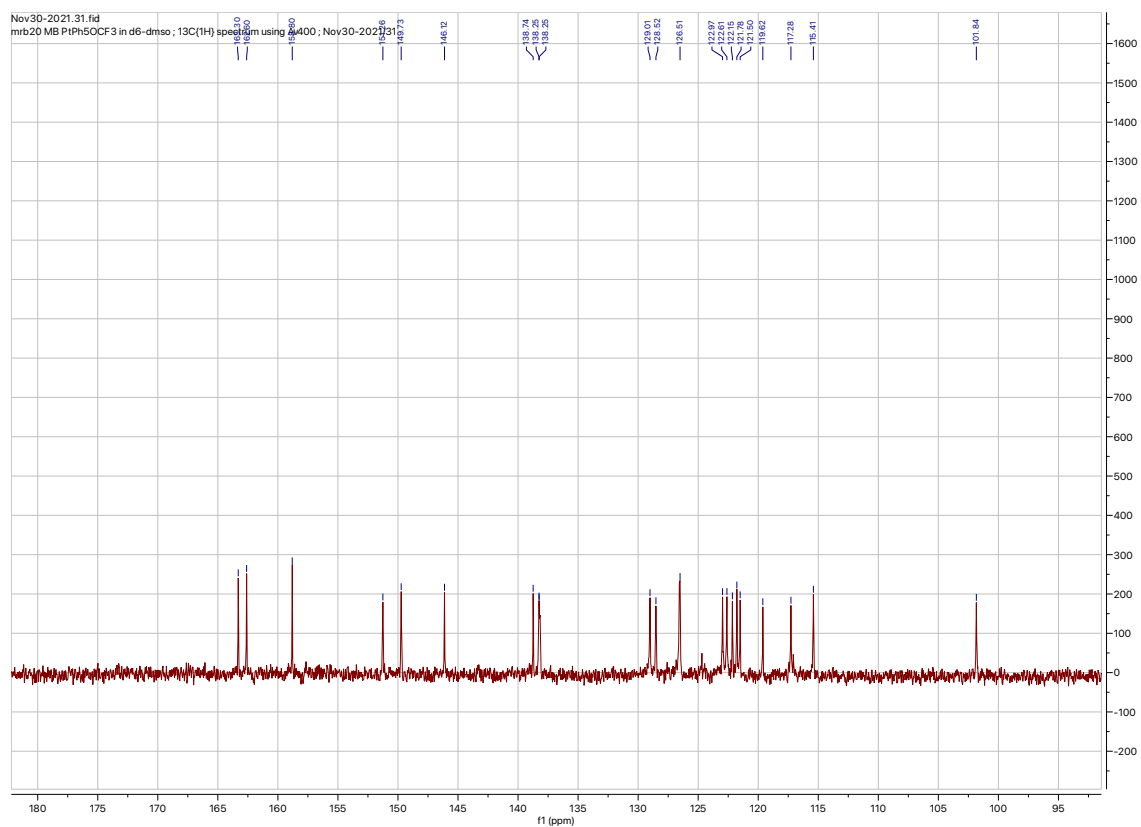


Figure A.84. ¹³C NMR spectrum of Pt_PPTMA_5OCF₃.

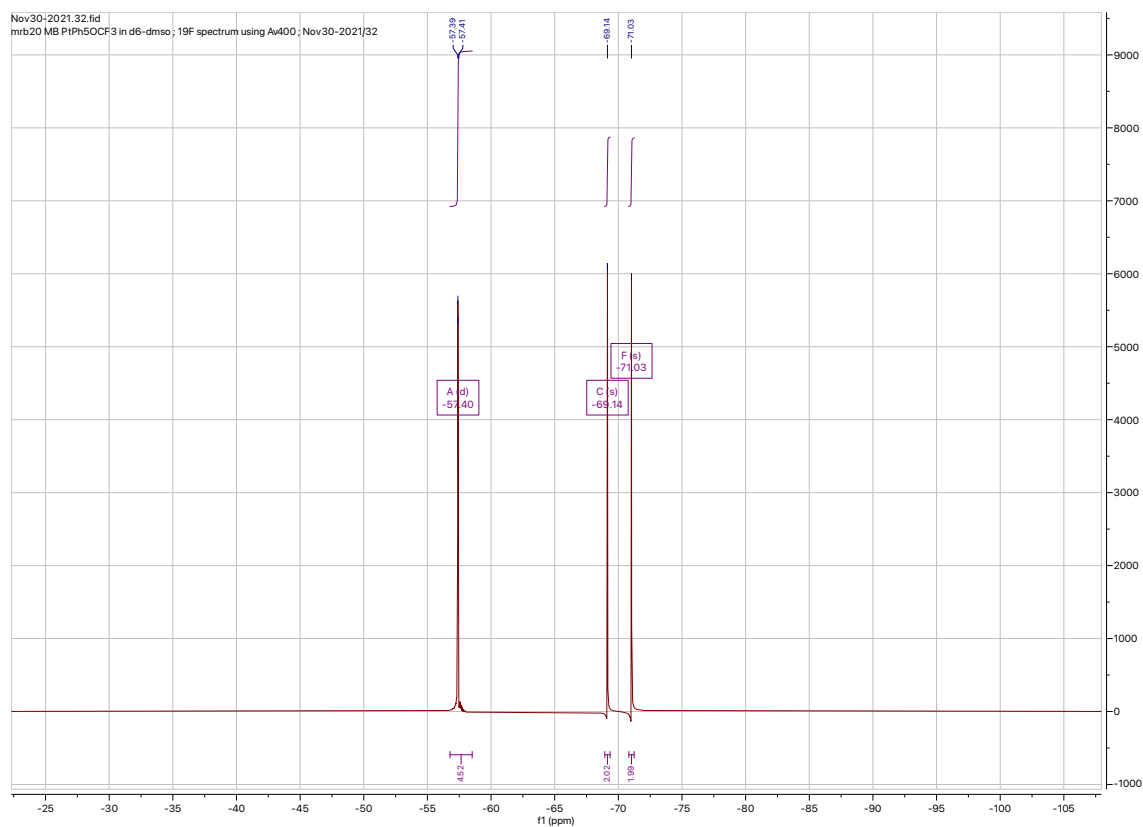


Figure A.85. ^{19}F NMR spectrum of Pt_PPTMA_5OCF_3 .

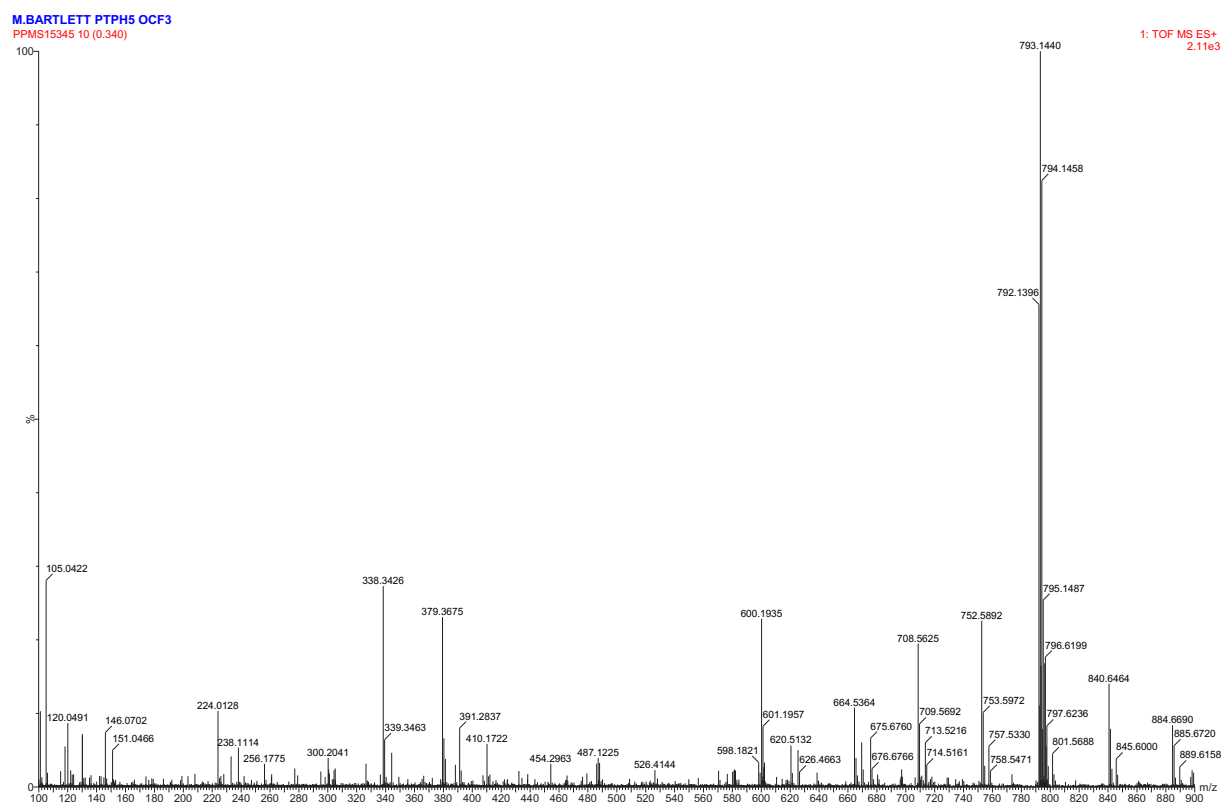


Figure A.86. HRMS of Pt_PPTMA_5OCF_3 .

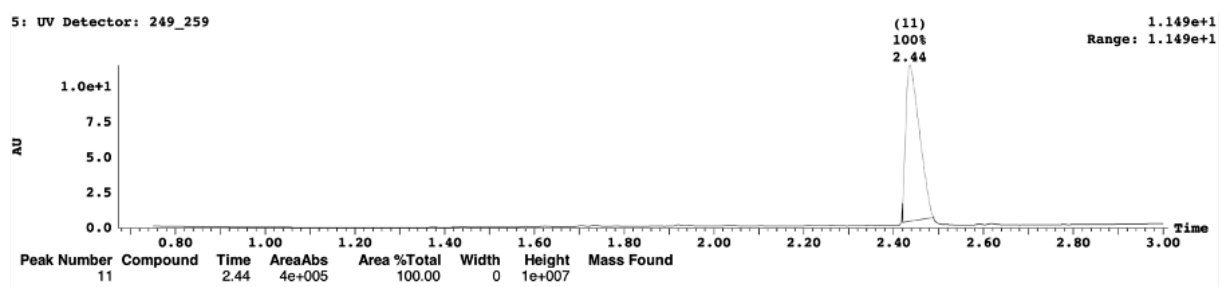


Figure A.87. LCMS UV trace of Pt_PPTMA_5OCF₃.

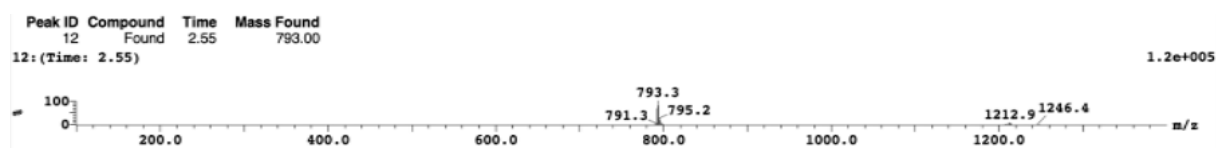
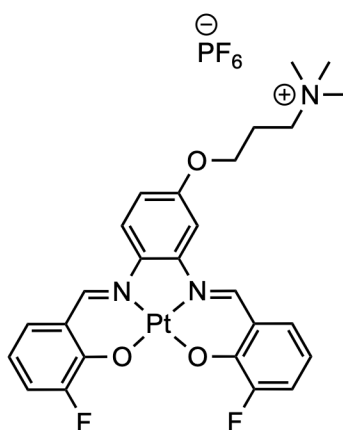


Figure A.88. Mass spec of Pt_PPTMA_5OCF₃ under LCMS UV peak at 2.44 minutes.

Pt_PPTMA_3F [hexafluorophosphate]:



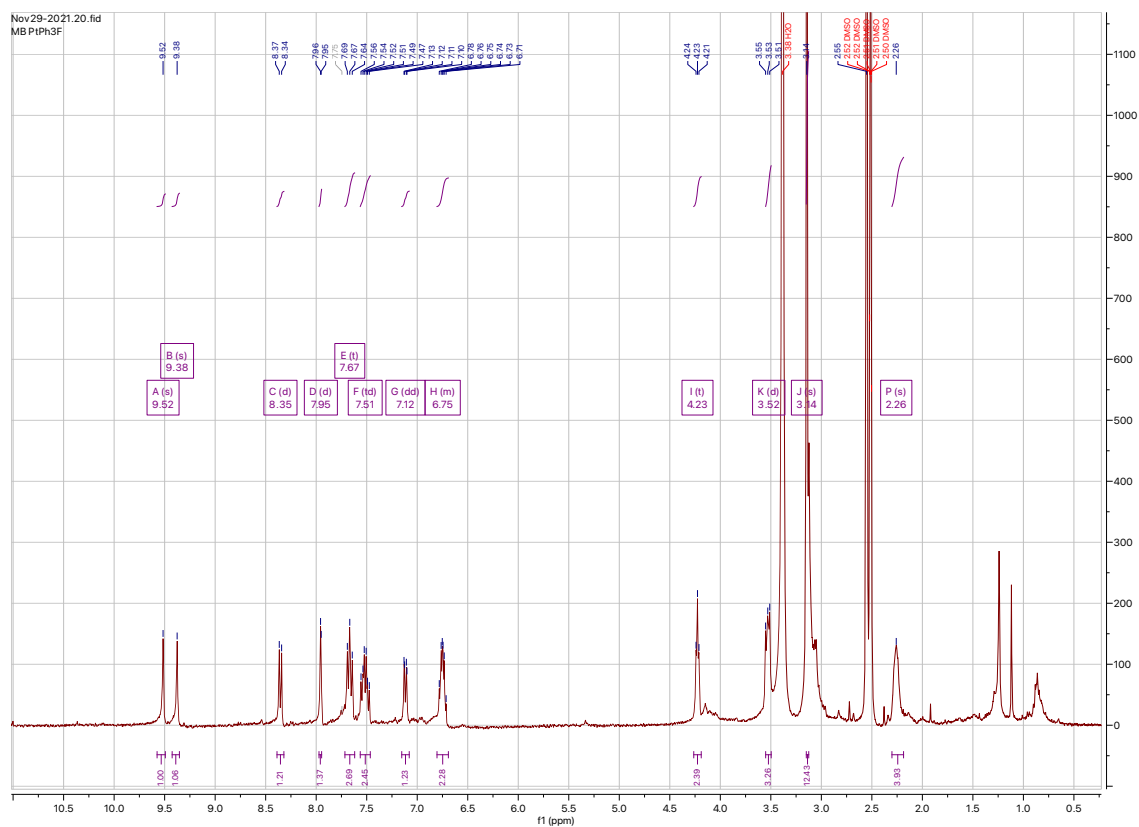


Figure A.89. Full ^1H NMR spectrum of Pt_PPTMA_3F.

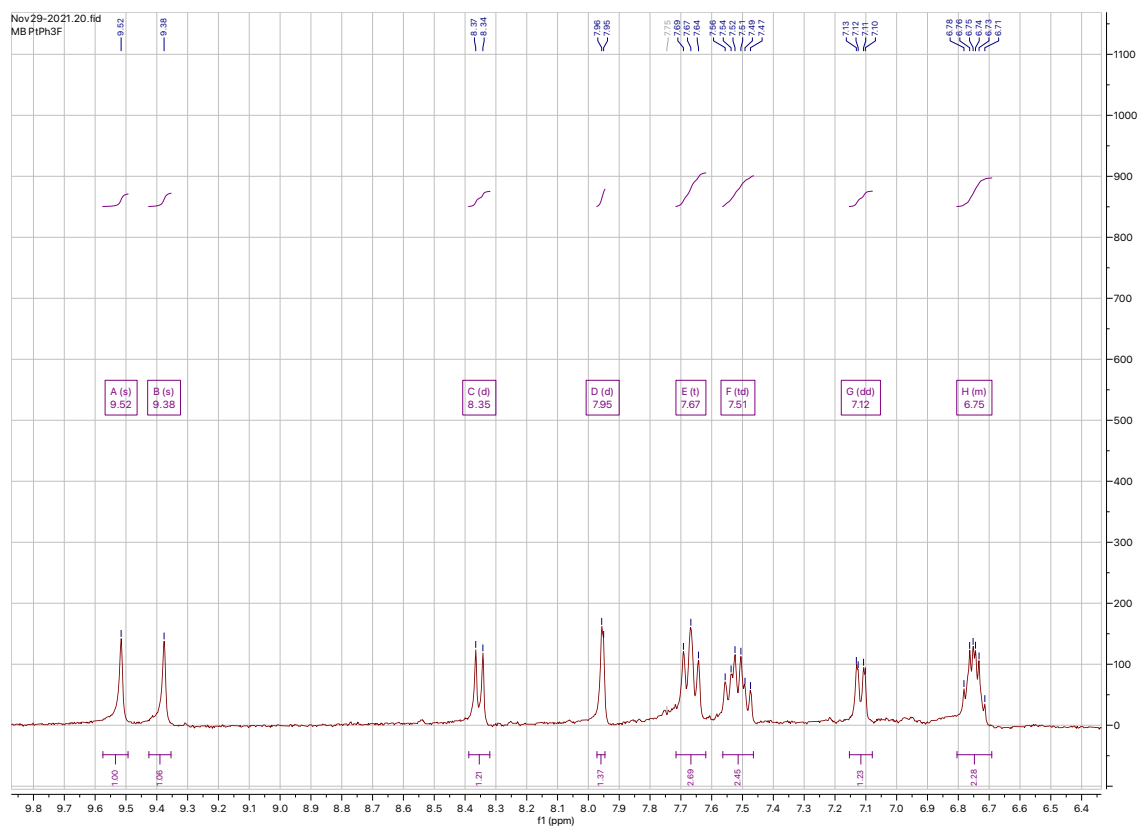


Figure A.90. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_3F.

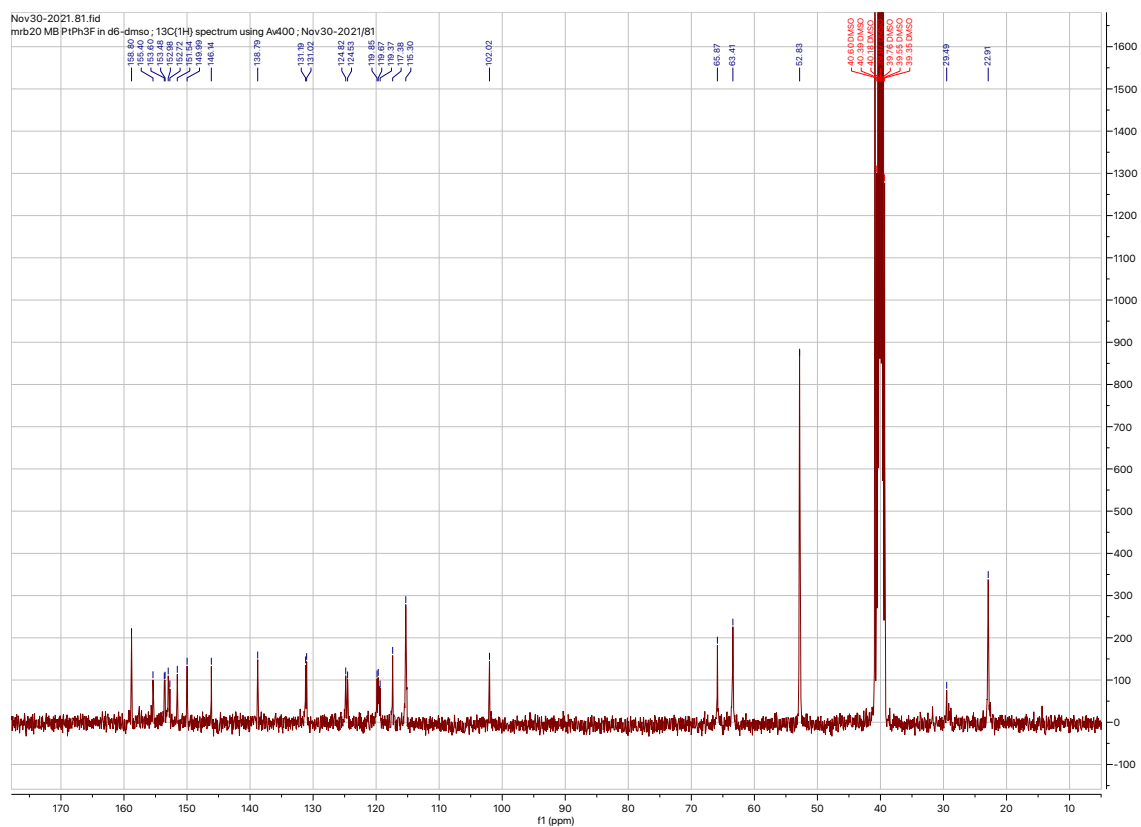


Figure A.91. ^{13}C NMR spectrum of Pt_PPTMA_3F.

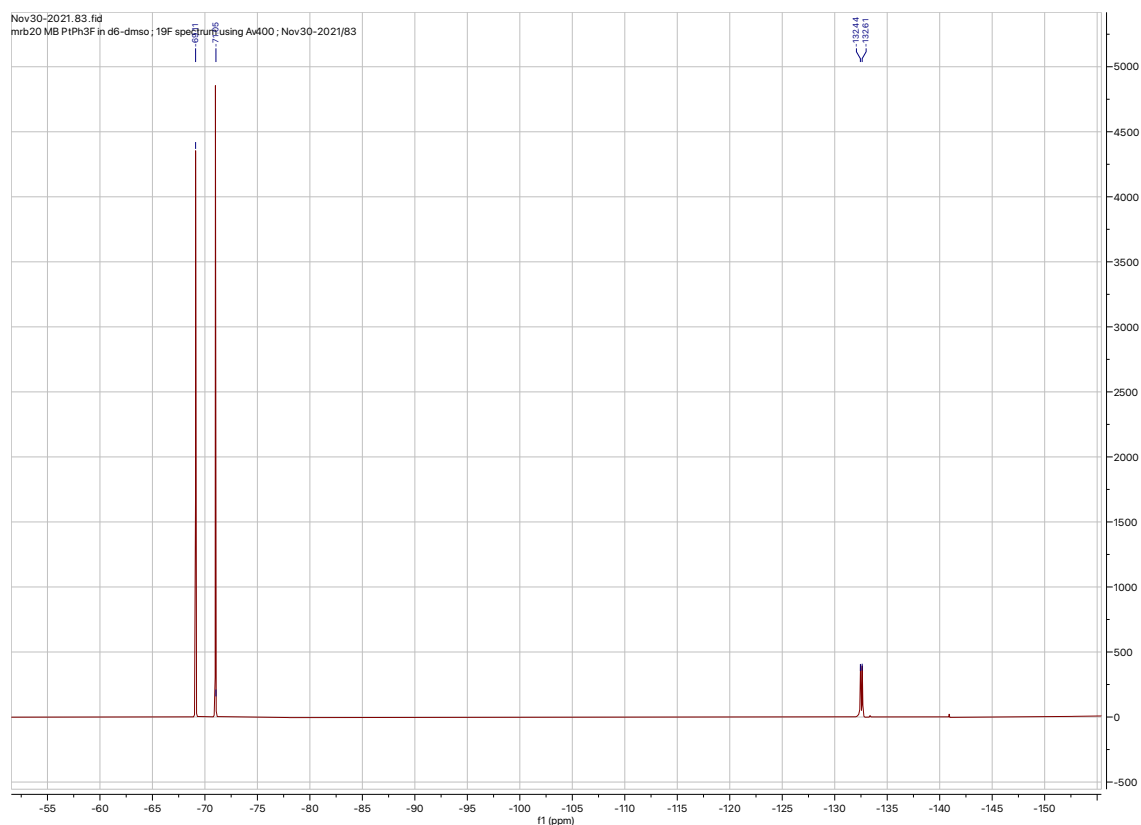


Figure A.92. ^{19}F NMR spectrum of Pt_PPTMA_3F.

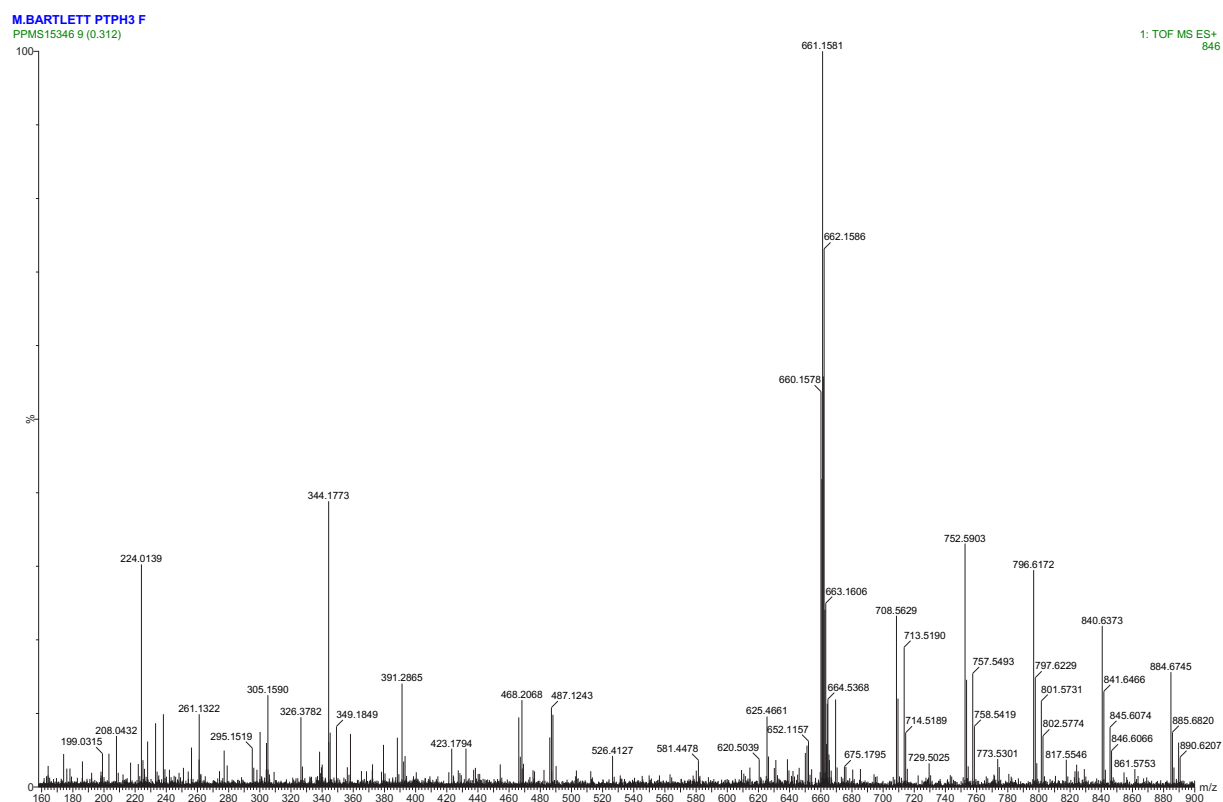


Figure A.93. HRMS of Pt_PPTMA_3F.

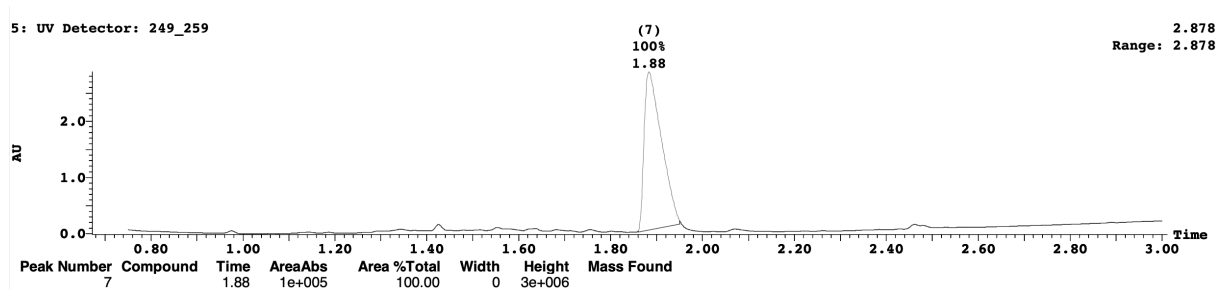


Figure A.94. LCMS UV trace of Pt_PPTMA_3F.

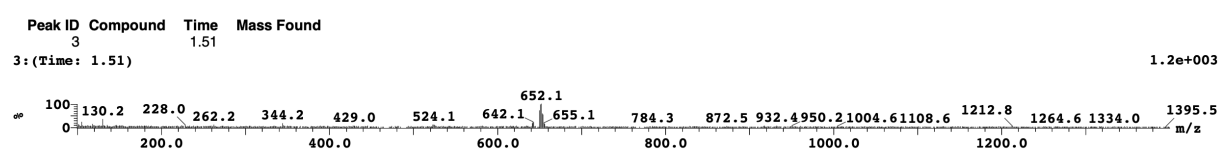
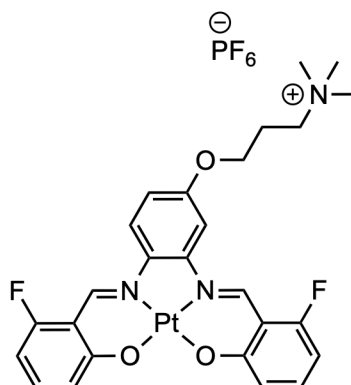


Figure A.95. Mass spec of Pt_PPTMA_3F under LCMS UV trace at 1.88 minutes.

Pt_PPTMA_6F [hexafluorophosphate]:



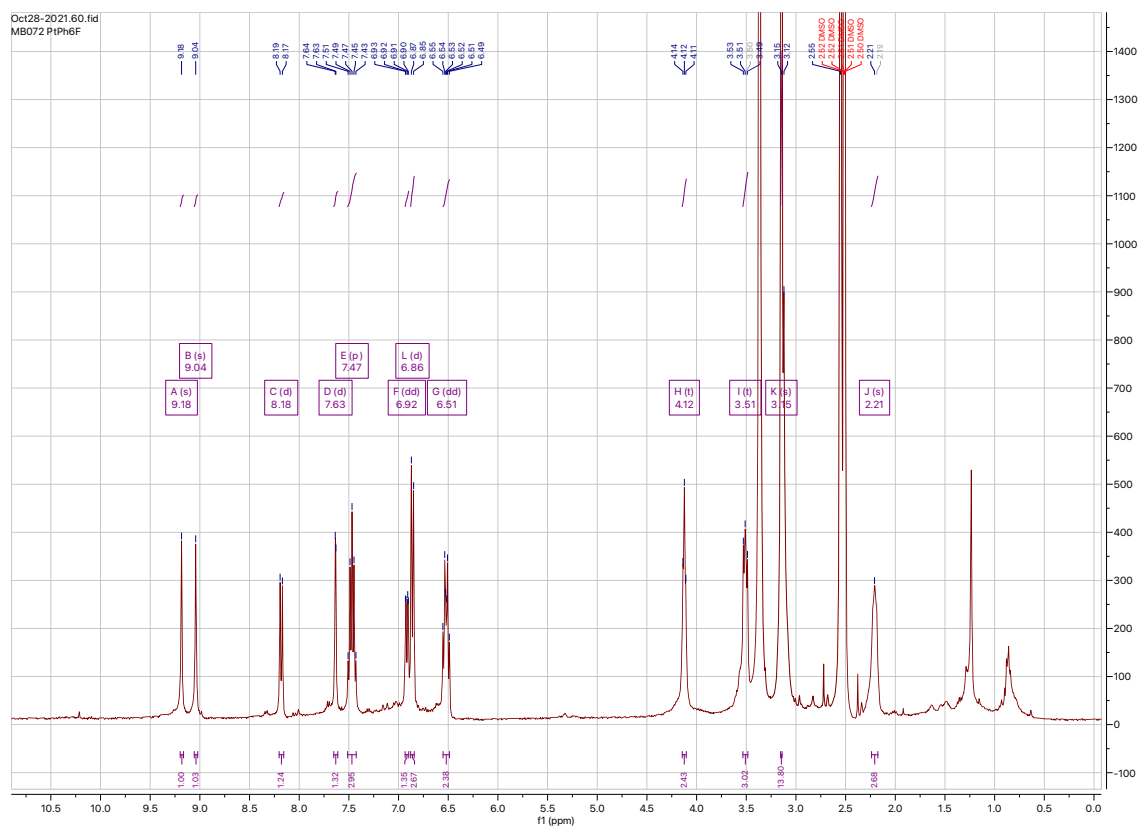


Figure A.96. Full ^1H NMR spectrum of Pt_PPTMA_6F.

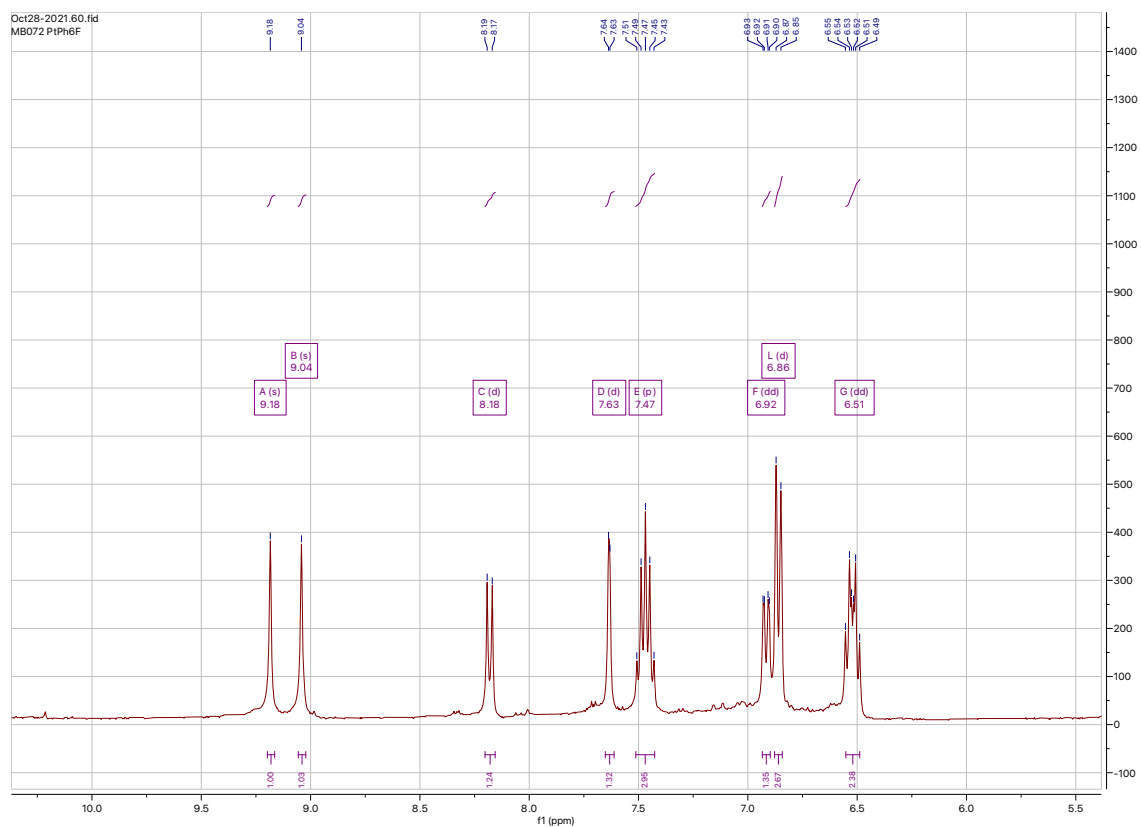


Figure A.97. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_6F.

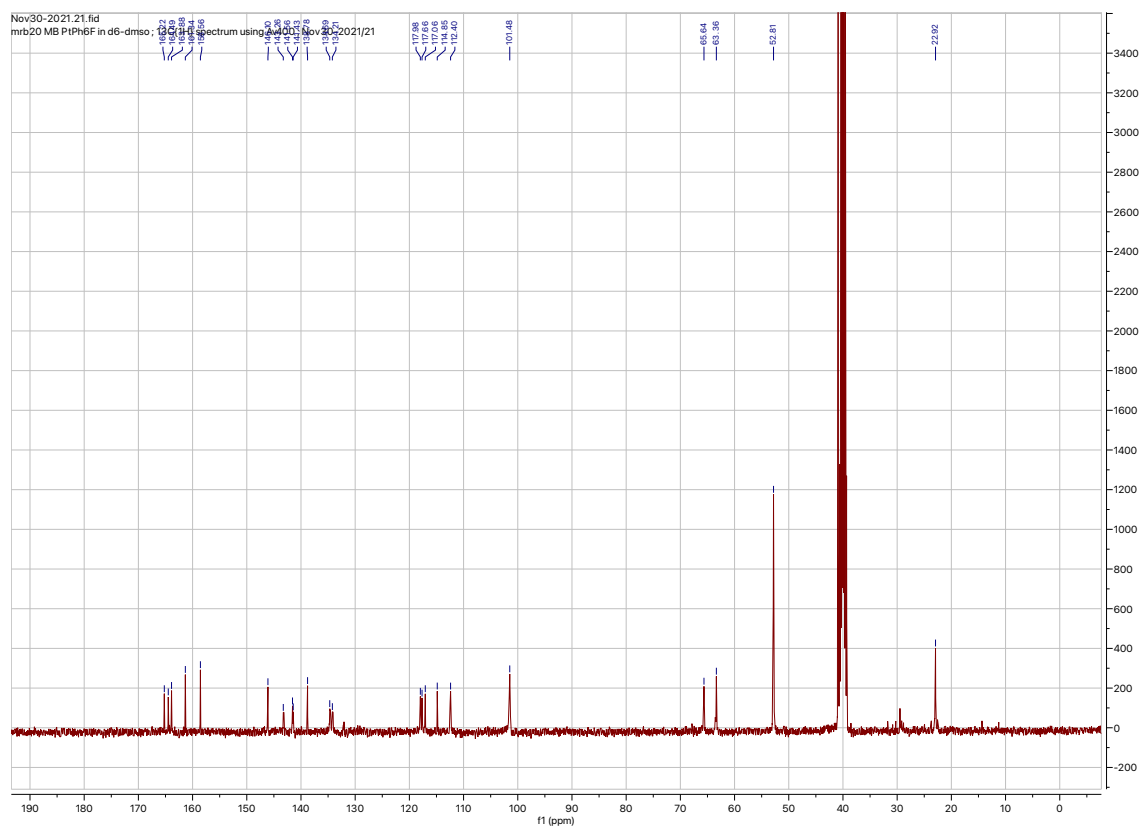


Figure A.98. ^{13}C NMR spectrum of Pt_PPTMA_6F.

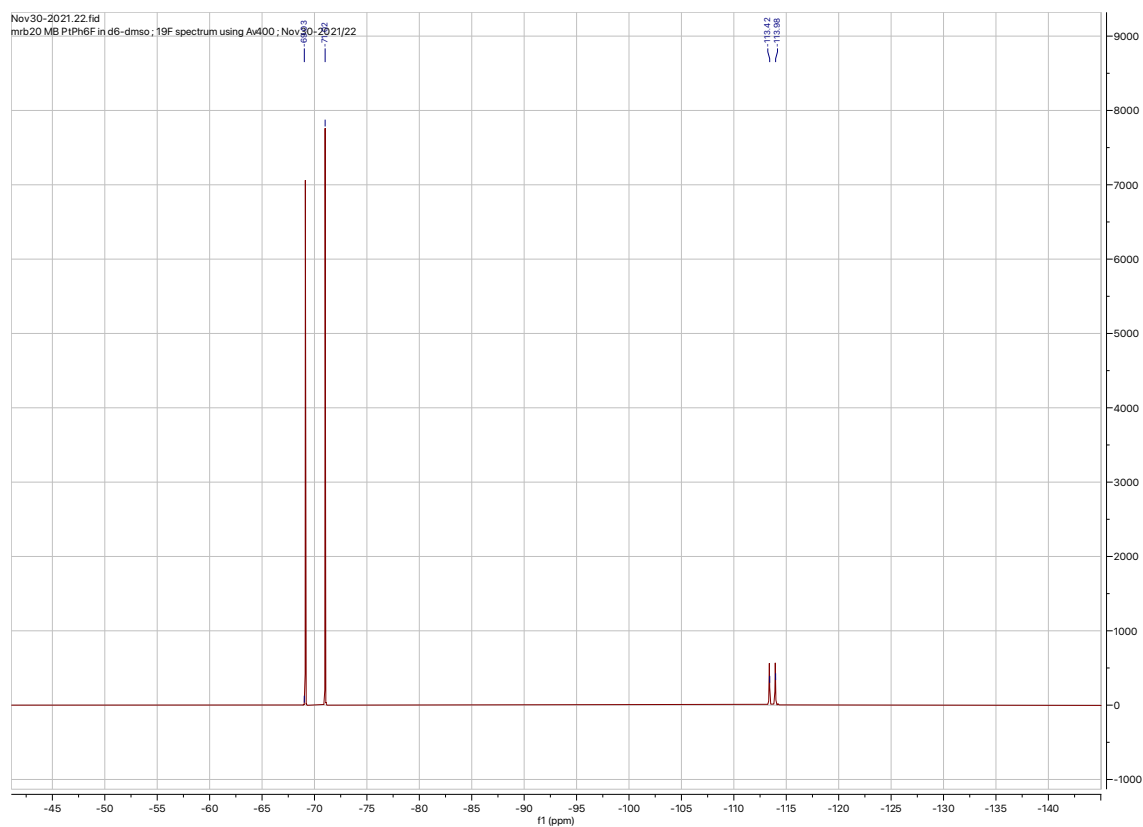


Figure A.99. ^{19}F NMR spectrum of Pt_PPTMA_6F.

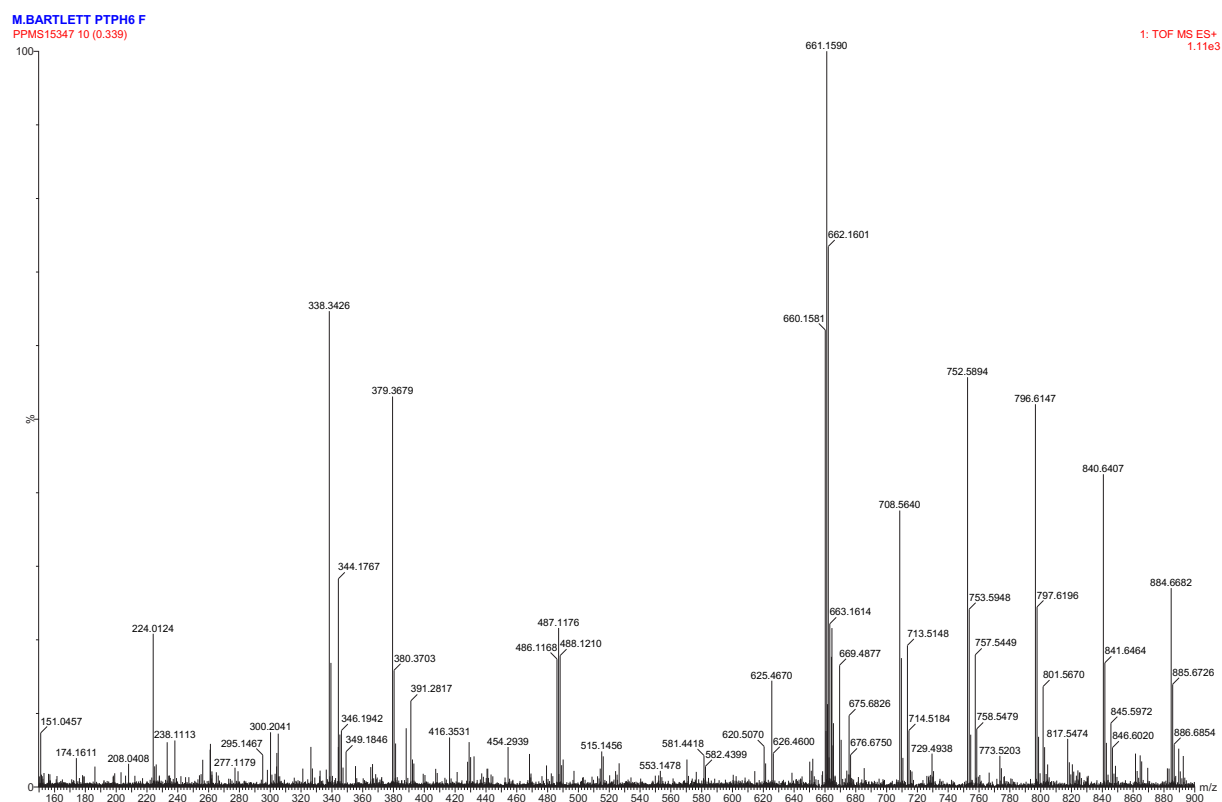


Figure A.100. HRMS of Pt_PPTMA_6F.

