Synthetic and Computational Studies Towards Novel Medical Imaging Agents for PET and MRI

SHANE ANGOH

Imperial College London

Department of Chemistry

January 2024

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Chemistry

Statement of Originality

The work reported in this thesis was completed between October 2019 and January 2024 at the Department of Chemistry, Imperial College London. The work reported is entirely my own, unless stated otherwise through cited reference or acknowledgement and has never been previously submitted for a degree at this or any other university.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution Non-Commercial No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

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.

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.

Abstract

Molecular imaging has a vital role in advancing our understanding of biological processes and treatment monitoring. This thesis explores the synthesis of manganese-based contrast agents (MnBCAs) for dual-modal PET-MRI; the development of predictive tools to aid in the design of both manganese and gallium complexes used in molecular imaging; as well as the modelling of the formation of ^{89}Zr-doxorubicin complexes inside liposomal formulations such as DOXIL *via* density functional theory (DFT).

Initial efforts were directed towards the synthesis of novel ligand families for PET-MRI MnBCAs, their subsequent complexation, and testing. Macrocyclic and acyclic ligands containing sulfonamides and amides were explored, with successful Mn^{2+} complexation achieved with acyclic picolyl amide ligands. DFT calculations were carried out to predict the thermodynamic stability constant of MnBCAs with Gaussian 16; UPW6B95D3/TZVP/IEFPCM(SMD) with two explicit solvent water molecules best predicted the stability constants of the chosen testing dataset (MAE: 1.84 log units, RMSE: 3.72 log units). Furthermore, the possible correlation between MRI-relevant properties and the DFT optimised geometry of 46 MnBCAs was investigated, with no significant correlation found. Machine learning models were assessed to consider features of the geometry, DFT calculated Gibbs energies, and other properties of the MnBCAs for stability constant prediction, and provided predictions which correlated well with experimental values (R^2 0.851, MAE: 0.50 log units, RMSE: 0.52 log units).

The identity of potential ^{89}Zr-doxorubicin complexes was modelled *via* DFT under simulated basic, neutral, and acidic conditions at the M06-2X/Sapporo-TZP-2012/6-31G(d,p)/IEFPCM level. For the acidic conditions inside DOXIL, modelling indicated that $Zr(oxinate)_4$ would almost entirely convert to $[Zr(oxinate)(doxorubicinate)_2]^+$ (97.2%) in the presence of doxorubicin.

Finally, a predictive model for the partition coefficient (log P) of gallium complexes with applications in myocardial perfusion imaging was developed. DFT calculations with PW6B95D3/TZVP/SMD predicted log P with excellent correlation to experimental values (R^2 0.829).

Acknowledgements

First of all, I would like to thank Prof. Nicholas Long for his unwavering encouragement and guidance throughout this PhD. I will always be grateful for the support given when I was most unwell, and your willingness to make one of the more unconventional work arrangements and projects within our mainly synthesis-focused group make it to the end. I would also like to extend my thanks to Drs Rafa T. M. de Rosales, Tom Bonasera, and Phil Murphy for their support and guidance in the direction of the project.

Thank you to the technical staff in the Department of Chemistry for their help in assisting my PhD research. Special thanks to Peter Haycock for an invaluable NMR service, Dr Lisa Haigh and Malgorzata Puchnarewicz for dealing with the countless sample submissions to mass spectrometry, and Dr Andrew White for solving the X-ray crystal structures in this thesis. I would especially like to thank the administrative staff of the Smart Medical Imaging CDT, and Imperial Chemistry Department for the huge amount of work put in to get this thesis through a pandemic, and their support during the most difficult parts of my research experience – namely Dr Valeria De Marco and Dr Mike Ray, whom I am immensely appreciative of their warmth and support both before and after getting long COVID.

My research experience has been enriched by the kindness and guidance of those who have helped me along the way. I would like to thank all members of the Long group, past and present, for making the lab and office welcoming and supportive places to be in, and allowing me to commandeer the lab speakers every day I was in. Special thanks to Tom and Jin, whom I started this journey with back in 2019, thanks for all of the laughs, sympathetic chats, and pushing me forward. To Brad and Ed, for some outrageous footballing opinions over the years, and letting me go on and on about Manchester United's troubles. To Saul and Chloe, for your mentorship over the last four years, and always allowing me to vent when things were not going as planned. To Ryan, for your valuable guidance and help with DFT. Additionally, I would like to thank the members of the de Rosales research group – Peter, Aish, and George, for their support and training in radiochemistry. Special thanks also to Faysal for training me on the relaxometer and our numerous discussions about MRI, academia, and life in general.