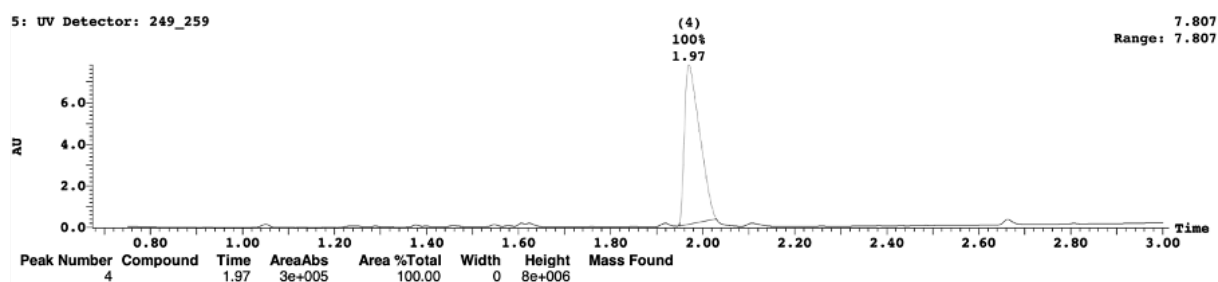


Figure A.101. LCMS UV trace of Pt_PPTMA_6F.

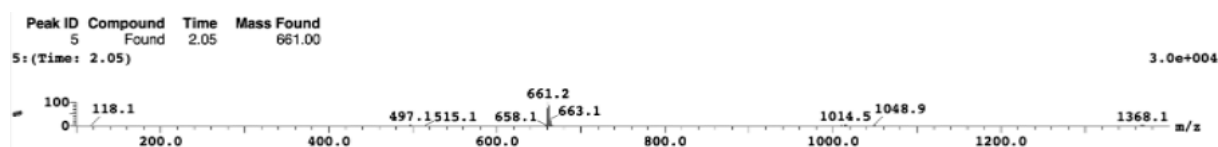
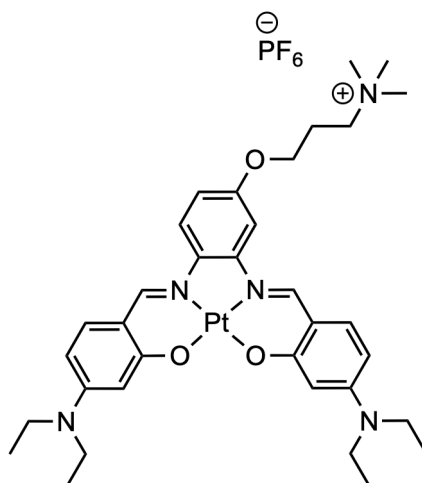


Figure A.102. Mass spec of Pt_PPTMA_6F under LCMS UV trace at 1.97 minutes.

Pt_PPTMA_4NEt₂ [hexafluorophosphate]:



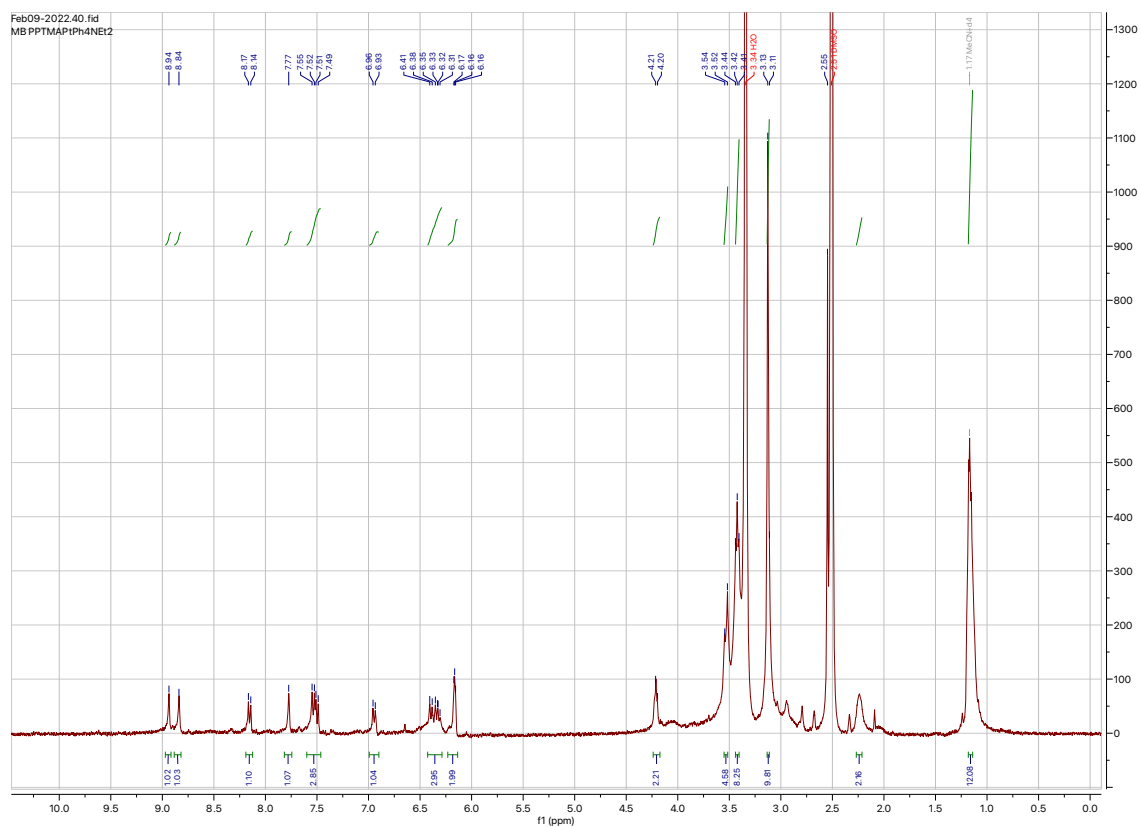


Figure A.103. Full ^1H NMR spectrum of Pt_PPTMA_4NEt₂.

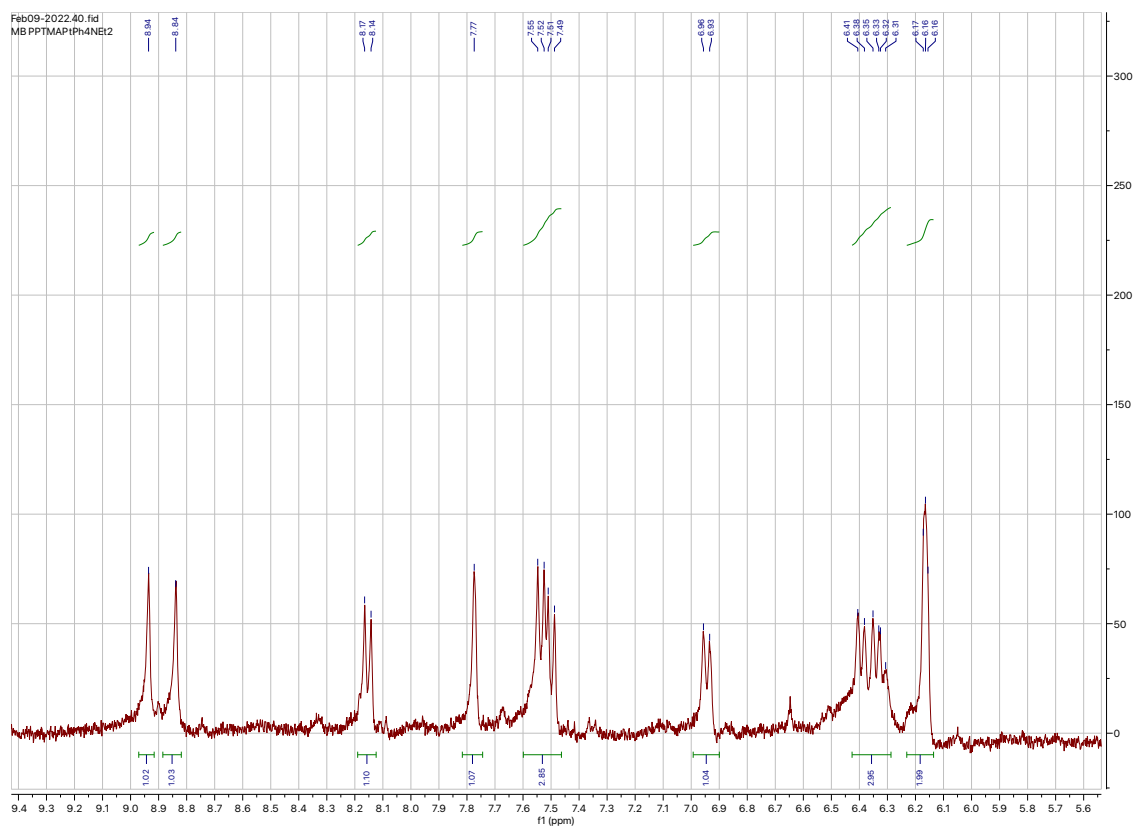


Figure A.104. Aromatic region of ^1H NMR spectrum of Pt_PPTMA_4NEt₂.

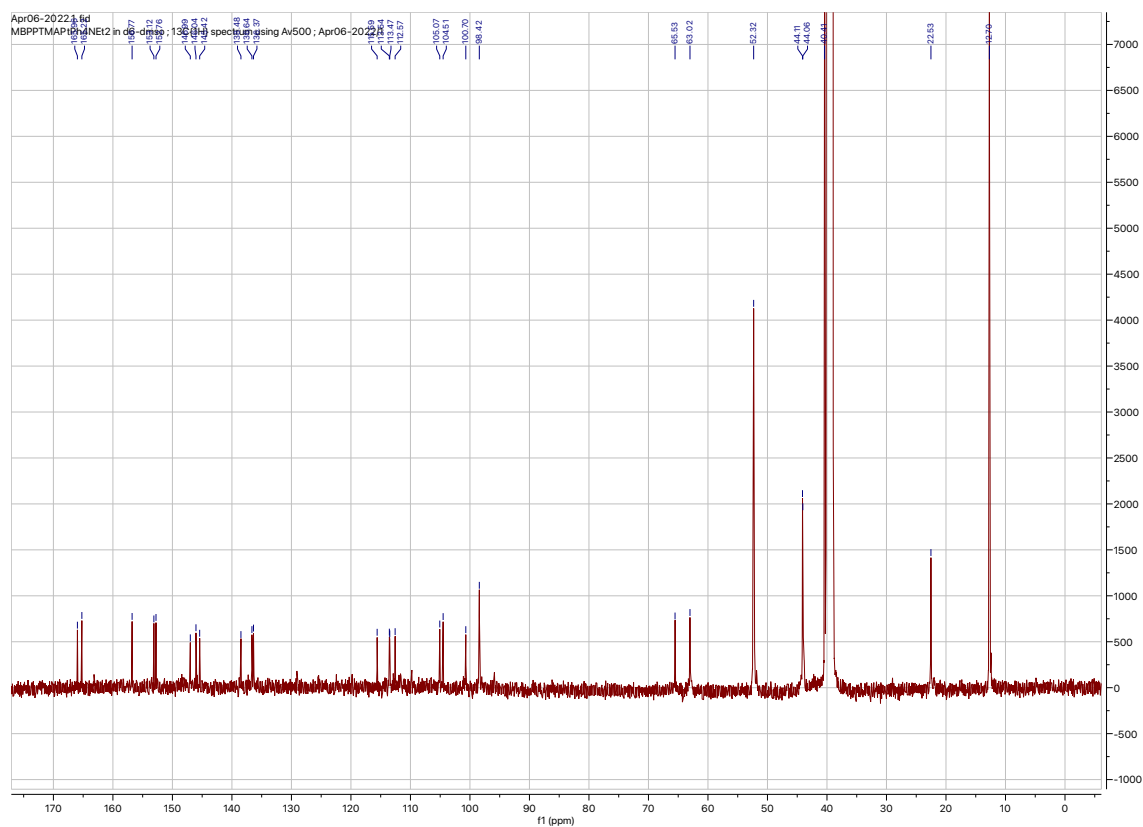


Figure A.105. ^{13}C NMR spectrum of Pt_PPTMA_4NEt₂.

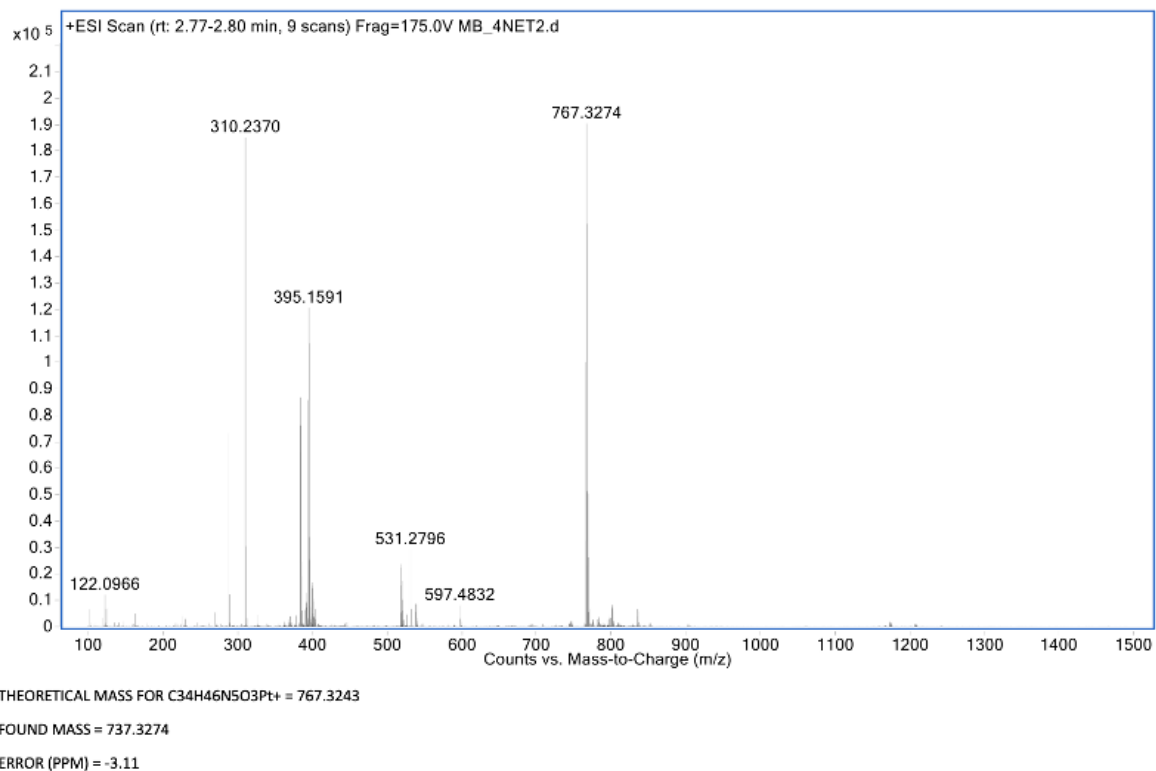


Figure A.106. HRMS of Pt_PPTMA_4NEt₂.

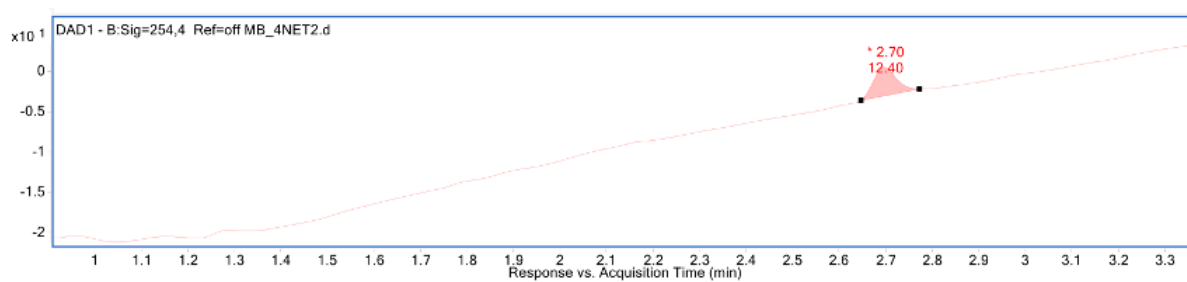


Figure A.107. LCMS UV trace of Pt_PPTMA_4NEt₂.

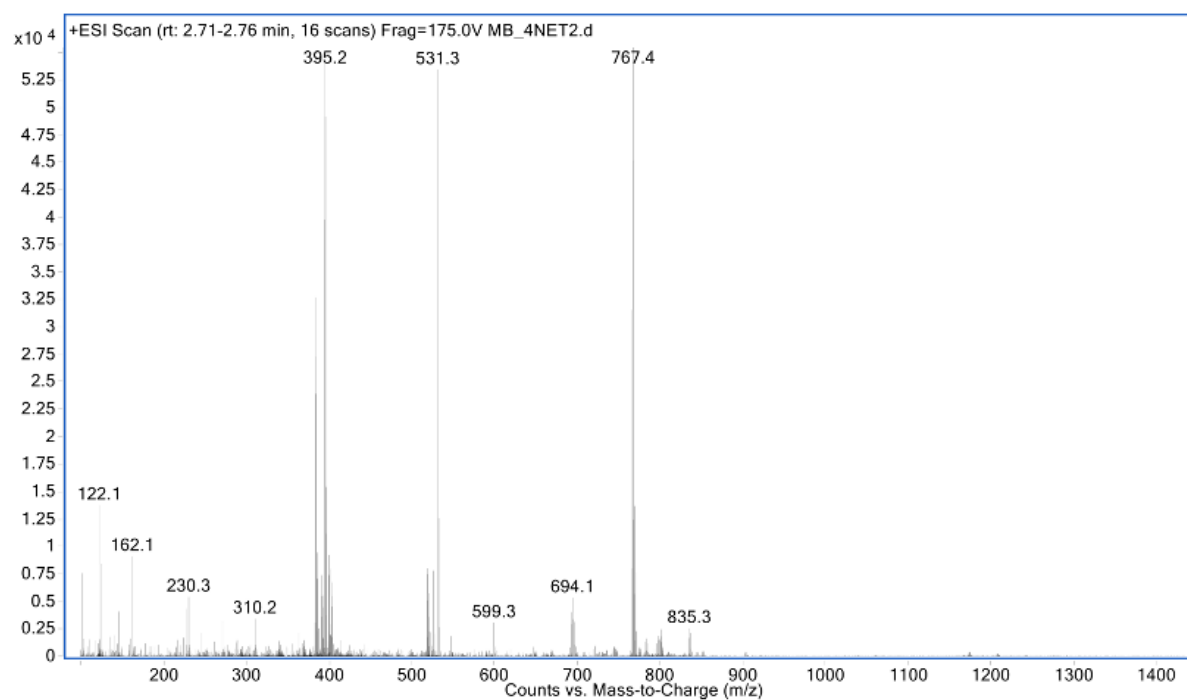
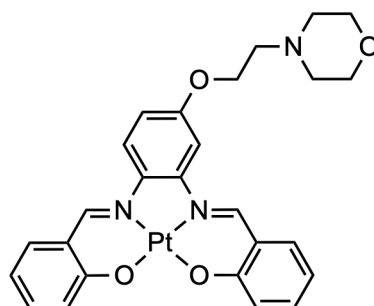


Figure A.108. Mass spec of Pt_PPTMA_4NEt₂ under LCMS UV trace at 2.70 minutes.

Pt_PEM_Ph:



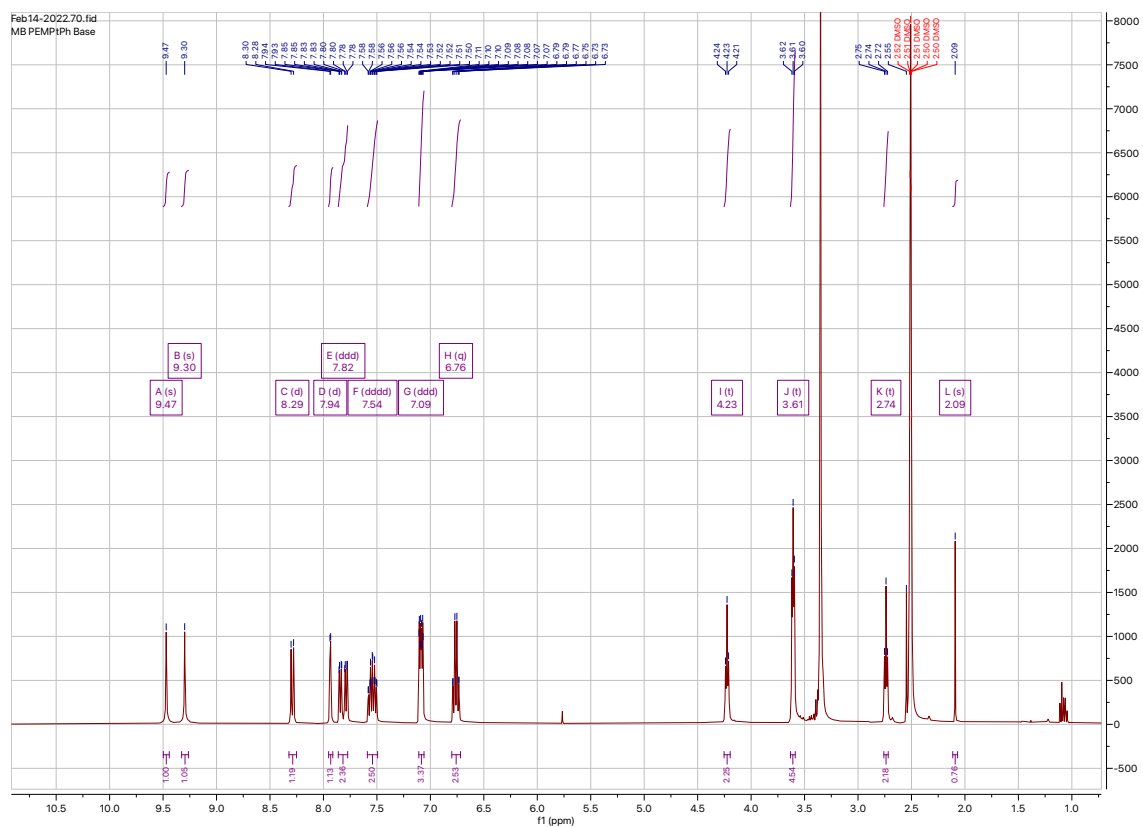


Figure A.109. Full ^1H NMR spectrum of Pt_PEM_Ph.

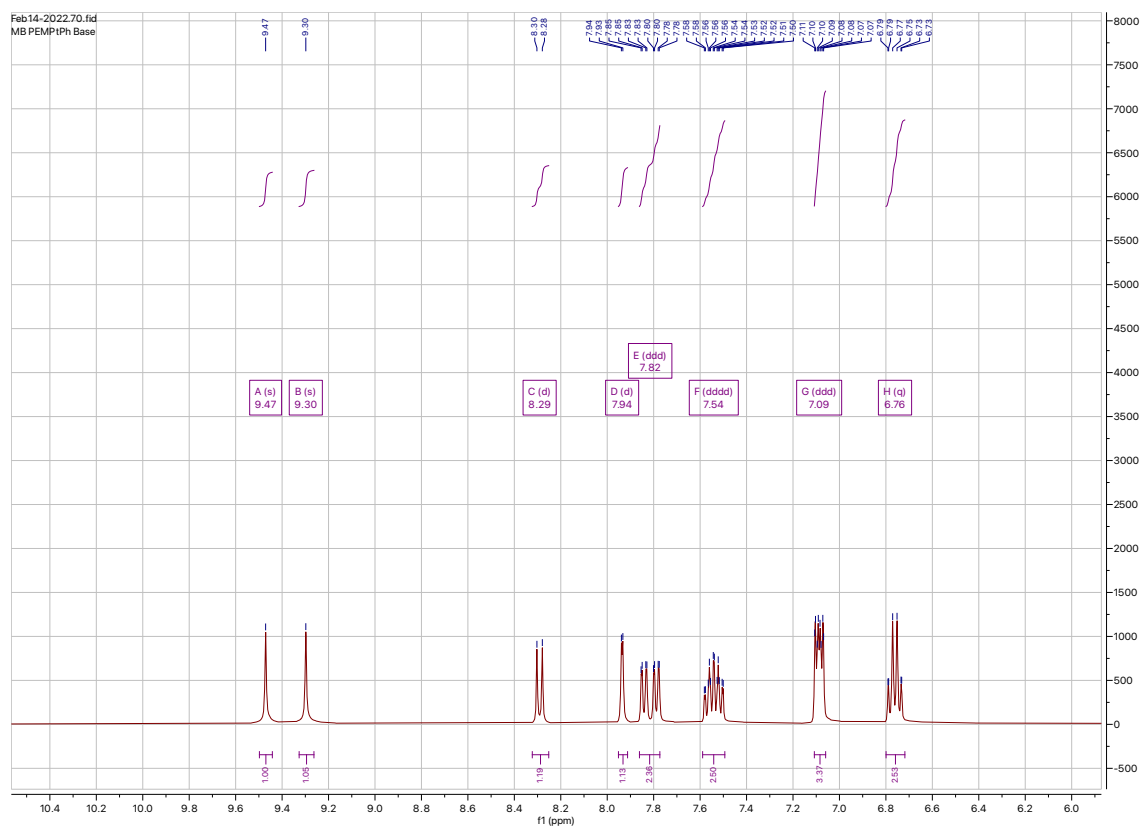
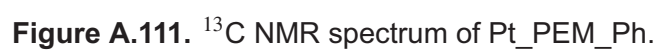


Figure A.110. Aromatic region of ^1H NMR spectrum of Pt_PEM_Ph.



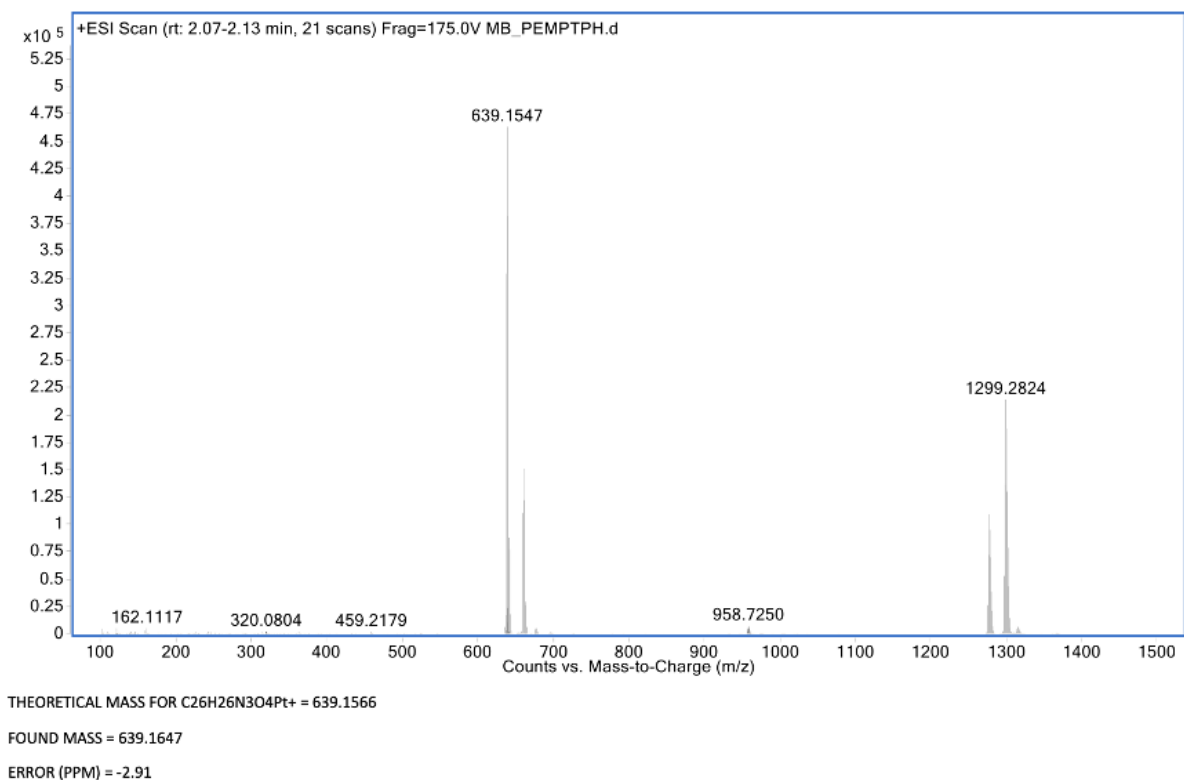


Figure A.112. HRMS of Pt_PEM_Ph.

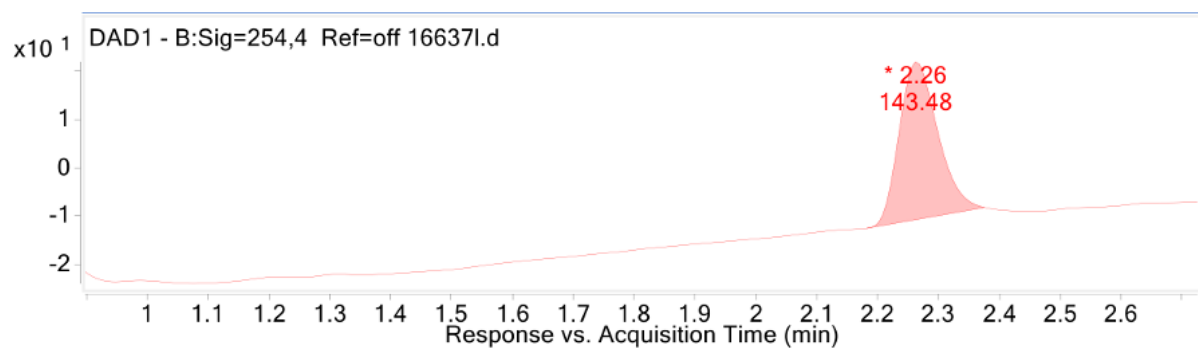


Figure A.113. LCMS UV trace of Pt_PEM_Ph.

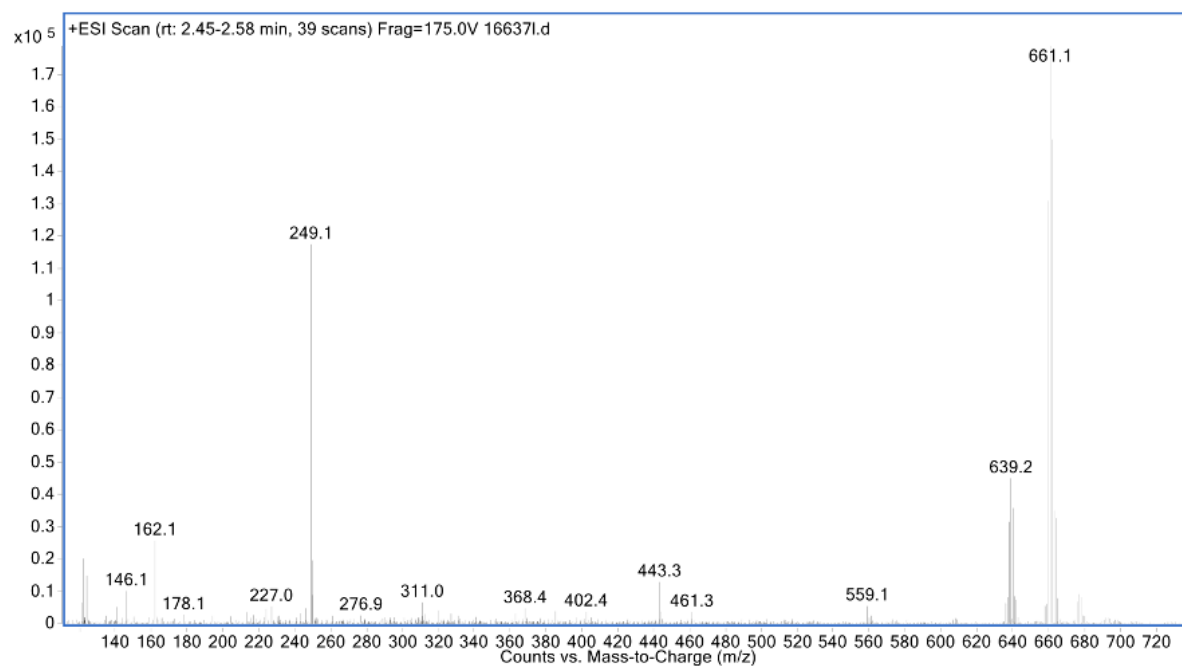
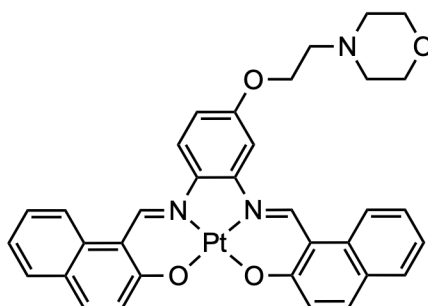


Figure A.114. Mass spec of Pt_PEM_Ph under LCMS UV trace at 2.26 minutes.

Pt_PEM_Naph:



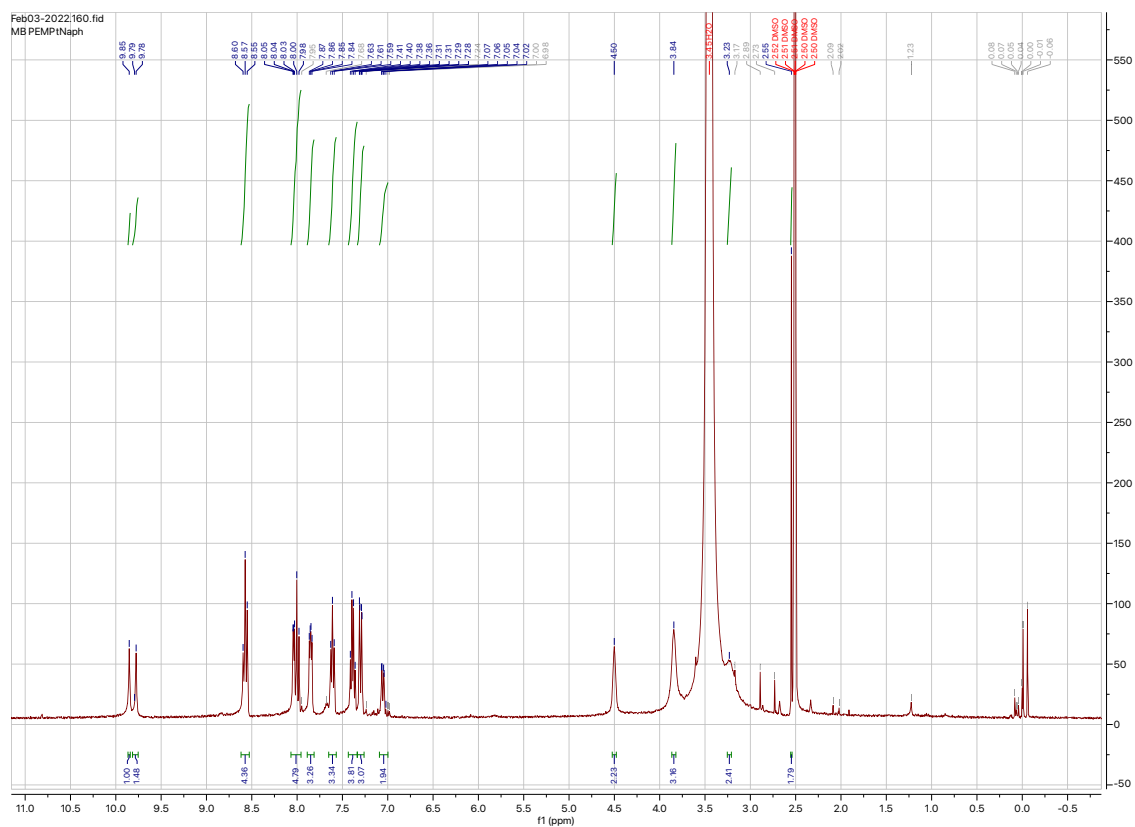


Figure A.115. Full ^1H NMR spectrum of Pt_PEM_Naph.

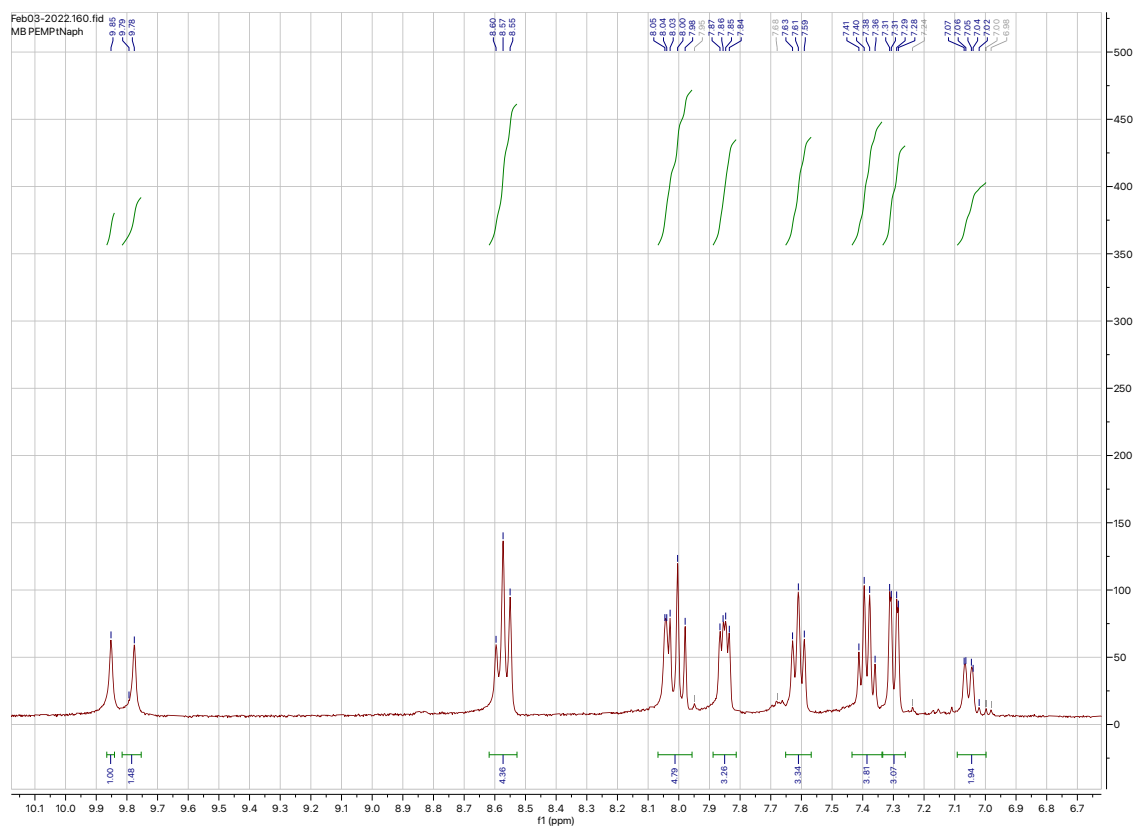


Figure A.116. Aromatic region of ^1H NMR spectrum of Pt_PEM_Naph.

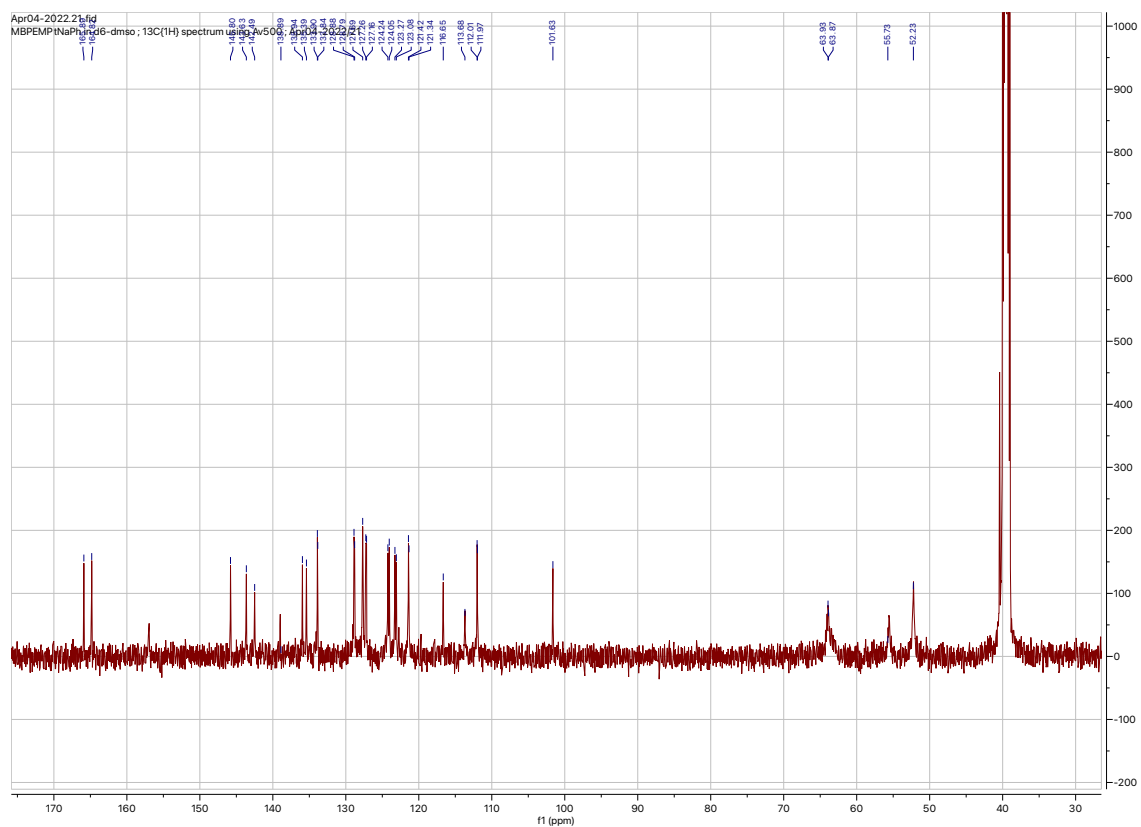


Figure A.117. ^{13}C NMR spectrum of Pt_PEM_Naph.

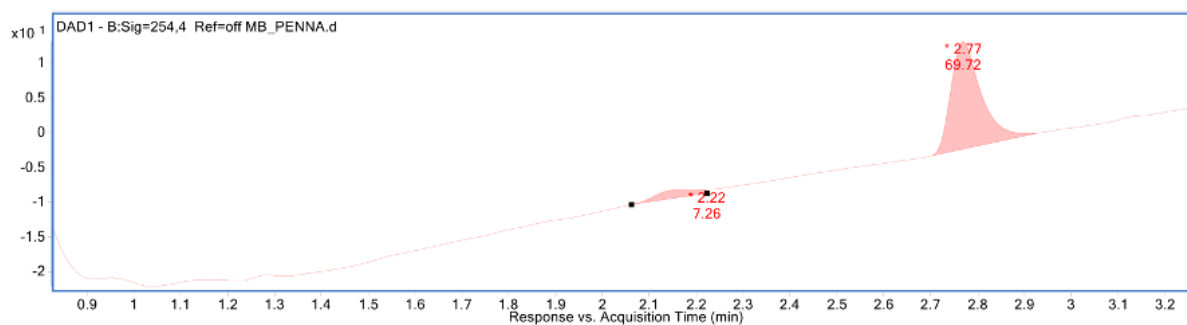


Figure A.118. LCMS UV trace of Pt_PEM_Naph.