Without exaggeration, this project would have never finished without the support from my family. There are not enough words to express how much your support has meant to me throughout what has been an incredibly challenging last few years. Mum, Dad, Laura, and Jessica – each one of you inspire me and continually shape me to become a better person, for this I am eternally grateful.

And finally, to Maddie. This success is ours, for without you, I surely would not have made it through. You have been my rock, you have helped me stand up after each setback without fail (metaphorically *and* literally), you have kept me smiling and loved throughout this journey. Thank you.

Table of Contents

Statement of Originality 2

Copyright Declaration 2

Abstract 3

Acknowledgements 4

Table of Contents 5

Abbreviations 11

Chapter 1: Introduction 13

1.1 Molecular imaging 14

1.1.1 Molecular imaging modalities 14

1.1.1.1 Magnetic Resonance Imaging 16

1.1.1.2 Positron Emission Tomography 19

1.1.1.3 Other imaging techniques 21

1.1.1.3.1 Single-Photon Emission Computed Tomography 21

1.1.1.3.2 X-ray/Computed Tomography 21

1.1.1.3.3 Ultrasound 22

1.2 Dynamic Contrast Enhanced MRI 23

1.3 Gadolinium Based Contrast Agents 27

1.3.1 Nephrogenic Systemic Fibrosis 29

1.4 Alternatives to gadolinium 31

1.4.1 Manganese-Enhanced Magnetic Resonance Imaging 31

1.4.2 Manganese-Based Contrast Agents 32

1.4.2.1 MnBCA design considerations 35

1.4.2.2 Mn^{3+} MnBCAs 36

1.5 PET and dual-modal imaging 37

1.5.1 PET imaging 37

1.5.2 Dual-modal imaging 38

1.6 DFT and analytics 40

1.7 Aims of this work 44

1.8 References 45
Chapter 2: Synthesis of Ligands for Manganese-Based Contrast Agents 52
2.1 Introduction 53
2.1.1 Acyclic ligands 55
2.1.2 Macrocyclic ligands 55
2.1.3 Phosphasalens 56
2.1.4 Sulfonamides and amides 58
2.2 Results and Discussion 59
2.2.1 Phosphasalens 59
2.2.1.1 Initial phosphasalen ligands and DFT evaluation 60
2.2.1.2 Synthesis 64
2.2.1.2.1 PS1 64
2.2.1.2.2 PS2 65
2.2.1.2.3 PS3 67
2.2.1.3 Limitations 68
2.2.2 Sulfonamides 70
2.2.2.1 Development of amine bridges 72
2.2.2.2 Attempted syntheses of 1,10-phenanthroline-2,9-disulfonyl dichloride 74
2.2.2.3 Attempted syntheses of pyridine-2-sulfonyl chloride 76
2.2.2.4 Sulfonamide formation reactions 76
2.2.2.4.1 SuFEx reactions 77
2.2.2.4.2 Methyl acetate sulfonamides 78
2.2.3 Amides 78
2.2.3.1 Attempted syntheses of macrocyclic amides 79
2.2.3.2 Acyclic amides 81
2.2.3.3 Picolyl amides 83
2.2.3.3.1 PicAm-1 complexation 84
2.2.3.3.2 Relaxivity measurement on MnPicAm-1 85
2.2.3.3.3 PicAm-2 complexation 87

2.3 Conclusion 88
2.3.1 Future work 89
2.3.2 COVID-19 disruption 92
2.4 References 93
Chapter 3: Can a DFT optimisation predict the key metrics for MnBCAs before they have been synthesised? 96
3.1 Introduction 97
3.1.1 Thermodynamic stability prediction 97
3.1.2 MnBCA DFT in literature 101
3.1.3 Machine learning 104
3.1.4 SAMPL challenges 107
3.1.5 Aims 108
3.2 Results 109
3.2.1 Computational methods 109
3.2.2 DFT log(K_{MnL}) calculation 110
3.2.2.1 Thermodynamic cycle 110
3.2.2.2 Testing dataset 111
3.2.2.3 Training dataset 112
3.2.2.4 Model chemistries 114
3.2.2.5 Optimisations 115
3.2.2.6 Model training 116
3.2.2.6.1 Clustering 122
3.2.2.7 Model testing 124
3.2.2.8 Model prediction 126
3.2.2.8.1 Bispidine-COOH 127
3.2.2.8.2 Isomer prediction 127
3.2.3 DFT geometry and MnBCA metrics 127
3.2.3.1 Hypotheses and model conditions 127
3.2.3.1.1 Evaluating model chemistries 129

3.2.3.1.2 Data collection 132
3.2.3.1.3 Success parameters 132
3.2.3.2 Continuous Shape Measure and log(K_{MnL}) 134
3.2.3.3 Bond length distribution and log(K_{MnL}) 141
3.2.3.4 Water parameters 145
3.2.3.4.1 Bond length distribution 145
3.2.3.4.2 Charge 148
3.2.3.4.3 Steric overlap 149
3.2.4 Leveraging the large dataset from Section 3.2.3 150
3.2.4.1 Clustering 153
3.2.5 Machine learning prediction of MnBCA KPIs 156
3.3 Discussion 159
3.4 Conclusion 161
3.4.1 Future work 163
3.5 References 164
Chapter 4: What Zr(IV) species is being formed within DOXIL? 169
4.1 Introduction 170
4.1.1 Liposomal drug delivery 170
4.1.2 Zr DFT studies 173
4.1.3 Aim of this work 177
4.2 Results and Discussion 178
4.2.1 Method 178
4.2.2 Benchmarking each model chemistry 178
4.2.2.1 Chosen basis sets and functionals to be evaluated 178
4.2.2.2 Evaluation parameters 179
4.2.3 Model conditions 182
4.2.3.1 Approximations 183
4.2.3.1.1 Fragments 183
4.2.3.1.2 Isomers/conformers 183

4.2.3.1.3 Mechanisms and pH ... 184
4.2.4 Reaction in a neutral environment ... 185
4.2.5 Reaction in an acidic environment ... 190
4.2.6 Reaction in a basic environment ... 193
4.2.7 Isomerisation ... 195
4.2.8 Saturating the 6-coordinate zirconium complexes ... 196
4.2.9 Increasing the temperature to 310.15 K ... 197
4.2.10 Limitations ... 198
4.3 Conclusion ... 200
4.3.1 Future work ... 201
4.4 References ... 202
Chapter 5: Can a DFT model make gallium chemists more efficient? ... 205
5.1 Introduction ... 206
5.1.1 Effect of anthracyclines on cardiomyocytes ... 206
5.1.2 Mitochondrial function ... 206
5.1.3 Imaging mitochondrial function ... 208
5.1.3.1 Gallium complexes ... 209
5.1.4 Computational efforts in finding log P ... 214
5.1.5 Aim of this work ... 218
5.2 Results ... 219
5.2.1 Computational methods ... 219
5.2.2 Evaluating DFT model chemistries ... 219
5.2.3 Success parameters ... 221
5.2.4 Simple linear regression ... 221
5.2.4.1 Criteria 1 – Positive or negative log P determination ... 221
5.2.4.2 Criteria 2 – Quantitative log P determination ... 223
5.2.5 Multivariate regression and neural network log P prediction ... 226
5.3 Discussion ... 228
5.4 Conclusions ... 229

5.4.1 Future work 230
5.5 References 232
Chapter 6: Conclusions and Future work 235
6.1 Conclusions 236
6.2 Future work 240
6.3 Final remarks 242
6.4 References 243
Chapter 7: Experimental Procedures 244
7.1 General considerations 245
7.2 Synthetic chemistry 246
7.2.1 Phosphasalens 246
7.2.1.1 Phosphasalen hydrolysis experiment 249
7.2.2 Sulfonamides 250
7.2.3 Amides 265
7.3 References 277
Chapter 8: Appendix 279
8.1 Compound characterisation data 280
8.2 Experimental data for crystallography data 319
8.3 List of MnBCA ligands 322
8.3.1 References 332
8.4 Optimised geometry coordinates of DFT complexes 333