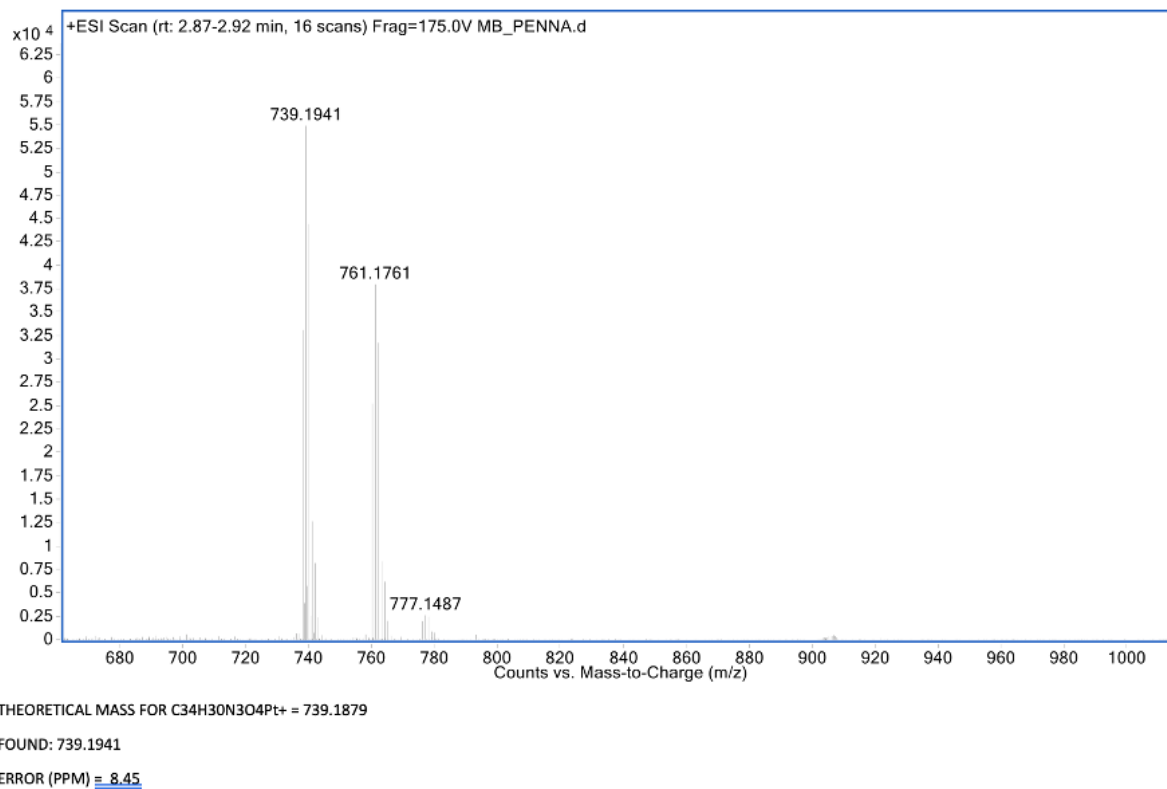
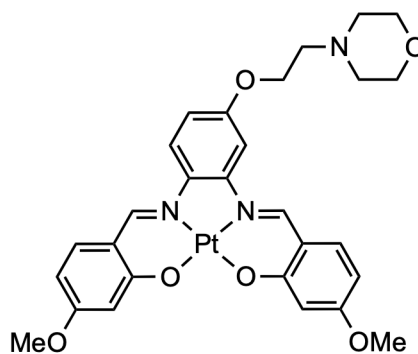


Figure A.119. HRMS of Pt_PEM_Naph under LCMS UV peak at 2.77 minutes.

Pt_PEM_4OMe:



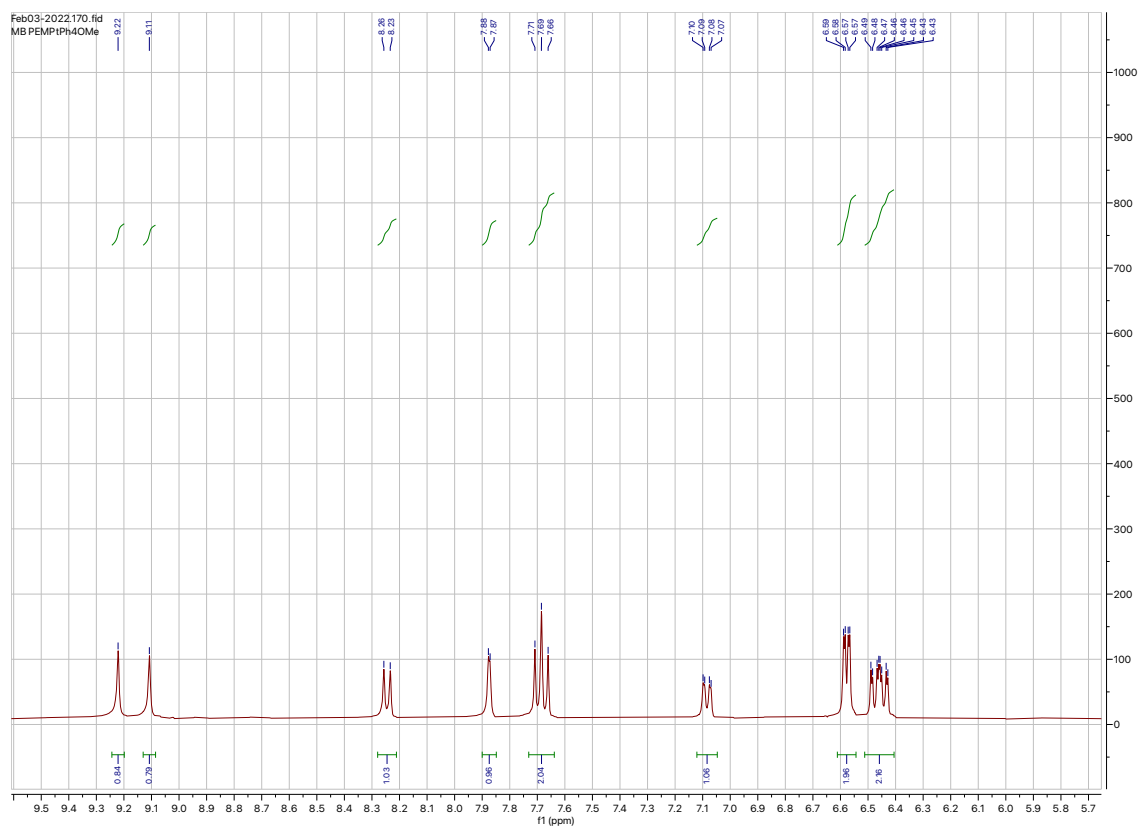
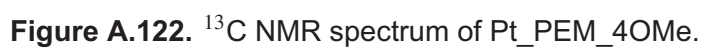
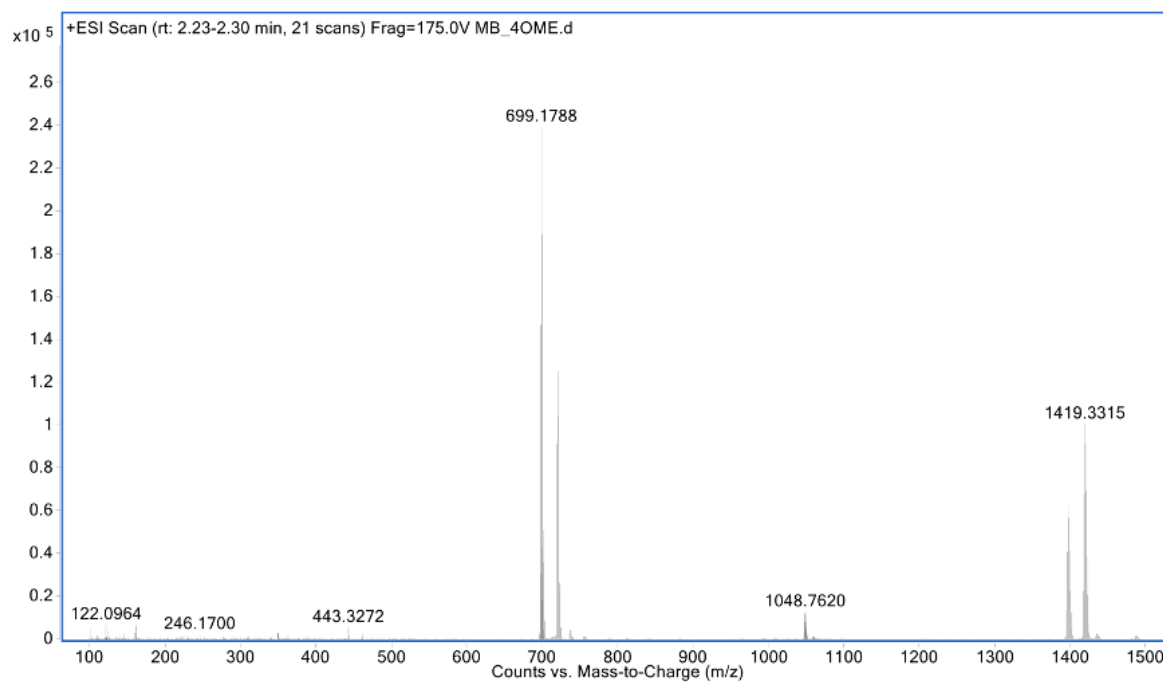


Figure A.121. Aromatic region of ^1H NMR spectrum of Pt_PEM_4OMe.





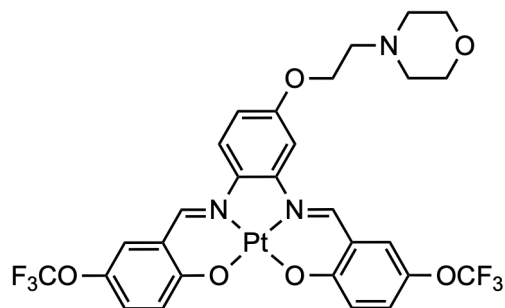
THEORETICAL MASS FOR C₂₈H₃₀N₃O₆Pt⁺ = 699.1777

FOUND MASS = 699.17888

ERROR (PPM) = 1.59

Figure A.124. HRMS of Pt_PEM_4OMe.

Pt_PEM_5OCF₃:



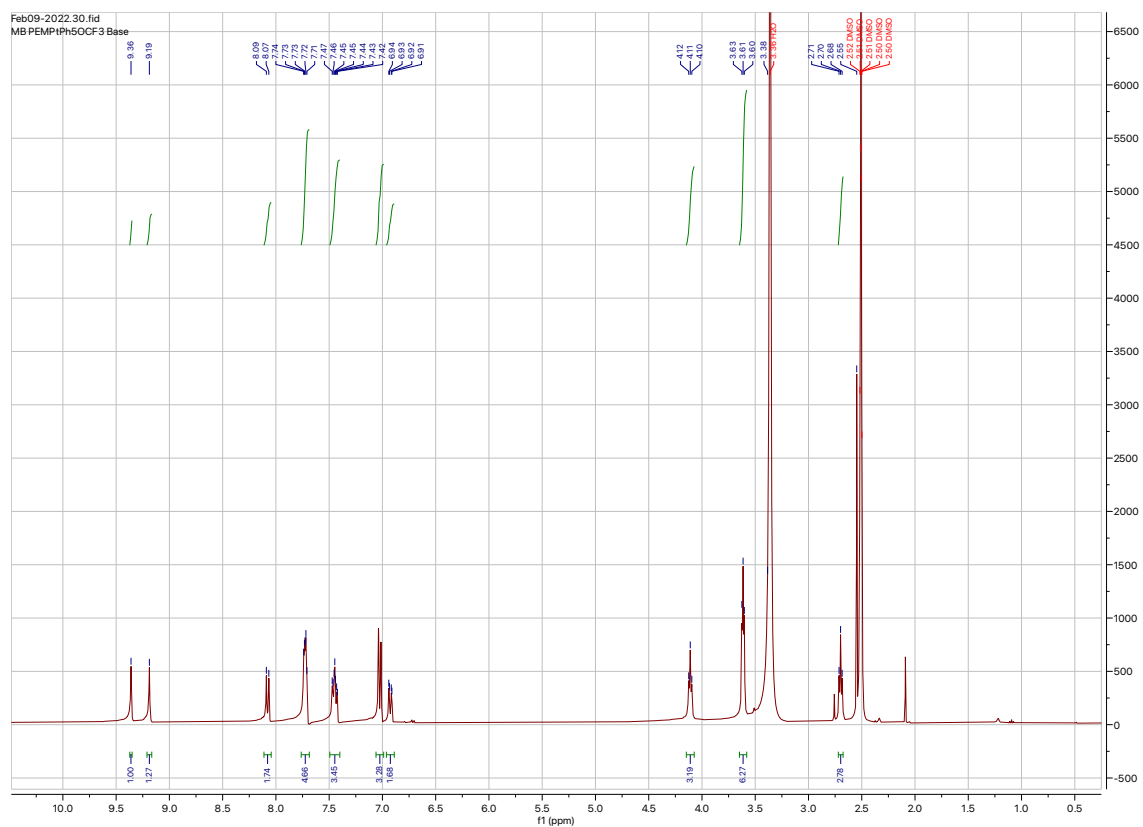


Figure A.125. Full ^1H NMR spectrum of Pt_PEM_5OCF_3 .

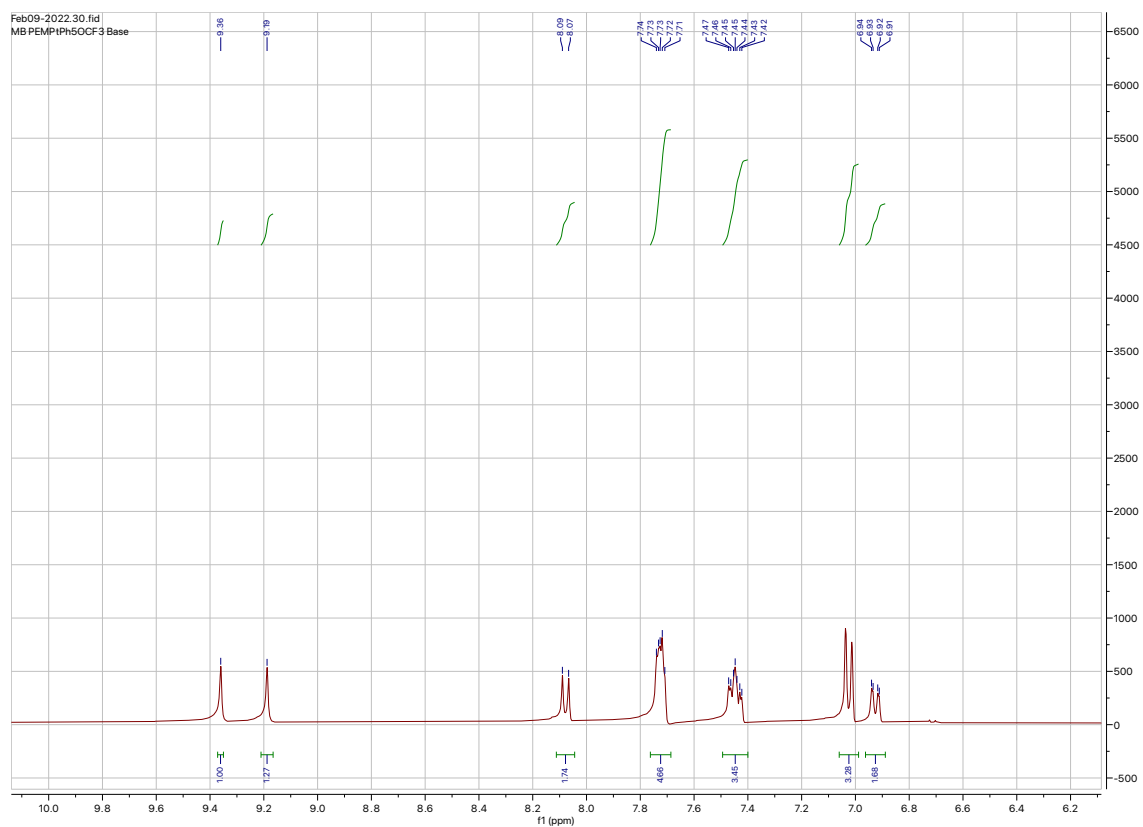


Figure A.126. Aromatic region of ^1H NMR spectrum of Pt_PEM_5OCF₃.

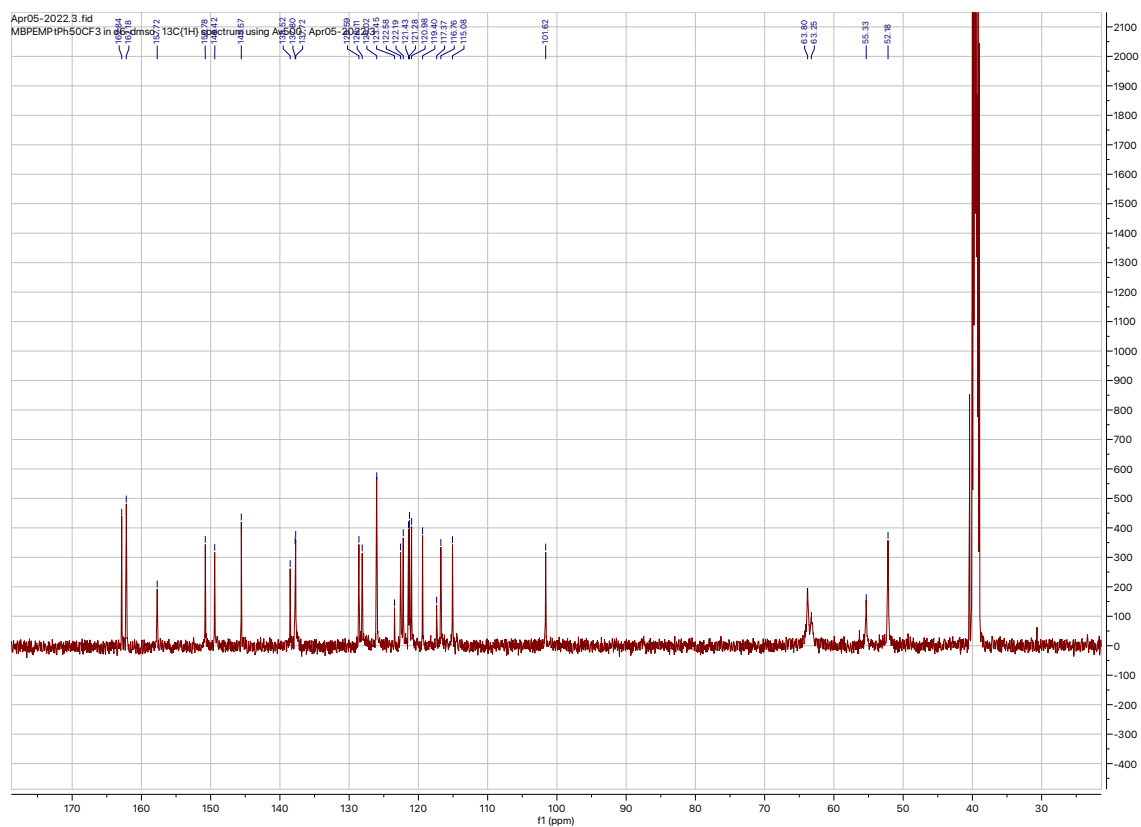


Figure A.127. ^{13}C NMR spectrum of Pt_PEM_5OCF₃.

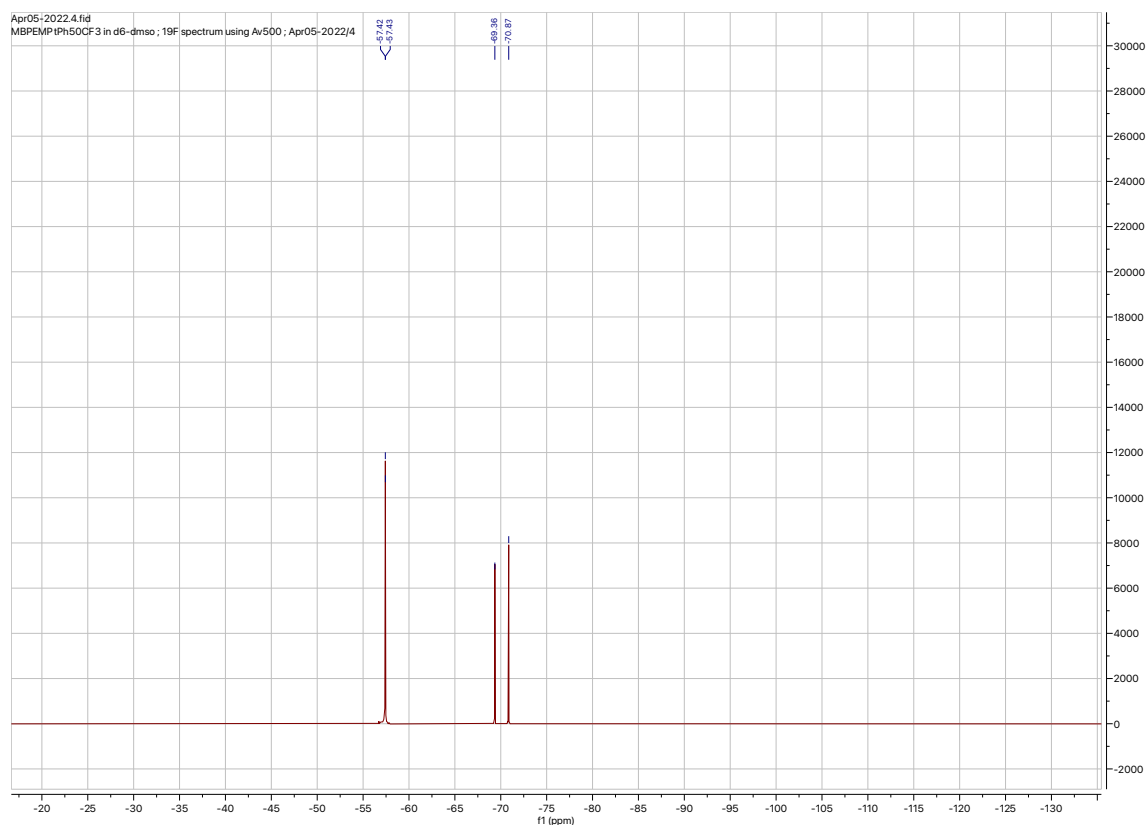


Figure A.128. ^{19}F NMR spectrum of Pt_PEM_5OCF_3 .

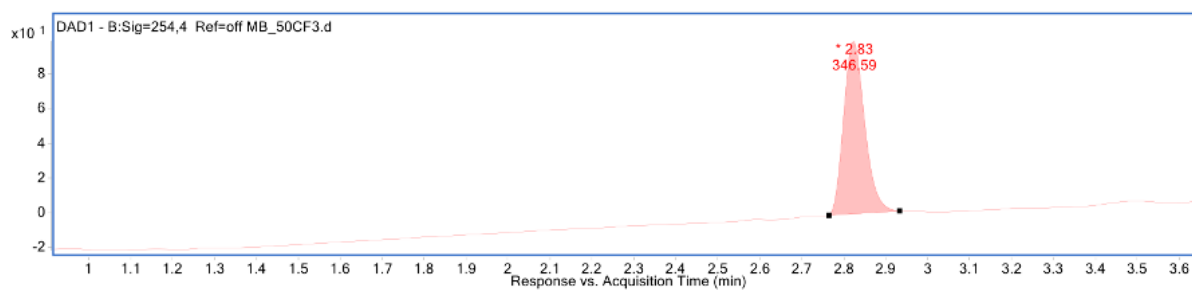


Figure A.129. LCMS UV trace of Pt_PEM_5OCF_3 .

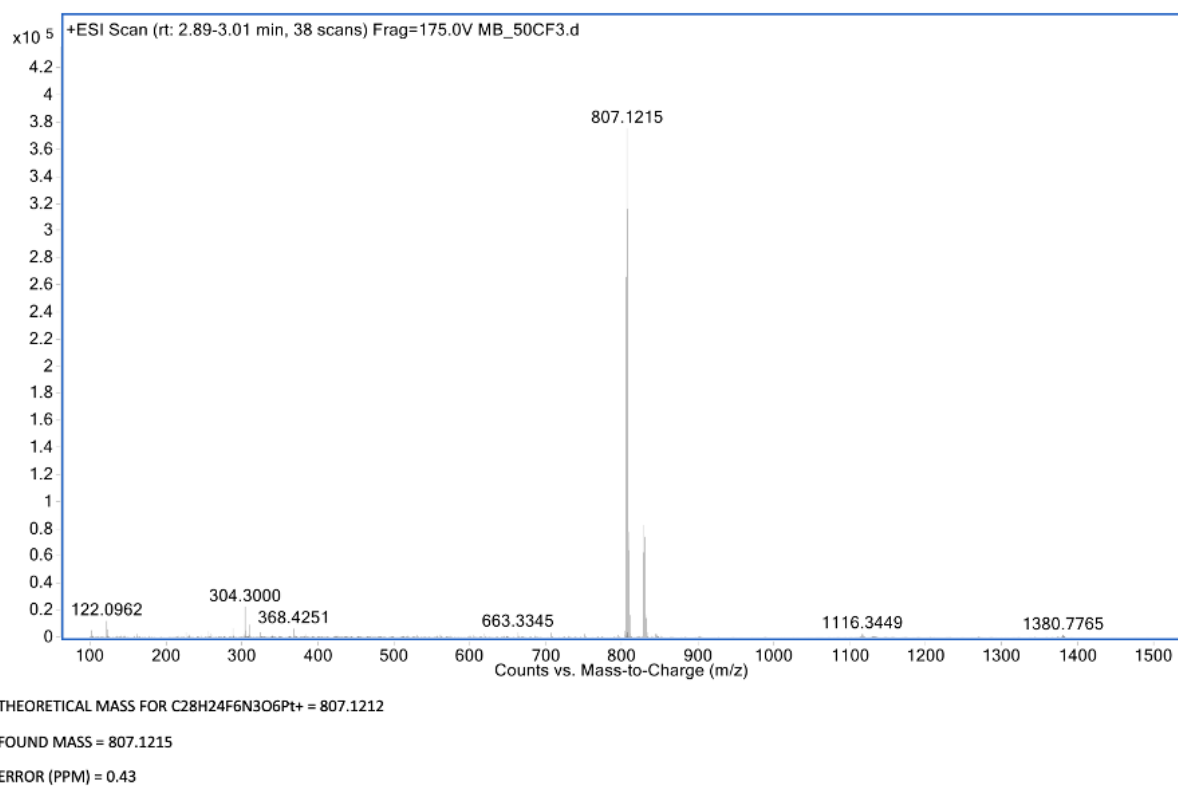
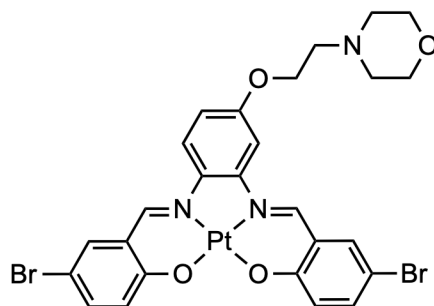


Figure A.130. HRMS of Pt_PEM_5OCF₃.

Pt_PEM_5Br:



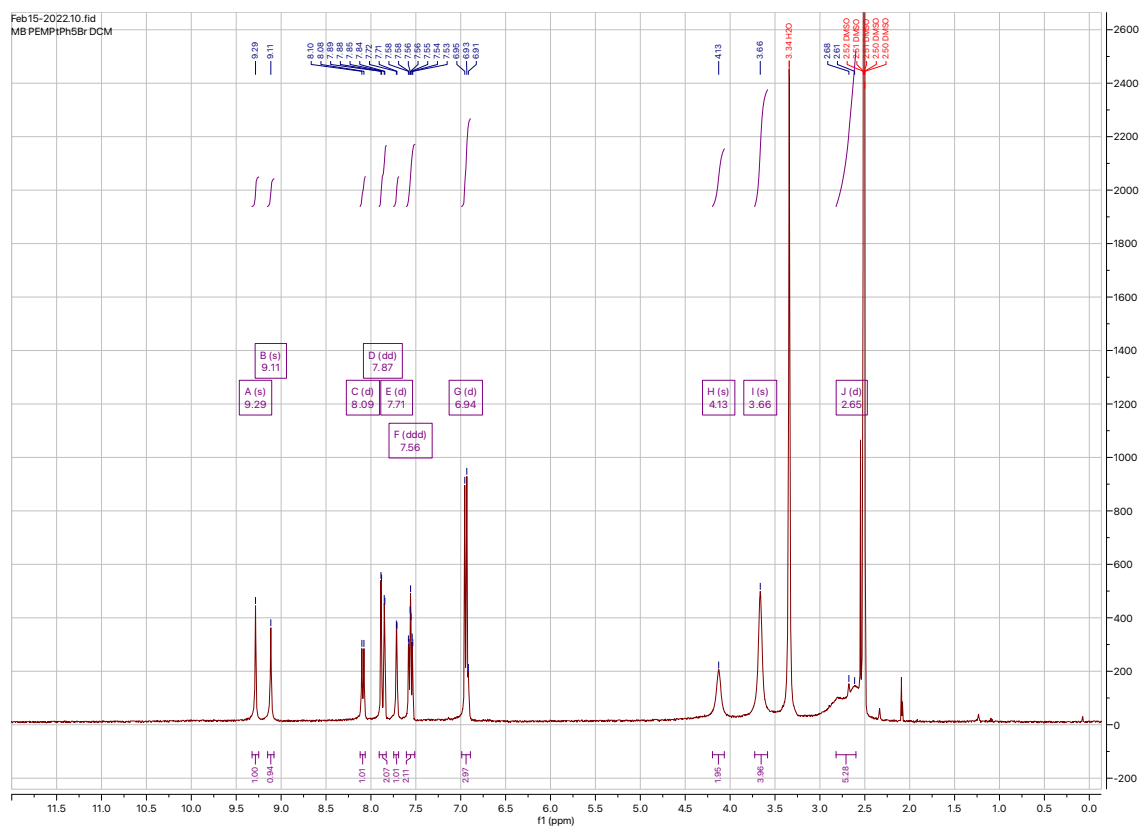


Figure A.131. Full ^1H NMR spectrum of Pt_PEM_5Br.

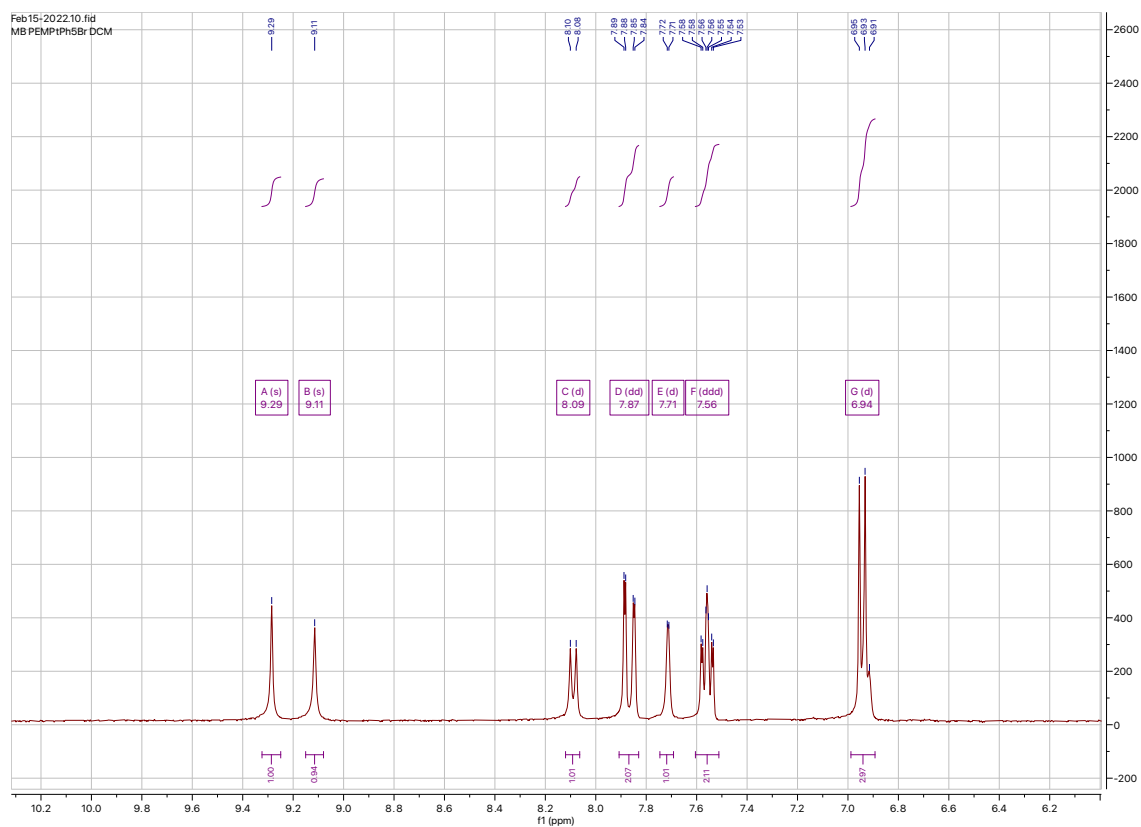


Figure A.132. Aromatic region of ^1H NMR spectrum of Pt_PEM_5Br.

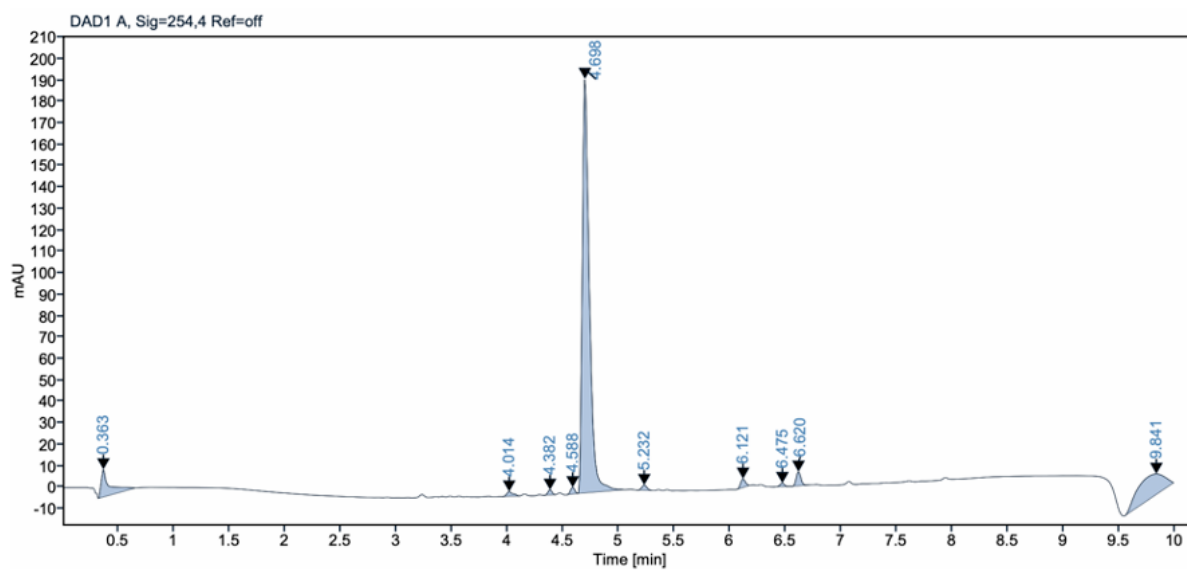


Figure A.135. LCMS UV trace of Pt_PEM_5Br.

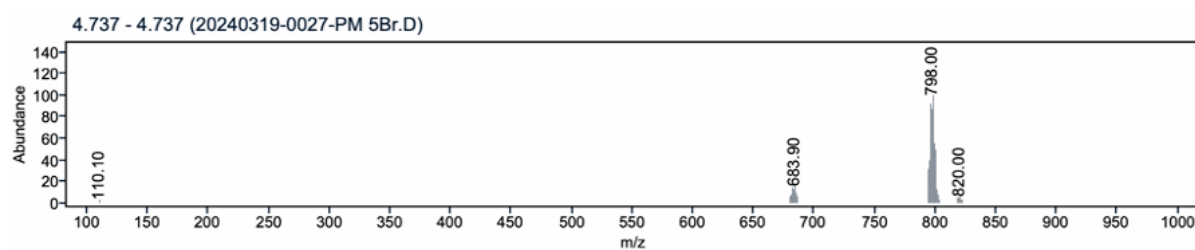
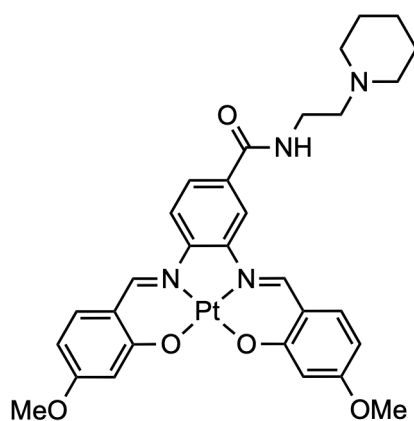


Figure A.136. Mass spec of Pt_PEM_5Br under LCMS UV peak at 4.70 minutes.

Pt_AEP_4OMe:



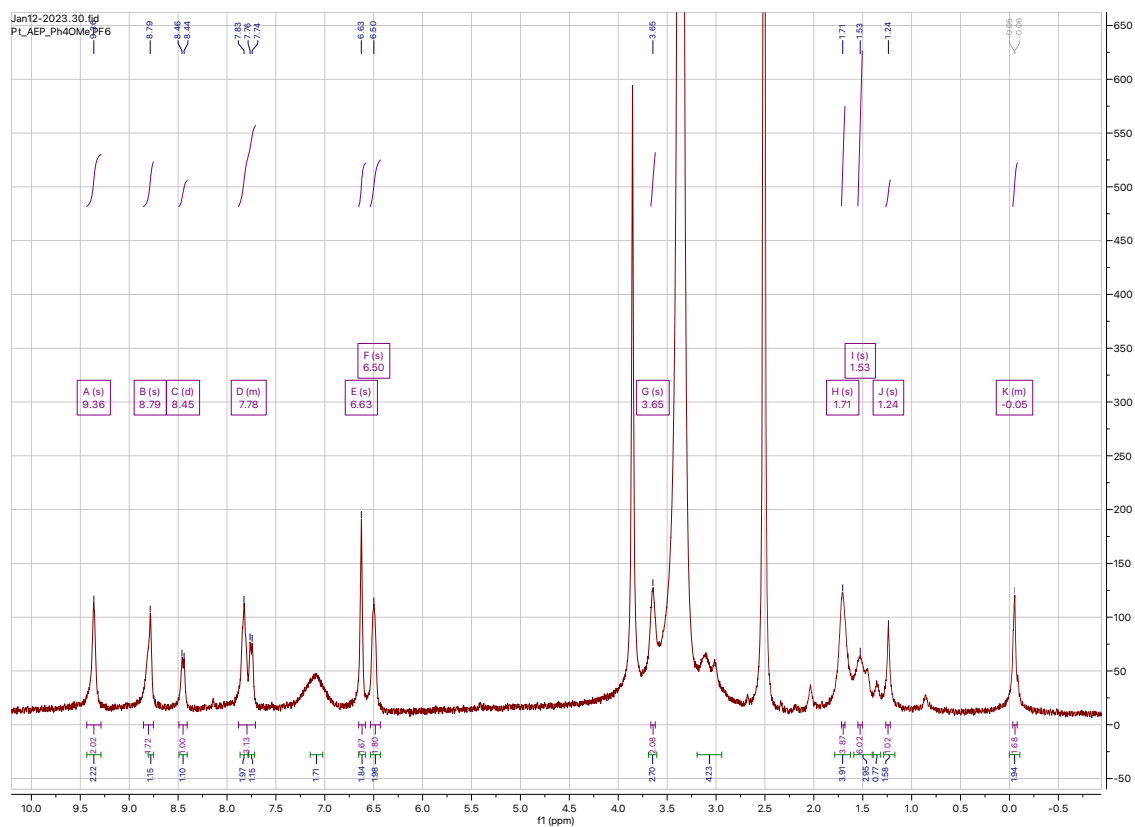


Figure A.137. Full ^1H NMR spectrum of Pt_AEP_4OMe.

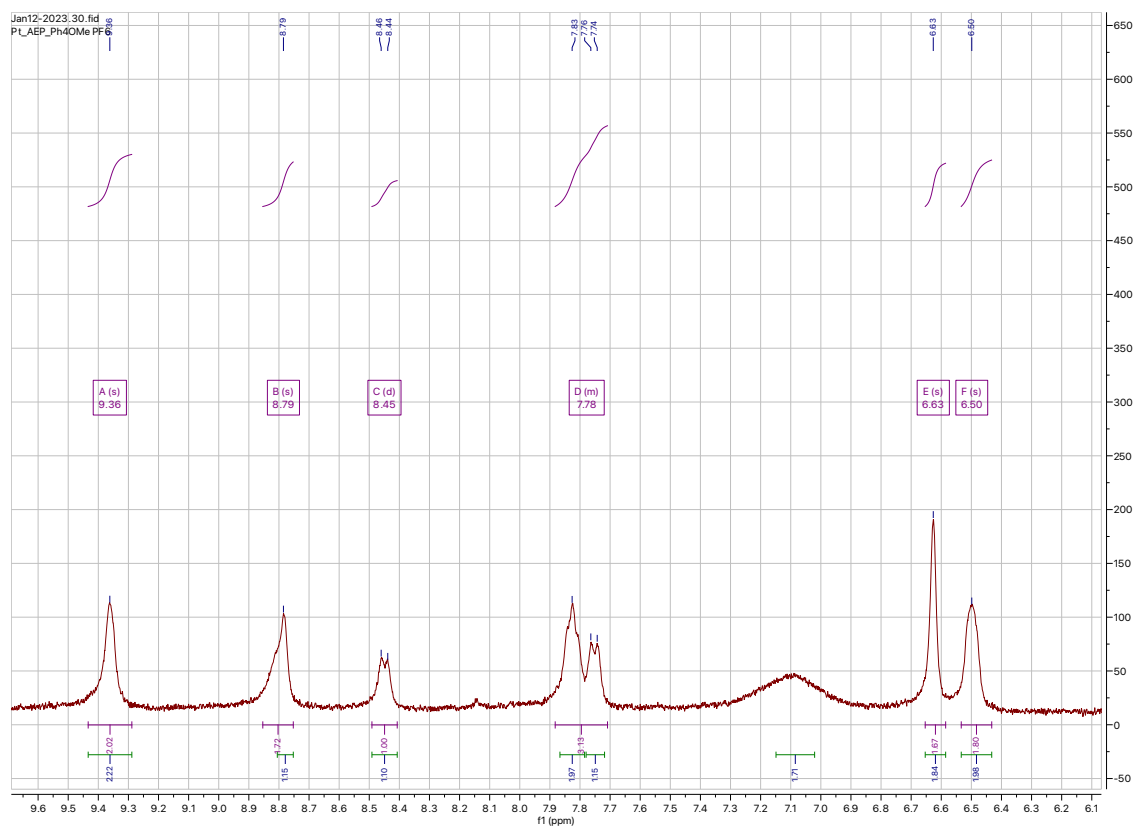


Figure A.138. Aromatic region of ^1H NMR spectrum of Pt_AEP_4OMe.

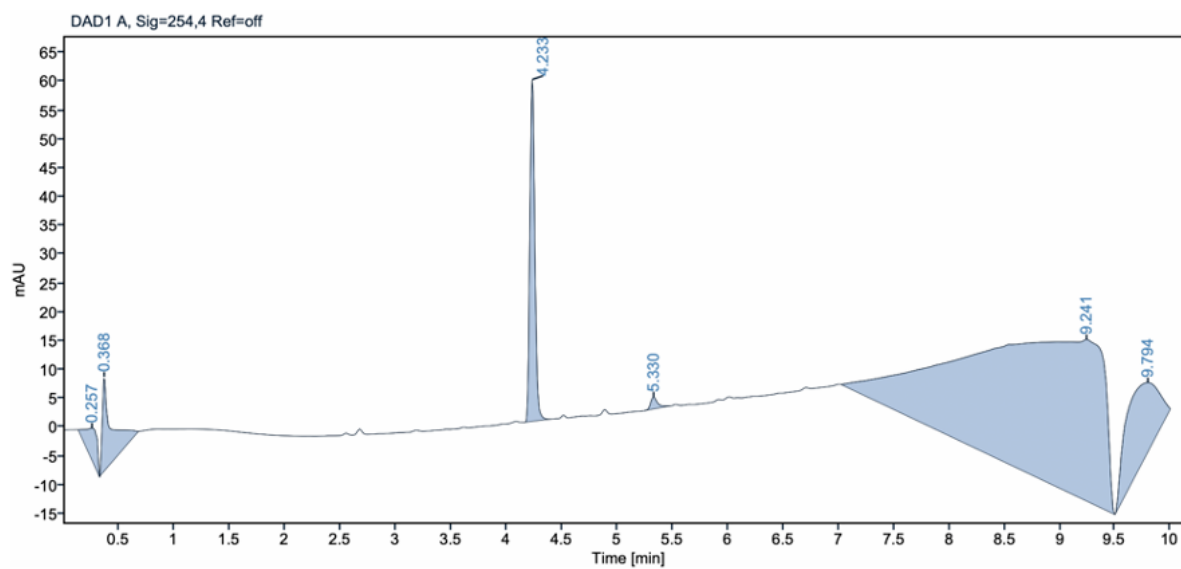


Figure A.139. LCMS UV trace of Pt_AEP_4OMe.

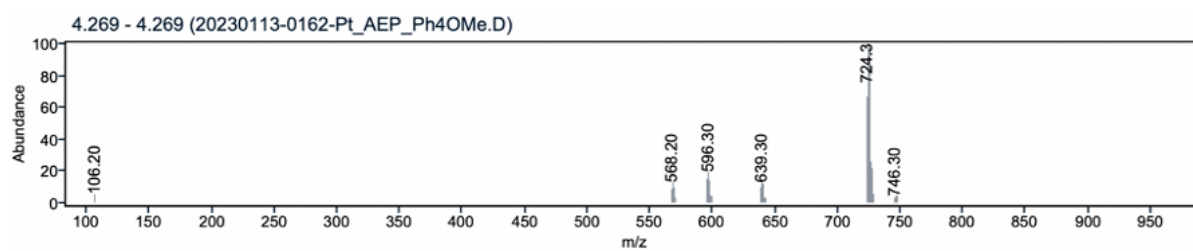


Figure A.140. Mass spec of Pt_AEP_4OMe under LCMS UV trace at 4.23 minutes.

Pt_PhBr_Ph:

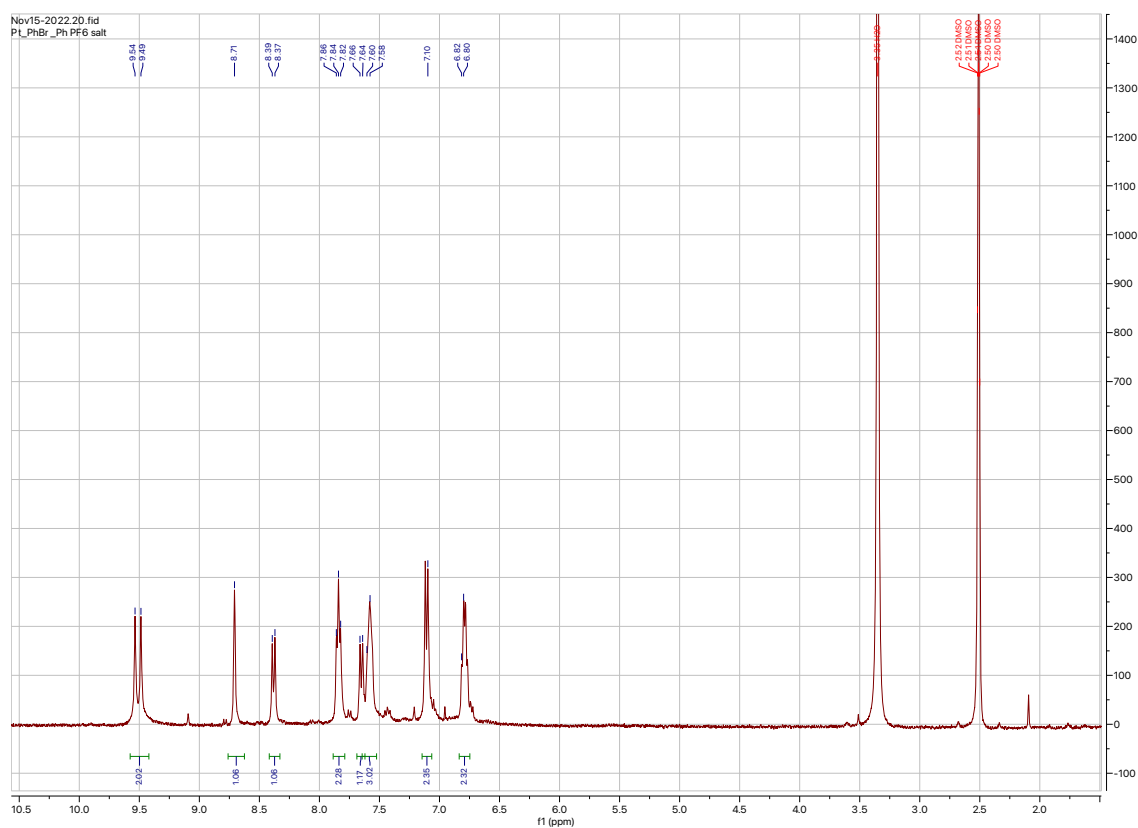
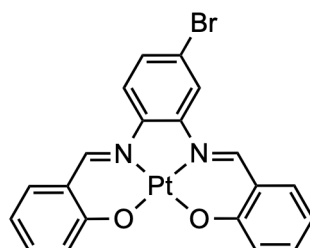


Figure A.141. Full ^1H NMR spectrum of Pt_PhBr_Ph.



Figure A.142. Aromatic region of ^1H NMR spectrum of Pt_PhBr_Ph.

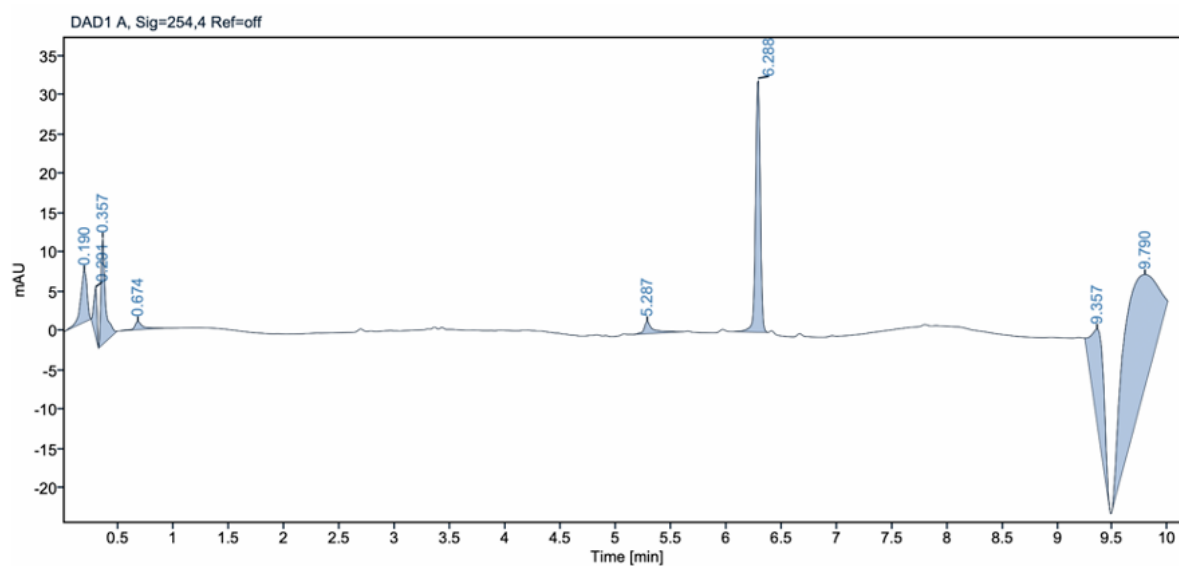


Figure A.143. LCMS UV trace of Pt_PhBr_Ph.

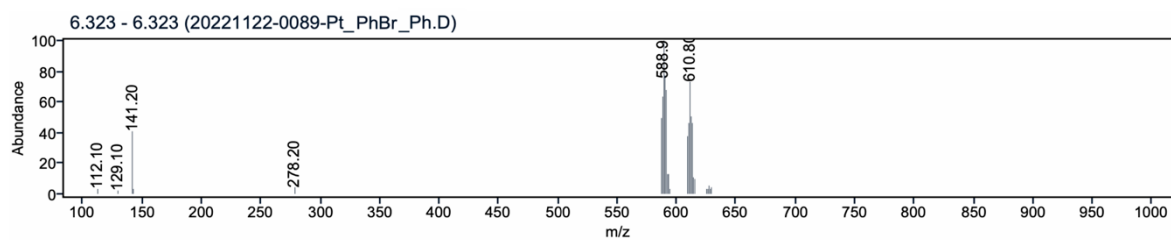


Figure A.144. Mass spec of Pt_PhBr_Ph under LCMS UV trace at 6.32 minutes. [MH⁺ = 589 m/z, M+Na+611 m/z)

A.2 Screening data

A.2.1 Photophysical properties

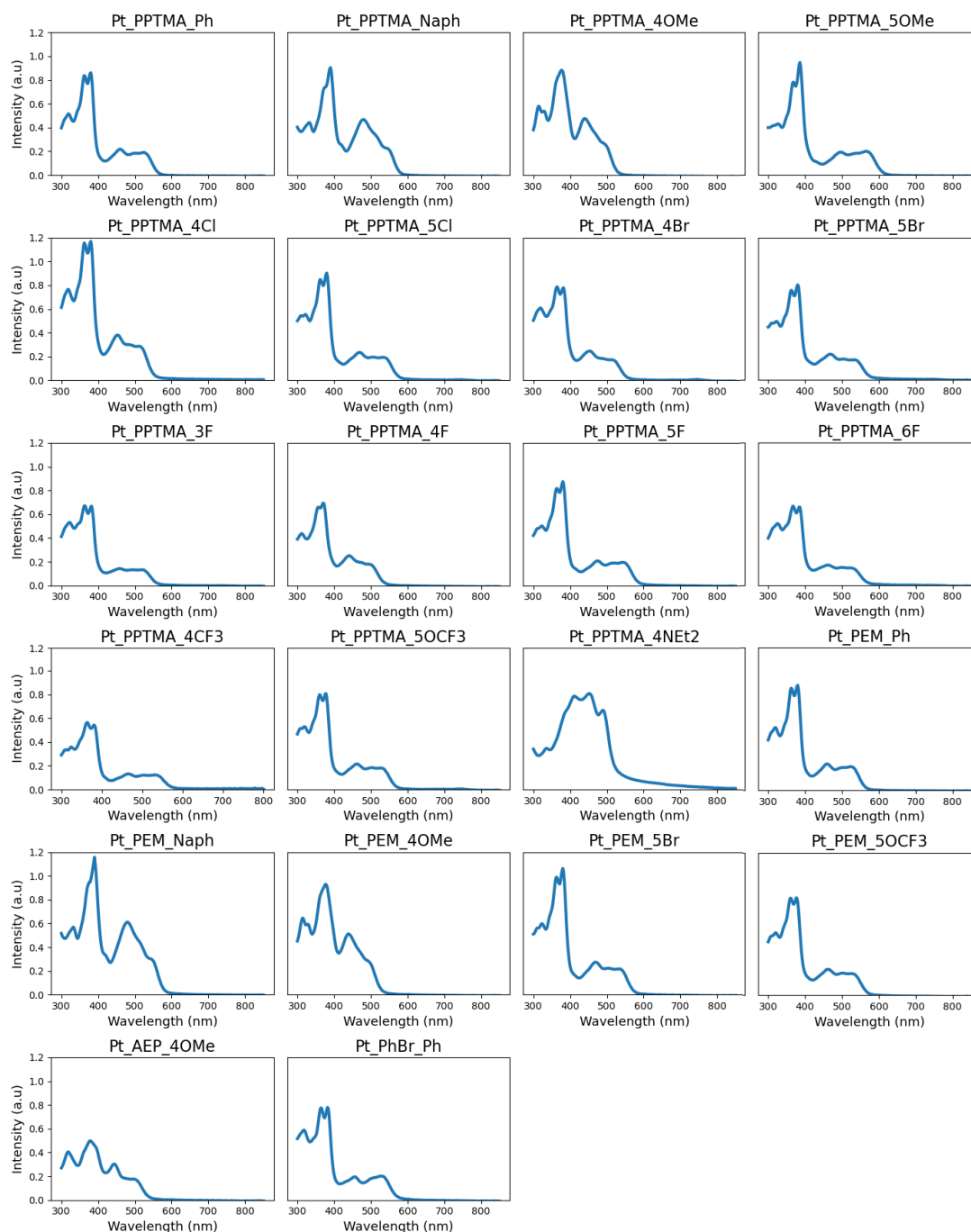


Figure A.145. Absorbance spectra for the full platinum(II) salen series in DMSO.

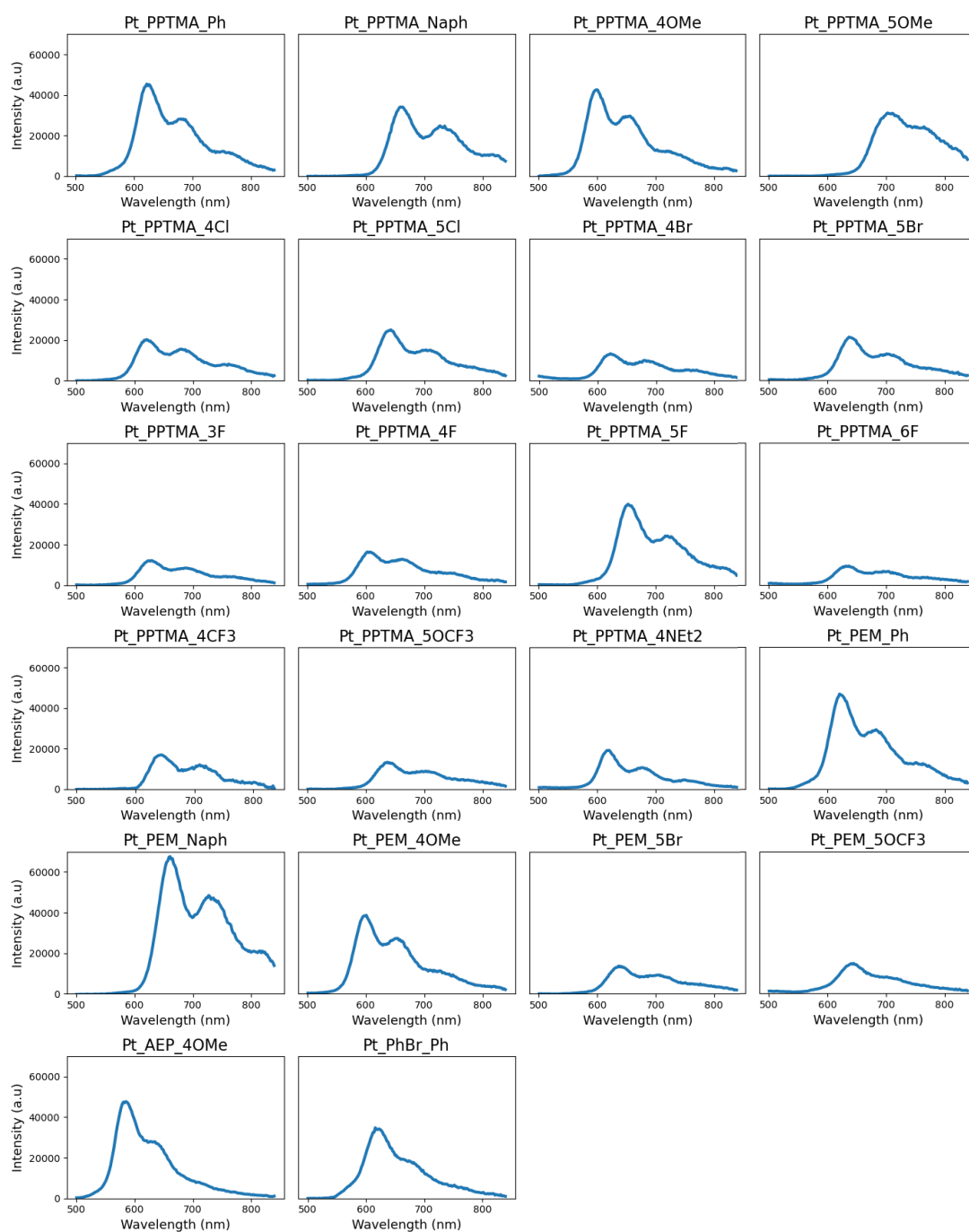


Figure A.146. Emission spectra for the full platinum(II) salen series in DMSO.

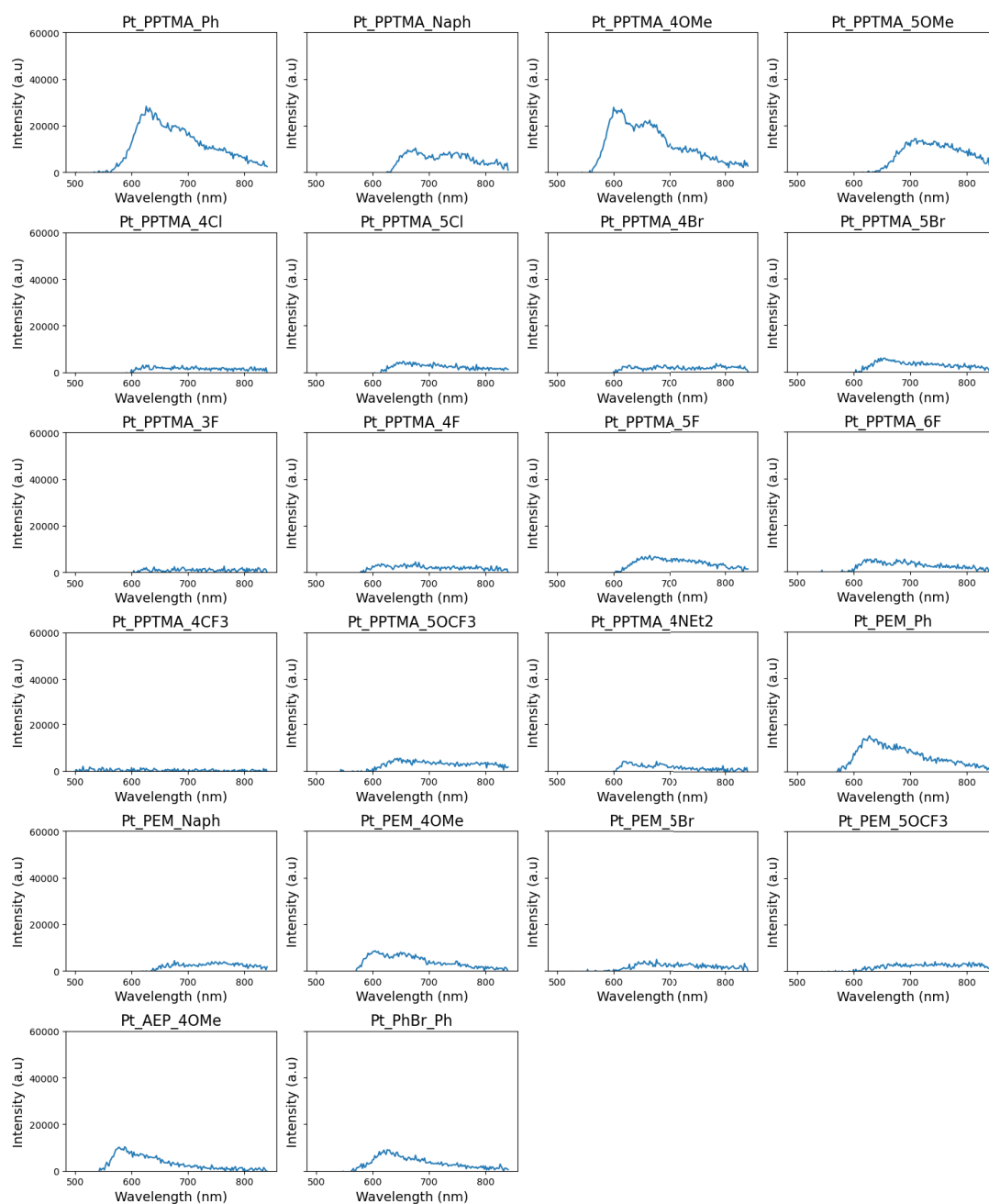


Figure A.147. Emission spectra for the full platinum(II) salen series in water.

Complex	Absorbance λ max (nm)	Emission λ max (nm)
Pt_PPTMA_Ph	320, 362, 380, 460	624, 682, 764
Pt_PPTMA_5Cl	322, 362, 380, 470	645, 706
Pt_PEM_Ph	320, 362, 380, 460	623, 686, 766
Pt_PPTMA_Naph	322, 364, 390, 480	661, 732
Pt_PPTMA_3F	322, 364, 382, 458	631, 690, 766
Pt_AEP_4OMe	320, 378, 444	641, 707
Pt_PPTMA_4Br	320, 364, 382, 454	624, 686, 764
Pt_PPTMA_4F	312, 356, 370, 444	605, 667, 753
Pt_PEM_4OMe	316, 376, 438	597, 651, 725
Pt_PEM_5Br	324, 360, 380, 468	639, 709,
Pt_PPTMA_5Br	324, 364, 380, 468	639, 701
Pt_PPTMA_5F	324, 364, 380, 474	655, 725, 821
Pt_PEM_Naph	332, 390, 480	659, 663, 819
Pt_PPTMA_4OMe	314, 376, 440	601, 655, 725
Pt_PPTMA_6F	326, 368, 384, 462	633, 697, 771
Pt_PEM_5OCF ₃	320, 360, 376, 464	647, 701
Pt_PPTMA_5OMe	328, 368, 386, 496	705, 767
Pt_PPTMA_4CF ₃	324, 364, 380, 466	642, 714
Pt_PPTMA_4Cl	318, 362, 380, 452	623, 685, 765
Pt_PPTMA_5OCF ₃	320, 360, 376, 462	643, 705
Pt_PPTMA_4NEt ₂	336, 412, 452, 490	623, 681, 753
Pt_PhBr_Ph	318, 364, 380, 455, 530	616, 633, 713

Table A.1. Peak absorbance and emission wavelengths for the platinum(II) salphen complexes in DMSO.

Complex	Peak Emission λ max (nm)	T ₁ to S ₀ energy gap (eV)
Pt_PPTMA_Ph	624	-2.01140745
Pt_PPTMA_5Cl	645	-1.94200186
Pt_PEM_Ph	623	-2.0138
Pt_PPTMA_Naph	661	-1.9412
Pt_PPTMA_3F	631	-1.9686
Pt_AEP_4OMe	587	-2.1594
Pt_PPTMA_4Br	624	-2.0417
Pt_PPTMA_4F	605	-2.1398
Pt_PEM_4OMe	603	-2.1754
Pt_PEM_5Br	639	-1.9492
Pt_PPTMA_5Br	639	-1.9469
Pt_PPTMA_5F	655	-1.8579
Pt_PEM_Naph	663	-1.936
Pt_PPTMA_4OMe	601	-2.1803
Pt_PPTMA_6F	639	-2.0401
Pt_PEM_5OCF ₃	647	-1.982
Pt_PPTMA_5OMe	705	-1.6750
Pt_PPTMA_4CF ₃	645	-1.9358
Pt_PPTMA_4Cl	623	-2.0521
Pt_PPTMA_5OCF ₃	639	-1.9808
Pt_PPTMA_4NEt ₂	623	-2.0759
Pt_PhBr_Ph	616	-1.9866

Table A.2. Table showing the peak emission wavelengths in DMSO and energy gap from the S₀ to T₁ transition used to calculate the Pearson correlation coefficient.

A.2.2 ROS generation plots

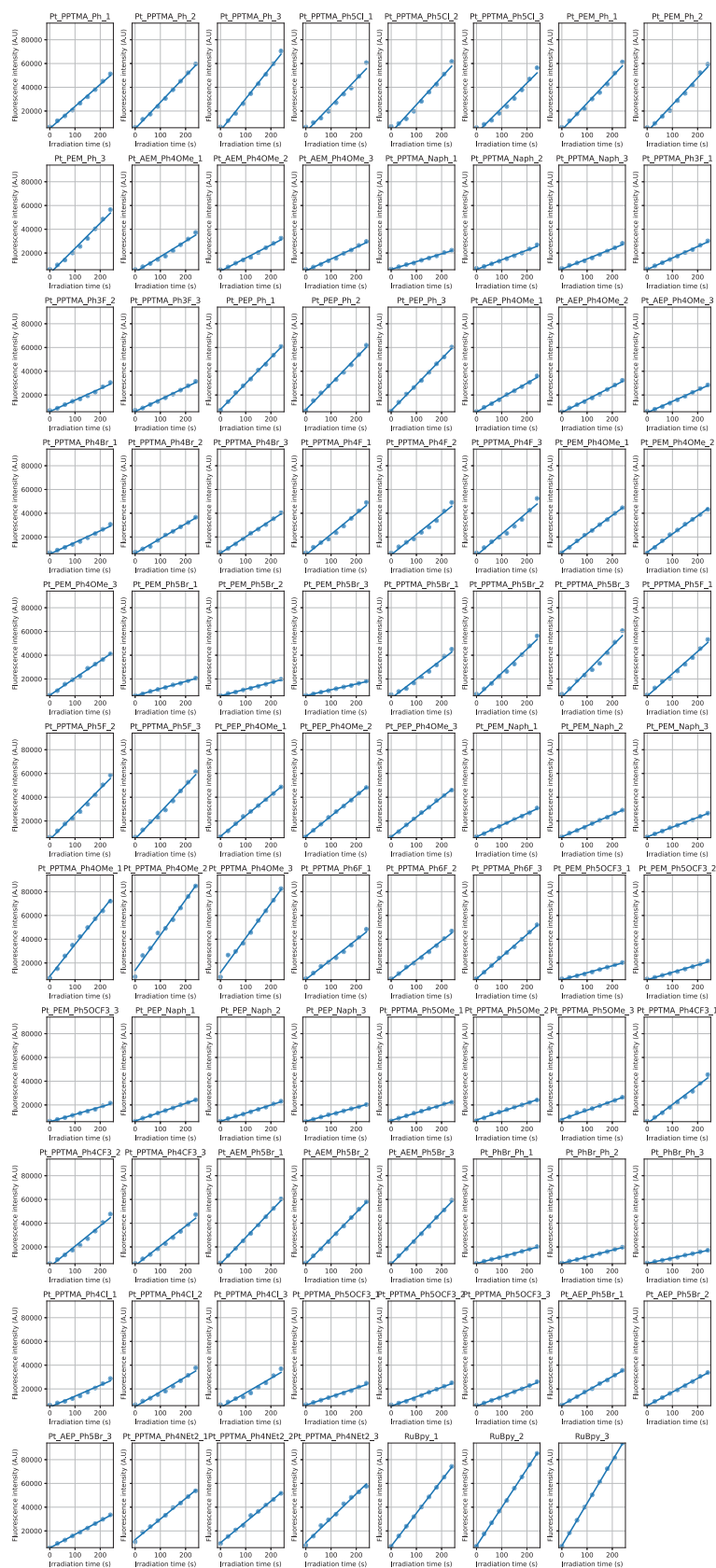


Figure A.148. Change in DHR-123 emission with each Pt(II)-Salphen complex in triplicate with increasing irradiation.

Complex	Relative DHR ROS Generation (%)	Error (%)
Pt_PPTMA_Ph	70.3927328	10.575306
Pt_PPTMA_Naph	24.230957	7.8707971
Pt_PPTMA_4Br	37.527905	9.7736019
Pt_PPTMA_5Br	60.383212	8.2831065
Pt_PPTMA_4OMe	89.108936	3.0848665
Pt_PPTMA_5OMe	22.201399	4.488752
Pt_PPTMA_4Cl	35.48243	9.4652445
Pt_PPTMA_5Cl	69.235978	2.778408
Pt_PPTMA_3F	30.735893	1.7935157
Pt_PPTMA_4F	54.583638	1.6097527
Pt_PPTMA_5F	65.35611	5.3445146
Pt_PPTMA_6F	53.792674	4.1825636
Pt_PPTMA_4CF ₃	51.876731	2.0914879
Pt_PPTMA_5OCF ₃	24.383467	3.7529399
Pt_PEM_Ph	69.157322	2.1181432
Pt_PEM_4OMe	47.30435	2.2796208
Pt_PEM_5OCF ₃	19.076263	2.0768512
Pt_AEP_4OMe	33.898476	7.0801676
Pt_PEM_5Br	17.273009	6.0091368
Pt_PEM_Naph	28.514294	5.3902209
Pt_PPTMA_4NEt ₂	57.807341	5.7272506
Pt_PhBr_Ph	16.481002	6.8784854
RuBpy	100	7.2796631

Table A.3. DHR-123 assay relative ROS generation percentage and errors.

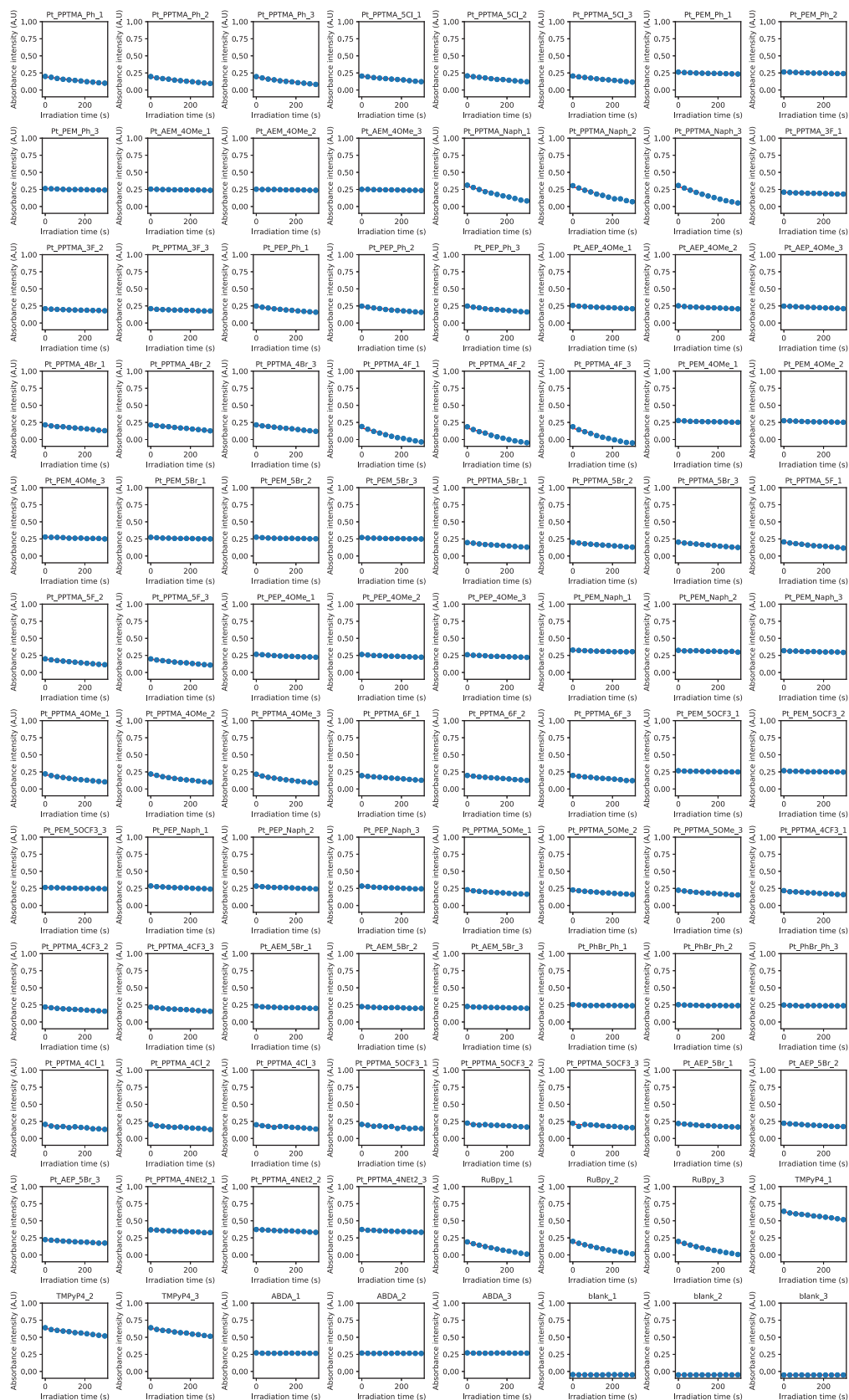


Figure A.149. Linear plots of change in ABDA absorbance with each Pt(II)-Salphen complex in triplicate with increasing intervals of irradiation.

Complex	Relative ABDA ROS Generation (%)	Error (%)
Pt_PPTMA_Ph	55.7082984	0.407
Pt_PPTMA_Naph	131.047779	0.393
Pt_PPTMA_4Br	47.1751886	0.406
Pt_PPTMA_5Br	38.6923722	0.503
Pt_PPTMA_4OMe	63.8222967	0.298
Pt_PPTMA_5OMe	35.2053646	0.324
Pt_PPTMA_4Cl	31.2824811	0.490
Pt_PPTMA_5Cl	46.4207879	0.306
Pt_PPTMA_3F	15.4568315	0.493
Pt_PPTMA_4F	125.481978	0.241
Pt_PPTMA_5F	46.7225482	0.111
Pt_PPTMA_6F	38.4073764	0.339
Pt_PPTMA_4CF ₃	30.9807209	0.239
Pt_PPTMA_5OCF ₃	28.8013412	0.724
Pt_PEM_Ph	12.9086337	0.376
Pt_PEM_4OMe	12.8248114	0.512
Pt_PEM_5OCF ₃	9.38809723	0.404
Pt_AEP_4OMe	21.9446773	0.594
Pt_PEM_5Br	10.4442582	0.444
Pt_PEM_Naph	12.0536463	0.614
Pt_PPTMA_4NEt ₂	21.6093881	0.393
Pt_PhBr_Ph	5.50	0.349
RuBpy	100	0.184

Table A.4. ABDA assay relative ROS generation percentage and errors.

A.2.3 DNA binding curves

A.2.3.1 HTG4 binding assay: hand prepared

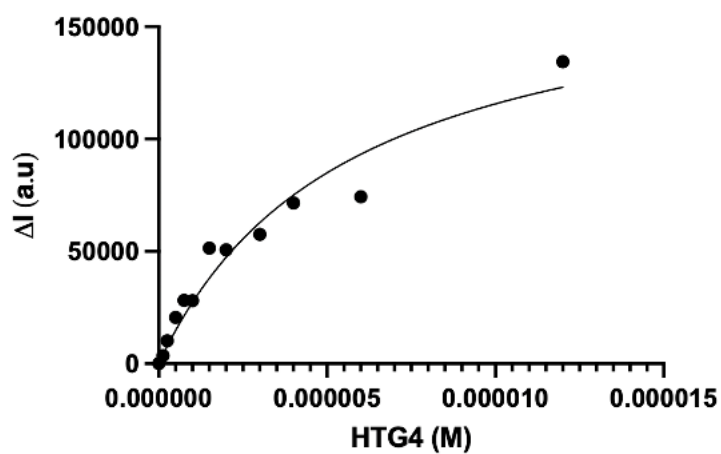


Figure A.150. Pt_PPTMA_Ph HTG4 DNA binding curve. Assay prepared by hand.

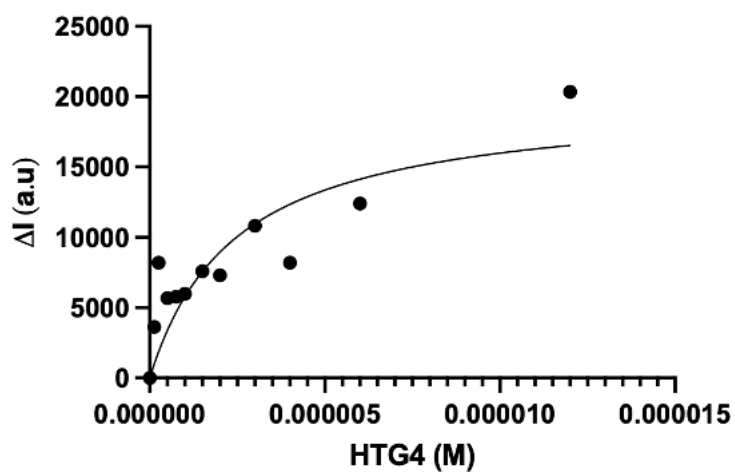


Figure A.151. Pt_PPTMA_Naph HTG4 DNA binding curve. Assay prepared by hand.

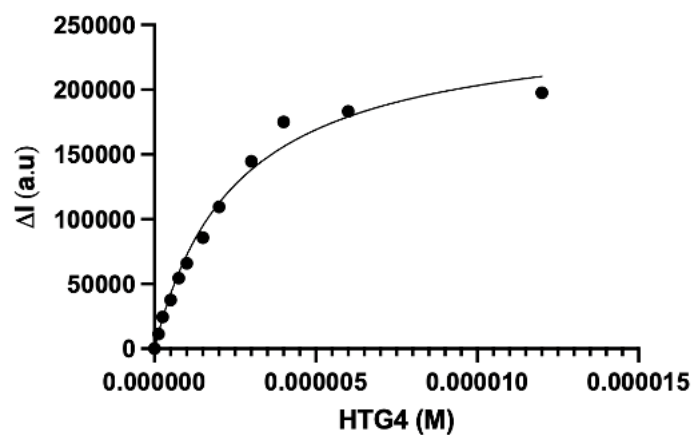


Figure A.152. Pt_PPTMA_4OMe HTG4 DNA binding curve. Assay prepared by hand.

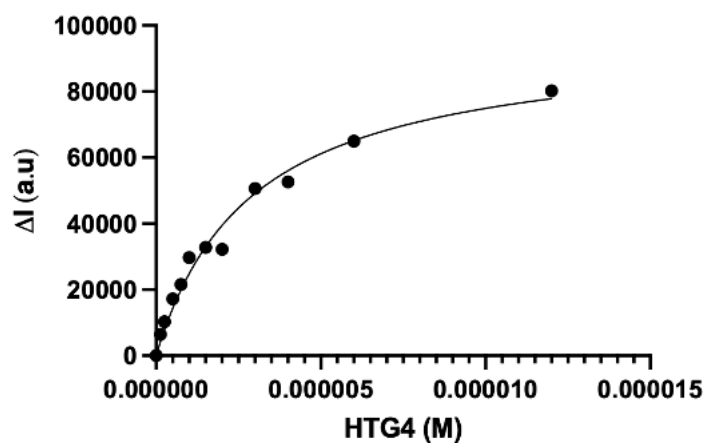


Figure A.153. Pt_PPTMA_4F HTG4 DNA binding curve. Assay prepared by hand.

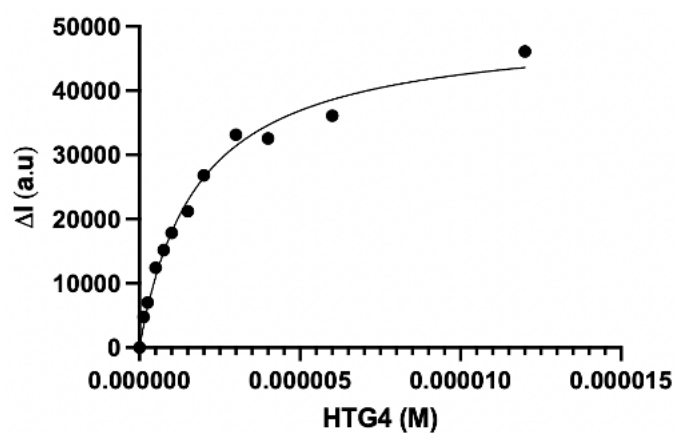


Figure A.154. Pt_PPTMA_5F HTG4 DNA binding curve. Assay prepared by hand.

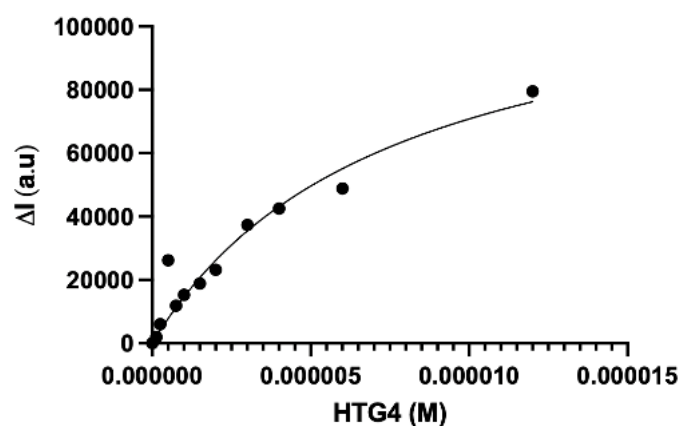


Figure A.155. Pt_PPTMA_5OMe HTG4 DNA binding curve. Assay prepared by hand.

	HTG4	
Complex	K_a (M^{-1})	% error
Pt_PPTMA_Ph	1.77E+05	27.3437293
Pt_PPTMA_Naph	4.08E+05	55.0352949
Pt_PPTMA_4OMe	3.96E+05	12.466785
Pt_PPTMA_5OMe	1.34E+05	34.8310089
Pt_PPTMA_4F	3.45E+05	13.9356896
Pt_PPTMA_5F	5.59E+05	10.9381853

Table A.5. Table contains K_a 's for trial complexes with HTG4 DNA. From hand prepared assays.

A.2.3.2 HTG4 binding assay: Opentron2 prepared

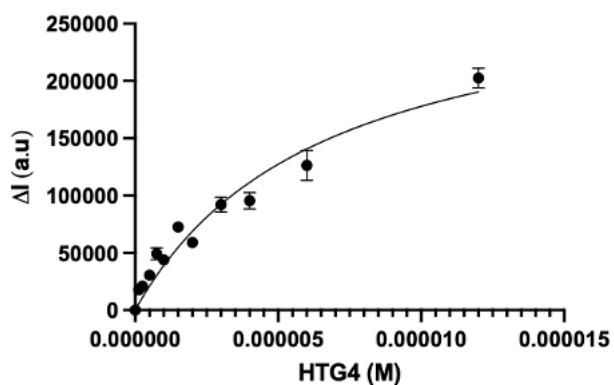


Figure A.156. Pt_PPTMA_Ph HTG4 DNA binding curve. Assay prepared by Opentron2.

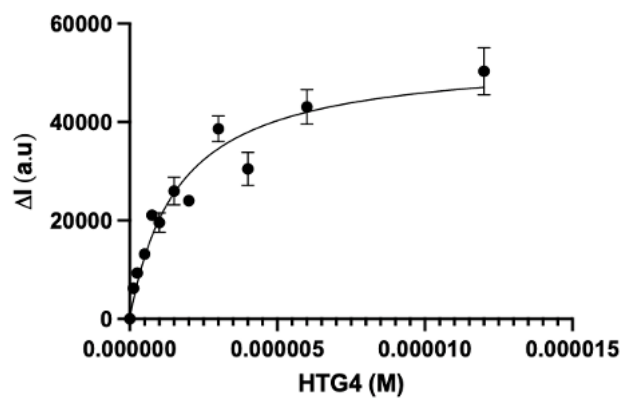


Figure A.157. Pt_PPTMA_5Br HTG4 DNA binding curve. Assay prepared by Opentron2.

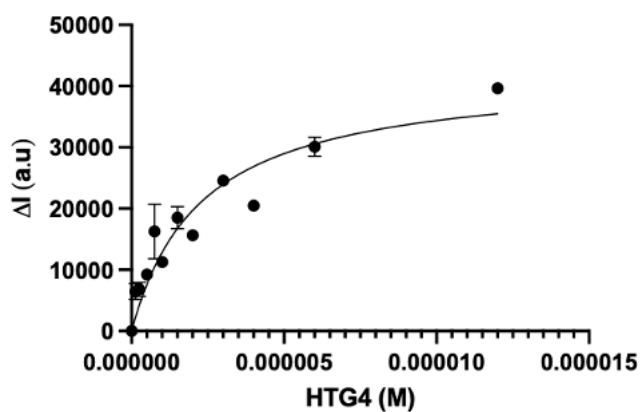


Figure A.158. Pt_PPTMA_Naph HTG4 DNA binding curve. Assay prepared by Opentron2.

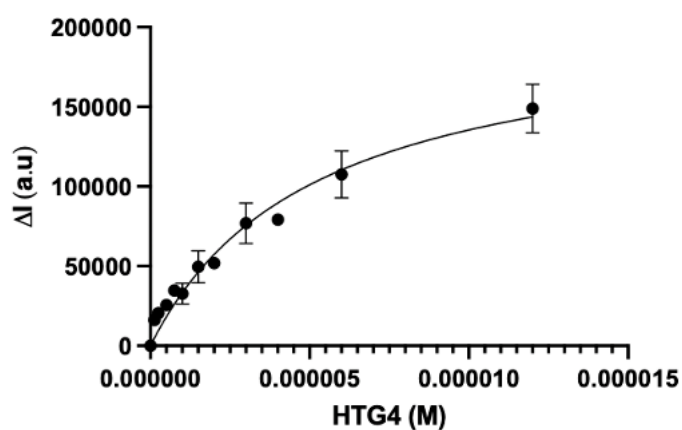


Figure A.159. Pt_PPTMA_4OMe HTG4 DNA binding curve. Assay prepared by Opentron2.