Abbreviations

ADP	Adenosine diphosphate
AI	Artificial intelligence
ATP	Adenosine triphosphate
BLD	Bond length distribution
Boc	*tert*-Butyloxycarbonyl protecting group
CFSE	Crystal field stabilisation energy
CT	Computed tomography
CyH	Cyclohexyl
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCE-MRI	Dynamic contrast enhanced MRI
DCM	Dichloromethane
DD	Dipole-dipole
DFO	Desferrioxamine B
DFT	Density functional theory
DMF	Dimethyl formamide
Dox	Doxorubicin
DPPM	1,1-Bis(diphenylphosphino)methane
DTPA	Diethylenetriaminepentaacetate
EDTA	Ethylenediaminetetraacetate
EMA	European Medicines Agency
ES	Electrospray
ETC	Electron transport chain
FDA	Food and Drug Administration
GBCA	Gadolinium based contrast agent
GBD	Global Burden of Diseases, Injuries, and Risk Factors Study
GGA	Generalised gradient approximations
HPC	High performance computing
HR-MS	High resolution-mass spectrometry
HSA	Human serum albumin
IEFPCM	Integral equation formalism polarisable continuum model
IS	Inner sphere
k_{ex}	Water exchange
KHMDS	Potassium bis(trimethylsilyl)amide

K_{MnL}	Equilibrium constant in the formation of MnBCAs
KPI	Key performance indicator
LC-MS	Liquid chromatography–mass spectrometry
MALDI	Matrix-assisted laser desorption/ionisation
ML	Machine learning
MnBCA	Manganese-based contrast agent
MO	Molecular orbital
MRI	Magnetic resonance imaging
NMR	Nuclear magnetic resonance
N,N'-CDI	1,1'-Carbonyldiimidazole
OS	Outer sphere
PCM	Polarisable continuum model
PEG	Polyethylene glycol
PES	Potential energy surface
PET	Positron emission tomography
PLA	PEGylated liposomal alendronate
Ph	Phenyl
Py	Pyridyl
QM	Quantum mechanical
RCY	Radiochemical yield
RPM	Revolutions per minute
SAMPL	Statistical assessment of the modelling of proteins and ligands
SCRF	Self-consistent reaction-field
SPE	Single-point energy
SPECT	Single-photon emission computed tomography
SS	Second sphere
SuFEx	Sulfur(VI) fluoride exchange
SVR	Support vector machine regression
THPN	Tetrakis(3-Hydroxy-4-Pyridinone)
US	Ultrasound
XO	Xylenol orange
XRD	X-ray diffraction

Chapter 1

Introduction

1.1 Molecular imaging

Molecular imaging allows clinicians to ‘see’ inside their patients, providing both qualitative and quantitative information of the biological processes under study to make well informed decisions on patients' care. The field spans the breadth of disciplines in science – in a typical use of a molecular imaging technique, scientists and doctors will need to understand biochemical processes and targets, synthesised an imaging agent, engineered the capture of electromagnetic (or sound) waves, processed huge amounts of data, and finally use the insights gained to understand more about the tissue under study.

In vivo molecular imaging can provide benefits over *in vitro* studies by getting a 'full picture' of the interaction of process/disease with the rest of the body i.e. PET pharmacokinetics of cellular processes; real-time imaging allows clinicians to understand the tissue whilst in-use i.e. angiography in dynamic cardiac MRI; and repeat imaging is possible compared to the potential destruction of tissue when analysed *in vitro.*[1–6]

1.1.1 Molecular imaging modalities

Each molecular imaging modality has its own set of strengths and weaknesses in a range of properties to assess a molecular imaging technique, such as spatial resolution, sensitivity, penetration, and cost. Several criteria are used to answer the typical questions set out when deciding on a modality:

- Can the patient tolerate the associated toxicity with the technique?
- Is the resolution of the technique detailed enough to provide clinically useful information?
- Is the technique sensitive enough to image the target under study?
- Do we need to quantify what we are imaging?
- Do we need the technique to be dynamic?
- Is the target at a suitable depth for the desired modality?
- Would complementary multi-modal imaging be useful?

Coupled with this, clinicians will also need to understand the patient and their needs to come to a decision on which type of scan they use to image a disease. The strengths and weaknesses of the available imaging techniques are detailed in Table 1.1.