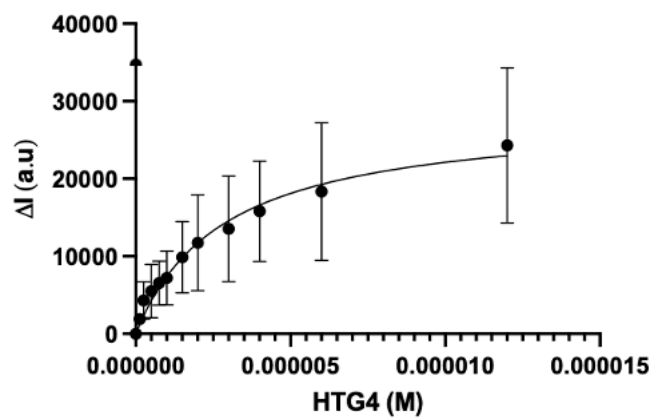


Figure A.160. Pt_PPTMA_5OCF₃ HTG4 DNA binding curve. Assay prepared by Opentron2.

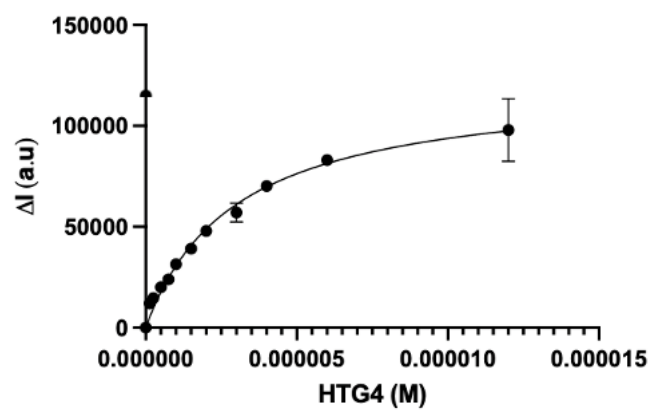


Figure A.161. Pt_PPTMA_4CF₃ HTG4 DNA binding curve. Assay prepared by Opentron2.

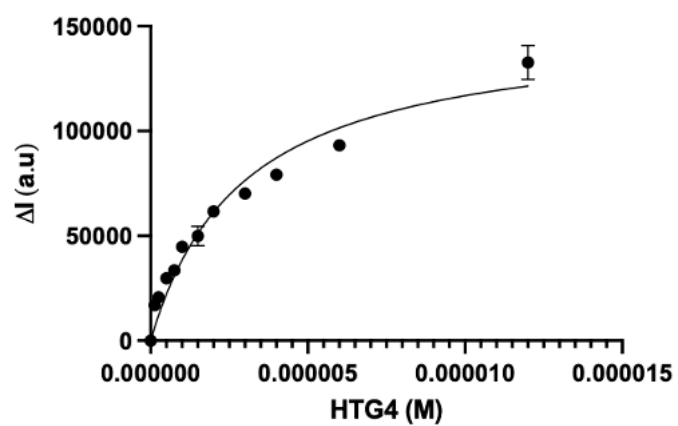


Figure A.162. Pt_PPTMA_4F HTG4 DNA binding curve. Assay prepared by Opentron2.

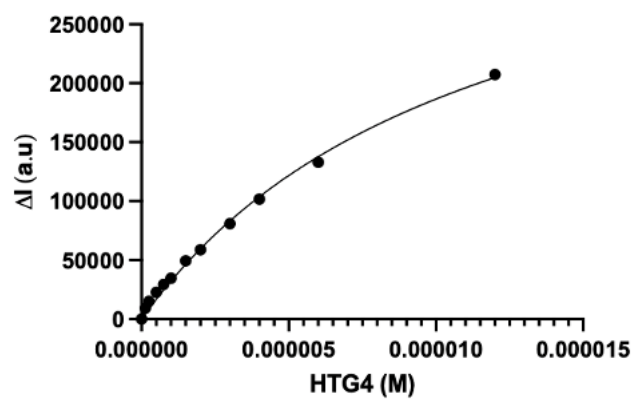


Figure A.163. Pt_PPTMA_3F HTG4 DNA binding curve. Assay prepared by Opentron2.

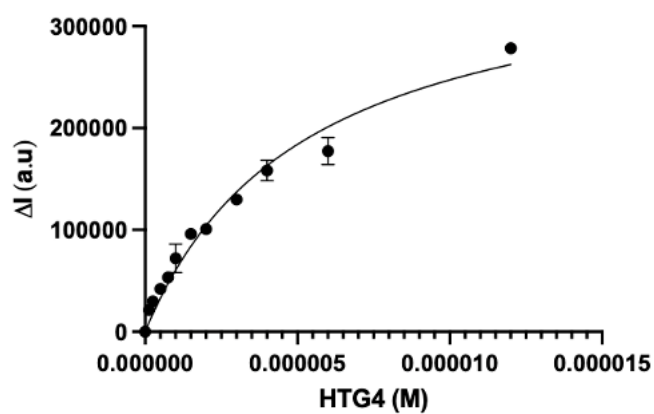


Figure A.164. Pt_PPTMA_5F HTG4 DNA binding curve. Assay prepared by Opentron2.

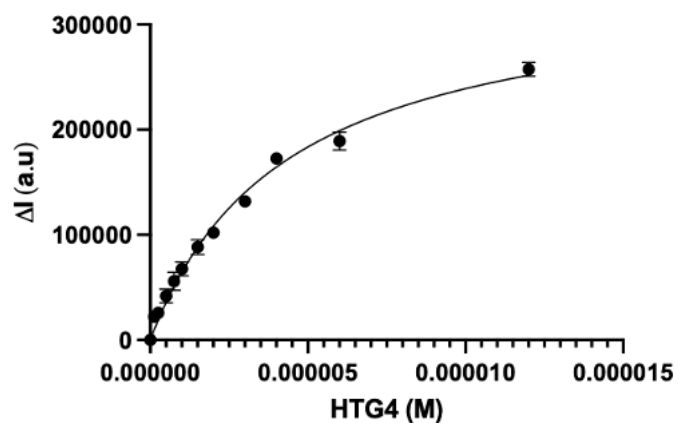


Figure A.165. Pt_PPTMA_6F HTG4 DNA binding curve. Assay prepared by Opentron2.

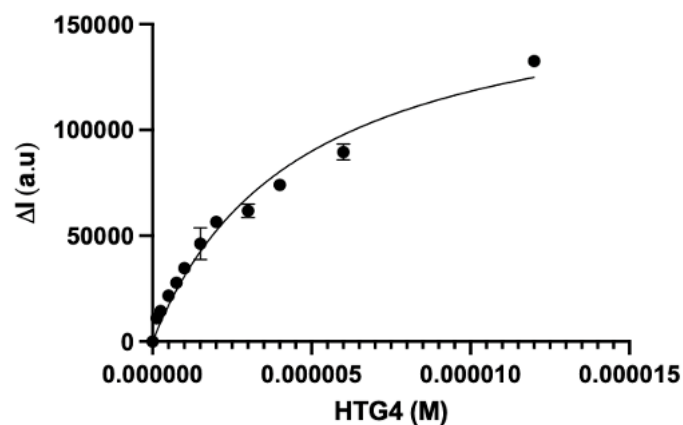


Figure A.166. Pt_PPTMA_4Br HTG4 DNA binding curve. Assay prepared by Opentron2.

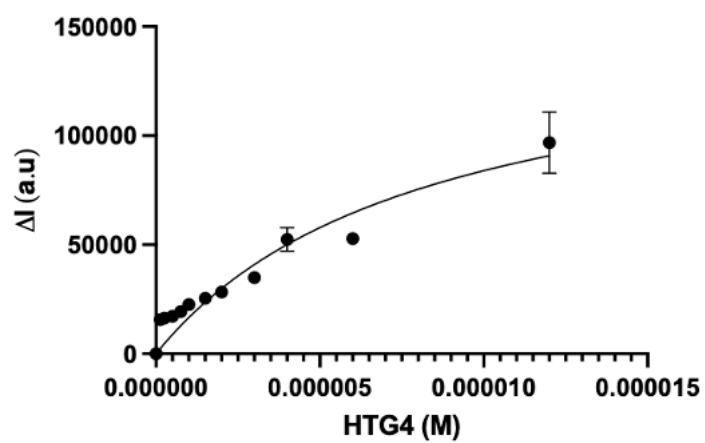


Figure A.167. : Pt_PPTMA_5OMe HTG4 DNA binding curve. Assay prepared by Opentron2.

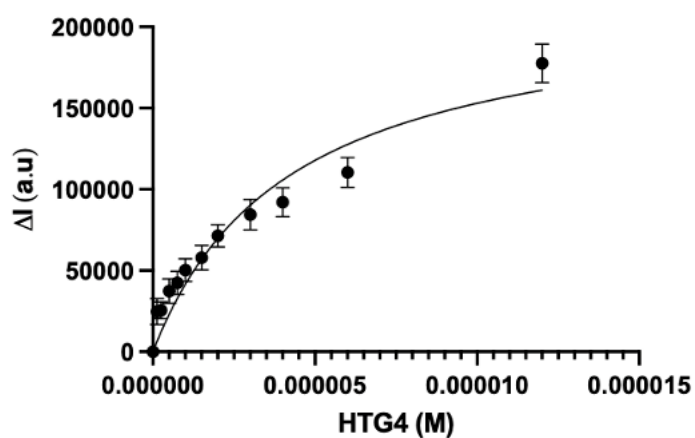


Figure A.168. Pt_PPTMA_4Cl HTG4 DNA binding curve. Assay prepared by Opentron2.

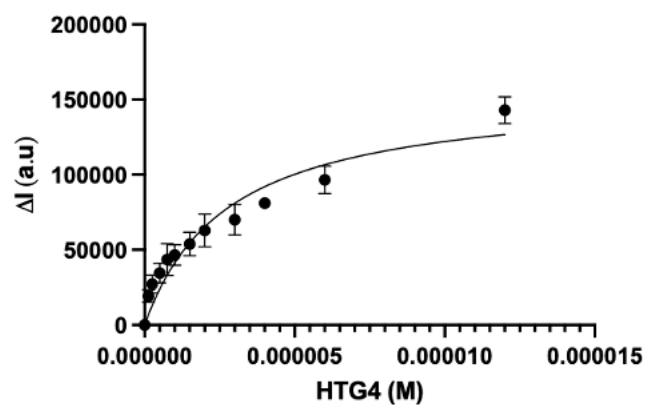


Figure A.169. Pt_PPTMA_5Cl HTG4 DNA binding curve. Assay prepared by Opentron2.

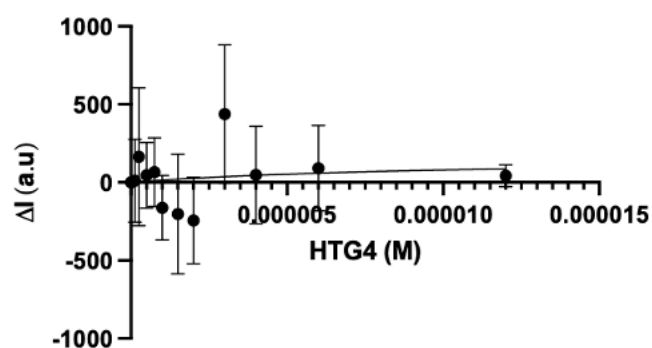


Figure A.170. Pt_PPTMA_4NEt₂ HTG4 DNA binding curve. Assay prepared by Opentron2.

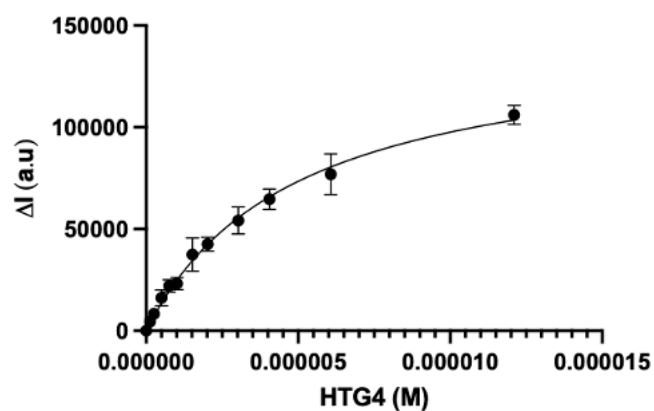


Figure A.171. Pt_PEM_Ph HTG4 DNA binding curve. Assay prepared by Opentron2.

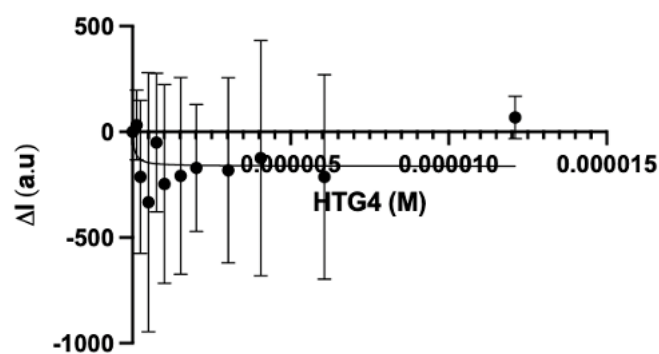


Figure A.172. Pt_PEM_Naph HTG4 DNA binding curve. Assay prepared by Opentron2.

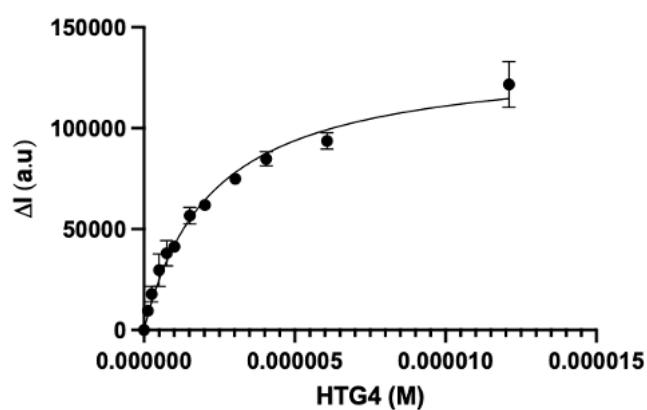


Figure A.173. Pt_PEM_4OMe HTG4 DNA binding curve. Assay prepared by Opentron2.

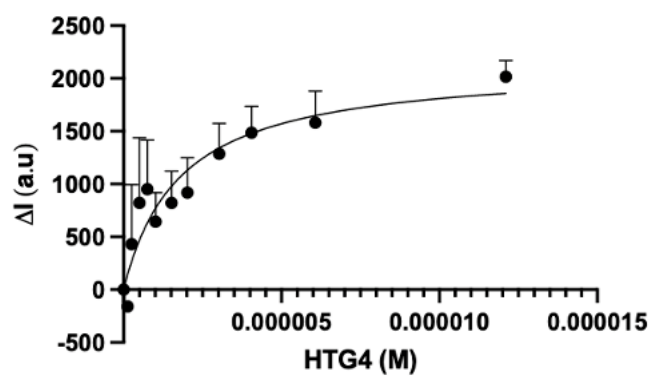


Figure A.174. Pt_PEM_5Br HTG4 DNA binding curve. Assay prepared by Opentron2.

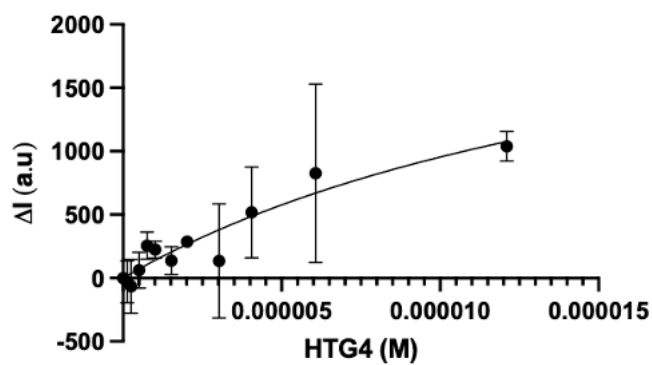


Figure A.175. Pt_PEM_5OCF₃ HTG4 DNA binding curve. Assay prepared by Opentron2.

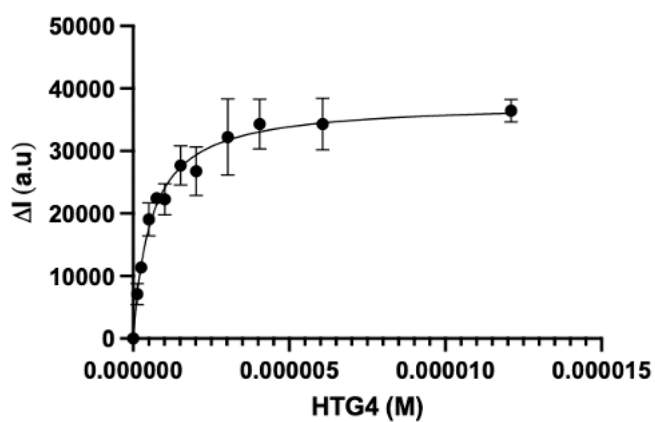


Figure A.176. Pt_AEP_4OMe HTG4 DNA binding curve. Assay prepared by Opentron2.

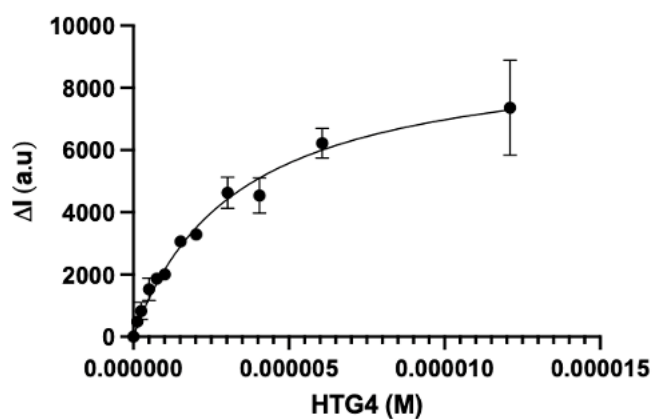


Figure A.177. Pt_PhBr_Ph HTG4 DNA binding curve. Assay prepared by Opentron2.

	HTG4	
Complex	K_a (M^{-1})	% error
Pt_PPTMA_Ph	1.52E+05	16.0088238
Pt_PPTMA_Naph	4.36E+05	18.420698
Pt_PPTMA_4OMe	1.93E+05	14.4079978
Pt_PPTMA_5OMe	1.22E+05	23.9136316
Pt_PPTMA_4F	3.40E+05	11.5106694
Pt_PPTMA_5F	1.90E+05	10.6816017
Pt_PPTMA_5Br	6.09E+05	14.0916943
Pt_PPTMA_5OCF ₃	3.51E+05	36.9581056
Pt_PPTMA_4CF ₃	3.36E+05	9.64129748
Pt_PPTMA_3F	8.76E+04	6.40111004
Pt_PPTMA_6F	2.31E+05	6.85513353
Pt_PPTMA_4Br	2.17E+05	9.98379403
Pt_PPTMA_4Cl	2.35E+05	17.1003942
Pt_PPTMA_5Cl	3.59E+05	17.2071696
Pt_PEM_5Br	5.59E+05	31.4016329
Pt_PEM_4OMe	4.45E+05	8.53824458
Pt_PEM_Ph	2.04E+05	9.44460387
Pt_AEP_4OMe	1.74E+06	12.1778813
Pt_PhBr_Ph	3.00E+05	14.4130169
Pt_PEM_Naph	0.00E+00	N/A
Pt_PEM_5OCF ₃	0.00E+00	N/A

Table A.6. Table contains K_a 's for trial complexes with HTG4 DNA. From OpenTron2 prepared assays.

A.2.3.3 High throughput screen: c-myc, HTelo, ckit-1, LTR(III), and CT-DNA

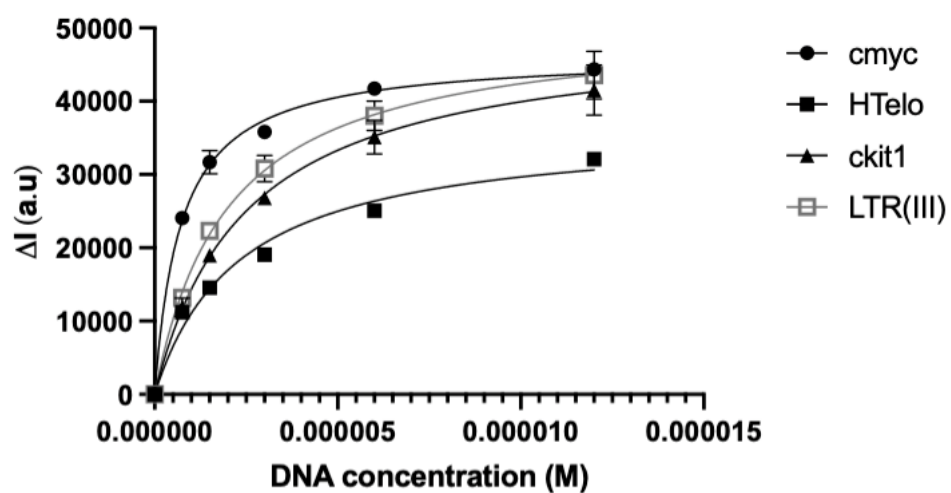


Figure A.178. Pt_PPTMA_Ph G4 DNA binding curves.

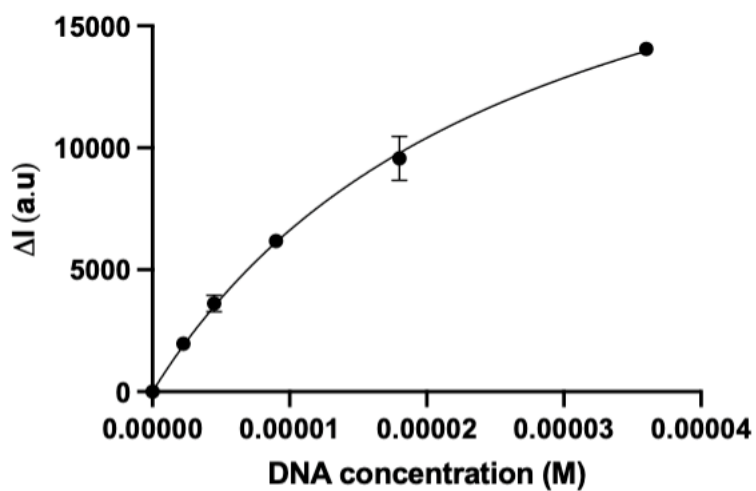


Figure A.179. Pt_PPTMA_Ph CT-DNA binding curve.

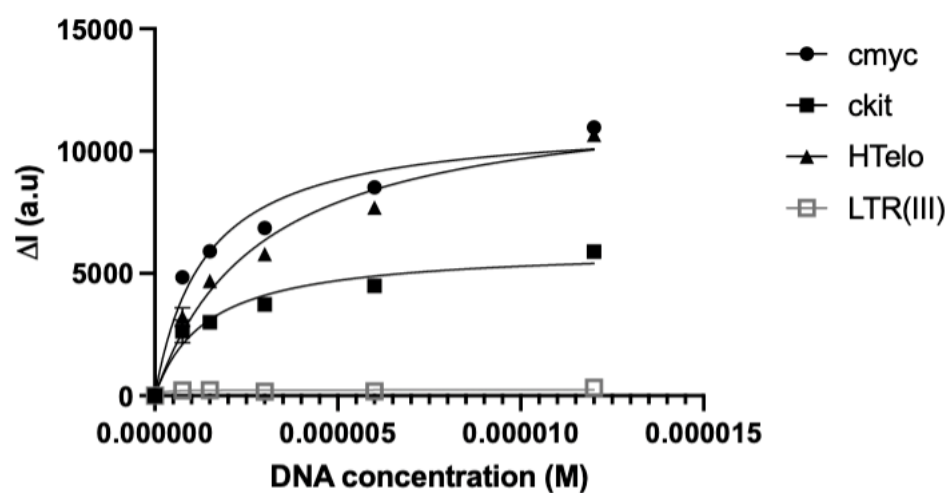


Figure A.180. Pt_PPTMA_Naph G4 DNA binding curves.

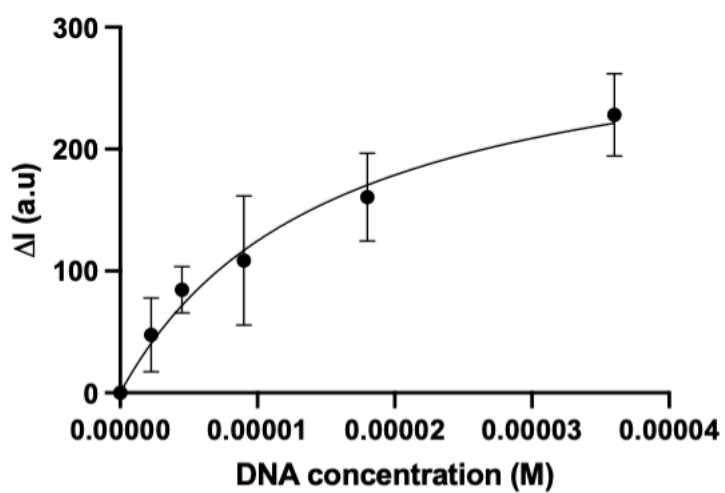


Figure A.181. Pt_PPTMA_Naph CT-DNA binding curve.

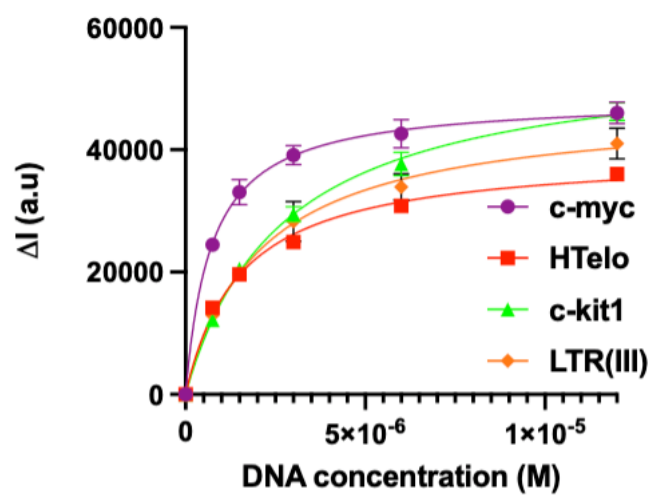


Figure A.182. Pt_PPTMA_4OMe G4 DNA binding curves.

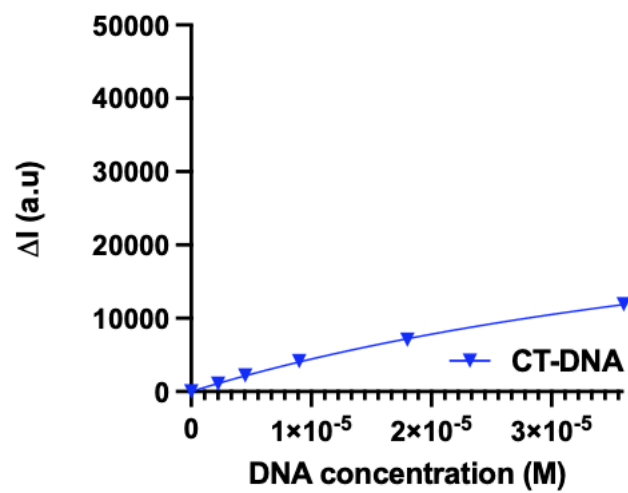


Figure A.183. Pt_PPTMA_4OMe CT-DNA binding curve.

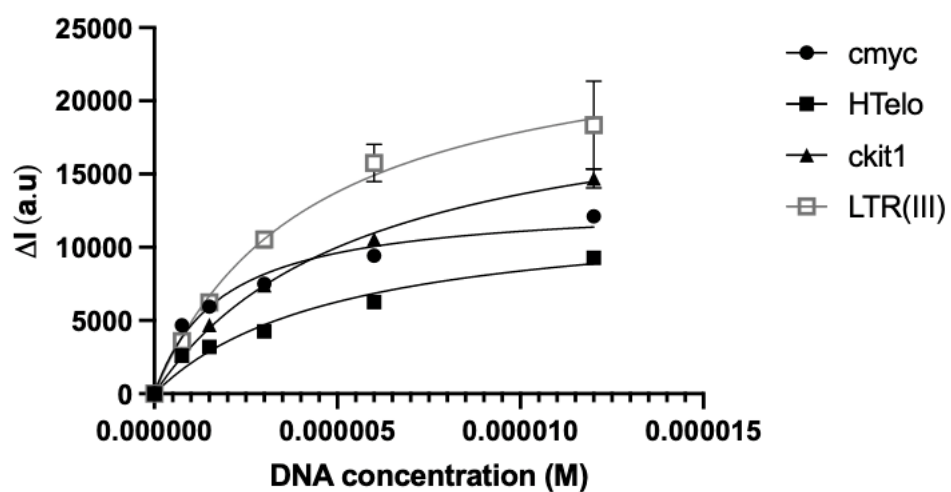


Figure A.184. Pt_PPTMA_5OMe G4 DNA binding curves.

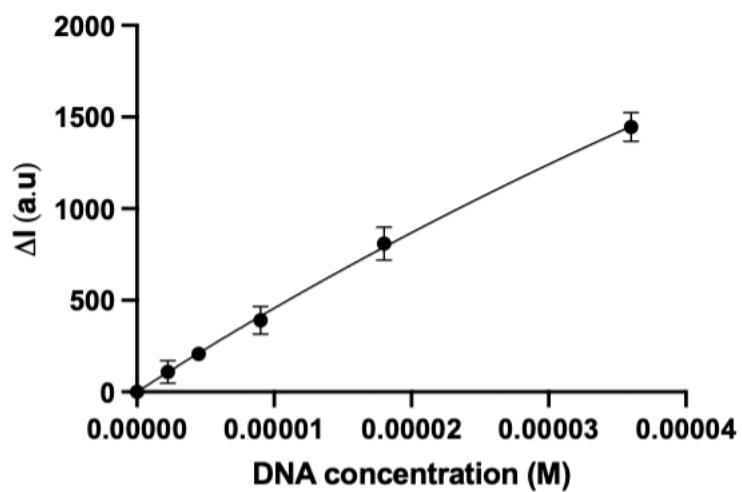


Figure A.185. Pt_PPTMA_5OMe CT-DNA binding curve.

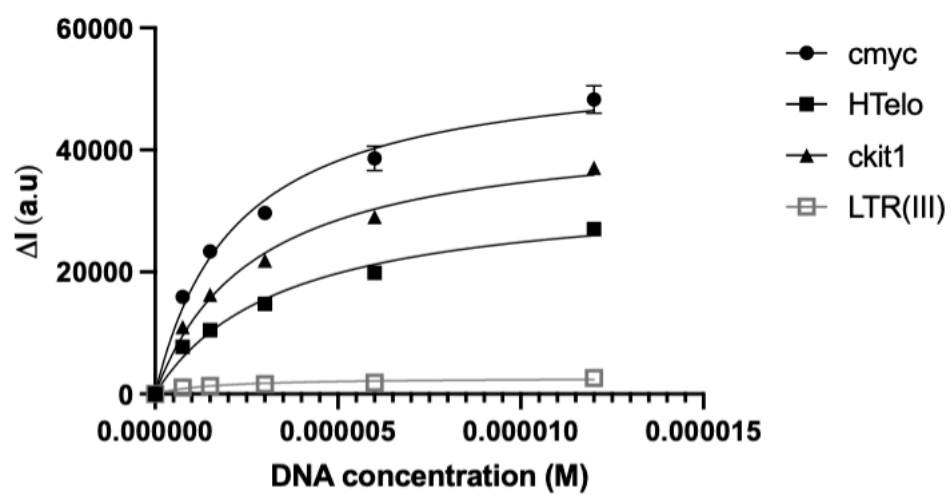


Figure A.186. Pt_PPTMA_4Br G4 DNA binding curves.

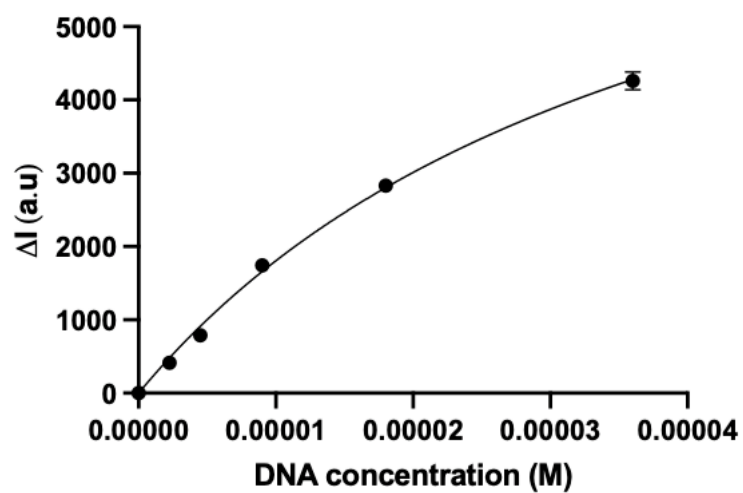


Figure A.187. Pt_PPTMA_4Br CT-DNA binding curve.

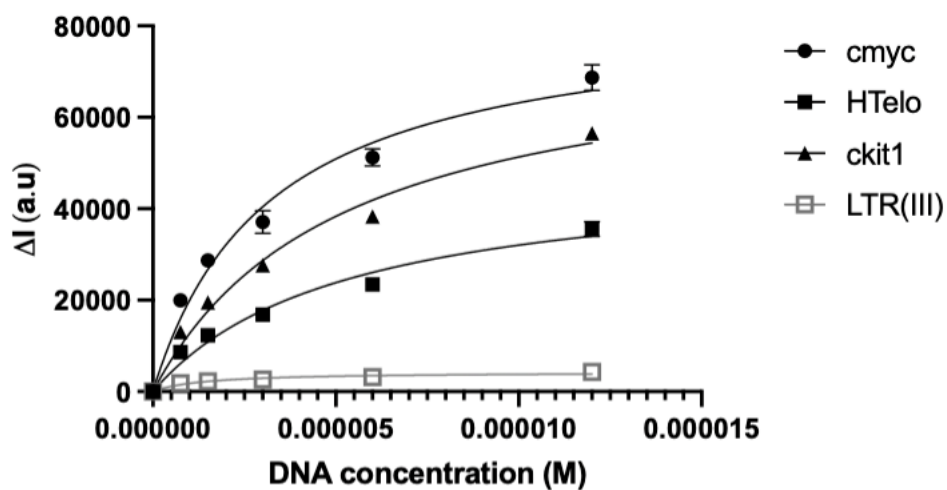


Figure A.188. Pt_PPTMA_5Br G4 DNA binding curves.

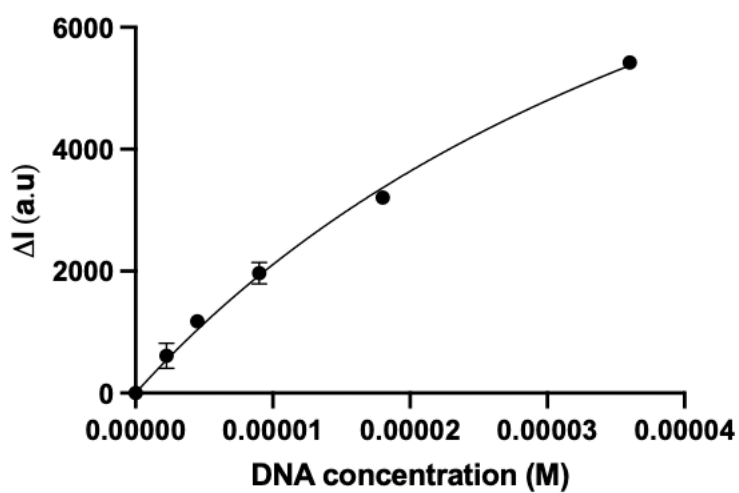


Figure A.189. Pt_PPTMA_5Br CT-DNA binding curve.

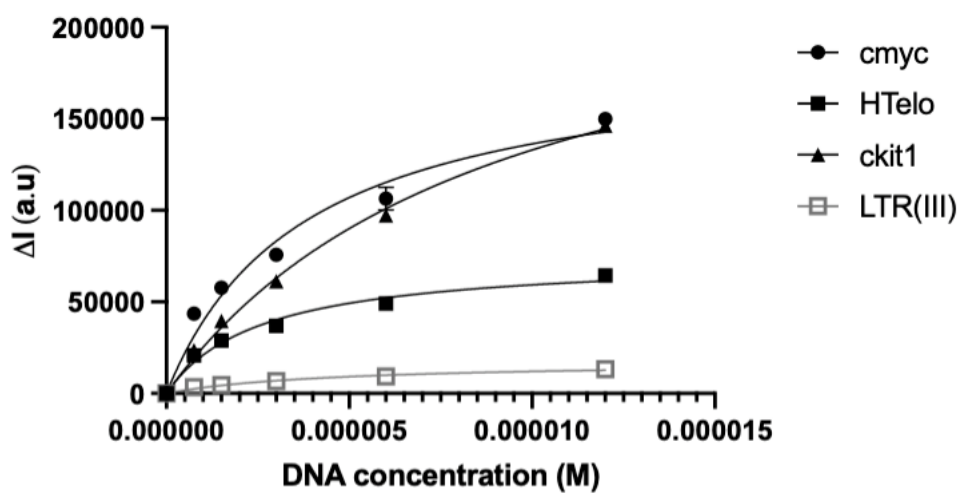


Figure A.190. Pt_PPTMA_4CF₃ G4 DNA binding curves.

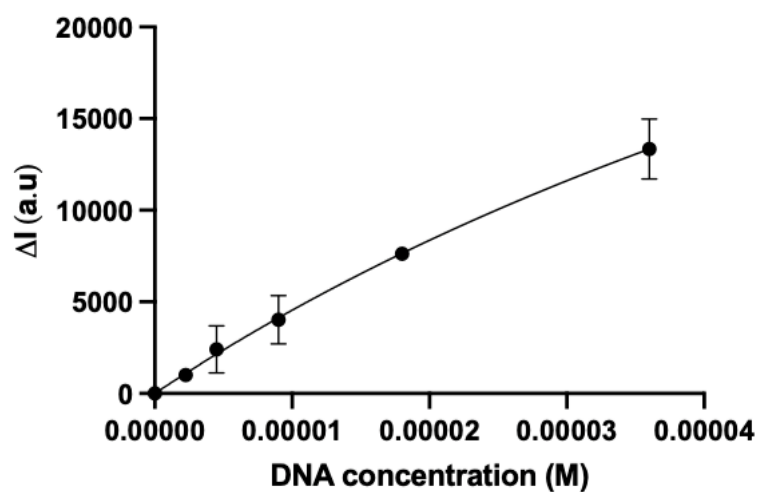


Figure A.191. Pt_PPTMA_4CF₃ CT-DNA binding curve.

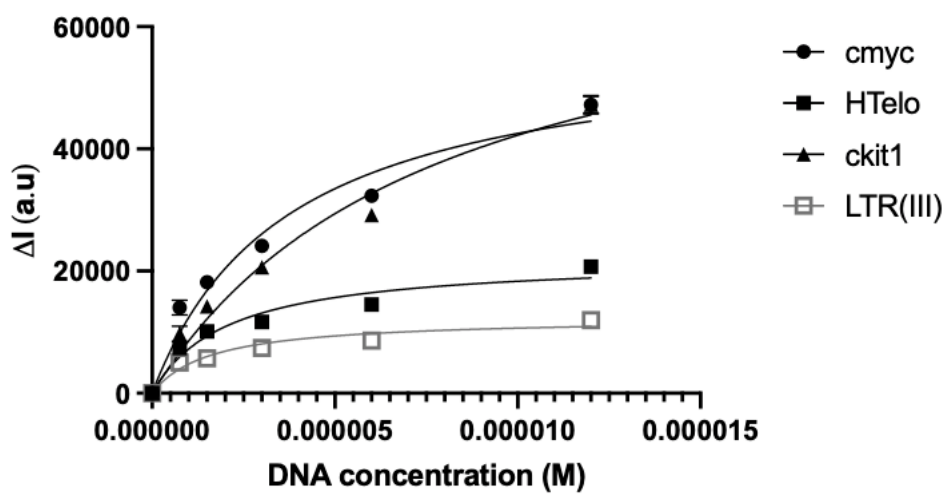


Figure A.192. Pt_PPTMA_5Cl G4 DNA binding curves.

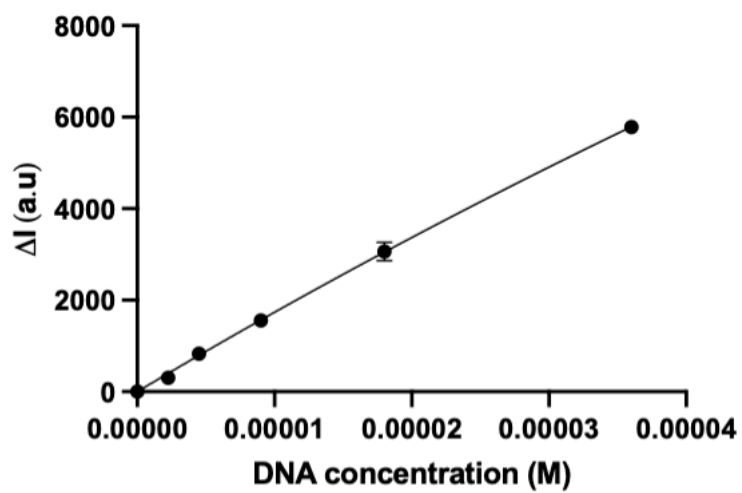


Figure A.193. Pt_PPTMA_5Cl CT-DNA binding curve.

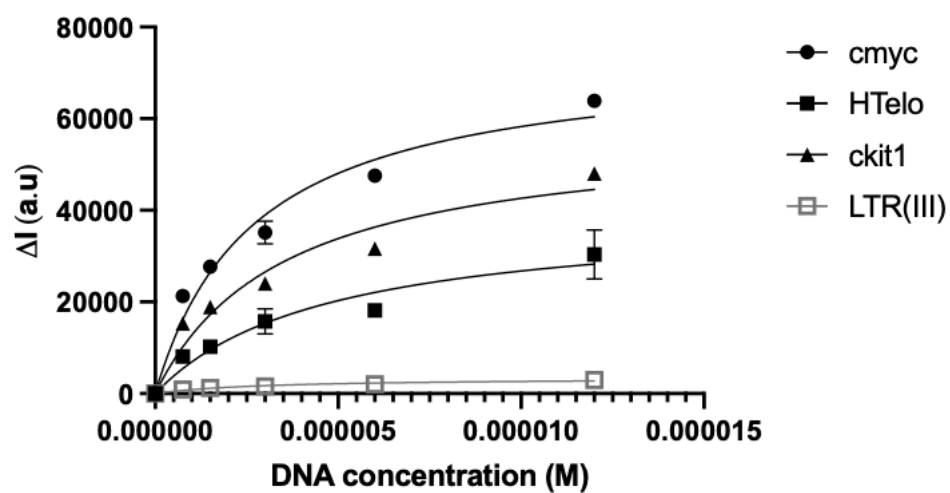


Figure A.194. Pt_PPTMA_5OCF₃ G4 DNA binding curves.

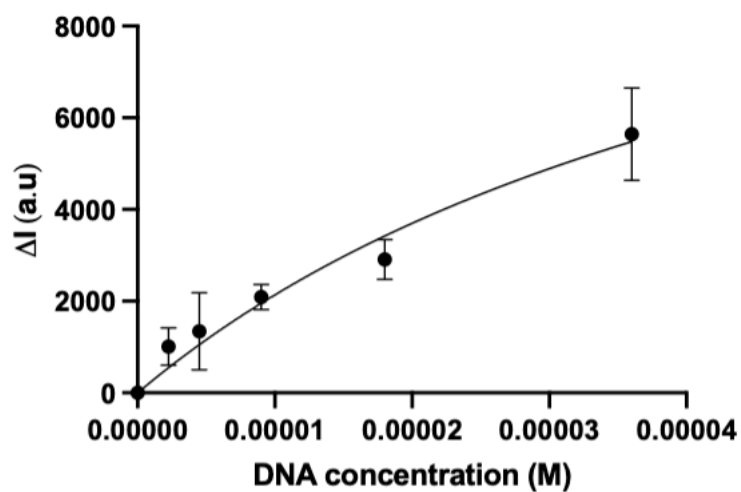


Figure A.195. Pt_PPTMA_5OCF₃ CT-DNA binding curve.