Table 1.1: Summary of evaluation metrics for non-invasive molecular imaging techniques.[1]

Modality	Spatial resolution	Sensitivity	Depth Penetration	Acquisition time	Cost	Safety
CT	0.5-1.0 mm	-	$> 10^{1}$ m	Minutes	££	Ionising radiation
MRI	~1 mm	10^{-3} M	$> 10^{1}$ m	Minutes to hours	£££	Relatively high dosage, no ionising radiation
PET	5+ mm	10^{-12} M	$> 10^{1}$ m	Seconds to minutes	£££	Ionising radiation
SPECT	8+ mm	10^{-11} M	$> 10^{1}$ m	Seconds to minutes	££	Ionising radiation
Ultrasound	Depth dependent, 0.01 mm to 1+ mm.	-	10^{-2} m	Seconds	£	Low hazard
Optical fluorescence	2+ mm	10^{-12} M	10^{-2} m	Seconds to minutes	£	Low hazard

1.1.1.1 Magnetic Resonance Imaging

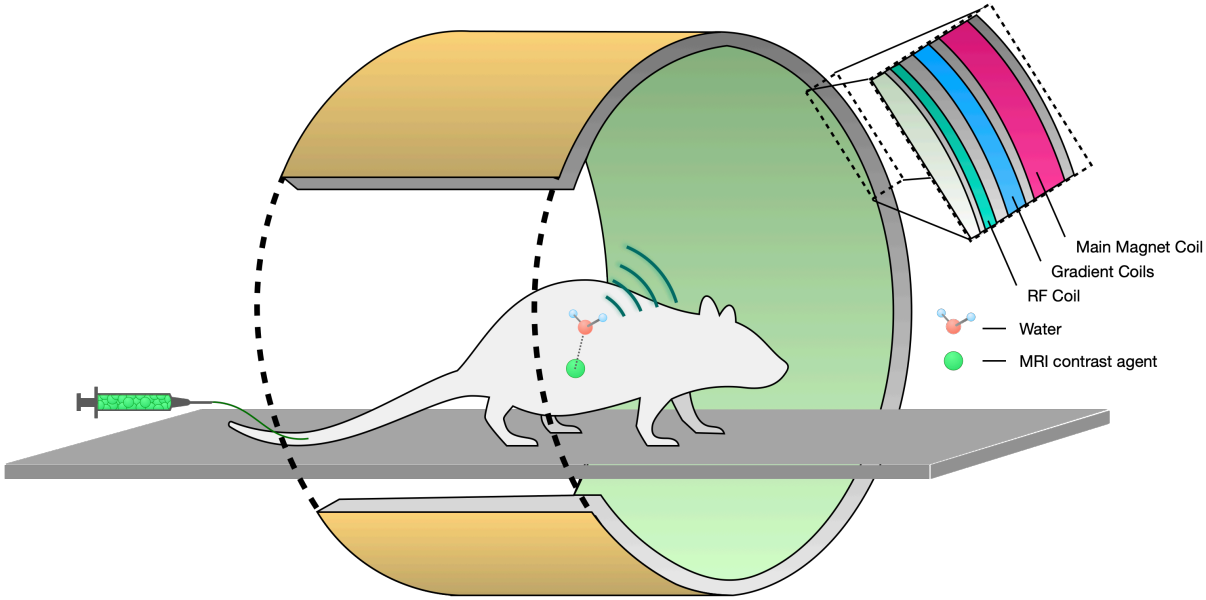


Figure 1.1: Schematic showing the basic make-up of a small animal MRI scanner: a main coil (pink) produces a homogenous magnetic field; gradient coils (blue) produce deviations in the magnetic field to localise the source of the signal in the X, Y, Z coordinate; a radio frequency (RF) coil (green) acts to knock the magnetic dipoles of the nuclei out of alignment with the main magnetic field. The signal generated from water nuclei from a test subject can be improved (contrast enhanced) with the use of an MRI contrast agent.

Magnetic resonance imaging (MRI) works with identical principles to nuclear magnetic resonance spectroscopy (NMR), but specifically measuring the response from hydrogen in water molecules. Nuclei produce magnetic fields (their magnetic moment/dipole) associated with their angular momentum (by g-factor, and by extension, gyromagnetic ratio). With the combination of an externally applied magnetic field and a radio frequency (RF) pulse, these magnetic moments can be flipped perpendicular to the field, and the resultant relaxation back to being parallel can be measured.[1]

The magnetic fields required to excite and measure these magnetic moment shifts are produced from a series of supercooled coils – as is the case in NMR, Figure 1.1. The main magnet coil produces the dominant applied magnetic field, the gradient coils alter the magnetic field to detect the signal in 3D space, and the RF coil provides an RF pulse to flip nuclei magnetic moments perpendicular to the main field, as well as detecting their RF signals upon relaxation. Compared to positron emission tomography (PET) or single-photon emission computed tomography (SPECT), MRI produces very good image resolution (accurate to ~1 mm) due to the limited distance water can move between the RF coil pulse and subsequent detection.