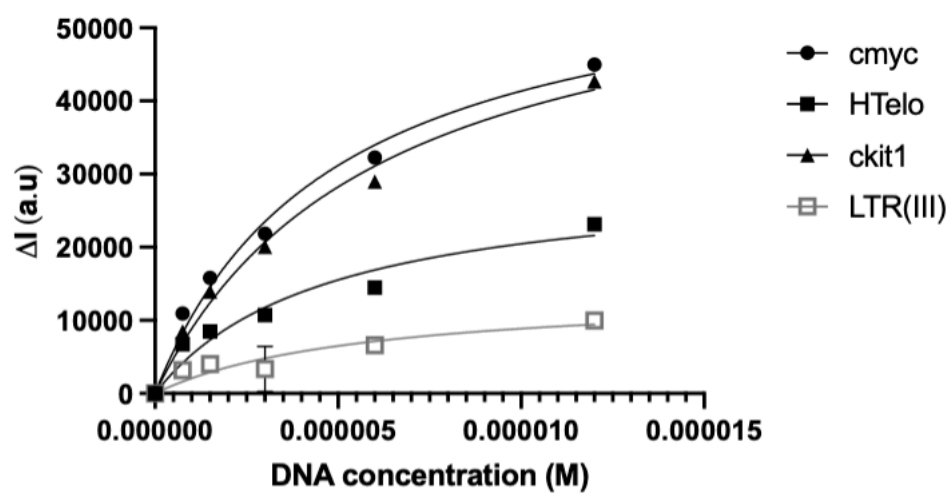


Figure A.196. Pt_PPTMA_4Cl G4 DNA binding curves.

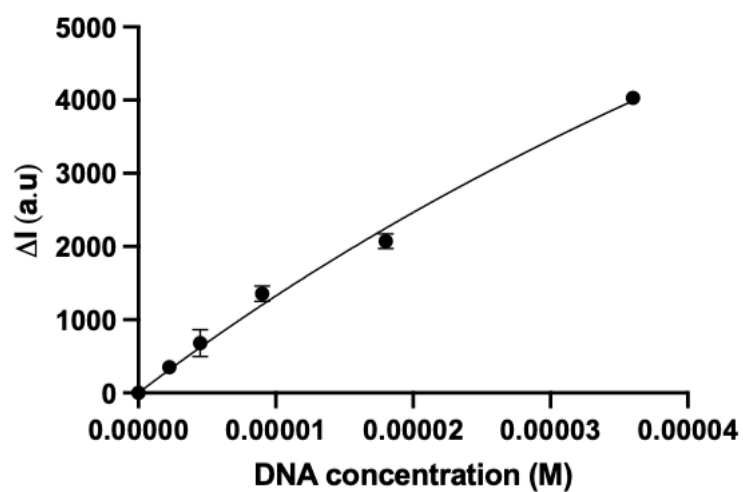


Figure A.197. Pt_PPTMA_4Cl CT-DNA binding curve.

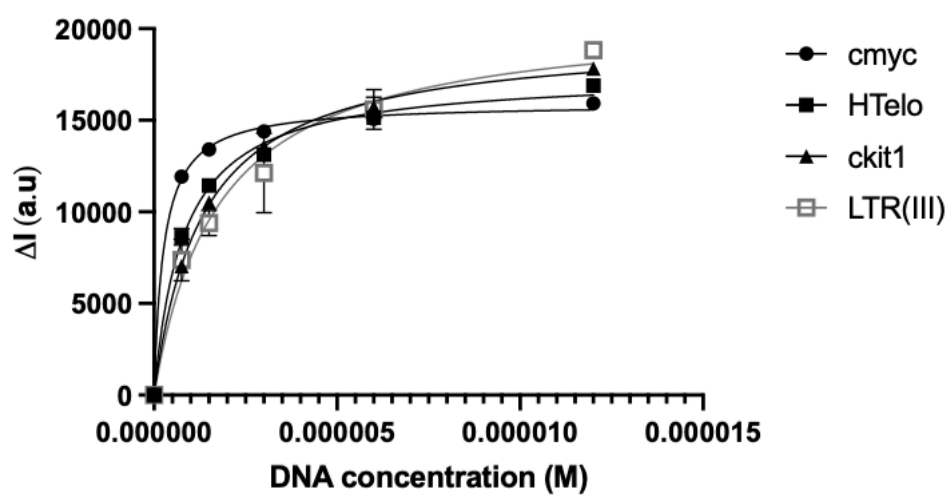


Figure A.198. Pt_PPTMA_3F G4 DNA binding curves.

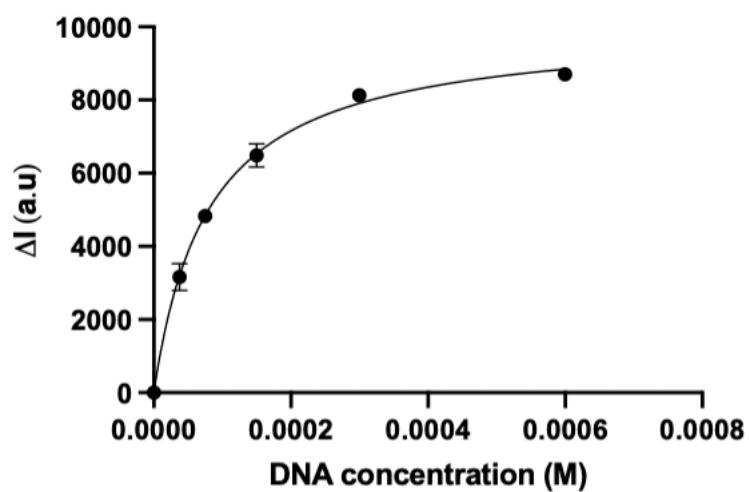


Figure A.199. Pt_PPTMA_3F CT-DNA binding curve.

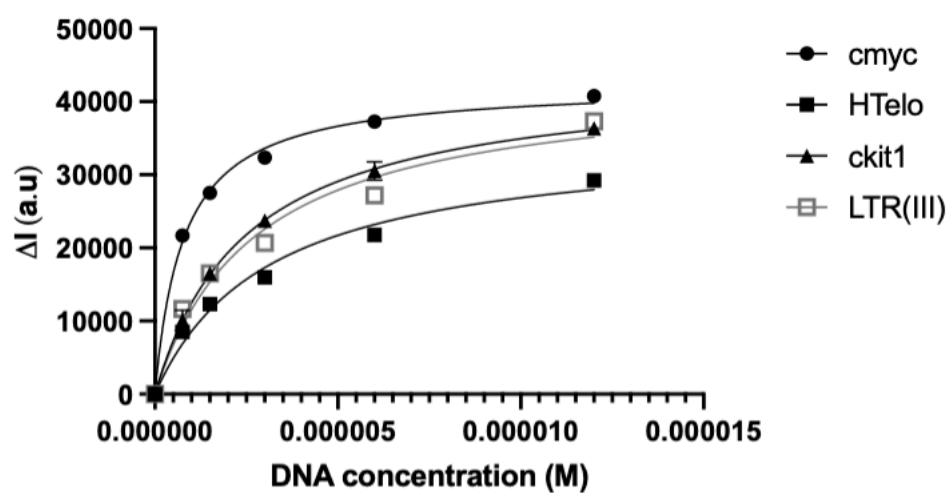


Figure A.200. Pt_PPTMA_4F G4 DNA binding curves.

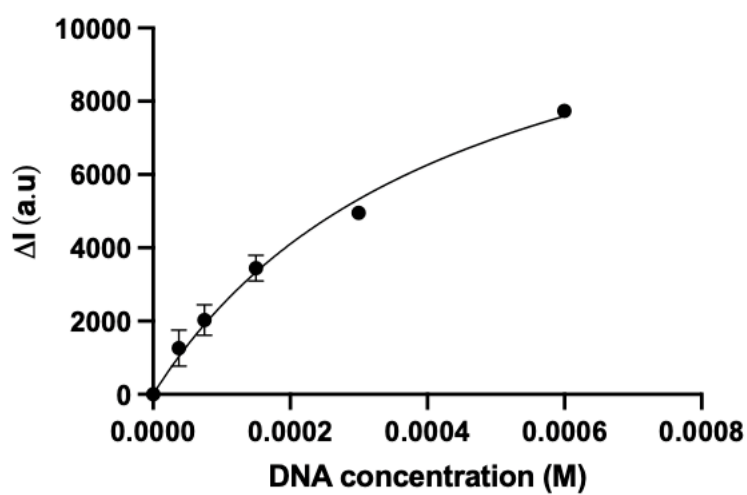


Figure A.201. Pt_PPTMA_4F CT-DNA binding curve.

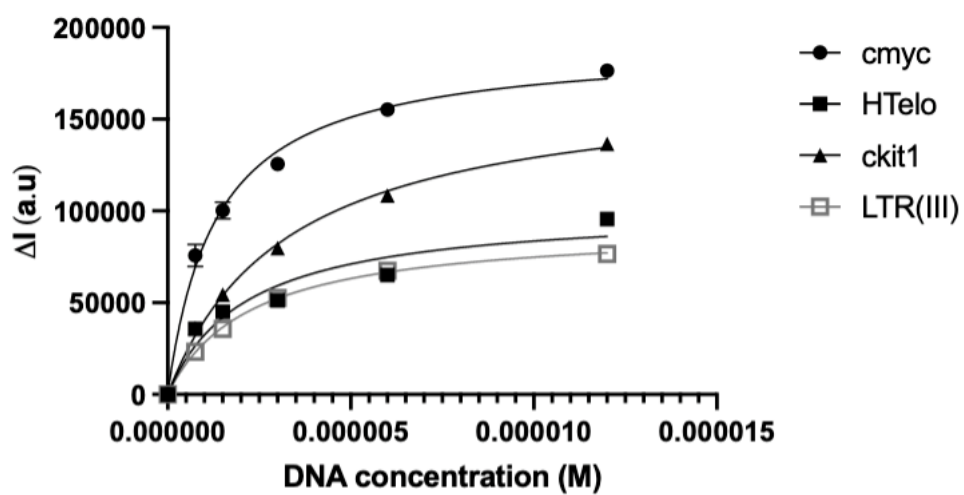


Figure A.202. Pt_PPTMA_5F G4 DNA binding curves.

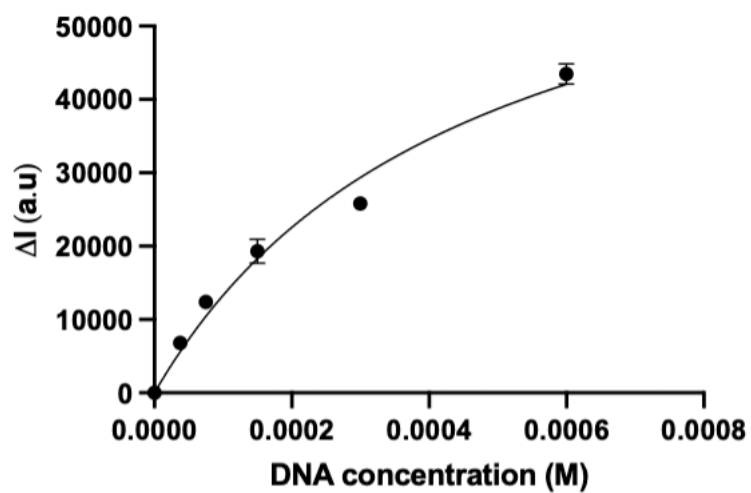


Figure A.203. Pt_PPTMA_5F CT-DNA binding curve.

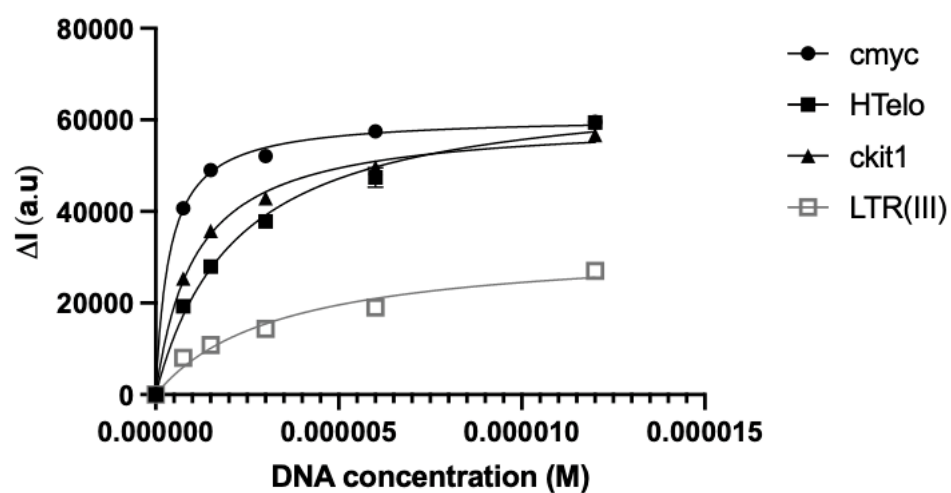


Figure A.204. Pt_PPTMA_6F G4 DNA binding curves.

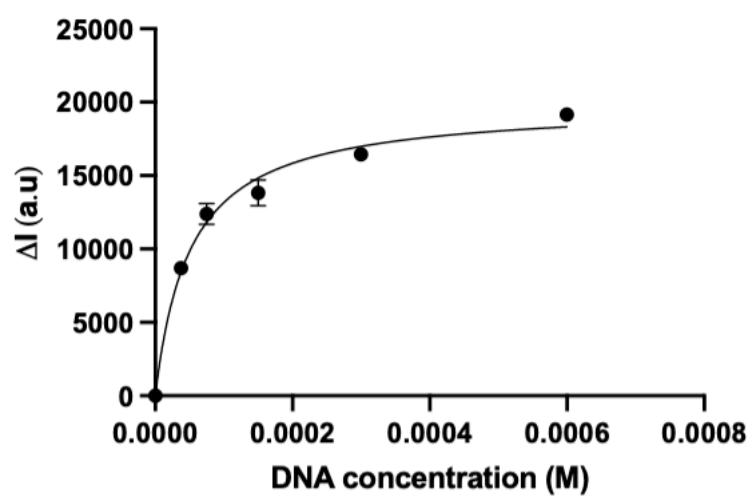


Figure A.205. Pt_PPTMA_6F CT-DNA binding curve.

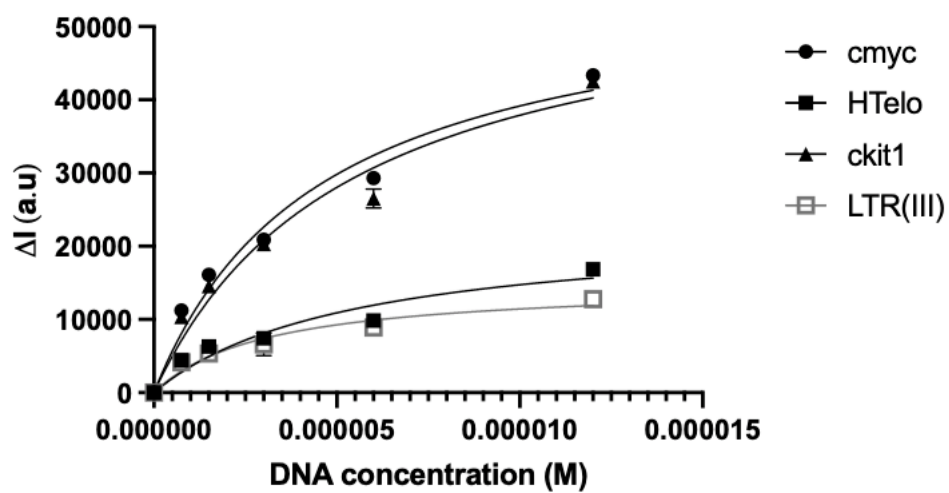


Figure A.206. Pt_PEM_Ph G4 DNA binding curves.

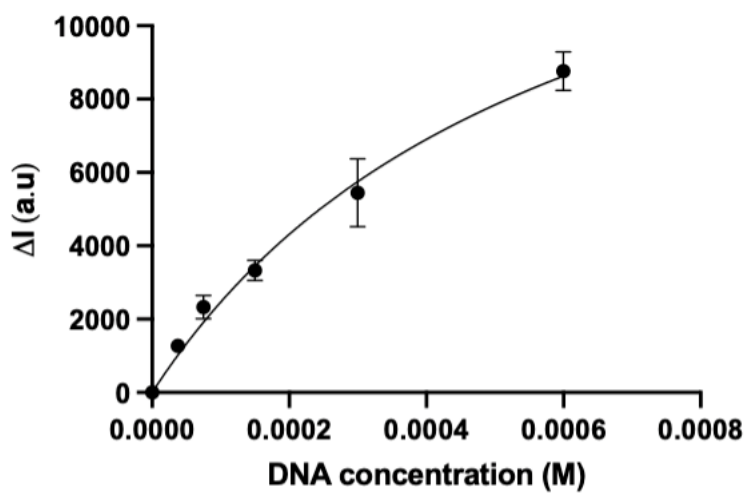


Figure A.207. Pt_PEM_Ph CT-DNA binding curve.

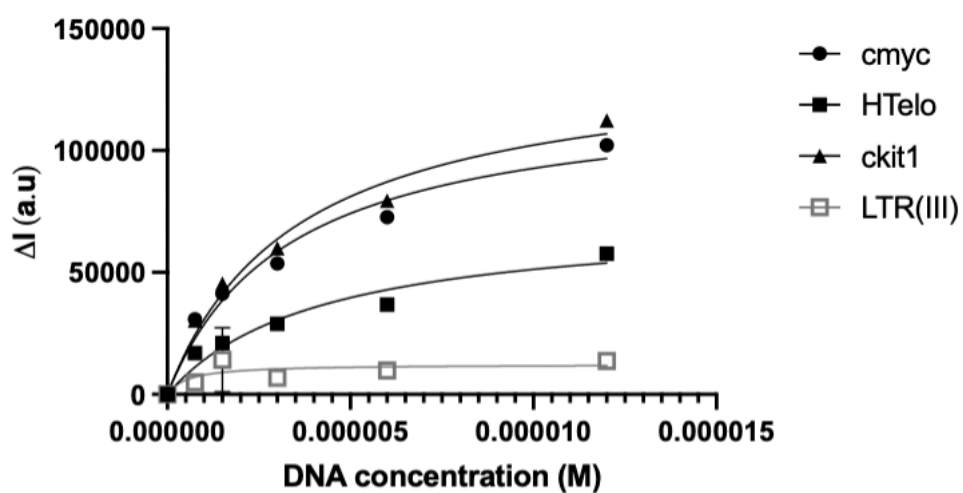


Figure A.208. Pt_PEM_4OMe G4 DNA binding curves.

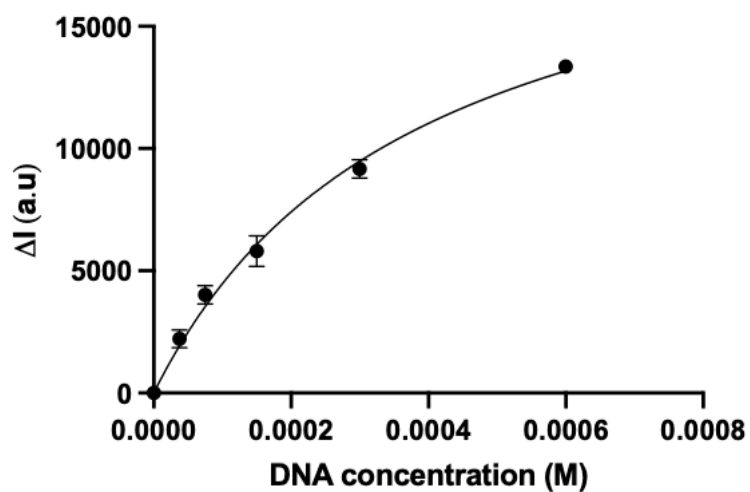


Figure A.209. Pt_PEM_4OMe CT-DNA binding curve.

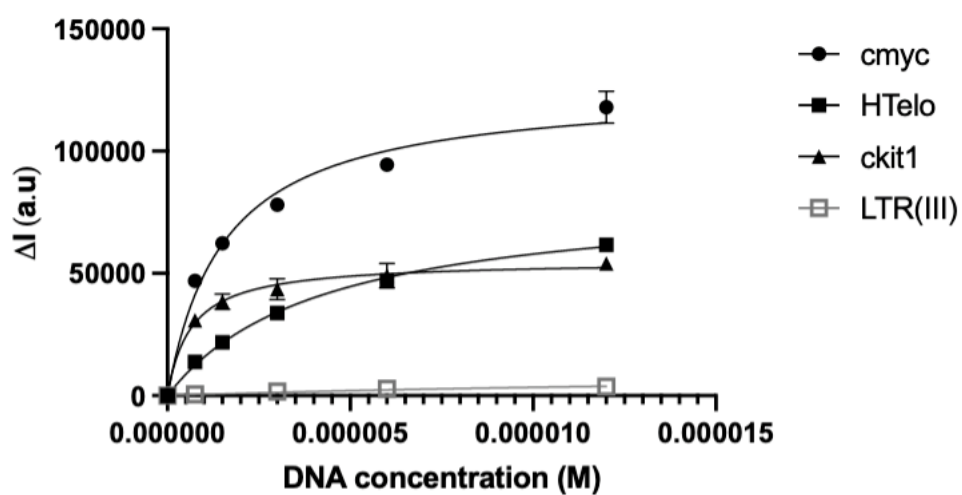


Figure A.210. Pt_PEM_Naph G4 DNA binding curves.

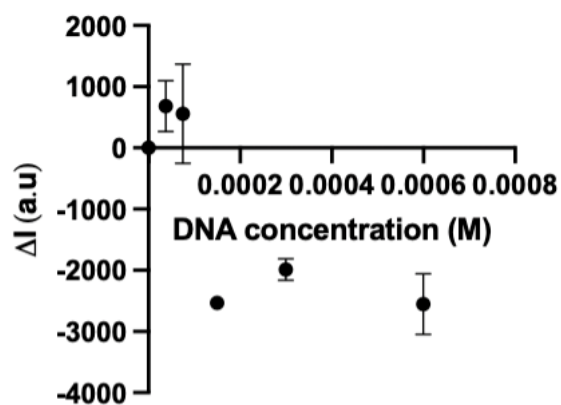


Figure A.211. Pt_PEM_Naph CT-DNA binding curve.

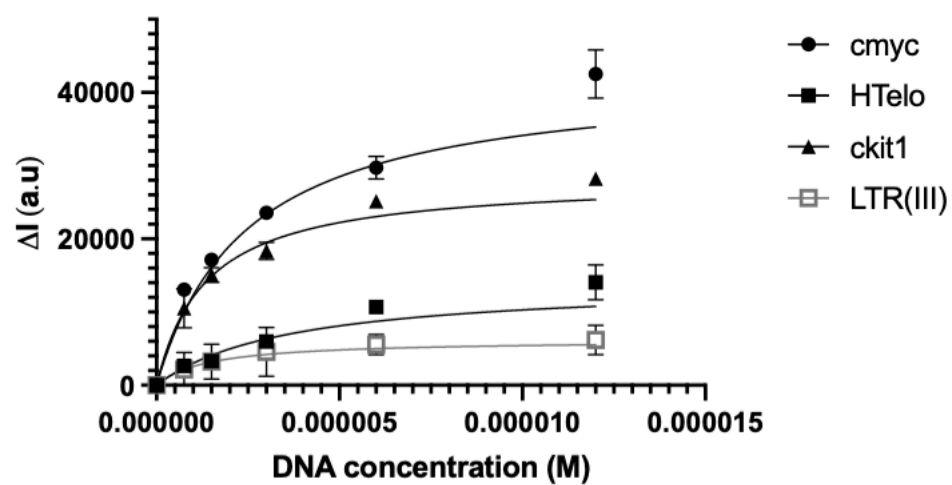


Figure A.212. Pt_PEM_5OCF₃ G4 DNA binding curves.

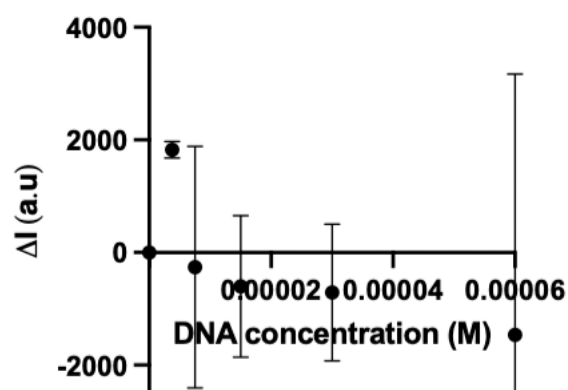


Figure A.213. Pt_PEM_5OCF₃ CT-DNA binding curve.

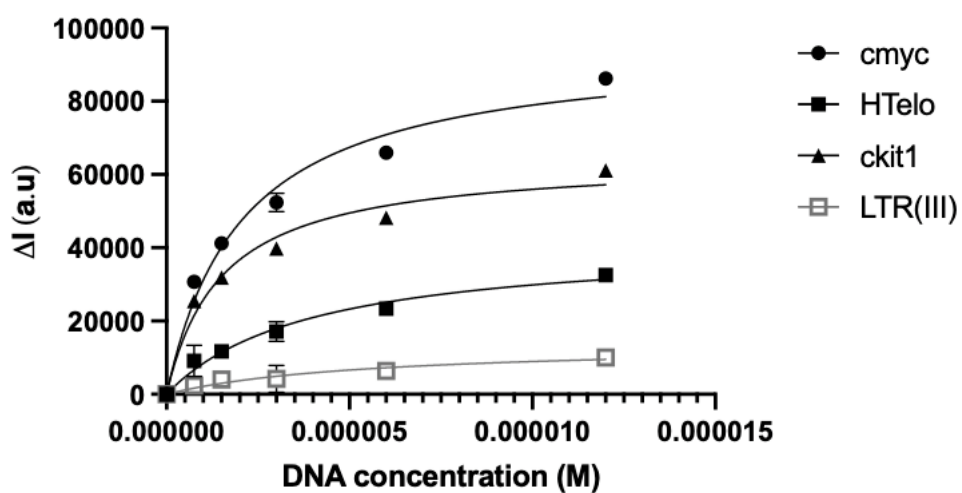


Figure A.214. Pt_PEM_5Br G4 DNA binding curves.

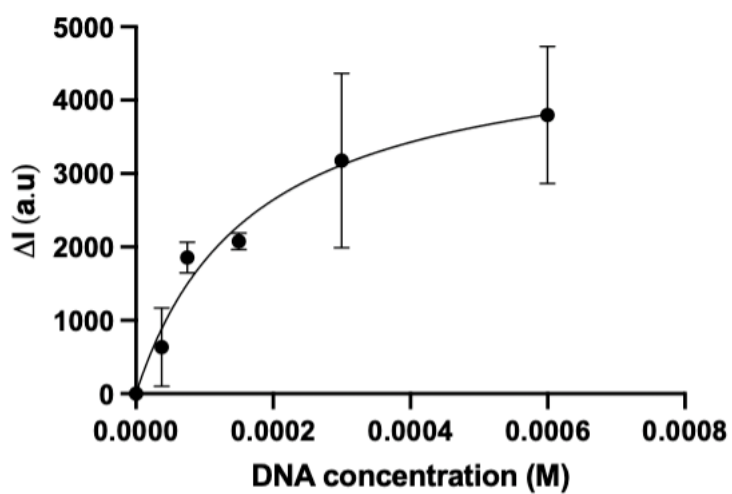


Figure A.215. Pt_PEM_5Br CT-DNA binding curve.

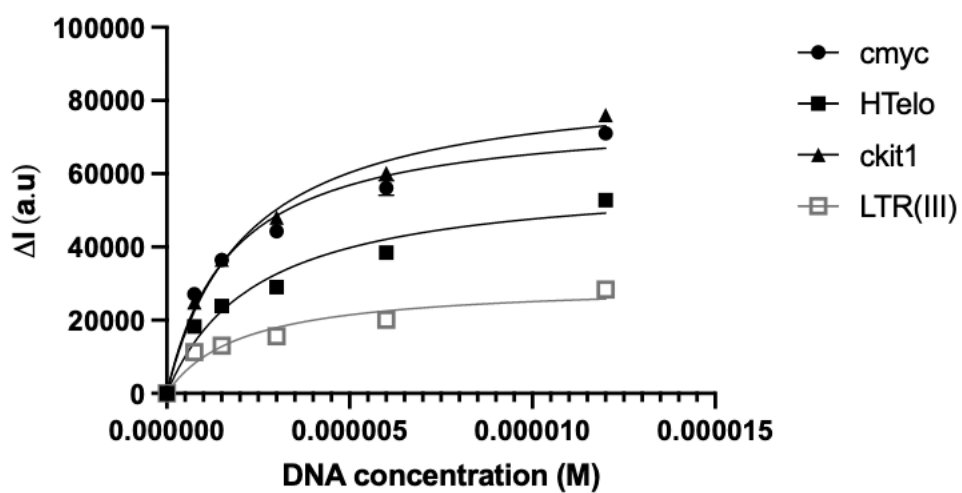


Figure A.216. Pt_AEP_4OMe G4 DNA binding curves.

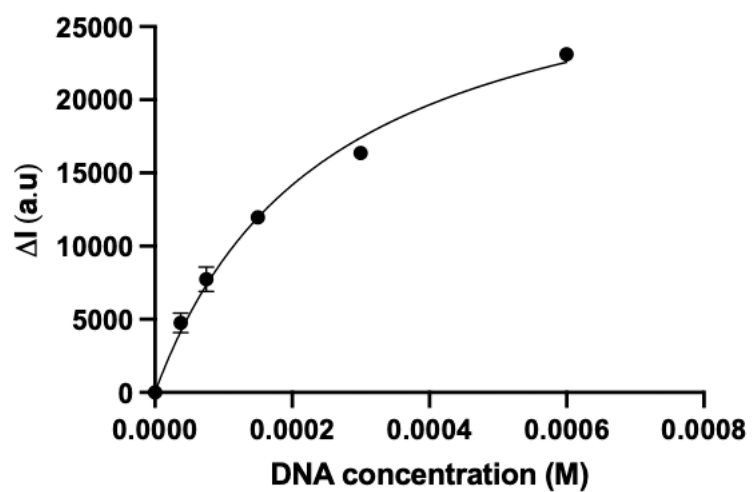


Figure A.217. Pt_AEP_4OMe CT-DNA binding curve.

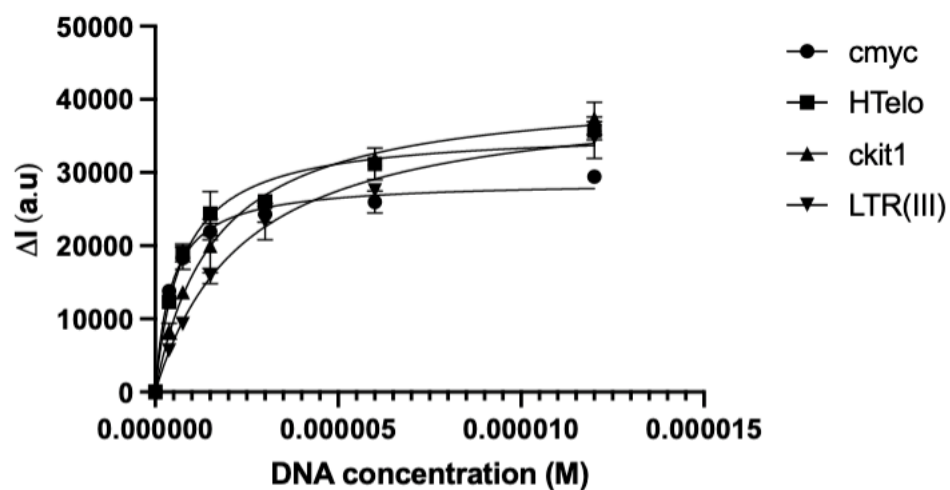


Figure A.218. Pt_PPTMA_4NEt₂ G4 DNA binding curves.

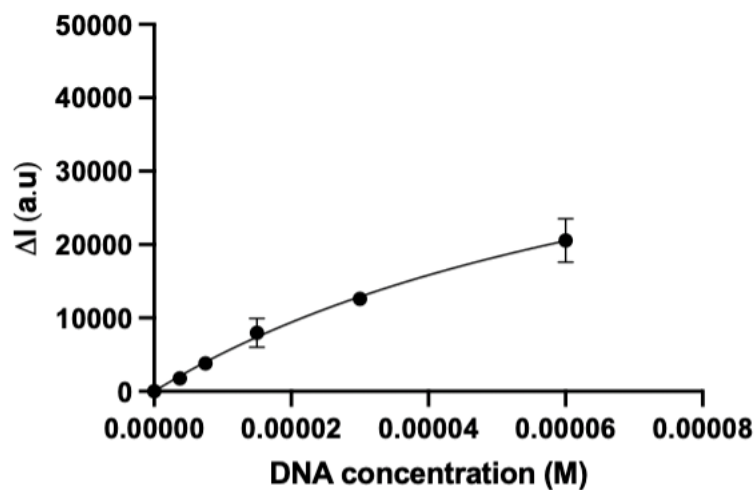


Figure A.219. Pt_PPTMA_4NEt₂ CT-DNA binding curve.

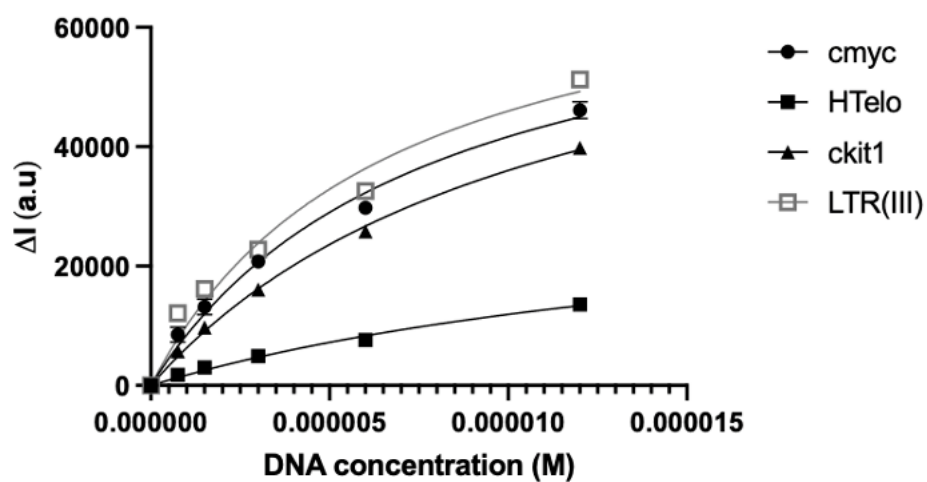


Figure A.220. Pt_PhBr_Ph G4 DNA binding curves.

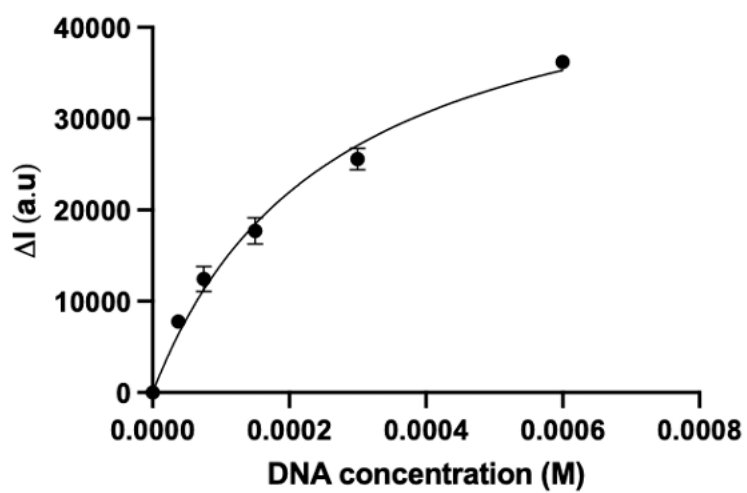


Figure A.221. Pt_PhBr_Ph CT-DNA binding curves.

	c-myc		HTelo		c-kit1		LTR(III)		CT-DNA	
Complex	Ka (M ⁻¹)	% error	Ka (M ⁻¹)	% error	Ka (M ⁻¹)	% error	Ka (M ⁻¹)	% error	Ka (M ⁻¹)	% error
Pt_PPTMA_3F	3.78E+06	6.36	1.22E+06	7.74	7.53E+05	4.31	5.67E+05	17.48	1.25E+04	6.14
Pt_PPTMA_6F	2.66E+06	7.01	4.53E+05	9.12	9.36E+05	7.17	2.98E+05	19.03	1.96E+04	11.78
Pt_PPTMA_4NEt ₂	2.28E+06	10.73	1.36E+06	13.48	6.17E+05	9.71	4.11E+05	11.04	1.16E+04	33.60
Pt_PEM_4OMe	3.01E+05	17.89	2.41E+05	24.62	2.84E+05	15.48	1.81E+06	140.00	2.61E+03	12.30
Pt_PEM_Naph	6.45E+05	13.51	2.36E+05	6.08	1.55E+06	15.39	No binding	N/A	No binding	N/A
Pt_PPTMA_4OMe	1.39E+06	8.33	6.53E+05	7.96	3.71E+05	6.40	4.84E+05	11.14	1.52E+04	6.61
Pt_PPTMA_Ph	1.38E+06	6.78	4.33E+05	13.74	4.03E+05	9.52	5.01E+05	8.84	3.73E+04	9.57
Pt_PPTMA_4F	1.28E+06	6.67	3.22E+05	13.81	3.98E+05	5.55	3.65E+05	17.11	2.25E+03	17.01
Pt_PPTMA_5F	7.55E+05	8.56	4.68E+05	26.92	3.06E+05	5.71	4.43E+05	7.22	2.18E+03	20.92
Pt_PPTMA_Naph	7.35E+05	19.71	7.04E+05	21.99	3.46E+05	17.58	No binding	N/A	6.54E+04	39.77
Pt_PPTMA_5Br	3.14E+05	15.00	1.90E+05	19.28	1.95E+05	15.13	7.16E+05	24.35	1.88E+04	14.85
Pt_PEM_5Br	4.85E+05	15.10	2.53E+05	22.10	6.65E+05	15.48	1.84E+05	53.58	5.90E+03	42.84
Pt_PPTMA_4Br	4.57E+05	11.71	2.89E+05	13.56	3.67E+05	9.75	6.33E+05	24.15	2.49E+04	10.18
Pt_PPTMA_5Cl	2.70E+05	20.53	4.88E+05	22.75	1.29E+05	20.80	6.29E+05	24.43	3.14E+03	35.33
Pt_AEP_4OMe	5.73E+05	14.68	4.06E+05	19.87	4.65E+05	9.47	4.89E+05	26.07	3.96E+03	10.43
Pt_PEM_5OCF ₃	3.23E+05	19.72	1.23E+05	24.37	5.67E+05	13.17	1.00E+00	72.96	No binding	N/A
Pt_PPTMA_5OMe	5.49E+05	15.61	1.97E+05	22.22	1.76E+05	7.77	2.37E+05	18.86	5.45E+03	51.02
Pt_PPTMA_4CF ₃	2.65E+05	18.33	3.90E+05	14.29	1.10E+05	9.29	2.21E+05	17.75	4.24E+03	154.05
Pt_PPTMA_5OCF ₃	3.70E+05	17.43	2.14E+05	33.67	2.85E+05	24.93	3.35E+05	59.16	4.18E+04	52.03
Pt_PEM_Ph	2.19E+05	18.86	1.84E+05	32.15	1.85E+05	22.66	3.27E+05	25.11	1.67E+03	24.08
Pt_PPTMA_4Cl	2.07E+05	13.38	2.14E+05	27.58	1.64E+05	12.59	1.74E+05	53.30	8.11E+03	34.79
Pt_PhBr_Ph	1.29E+05	14.8228703	5.36E+04	20.4844051	9.03E+04	8.94966353	1.51E+05	21.2149307	4.05E+02	64.3634807

Table A.7. Table of all binding affinities (Ka) and percentage error (% error) for complexes against 5 DNA topologies.

A.2.4 DNA damage assays

A.2.4.1 pUC19 gels

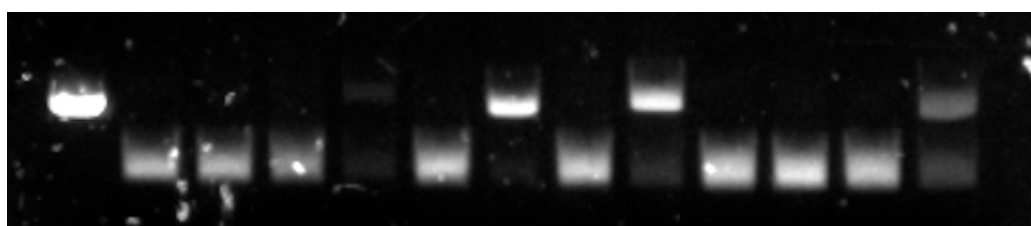


Figure A.222. Image of the gel electrophoresis plasmid damage assay repeat 2. The contents of each lane are: 1) Eco31I. 2) pUC19 dark. 3) pUC19 irradiated. 4) Pt_PPTMA_4OMe dark. 5) Pt_PPTMA_4OMe irradiated. 6) Pt_PEM_4OMe dark. 7) Pt_PEM_4OMe irradiated. 8) Pt_AEP_4OMe dark. 9) Pt_AEP_4OMe irradiated. 10) Pt_PEM_5Br dark. 11) Pt_PEM_5Br irradiated. 12) Ru(Bpy)₃ dark. 13) Ru(Bpy)₃ irradiated.

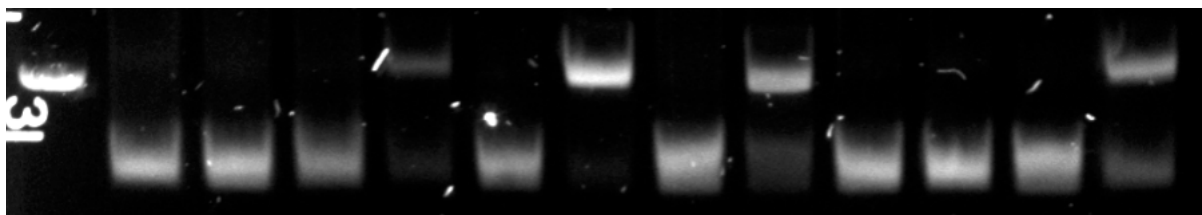


Figure A.223. Image of the gel electrophoresis plasmid damage assay repeat 3. The contents of each lane are: 1) Eco31I. 2) pUC19 dark. 3) pUC19 irradiated. 4) Pt_PPTMA_4OMe dark. 5) Pt_PPTMA_4OMe irradiated. 6) Pt_PEM_4OMe dark. 7) Pt_PEM_4OMe irradiated. 8) Pt_AEP_4OMe dark. 9) Pt_AEP_4OMe irradiated. 10) Pt_PEM_5Br dark. 11) Pt_PEM_5Br irradiated. 12) Ru(Bpy)3 dark. 13) Ru(Bpy)3 irradiated.

A.2.4.2 Quench gels

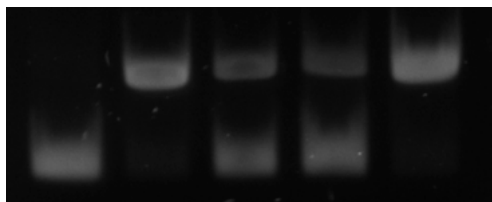


Figure A.224. Image of the gel electrophoresis plasmid damage assay repeat 1. The contents of each lane were irradiated and contain: 1) pUC19. 2) Pt_PEM_4OMe. 3) Pt_PEM_4OMe + NaN₃. 4) Pt_PEM_4OMe + KI. 5) Pt_PEM_4OMe + DMSO.

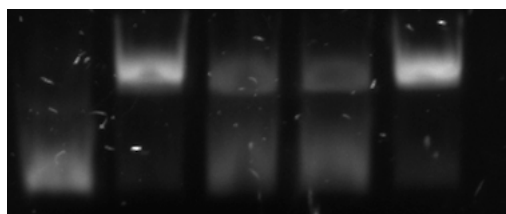


Figure A.225. Image of the gel electrophoresis plasmid damage assay repeat 2. The contents of each lane were irradiated and contain: 1) pUC19. 2) Pt_PEM_4OMe. 3) Pt_PEM_4OMe + NaN₃. 4) Pt_PEM_4OMe + KI. 5) Pt_PEM_4OMe + DMSO.

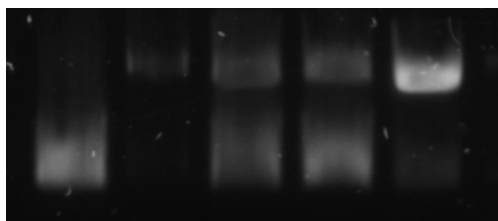


Figure A.226. Image of the gel electrophoresis plasmid damage assay repeat 3. The contents of each lane were irradiated and contain: 1) pUC19. 2) Pt_PEM_4OMe. 3) Pt_PEM_4OMe + NaN₃. 4) Pt_PEM_4OMe + KI. 5) Pt_PEM_4OMe + DMSO.

A.2.4.3 G4 damage circular dichroism

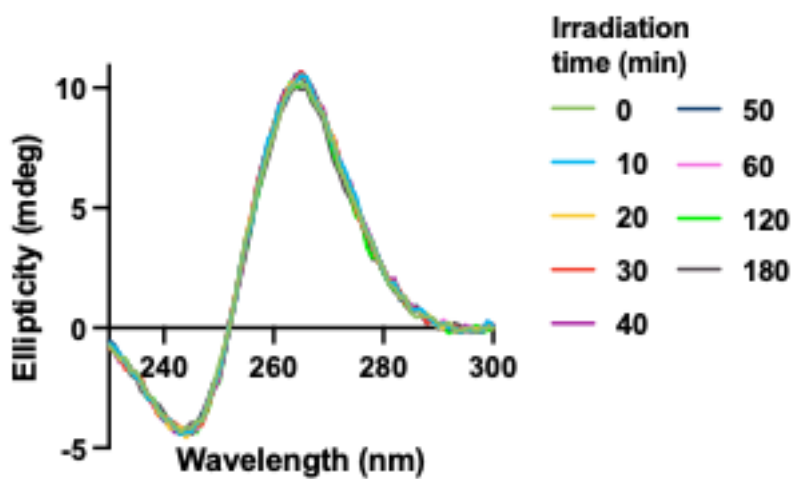


Figure A.227. Full ellipticity plots of c-myc G4 DNA with Pt_PPTMA_4OMe with increasing time without irradiation.

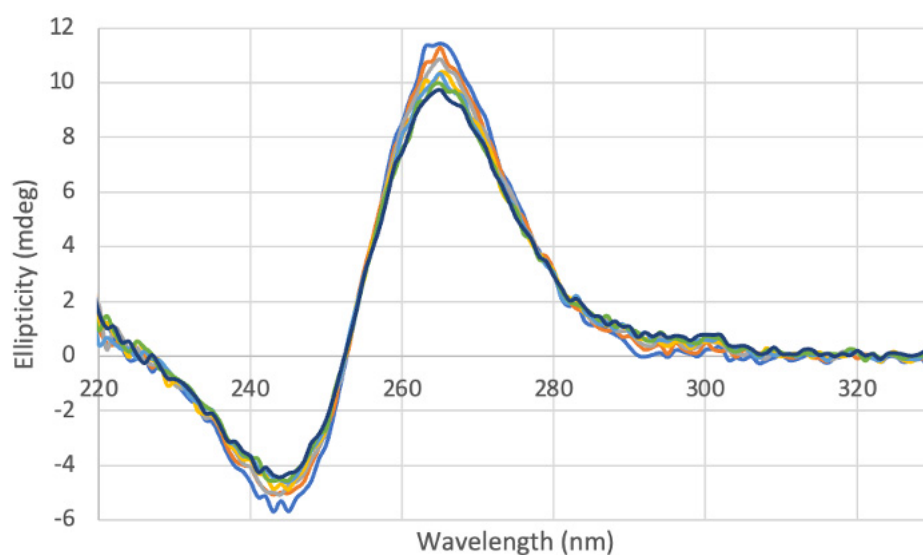


Figure A.228. Full ellipticity plots of c-myc G4 DNA with Pt_PEM_4OMe with increasing intervals of irradiation.

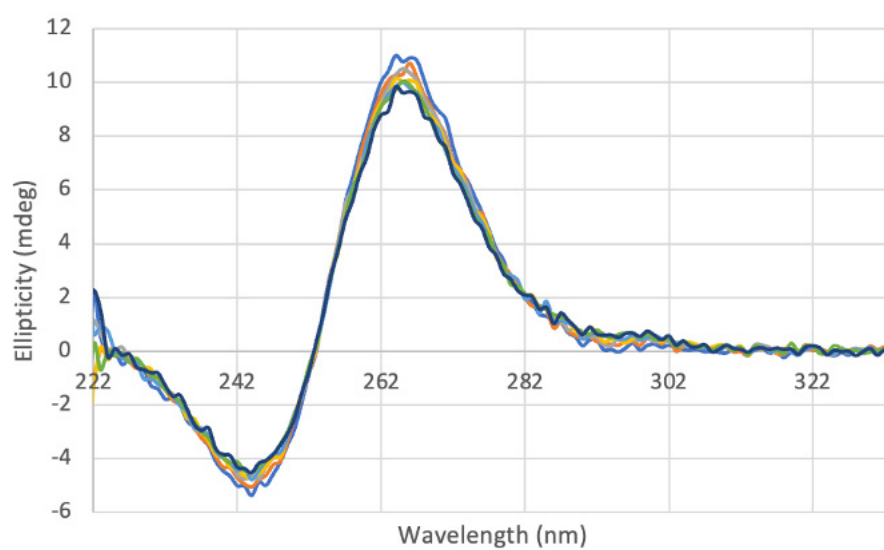


Figure A.229. Full ellipticity plots of c-myc G4 DNA with Pt_AEP_4OMe with increasing intervals of irradiation.

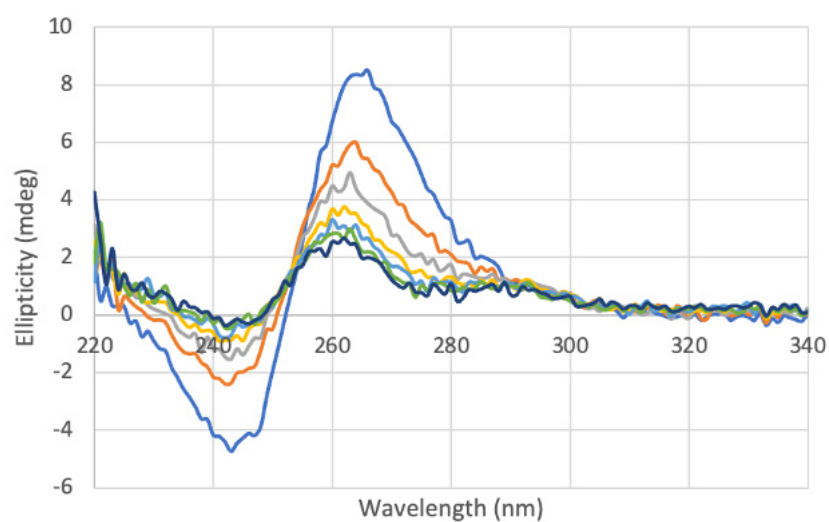


Figure A.230. Full ellipticity plots of c-myc G4 DNA with TMPyP4 with increasing intervals of irradiation.

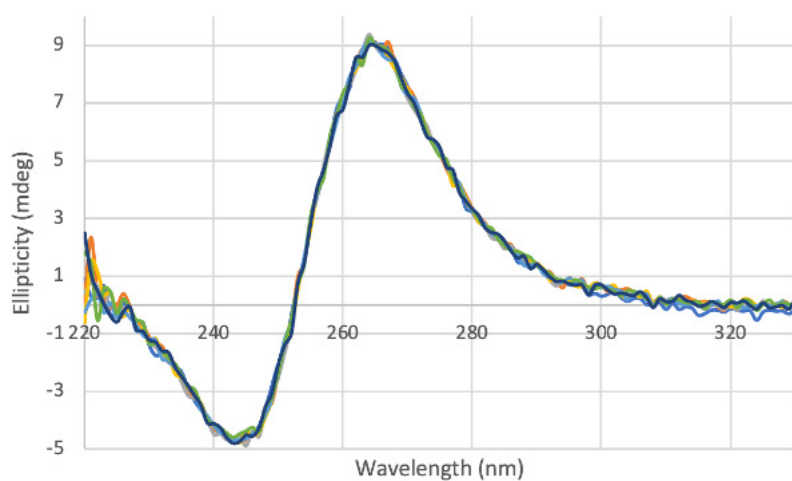


Figure A.231. Full ellipticity plots of c-myc G4 DNA with positive control TMPyP4 with increasing time without irradiation.

A.2.5 Phototoxicity IC₅₀ values

Complex	Mean IC ₅₀ (μM)	Standard Error Mean (μM)
Pt_PPTMA_Naph	0.123	0.00759
Pt_PPTMA_4CF ₃	7.77	1.29
Pt_PPTMA_5OCF ₃	1.03	0.126
Pt_PEM_Ph	0.641	0.0168
Pt_PEM_4OMe	0.294	0.0137
Pt_PEM_Naph	0.21	0.00856
Pt_PEM_5OCF ₃	1.71	0.154
Pt_PEM_5Br	4.51	0.69
Pt_AEP_4OMe	2.57	0.231
Pt_PPTMA_4NEt ₂	2.01	0.0965
Pt_PPTMA_4Br	3.18	0.284
Pt_PPTMA_5Br	1.59	0.16
Pt_PPTMA_4OMe	24.3	8.42
Pt_PPTMA_4Cl	4.22	0.386
Pt_PPTMA_5Cl	2.96	0.316
Pt_PhBr_Ph	8.80	1.70

Table A.8. Platinum(II) salphen complexes mean phototoxicity IC₅₀ values and their standard error against HeLa cells.

A.2.6 High throughput imaging screen

All 96 well plates have the same format. Each row contains 4 compounds at 3 concentrations lowest to highest from left to right. The layout of the plates is detailed in table A.9 below. Compounds with an (X) were impure.

	1 + 2 + 3	4 + 5 + 6	7 + 8 + 9	10 + 11 + 12
A	Pt_PPTMA_Ph	Pt_PPTMA_5Cl	Pt_PEM_Ph	Pt_AEM_4OMe (X)
B	Pt_PPTMA_Naph	Pt_PPTMA_3F	Pt_PEP_Ph (X)	Pt_AEP_4OMe
C	Pt_PPTMA_4Br	Pt_PPTMA_4F	Pt_PEM_4OMe	Pt_PEM_5Br
D	Pt_PPTMA_5Br	Pt_PPTMA_5F	Pt_PEP_4OMe (X)	Pt_PEM_Naph
E	Pt_PPTMA_4OMe	Pt_PPTMA_6F	Pt_PEM_5OCF ₃	Pt_PEP_Naph (X)
F	Pt_PPTMA_5OMe	Pt_PPTMA_4CF ₃	Pt_AEM_5Br (X)	Pt_PhBr_Ph
G	Pt_PPTMA_4Cl	Pt_PPTMA_5OCF ₃	Pt_AEP_5Br (X)	Pt_PPTMA_4NEt ₂
H	Cells only	Cells only	Cells only	Cells only

Table A.9. Table showing the order of complexes within the 96 well plate fluorescent microscopy images. The top row indicates which columns within the 96 well plate contain the experimental concentration range for each compound

A.2.6.1 Concentrations 2.5 μ M, 5 μ M and 10 μ M

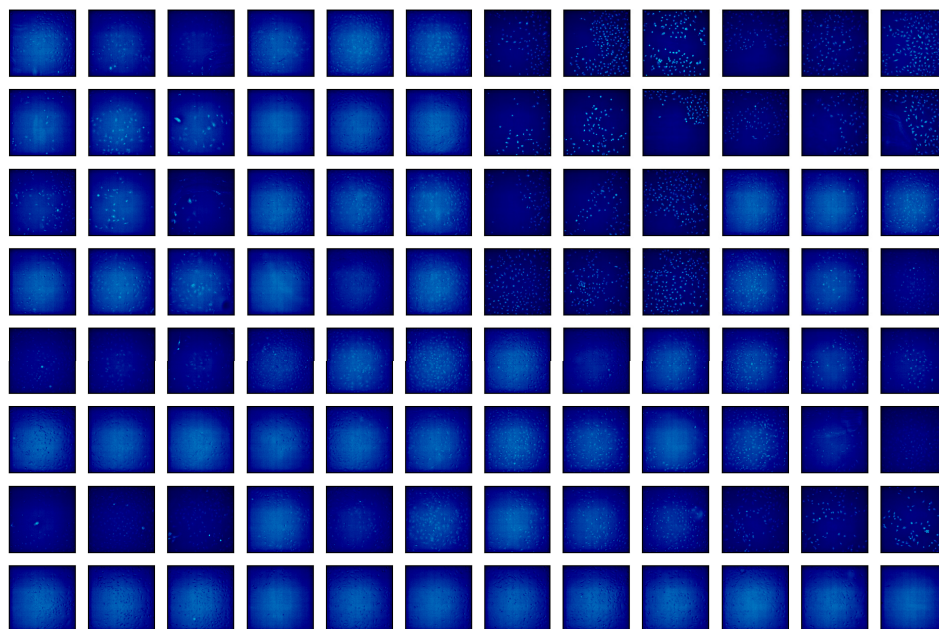


Figure A.232. 96 well plate of complexes at 2.5, 5 and 10 μ M after 3 hours of incubation.

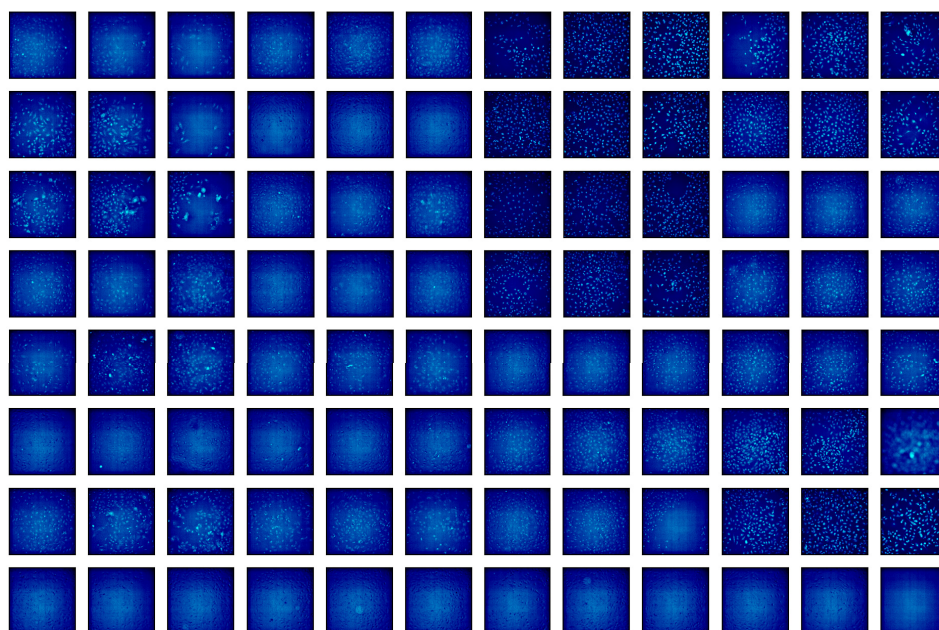


Figure A.233. 96 well plate of complexes at 2.5, 5 and 10 μ M after 6 hours of incubation.

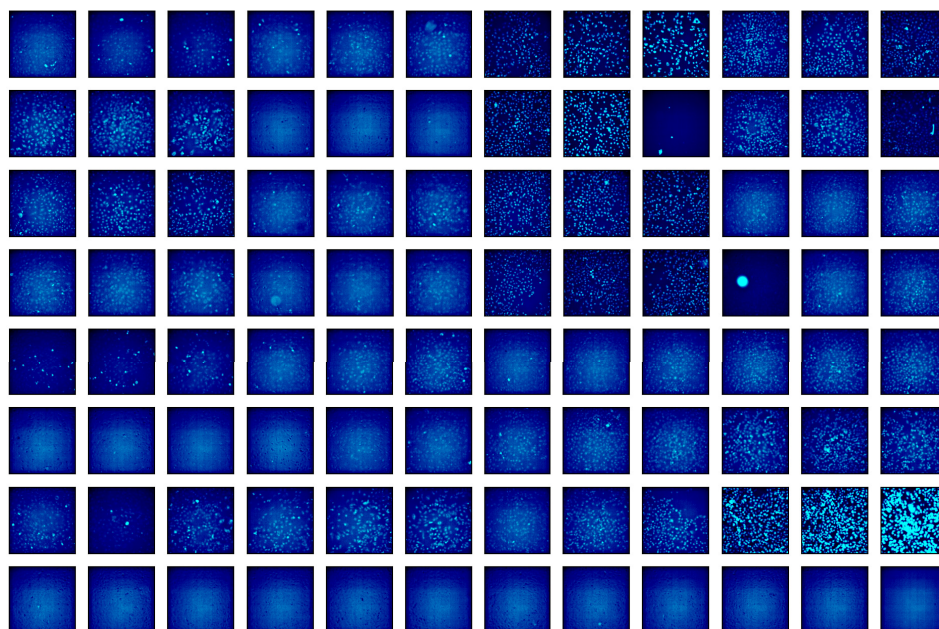


Figure A.234. 96 well plate of complexes at 2.5, 5 and 10 μ M after 24 hours of incubation.

A.2.6.2 Concentrations 6.25 μ M, 12.5 μ M and 25 μ M

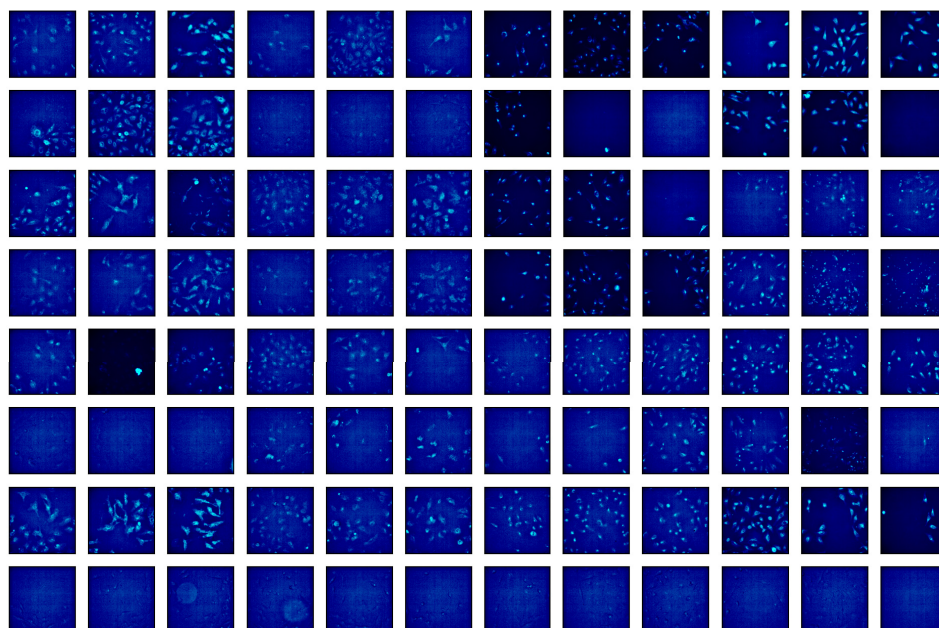


Figure A.235. 96 well plate of complexes at 6.25, 12.5 and 25 μ M after 3 hours of incubation.

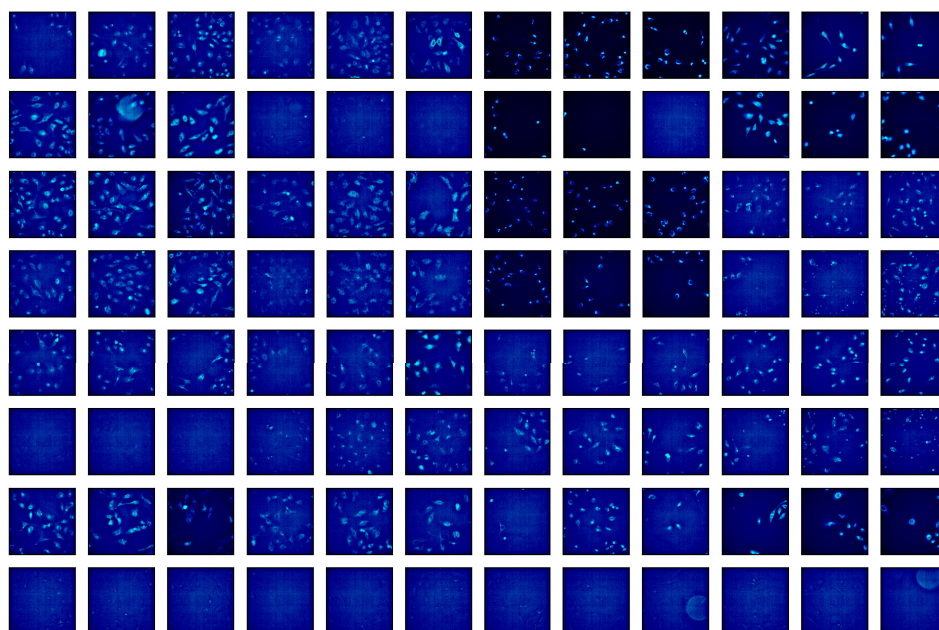


Figure A.236. 96 well plate of complexes at 6.25, 12.5 and 25 μ M after 6 hours of incubation.

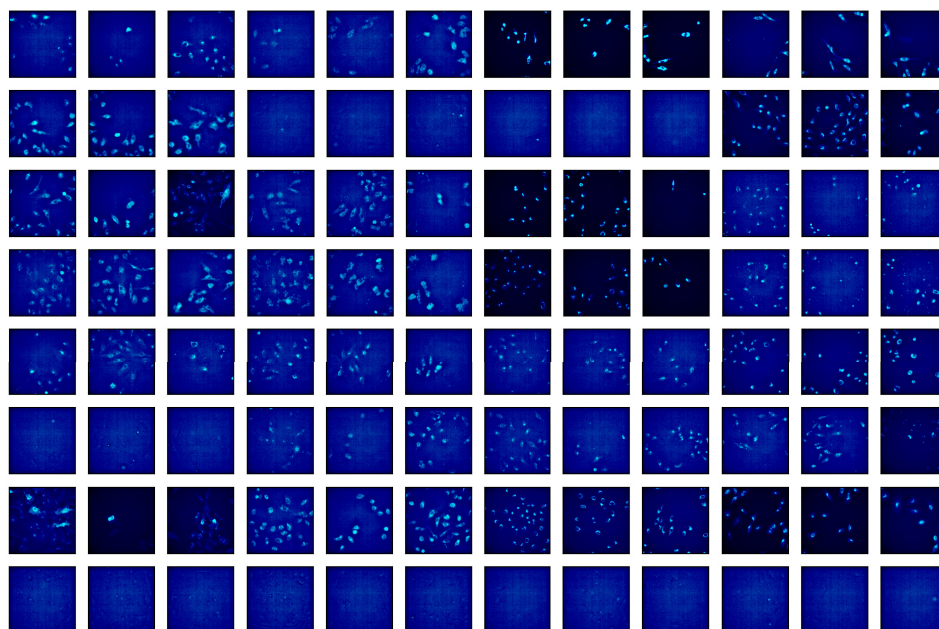


Figure A.237. 96 well plate of complexes at 6.25, 12.5 and 25 μ M after 24 hours of incubation.

A.2.6.3 Concentrations 25 μ M, 50 μ M and 100 μ M

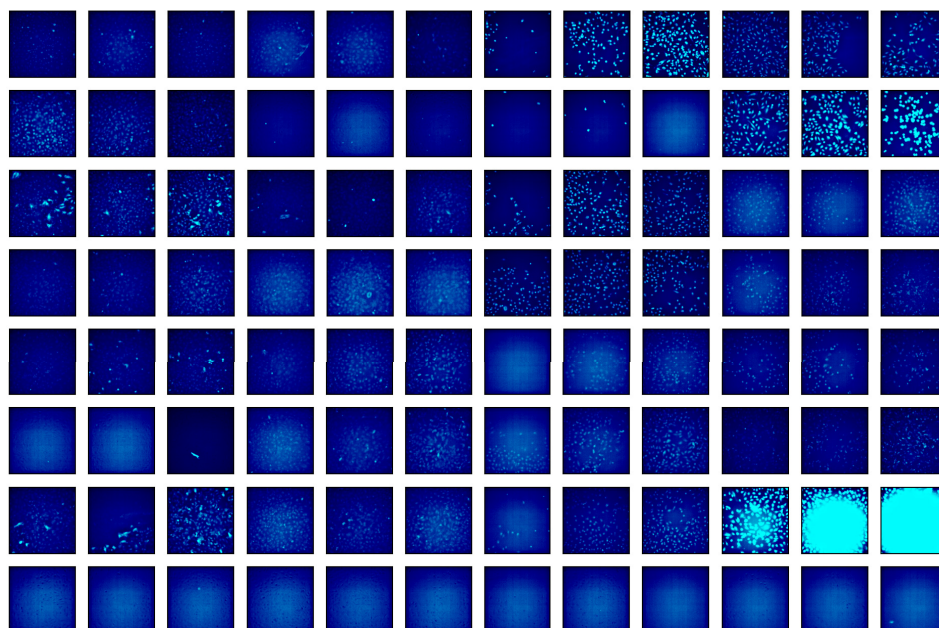


Figure A.238. 96 well plate of complexes at 25, 50 and 100 μ M after 6 hours of incubation.

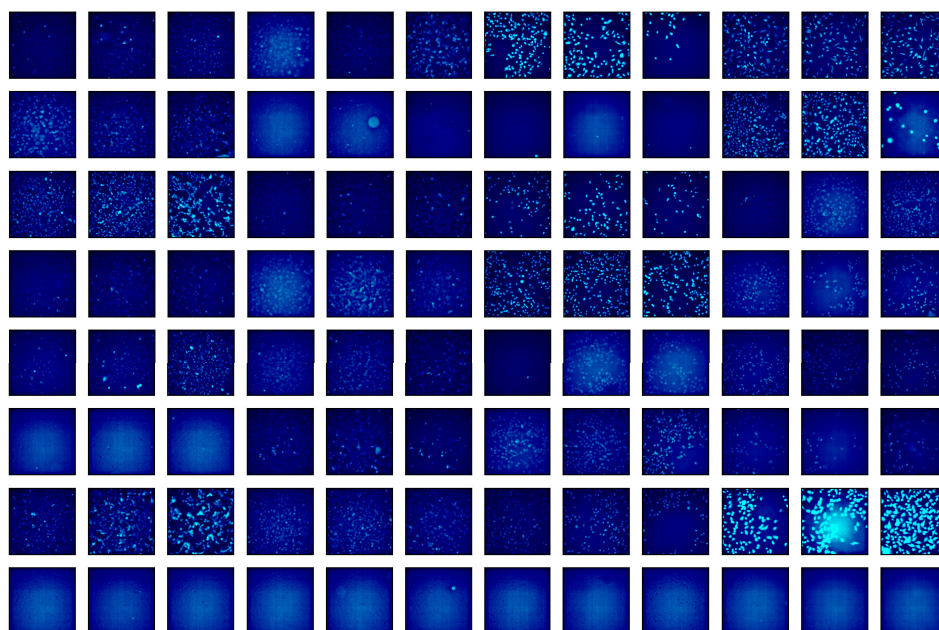


Figure A.239. 96 well plate of complexes at 25, 50 and 100 μM after 24 hours of incubation.

A.3 Computational data

A.3.1 Molecular descriptors

General Descriptor Type	Descriptor Class Name	Description
Topological	ABCIndex	Atom-Bond Connectivity index, measures molecular connectivity based on atom degrees and bond orders.
	AdjacencyMatrix	Represents molecular structure as a matrix of connected atoms.
	BalabanJ	Measures the average distance sum connectivity, related to molecular branching.

BaryszMatrix	Represents atomic properties and their topological relationships.
BertzCT	Complexity index based on bond types and symmetry.
Chi	Molecular connectivity indices, describe molecular branching and size.
DetourMatrix	Represents longest paths between atom pairs.
DistanceMatrix	Represents shortest paths between atom pairs.
EccentricConnectivityIndex	Describes atom centrality within the molecular graph.
ExtendedTopochemicalAtom	Characterizes atom environments based on connectivity.
MolecularDistanceEdge	Describes molecular shape based on distance sums.
InformationContent	Describes complexity based on symmetry and atom types.
PathCount	Counts paths of various lengths in the molecular graph.
TopologicalCharge	Describes charge distribution based on molecular topology.
TopologicalIndex	Various indices describing molecular shape and connectivity.

	VertexAdjacencyInformation	Information content of the adjacency matrix.
	WalkCount	Counts walks of various lengths in the molecular graph.
	WienerIndex	Describes molecular branching and shape.
	ZagrebIndex	Connectivity indices based on vertex degrees.
Physico-chemical	Autocorrelation	Describes the distribution of atomic properties along the molecular structure.
	AcidBase	Describes acid-base properties of the molecule.
	CPSA	Charged Partial Surface Area descriptors, relate to molecular polarity.
	Lipinski	Calculates Lipinski's Rule of Five for drug-likeness.
	LogS	Estimates aqueous solubility.
	McGowanVolume	Estimates molecular volume.
	Polarizability	Describes molecule's ability to form instantaneous dipoles.
	SLogP	Estimates octanol-water partition coefficient.
	TopoPSA	Topological Polar Surface Area, relates to drug absorption.

	VdwVolumeABC	Estimates van der Waals volume.
	HydrogenBond	Estimates hydrogen bonding capacity.
	Weight	Molecular weight and related properties.
Electronic	BCUT	Describes atomic properties and their relationships.
	EState	Electrotopological State indices, describe electronic and topological atom states.
	MoRSE	Represents 3D structure based on electron diffraction.
	MoeType	Various descriptors from the Molecular Operating Environment.
Structural	Aromatic	Describes aromatic features of the molecule.
	AtomCount	Counts atoms of different types.
	BondCount	Counts bonds of different types.
	CarbonTypes	Describes different carbon environments.
	Constitutional	Basic molecular properties like atom counts and molecular weight.
	Framework	Describes molecular scaffolds.

Geometric	RingCount	Counts rings of different sizes and types.
	RotatableBond	Counts rotatable bonds, related to molecular flexibility.
	MolecularId	Unique identifiers for molecules.
	GeometricalIndex	Describes 3D shape of molecules.
	GravitationallIndex	Describes mass distribution in 3D space.
	MomentOfInertia	Describes mass distribution around rotational axes.
	FragmentComplexity	Describes molecular complexity based on fragments.
	KappaShapeIndex	Describes molecular shape.
	PBF	Plane of Best Fit, describes 3D molecular shape.

Table A.11. Molecular descriptor classes within the Mordred descriptor package.

A.3.2 Dimensionality reduction

A.3.2.1 Equations

t-SNE dimensionality reduction:

1. Compute pairwise similarities in high-dimensional space The conditional probability that point x_i selects point x_j as a neighbour, given that neighbours are picked in proportion to their Gaussian distributed distance, is defined as:

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)} \quad (\text{A.1})$$

Where:

- x_i : i -th data point in the high-dimensional space.
- x_j : j -th data point in the high-dimensional space.
- $\|x_i - x_j\|^2$: Squared Euclidean distance between points x_i and x_j .
- σ_i : Standard deviation of the Gaussian distribution centered at point x_i (influenced by the perplexity hyperparameter).
- k : An index variable used in the summation, representing all other data points except for i .

2. Compute pairwise similarities in low-dimensional space The similarity between points y_i and y_j in the low-dimensional space is defined as:

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

Where:

- q_{ij} : Similarity between points y_i and y_j in the low-dimensional space.
- y_i : i -th data point in the low-dimensional space (after applying t-SNE).
- y_j : j -th data point in the low-dimensional space.
- $\|y_i - y_j\|^2$: Squared Euclidean distance between points y_i and y_j .
- k : Index variable used in the summation, representing all other points in the low-dimensional space.
- l : Index variable used in the summation, representing all other points in the low-dimensional space.

3. Minimize the Kullback-Leibler (KL) divergence The final step in t-SNE is to minimize the Kullback-Leibler divergence, which measures the difference between the high-dimensional and low-dimensional pairwise similarity distributions. This is typically done using gradient descent.

Where:

- C : Cost function to be minimized, representing the Kullback-Leibler divergence between the distributions P and Q .
- $\text{KL}(P \parallel Q)$: Kullback-Leibler divergence from distribution P to distribution Q .
- p_{ij} : Symmetrized version of the conditional probabilities, defined as:

$$p_{ij} = \frac{p_{(j|i)} + p_{(i|j)}}{2n}$$

- n : Total number of data points.

4. Optimize the low-dimensional embedding using gradient descent To minimize the KL divergence, the low-dimensional embedding is optimized using gradient descent. The gradient of the cost function C with respect to the low-dimensional point y_i is given by:

$$\frac{\partial C}{\partial y_i} = 4 \sum_{j \neq i} (p_{(j|i)} - q_{(j|i)}) (1 + \|y_i - y_j\|^2)^{-1} (y_i - y_j) \quad (\text{A.3})$$

Where:

- $\frac{\partial C}{\partial y_i}$: Gradient of the cost function C with respect to the low-dimensional point y_i . It indicates how the position of point y_i should be adjusted to minimize the cost.
- p_{ij} : Probability that point x_i would pick point x_j as a neighbour in the high-dimensional space. It is calculated using a Gaussian distribution.
- q_{ij} : Probability that point y_i would pick point y_j as a neighbour in the low-dimensional space. It is calculated using a Student's t -distribution.
- y_i : Low-dimensional representation of the i -th data point.
- y_j : Low-dimensional representation of the j -th data point.
- $\|y_i - y_j\|^2$: Squared Euclidean distance between the low-dimensional points y_i and y_j .
- $\sum_j$: Summation of all data points j in the low-dimensional space, allowing the gradient to account for the influence of all other points on the position of y_i .

PCA

1. PCA Elbow: The explained variance for the i -th principal component is given by:

$$\text{Explained Variance}_i = \frac{\lambda_i}{\sum_{j=1}^m \lambda_j} \quad (\text{A.4})$$

Where:

- λ_i : Eigenvalue for the i -th principal component.
- m : Total number of principal components.

2. Standardization: To standardize the dataset, use the formula:

$$Z = \frac{X - \mu}{\sigma} \quad (\text{A.5})$$

Where:

- X : Dataset.
- μ : Mean of the feature.

- σ : Standard deviation of the feature.

3. Covariance Matrix: The covariance matrix C is used to understand how features vary together:

$$C = \frac{1}{n-1} Z^T Z \quad (\text{A.6})$$

Where:

- n : Number of samples.
- Z^T : Transposed matrix of Z .

4. Eigenvalues and Eigenvectors of C : The eigenvalue equation for matrix C is:

$$Cv = \lambda v \quad (\text{A.7})$$

Where:

- v : Eigenvector of C .
- λ : Eigenvalue.

5. Sorting Eigenvalues and Eigenvectors to Identify Principal Components: Select the top k dimensions by picking the top k eigenvectors corresponding to the largest k eigenvalues to form the matrix V_k of the top k eigenvectors.

6. Projection onto Principal Components: To obtain the transformed dataset Y , project the original data X onto the principal components:

$$Y = XV_k \quad (\text{A.8})$$

UMAP Dimensionality Reduction

1. Compute pairwise distances for Euclidean distance between points x_i and x_j in the k -th dimension:

$$d(x_i, x_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2} \quad (\text{A.9})$$

Where:

- x_{ik} : k -th coordinate of point x_i .
- x_{jk} : k -th coordinate of point x_j .

2. Construct fuzzy topological representation: a) Find the distance to the nearest neighbour for each point:

$$\rho_i = \min_{j \neq i} d(x_i, x_j) \quad (\text{A.10})$$

b) Compute the local connectivity for each point:

$$\sigma_i = \min_{\sigma} \sum_j \exp \left(-\frac{d(x_i, x_j) - \rho_i}{\sigma} \right) = \log_2(k) \quad (\text{A.11})$$

Where:

- σ_i : Bandwidth parameter for point x_i .
- k : Number of nearest neighbours.

c) Compute the fuzzy set membership strength:

$$v_{ij} = \exp \left(-\frac{d(x_i, x_j) - \rho_i}{\sigma_i} \right) \quad (\text{A.12})$$

d) Symmetrize the fuzzy set membership between points x_i and x_j :

$$w_{ij} = w_{ji} = v_{ij} + v_{ji} - v_{ij} \cdot v_{ji} \quad (\text{A.13})$$

3. Construct the low-dimensional embedding:

a) Initialize the low-dimensional coordinates Y .

b) Optimize the embedding using stochastic gradient descent:

$$Y_{t+1} = Y_t - \eta \nabla C(Y_t) \quad (\text{A.14})$$

Where:

- Y_t : Low-dimensional matrix embedding at iteration t .
- η : Learning rate.
- $C(Y_t)$: Cross-entropy between the high and low-dimensional fuzzy topological representations.

4. Repeat step 3b until convergence or for a fixed number of iterations.

Parameters:

- **n_neighbors**: Controls the size of local neighbourhoods (similar to k in k-NN). Higher values result in more global structure preservation, influencing σ_i .
- **min_dist**: Minimum distance between points in the low-dimensional space. Affects the compactness of the resulting embedding.
- **n_components**: Number of dimensions in the target embedding.

A.3.2.2 UMAP plots

Fingerprint UMAP

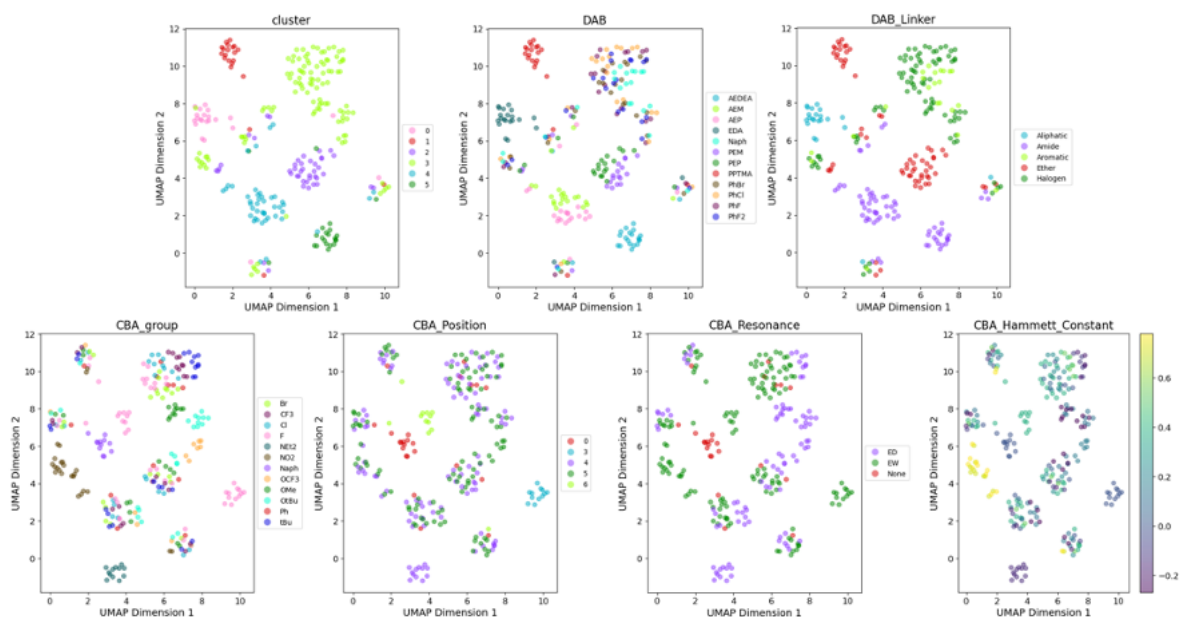


Figure A.241. 2D UMAP reduced fingerprint descriptors (n = 2) coloured from right to left as designated: cluster) coloured by assigned k-means cluster. DAB) coloured by the DAB group. DAB_Linker) coloured by the nature of the atom connecting the functionality to the DAB. CBA_group) coloured by the CBA functionality atom type without position information. CBA_Position) coloured by the position of the CBA functionality. CBA_Resonance) coloured by the electron donating or electron donating CBA resonance property. CBA_Hammett_Constants) coloured by the relative Hammett constant of the substituent on the CBA functionality with meta/para positions measured from the platinum bound oxygen.

Mordred UMAP

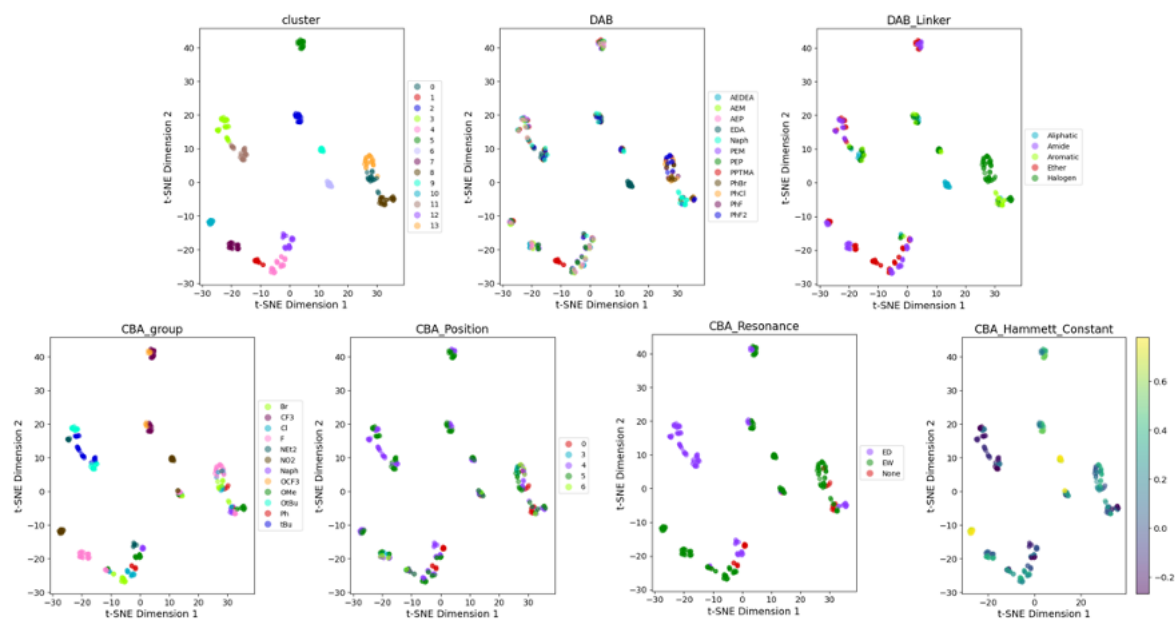


Figure A.242. Figure 15: 2D t-SNE plots of UMAP reduced Mordred descriptors ($n = 18$) coloured from right to left as designated: cluster) coloured by assigned k-means cluster. DAB) coloured by the DAB group. DAB_Linker) coloured by the nature of the atom connecting the functionality to the DAB. CBA_group) coloured by the CBA functionality atom type without position information. CBA_Position) coloured by the position of the CBA functionality. CBA_Resonance) coloured by the electron donating or electron donating CBA resonance property. CBA_Hammett_Constants) coloured by the relative Hammett constant of the substituent on the CBA functionality with meta/para positions measured from the platinum bound oxygen.

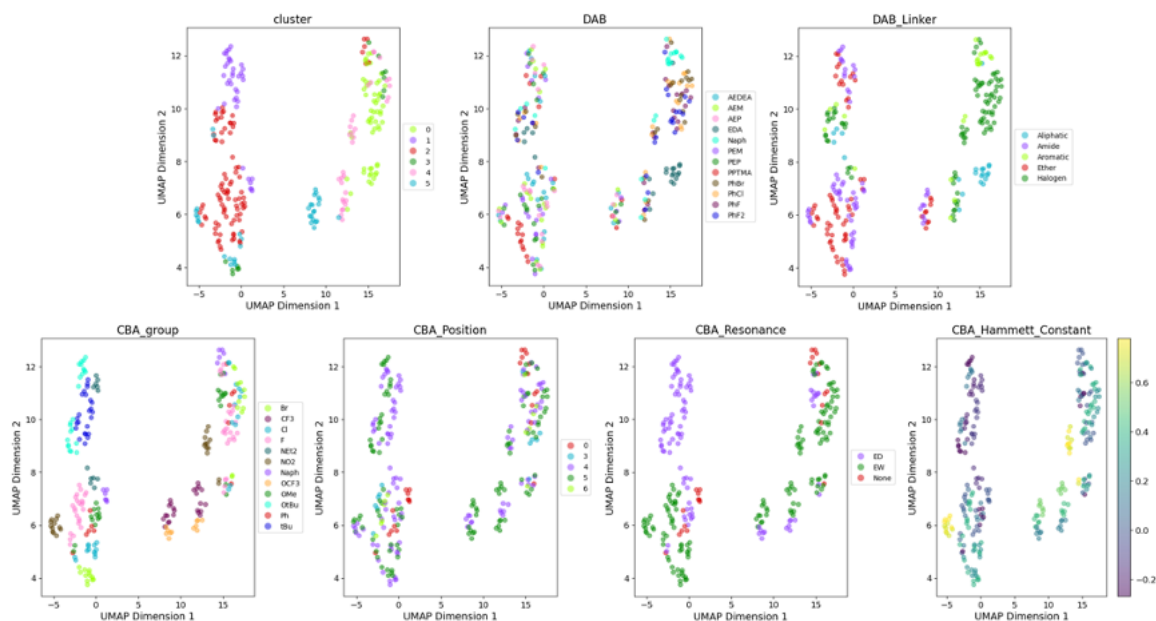


Figure A.243. 2D UMAP reduced Mordred descriptors ($n = 2$) coloured from right to left as designated: cluster) coloured by assigned k-means cluster. DAB) coloured by the DAB group. DAB_Linker) coloured by the nature of the atom connecting the functionality to the DAB. CBA_group) coloured by the CBA functionality atom type without position information. CBA_Position) coloured by the position of the CBA functionality. CBA_Resonance) coloured by the electron donating or electron donating CBA resonance property. CBA_Hammett_Constants) coloured by the relative Hammett constant of the substituent on the CBA functionality with meta/para positions measured from the platinum bound oxygen.