When the magnetic field is applied, the spins of the nuclei can either align with the field (low energy state), or against the field (higher energy state). A system with a non-0 K temperature, will have spins occupying both states – Figure 1.2 – with a distribution following the Boltzmann distribution, Equation 1.2.[5]

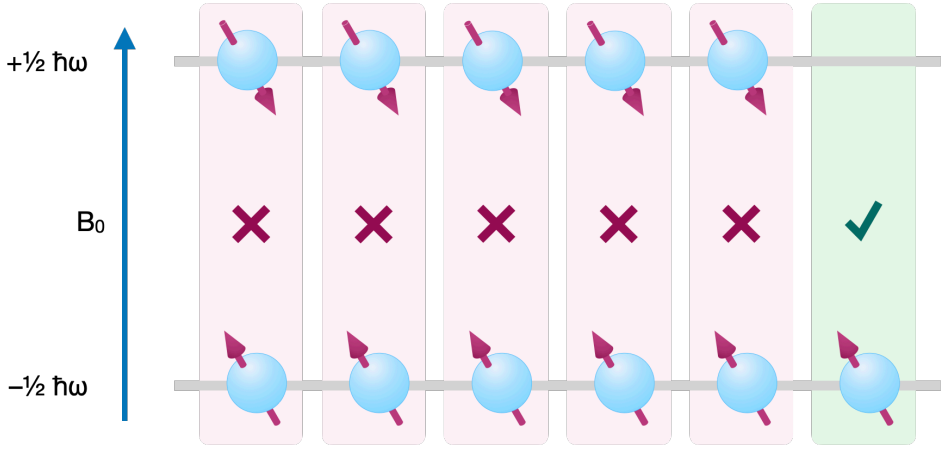


Figure 1.2: Diagram of the alignment of spins with or against an applied magnetic field (left most blue arrow). In a sample at thermal equilibrium both aligned and opposite spin aligned nuclei will be populated, with a very small surplus in the lower energy state (MRI active green box).

All of the higher energy anti-parallel spins will have their magnetic moments cancelled by an equal number of lower energy parallel spin magnetic moments. Only the resulting nuclei at the lower energy without a 'partner' will produce a measurable signal. This is known as the *polarisation* and is represented by Equation 1.1.

$$P = \frac{N_L - N_H}{N_L + N_H}$$

$$P = \frac{\gamma \hbar B_0}{2k_b T}$$

Equation 1.1: Polarisation equations, where γ is the gyromagnetic ratio, ħ is the reduced Planck's constant, and B_0 is the magnetic field strength.

$$\frac{N_H}{N_L} = e^{-\frac{\Delta E}{k_b T}}$$

Equation 1.2: Ratio of two populations via Boltzmann distribution, where ΔE is the energy difference between lower and higher states, k_b is Boltzmann constant, T is temperature.

For a clinical 1.5 T scanner, scanning hydrogens at 37.5 °C, provides a polarisation of only 4.93×10^{-6}, or that only ~5 ^{1}Hs per million will produce a signal (Equation 1.2). This means that very little signal per given tissue volume is produced compared to alternative techniques, and hence, the sensitivity is

low. To address this, contrast agents can be used that shorten the water relaxivity and provide images with increased in contrast. These are typically gadolinium-based contrast agents (GBCAs), or less frequently super paramagnetic iron oxide nanoparticle based (SPIONs).[7,8]