A.3.3 Machine learning algorithm equations

Linear Regression:

The coefficients β are estimated as follows:

$$\beta = (X^T X)^{-1} X^T y \quad (\text{A.15})$$

Where:

- X : Matrix of input features (including a column of ones for the intercept).
- y : Vector of observed values.
- X^T : Transpose of the matrix X .
- $(X^T X)^{-1}$: Inverse of the matrix product $X^T X$.

Ridge Regression:

The coefficient estimation is given by:

$$\beta = (X^T X + \lambda I)^{-1} X^T y \quad (\text{A.16})$$

Where:

- X : Matrix of input features (including a column of ones for the intercept).
- y : Vector of observed values.
- I : Identity matrix with dimensions $(p + 1) \times (p + 1)$, where p is the number of independent variables.
- λ : Regularization parameter controlling the strength of the regularization.

Support Vector Machine (SVM):

The SVM equation dependencies are:

$$\sum_{n=1}^N (\alpha_n - \alpha_n^*) = 0 \quad (\text{A.17})$$

Where:

- α_n : Lagrange multipliers corresponding to the support vectors.
- α_n^* : Lagrange multipliers corresponding to the support vectors.
- C : Regularization parameter.

The final function used to predict new values is:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) G(x_i, x) + b \quad (\text{A.18})$$

Where:

- b : Bias term or intercept that shifts the regression function.
- $G(x_i, x)$: Kernel function.
- α_i : Lagrange multiplier for the i -th support vector.

Scaling with StandardScaler:

A pre-requisite of the SVM model is that input data is required to be scaled, as the model assumes all features are centered around 0. The scaling is performed using:

$$z = \frac{x - u}{s} \quad (\text{A.19})$$

Where:

- x : Raw input feature.
- μ : Mean of the training samples.
- s : Standard deviation of the training samples.

A.3.4 Extended DFT and TDDFT energy prediction task model performances

HOMO prediction

HOMO	Mordred		Fingerprint	
CV	R^2	MAE (eV)	R^2	MAE (eV)
Linear	No fit	0.553 ± 0.154	No fit	$2.36 \times 10^{10} \pm 2.89 \times 10^{10}$
Ridge	0.105 ± 0.413	0.134 ± 0.0224	0.766 ± 0.0529	0.0743 ± 0.00890
Lasso	0.754 ± 0.0664	0.0729 ± 0.0103	No fit	0.164 ± 0.0158
RF	0.715 ± 0.0746	0.0831 ± 0.0140	0.713 ± 0.0658	0.0776 ± 0.0096
XGB	0.7120 ± 0.0735	0.0767 ± 0.0104	0.734 ± 0.0519	0.0747 ± 0.0093
SVM	0.765 ± 0.0731	0.0641 ± 0.0081	0.731 ± 0.0854	0.0657 ± 0.0081
NN	No fit	4.45 ± 0.571	0.395 ± 0.322	0.109 ± 0.0136

Table A.12. Table of ML model performances for predicting HOMO orbital energy. The optimised model parameters are as follows. Mordred: Ridge – alpha: 10. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 300. XGB – learning rate: 0.1, max depth: 3. SVM: C: 1, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. NN: batch size: 32, epochs: 300, dropout rate: 0.05, learning rate: 0.01. Fingerprints: Ridge – alpha: 1. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 100. XGB – learning rate: 0.1, max depth: 3. SVM - C': 10, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. NN – batch size: 128, epochs: 400, dropout rate: 0.05, learning rate: 0.01. Unspecified parameters were default.

HOMO	Mordred		Fingerprint	
Validation	External R ²	External MAE (eV)	External R ²	External MAE (eV)
Linear	No fit	13.7	No fit	2.61 x 10 ¹¹
Ridge	No fit	1.65	No fit	0.112
Lasso	No fit	0.121	No fit	0.148
RF	0.418	0.0714	0.626	0.0633
XGB	0.558	0.0696	0.701	0.0595
SVM	0.729	0.0588	0.456	0.0793
NN	No fit	5840	No fit	5.32

Table A.13. ML model validation performances for predicting HOMO energy. Parameters were kept the same as in table A.12.

LUMO prediction

LUMO	Mordred		Fingerprint	
CV	R ²	MAE (eV)	R ²	MAE (eV)
Linear	0.314 ± 0.971	0.106 ± 0.0457	No fit	5.18 x 10 ⁹ ± 5.14 x 10 ⁹
Ridge	0.955 ± 0.036	0.0319 ± 0.006	0.9782 ± 0.0087	0.0292 ± 0.0041
Lasso	0.952 ± 0.032	0.0367 ± 0.0062	No fit	0.2136 ± 0.0288
RF	0.860 ± 0.0855	0.0683 ± 0.0172	0.952 ± 0.0411	0.0365 ± 0.0079
XGB	0.881 ± 0.071	0.0565 ± 0.0123	0.961 ± 0.0297	0.0333 ± 0.0072
SVM	0.986 ± 0.0056	0.024 ± 0.0026	0.987 ± 0.0066	0.0207 ± 0.0034
NN	No fit	18.2 ± 2.29	0.984 ± 0.0061	0.0267 ± 0.0028

Table A.14. ML model performances for predicting LUMO energy. The optimised model parameters are as follows. Mordred: Ridge – alpha: 10. Lasso – alpha: 0.1. RF – max depth: 10, n_estimators: 200. XGB – learning rate: 0.1, max depth: 3. SVM - C: 0.1, degree: 2, epsilon: 0.01, gamma: 'scale', kernel: 'linear'. NN – batch size: 32, epochs: 200, dropout rate: 0.1, learning rate: 0.005. Fingerprint: Ridge – alpha: 0.1. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 100. XGB – learning rate: 0.1, max depth: 3. SVM - C: 10, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. NN – batch size: 32, epochs: 300, dropout rate: 0.05, learning rate: 0.001. Unspecified parameters were default.

LUMO	Mordred		Fingerprint	
Validation	External R ²	External MAE (eV)	External R ²	External MAE (eV)
Linear	No fit	7.87	No fit	5.79 x 10 ¹⁰
Ridge	No fit	0.261	0.645	0.0827
Lasso	0.172	0.124	No fit	0.179
RF	0.686	0.0731	0.312	0.111
XGB	0.871	0.0472	0.585	0.0837
SVM	0.388	0.0970	0.596	0.0911
NN	No fit	4370	No fit	2.41

Table A.15. ML model validation performances for predicting LUMO orbital energy. Parameters were kept the same as in table A.14.

HOMO-LUMO gap (HLG) prediction

HLG	Mordred		Fingerprint	
CV	R ²	MAE (eV)	R ²	MAE (eV)
Linear	No fit	0.878 ± 0.418	No fit	1.00 x 10 ¹⁰ ± 1.72 x 10 ¹⁰
Ridge	No fit	0.153 ± 0.0299	0.735 ± 0.0911	0.0813 ± 0.0104
Lasso	0.674 ± 0.0985	0.0919 ± 0.0137	No fit	0.17 ± 0.0269
RF	0.599 ± 0.1006	0.101 ± 0.014	0.718 ± 0.0711	0.0754 ± 0.0088
XGB	0.503 ± 0.185	0.106 ± 0.0195	0.688 ± 0.0922	0.0807 ± 0.0112
SVM	0.639 ± 0.113	0.100 ± 0.0162	0.746 ± 0.118	0.0748 ± 0.0156
NN	No fit	13.4 ± 1.49	0.920 ± 0.0278	0.0472 ± 0.0043

Table A.16. ML model performances for predicting HOMO LUMO gap energy. The optimised model parameters are as follows. Mordred: Ridge – alpha: 10. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 100. XGB – learning rate: 0.1, 'max depth': 3. SVM - C: 1, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. NN – batch size: 128, epochs: 100, dropout rate: 0.2, learning rate: 0.005. Fingerprint: Ridge – alpha: 1. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 100. XGB – learning rate: 0.1, max depth: 3. SVM - C: 10, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. NN – batch size: 32, epochs: 100, dropout rate: 0.05, learning rate: 0.001. Unspecified parameters were default.

HLG	Mordred		Fingerprint	
Validation	External R ²	External MAE (eV)	External R ²	External MAE (eV)
Linear	No fit	12.5	No fit	1.97 x 10 ¹⁰
Ridge	No fit	1.21	0.514	0.0952
Lasso	No fit	0.127	No fit	0.113
RF	0.504	0.0691	0.632	0.0645
XGB	0.555	0.0795	0.806	0.0429
SVM	0.574	0.0807	0.714	0.0615
NN	No fit	9900	No fit	2.95

Table A.17. ML model validation performances for predicting HOMO LUMO gap energy. Parameters were kept the same as in table A.16.

S₁ to S₀ transition energy prediction

S1 to S0	Mordred		Fingerprint	
CV	R ²	MAE (eV)	R ²	MAE (eV)
Linear	No fit	0.243 ± 0.082	No fit	9.15 x 10 ⁹ ± 1.04 x 10 ¹⁰
Ridge	0.604 ± 0.229	0.0742 ± 0.0162	0.840 ± 0.0926	0.0466 ± 0.00710
Lasso	0.762 ± 0.112	0.0536 ± 0.0087	No fit	0.126 ± 0.0209
RF	0.707 ± 0.114	0.0597 ± 0.0124	0.749 ± 0.191	0.0443 ± 0.0109
XGB	0.703 ± 0.160	0.0604 ± 0.0139	0.724 ± 0.188	0.0448 ± 0.0121
SVM	0.817 ± 0.109	0.0473 ± 0.0092	0.840 ± 0.119	0.0395 ± 0.0086
NN	No fit	2.348 ± 0.282	0.733 ± 0.0847	0.075 ± 0.0107

Table A.18. ML model performances for predicting S1 to S0 transition energy. The optimised model parameters are as follows. Mordred: Ridge – alpha: 10. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 300. XGB – learning rate: 0.1, max depth: 3. SVM - C: 0.1, degree: 2, epsilon: 0.01, gamma: 'scale', kernel: 'linear'. NN – batch size: 64, epochs: 400, dropout rate: 0.05, learning rate: 0.01. Fingerprint: Ridge – alpha: 1. Lasso – alpha: 0.1. RF – max depth: 10, n_estimators: 100. XGB – learning rate: 0.1, max depth: 3. SVM – C: 10, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. NN – batch size: 128, epochs: 300, dropout rate: 0.05, learning rate: 0.005. Unspecified parameters were default.

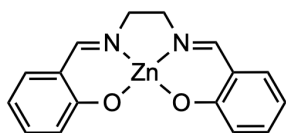
S1 to S0	Mordred		Fingerprint	
Validation	External R ²	External MAE (eV)	External R ²	External MAE (eV)
Linear	No fit	2.274	No fit	8.75 x 10 ¹⁰
Ridge	No fit	0.911	0.857	0.0431
Lasso	0.449	0.0919	0.00	0.117
RF	0.669	0.0611	0.829	0.0471
XGB	0.655	0.0713	0.874	0.0397
SVM	0.246	0.110	0.844	0.0413
NN	No fit	5260	No fit	2.49

Table A.19. ML model validation performances for predicting S1 to S0 transition energy. Parameters were kept the same as in table A.18.

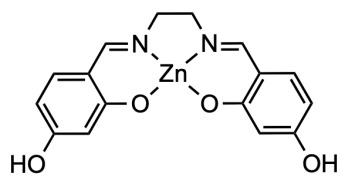
A.3.5 Transfer learning

A.3.5.1 Structures of zinc(II) salphen and platinum(II) NNC complexes used in transfer learning

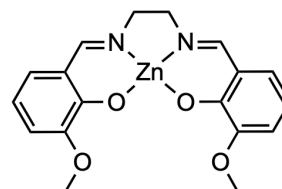
Zinc(II) salphen complexes



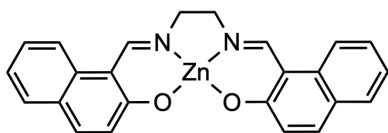
Zn_EDA_Ph



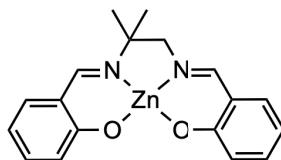
Zn_EDA_4OH



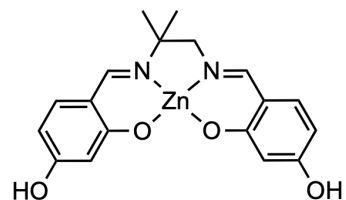
Zn_EDA_3OMe



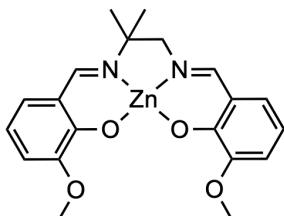
Zn_EDA_Naph



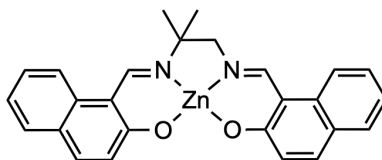
Zn_EDAC2H6_Ph



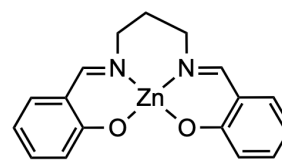
Zn_EDAC2H6_4OH



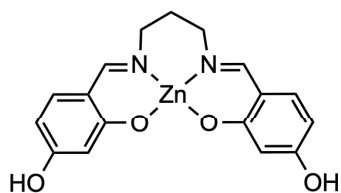
Zn_EDAC2H6_3OMe



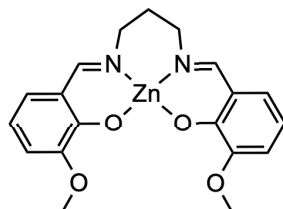
Zn_EDAC2H6_Naph



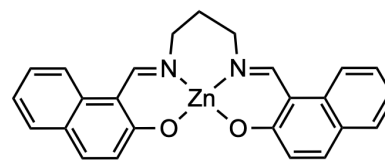
Zn_PDA_Ph



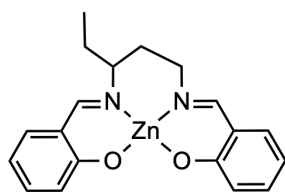
Zn_PDA_4OH



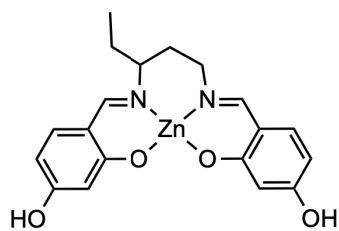
Zn_PDA_3OMe



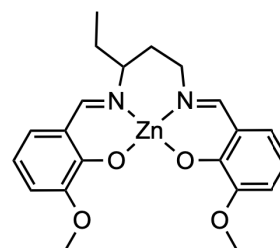
Zn_PDA_Naph



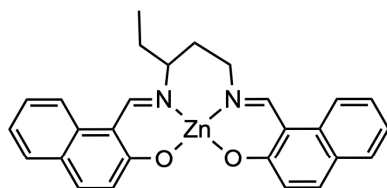
Zn_PDAC2H5_Ph



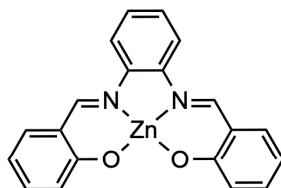
Zn_PDAC2H5_4OH



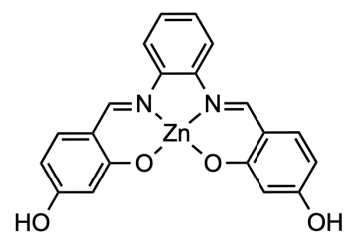
Zn_PDAC2H5_3OMe



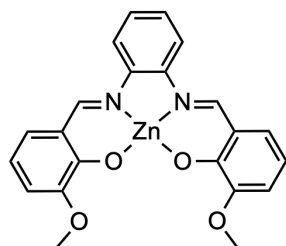
Zn_PDAC2H5_Naph



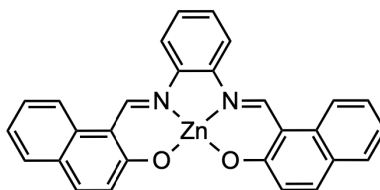
Zn_Ph_Ph



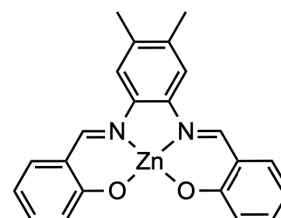
Zn_Ph_4OH



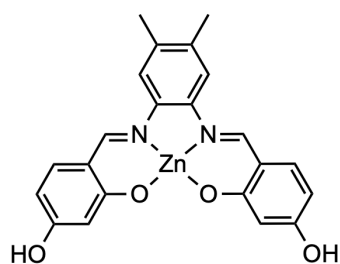
Zn_Ph_3OMe



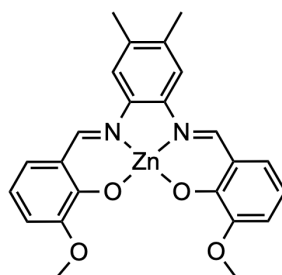
Zn_Ph_Naph



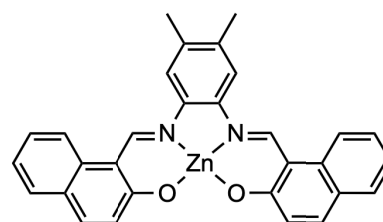
Zn_PhMe2_Ph



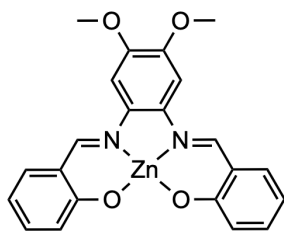
Zn_PhMe2_4OH



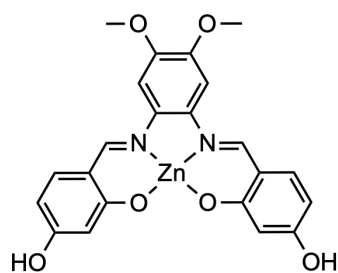
Zn_PhMe2_3OMe



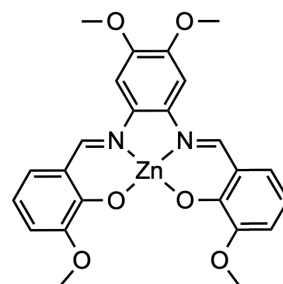
Zn_PhMe2_Naph



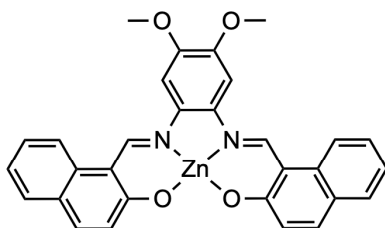
Zn_PhOMe2_Ph



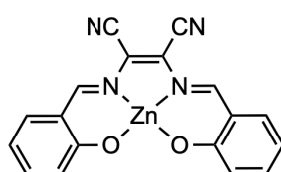
Zn_PhOMe2_4OH



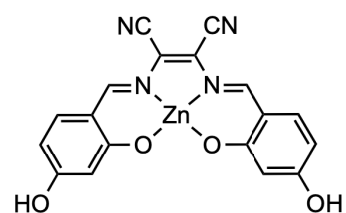
Zn_PhOMe2_3OMe



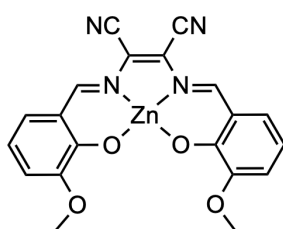
Zn_PhOMe2_Naph



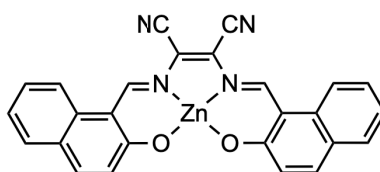
Zn_CN2_Ph



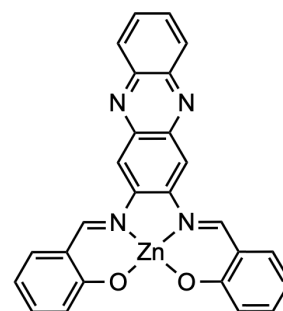
Zn_CN2_4OH



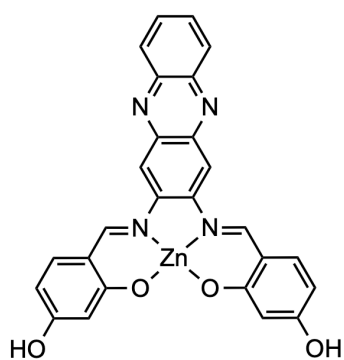
Zn_CN2_3OMe



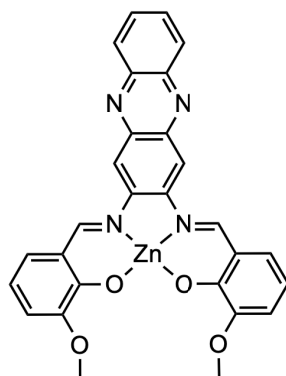
Zn_CN2_Naph



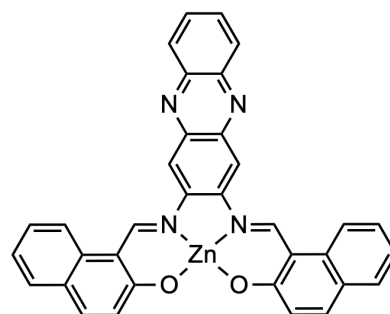
Zn_Phaz_Ph



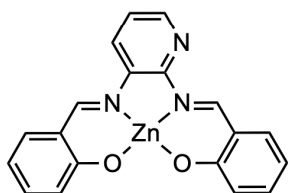
Zn_Phaz_4OH



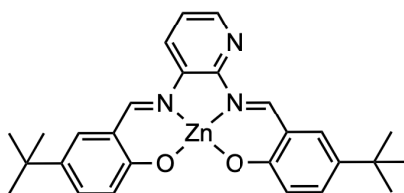
Zn_Phaz_3OMe



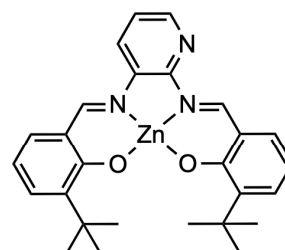
Zn_Phaz_Naph



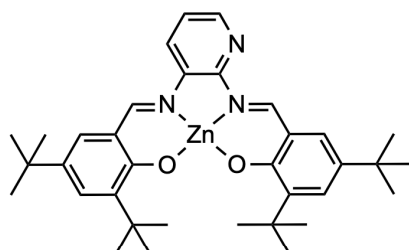
Zn_Py_Ph



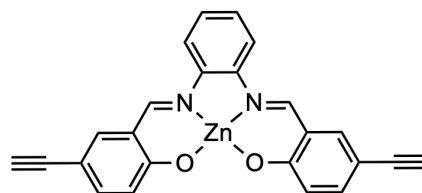
Zn_Py_5tBu



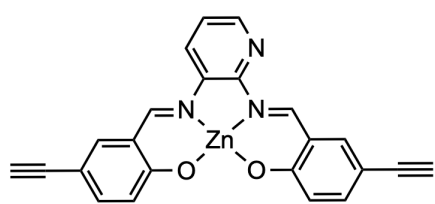
Zn_Py_3tBu



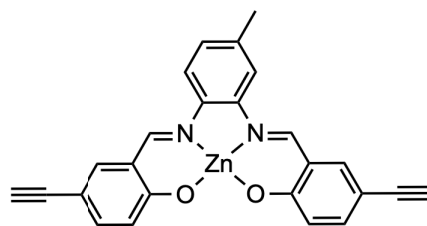
Zn_Py_5-3tBu



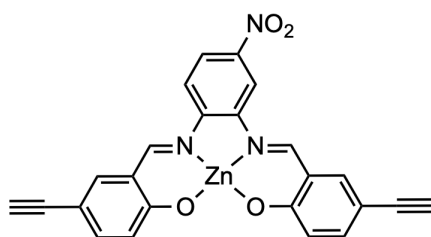
Zn_Ph_5Alkyne



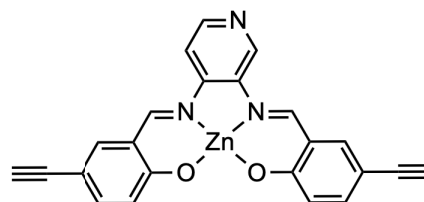
Zn_Py_5Alkyne



Zn_PhMe_5Alkyne



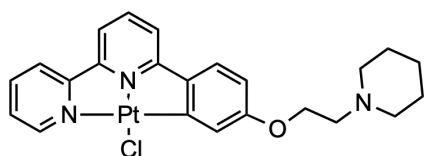
Zn_PhNO2_5Alkyne



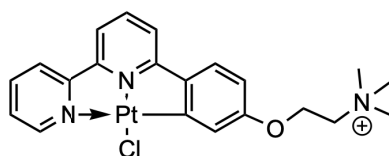
Zn_p-mPy_5Alkyne

Figure A.244. Structures of the zinc(II) salen complexes used within the transfer learning section of the computational chapter.

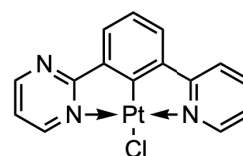
Platinum(II) NNC complexes



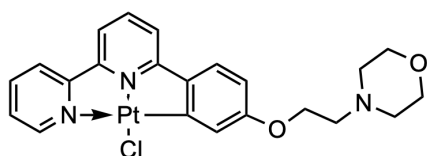
NNC_3PEP



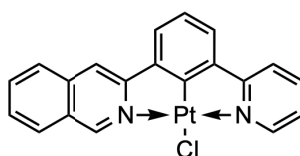
NNC_3PETMA



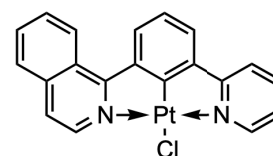
NCN_1NN



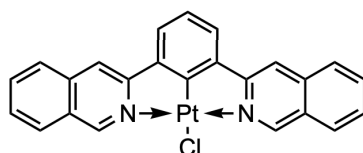
NNC_3PEM



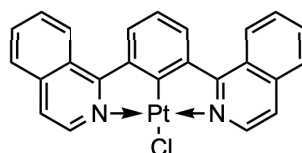
NCN_1mNaph



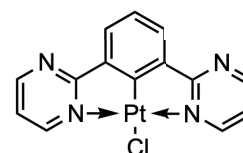
NCN_1pNaph



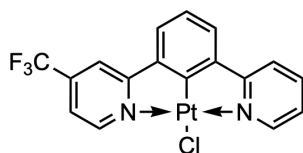
NCN_1mNaph_3mNaph



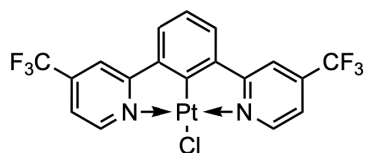
NCN_1pNaph_3pNaph



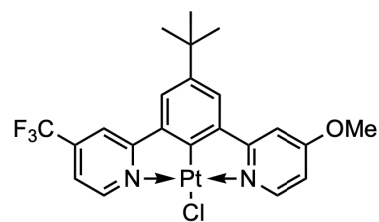
NCN_1NN_3NN



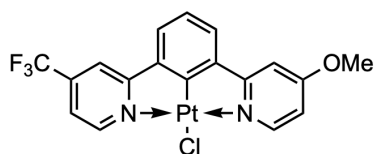
NCN_1CF3



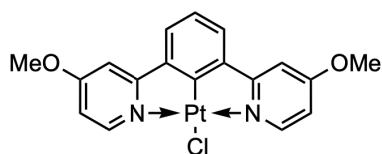
NCN_1CF3_3CF3



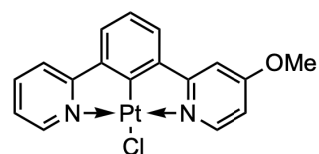
NCN_1CF3_3OMe_2tBu



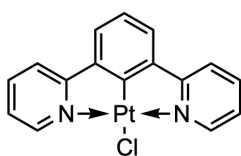
NCN_1CF3_3OMe



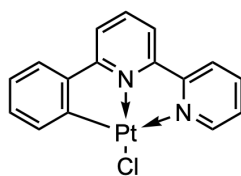
NCN_1OMe_3OMe



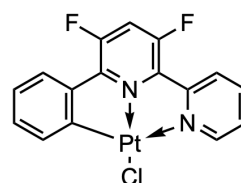
NCN_1OMe



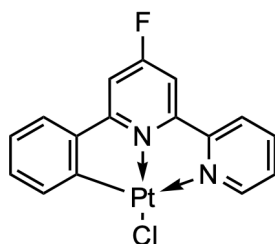
NCN



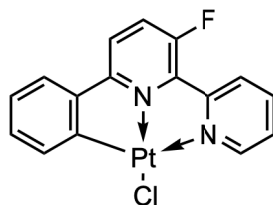
CNN



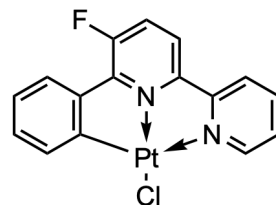
CNN_2-2F4F



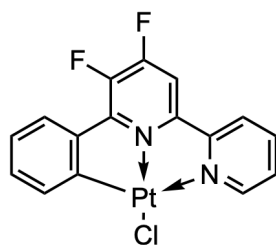
CNN_2-3F



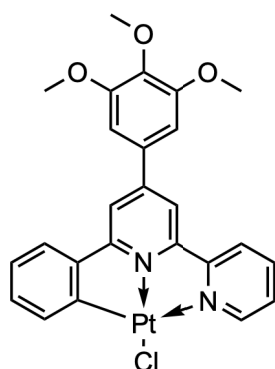
CNN_2-4F



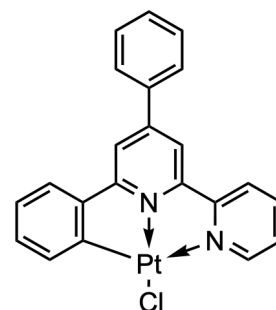
CNN_2-2F



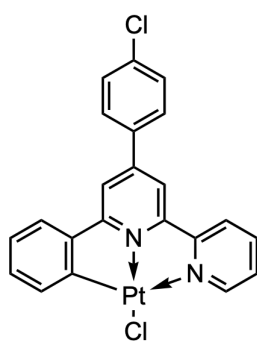
CNN_2-2F3F



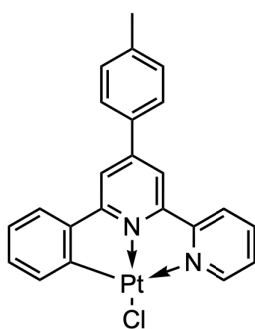
CNN_2Ph345OMe3



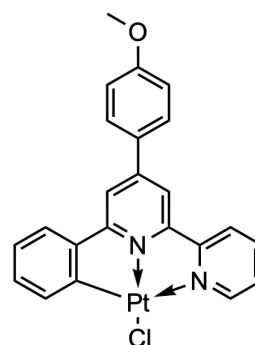
CNN_2Ph



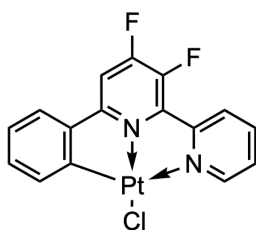
CNN_2PhpCl



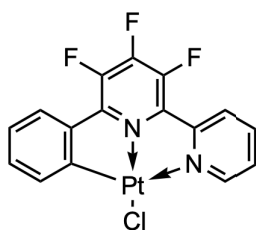
CNN_2Ph4Me



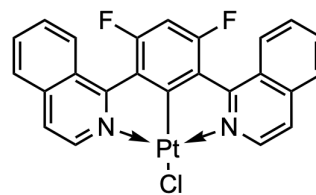
CNN_2Ph4OMe



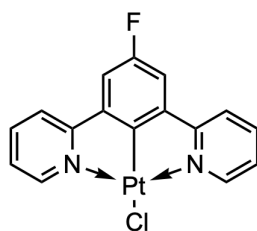
CNN_2-3F4F



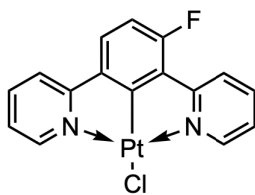
CNN_2-2F3F4F



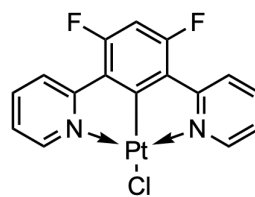
NCN_1pNaph_2-25F2_3pNaph



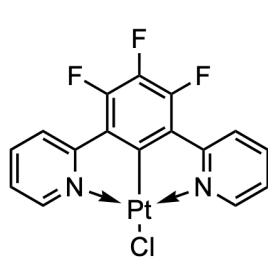
NCN_2-3F



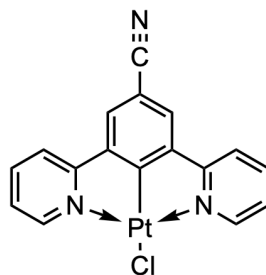
NCN_2-4F



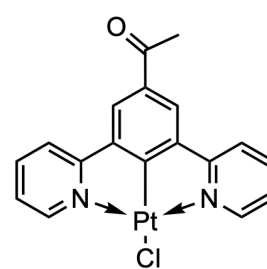
NCN_2-2F4F



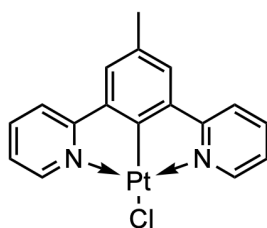
NCN_2-2F3F4F



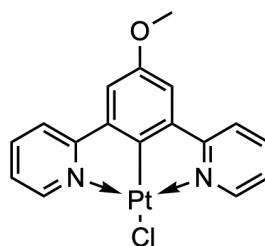
NCN_2-3CN



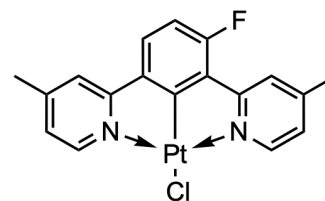
NCN_2-3COCH3



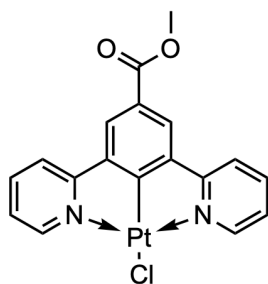
NCN_2-3Me



NCN_2-3OMe



NCN_1pMe_2mF_3pMe



NCN_2pCOOMe

Figure A.245. Structures of the platinum(II) NNC complexes used within the transfer learning section of the computational chapter.

A.3.5.2 TL models

Zinc dataset only training

Zinc(II) Salphen	Mordred		Fingerprint	
Experimental Model CV	R ²	MAE (eV)	R ²	MAE (eV)
Ridge	No fit	0.206 ± 0.179	0.835 ± 0.0831	0.0802 ± 0.0163
Lasso	No fit	0.156 ± 0.0677	No fit	0.229 ± 0.0719
RF	0.629 ± 0.187	0.113 ± 0.0342	0.808 ± 0.157	0.0735 ± 0.0399
XGB	0.493 ± 0.250	0.137 ± 0.0398	0.825 ± 0.216	0.0706 ± 0.0429
SVM	0.825 ± 0.117	0.0815 ± 0.0342	0.836 ± 0.170	0.0651 ± 0.0191

Table A.20. ML model performances for predicting zinc(II) salphen emission energy. Models were trained on the zinc dataset only. The optimised model parameters are as follows. Mordred: Ridge – alpha: 10. Lasso – alpha: 10. RF – max depth: 15, n_estimators: 300. XGB – learning rate: 0.1, max depth: 7. SVM – C: 10, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. Fingerprint: Ridge – alpha: 1. Lasso – alpha: 0.1. RF – max depth: 10, n_estimators: 200. XGB – learning rate: 0.1, max depth: 3. SVM – C: 10, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. Unspecified parameters were default.

NNC dataset only training

Platinum(II) NNC	Mordred		Fingerprint	
Experimental Model CV	R ²	MAE (eV)	R ²	MAE (eV)
Ridge	No fit	0.752 ± 1.01	0.708 ± 0.159	0.0701 ± 0.0204
Lasso	No fit	0.120 ± 0.0555	No fit	0.161 ± 0.0257
RF	0.737 ± 0.152	0.0689 ± 0.0182	0.528 ± 0.375	0.0814 ± 0.0267
XGB	0.723 ± 0.241	0.0599 ± 0.0173	0.205 ± 0.897	0.0947 ± 0.0608
SVM	0.379 ± 0.441	0.0927 ± 0.0279	0.612 ± 0.233	0.0809 ± 0.0192

Table A.21. ML model performances for predicting platinum(II) NNC emission energy. Models were trained on the platinum(II) NNC dataset only. The optimised model parameters are as follows. Mordred: Ridge – alpha: 10. Lasso – alpha: 1. RF – max depth: 15, n_estimators: 100. XGB – learning rate: 0.1, max depth: 3. SVM – C: 1, degree: 2, epsilon: 0.01, gamma: 'auto', kernel: 'rbf'. Fingerprint: Ridge – alpha: 1. Lasso – alpha: 0.1. RF – max depth: 15, n_estimators: 200. XGB – learning rate: 0.1, max depth: 5. SVM – C: 0.1, degree: 2, epsilon: 0.01, gamma: 'scale', kernel: 'linear'. Unspecified parameters were default.

NNC transfer learning model performance

Original Prediction Task	TL Model (CV)	R ²	MAE (eV)
HOMO LUMO gap	Mordred + RF	0.257 ± 0.350	0.113 ± 0.0352
HOMO LUMO gap	Mordred + XGB	0.583 ± 0.299	0.0692 ± 0.0214
HOMO LUMO gap	Mordred + SVM	0.369 ± 0.219	0.102 ± 0.0293
HOMO LUMO gap	Fingerprint + Ridge	0.632 ± 0.174	0.0723 ± 0.0210
HOMO LUMO gap	Fingerprint + RF	0.544 ± 0.306	0.0746 ± 0.0291
HOMO LUMO gap	Fingerprint + XGB	0.497 ± 0.442	0.0751 ± 0.0321
HOMO LUMO gap	Fingerprint + SVM	0.509 ± 0.136	0.0933 ± 0.0245
S1 to S0 energy	Mordred + RF	0.769 ± 0.119	0.0594 ± 0.0173
S1 to S0 energy	Mordred + XGB	0.583 ± 0.299	0.0692 ± 0.0214
S1 to S0 energy	Fingerprint + Ridge	0.632 ± 0.174	0.0723 ± 0.0210
S1 to S0 energy	Fingerprint + RF	0.526 ± 0.320	0.0759 ± 0.0288
S1 to S0 energy	Fingerprint + XGB	0.497 ± 0.442	0.0751 ± 0.0321
S1 to S0 energy	Fingerprint + SVM	0.509 ± 0.136	0.0933 ± 0.0245
T1 to S0 energy	Mordred + RF	0.257 ± 0.350	0.113 ± 0.0352
T1 to S0 energy	Fingerprint + Ridge	0.318 ± 0.463	0.120 ± 0.0463
T1 to S0 energy	Fingerprint + RF	0.156 ± 0.612	0.121 ± 0.0559
T1 to S0 energy	Fingerprint + XGV	0.244 ± 0.589	0.122 ± 0.0561
T1 to S0 energy	Fingerprint + SVM	0.342 ± 0.401	0.123 ± 0.0407

Table A.22. NNC transfer learning full results table.

A.3.6 DA model performances

Zinc(II) Salphen		
S1 to S0: Mordred + Lasso	R ²	MAE (eV)
Domain adaptation model	0.585 ± 0.119	0.131 ± 0.0350
Pure TL model	No fit	0.189 ± 0.0935
Direct prediction from original model	No fit	1.7843
Model trained only on zinc dataset	No fit	0.156 ± 0.0677
S1 to S0: Mordred + XGB	R ²	MAE (eV)
Domain adaptation model	0.411 ± 0.541	0.131 ± 0.0439
Pure TL model	0.377 ± 0.298	0.144 ± 0.0529
Direct prediction from original model	0.6225	0.1348
Model trained only on zinc dataset	0.493 ± 0.250	0.137 ± 0.0398

Table A.23. ML performances for the task of predicting zinc(II) salphen emission energy using the Mordred descriptors. The table contains models trained by domain adaptation, pure TL, direct prediction from the original platinum(II) salphen model and the models trained exclusively on the zinc dataset for comparison of performances.

A.3.7 Feature importance

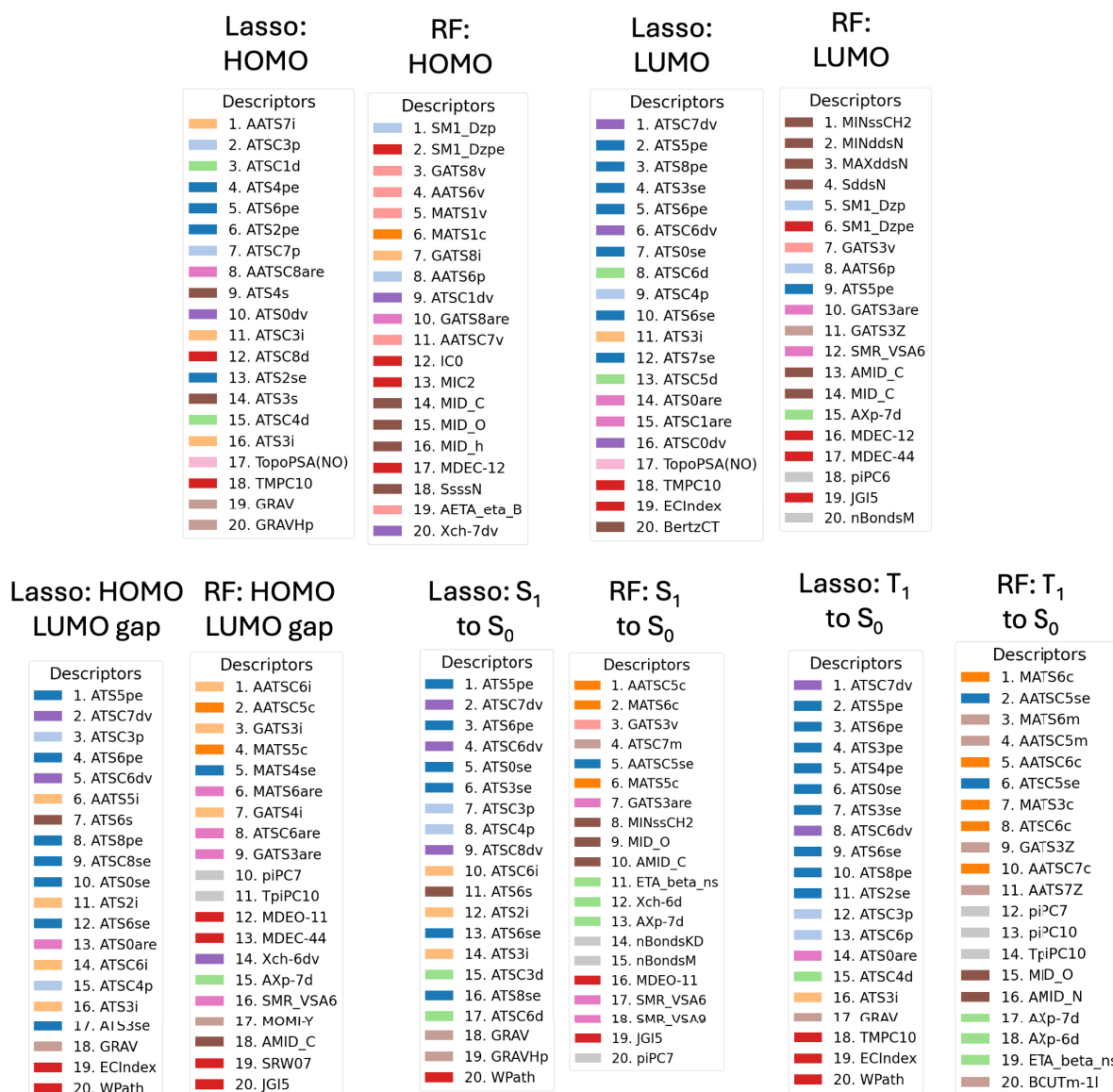


Figure A.246. Labelled for the important features for each DFT and TDDFT property determined by lasso and RF regression models. See Mordred github for descriptor definitions: <https://mordred-descriptor.github.io/documentation/master/descriptors.html>