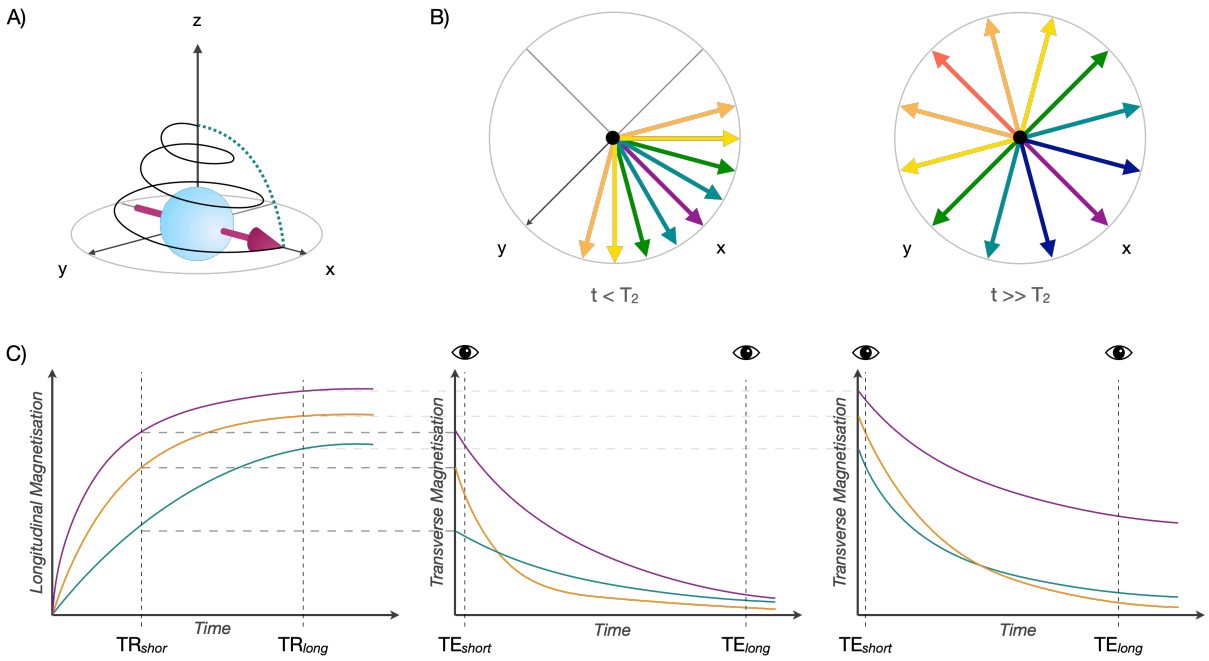


Figure 1.3: Diagrams explaining longitudinal and transverse relaxation in MRI. A) T_1 *(longitudinal) relaxation measures the rate of the population to return to being aligned with the magnetic field in the z-axis (precession is indicated with a solid black line). B)* T_2 *(transverse) relaxation measures the rate at which the population's precessions fall out of phase with each other. C) Both of these effects are tissue dependent (each line is from a different tissue), and can be exploited by altering the repetition and echo times (TR and TE), left: the net longitudinal magnetisation against time for three different tissues, centre: decay of transverse magnetisation at short TR, right: decay of transverse magnetisation at long TR.*[5]

Figure 1.3 details how the return to alignment after the RF pulse knocks the magnetic moment transverse to B_0 can be exploited for MRI images. After the main magnetic field (B_0, in this case along the z-axis) is applied, a perpendicular RF pulse (B_1) knocks the magnetic moment into the transverse (x-y plane) Figure 1.3A. The magnetic moment then precesses back to being aligned with the main magnetic field, and back to M_0 in the z-axis. The precession rate is dictated by the Larmor frequency, and the rate of the z-component of this precession returning to M_0 is T_1 relaxation.[5]

After the RF pulse has knocked all of the ^{1}H magnetic dipoles into the transverse plane, imperfections in B_0 and the effect of neighbouring ^{1}H magnetic fields cause the precession frequencies to fall out of phase with each other, Figure 1.3B. This relationship also follows an exponential decay relationship defined by T_2. Figure 1.3B (left) – at time < T_2, a signal is produced along the x-axis as significant decay has not occurred; Figure 1.3B (right) – at time >> T_2, the precessions are out of phase with each other reducing the transverse component of the signal to 0.

There are two types of relaxation of the magnetic moments – longitudinal (magnetic moment relaxing back to parallel, or T_1) and transverse (the decay of all of the precession being in phase with each other, or T_2). The intrinsic properties of each type of tissue provides different relaxation times.

If, after the first RF pulse, we apply another pulse at a repetition time (TR) and then measure signal at echo time (TE), we can produce signals that will exploit the different relaxation times for each tissue. This is the basis of T_1 (TR_{short} + TE_{short}) and T_2 (TR_{long} + TE_{long}) weighted images, Figure 1.3C (centre and right). In T_1 weighted images, tissues with a fast T_1 relaxation show up as brighter on the image – Figure 1.3C (left, purple line) – and in T_2 weighted images, tissues with a fast T_2 relaxation produce darkening on the image Figure 1.3C (right, orange line). TR_{short} + TE_{long} produces essentially no signal, and TR_{long} + TE_{short} shows the inherent difference each tissue has in producing M_0 – related to factors seen in polarisation, Table 1.2.

Table 1.2: Brain tissue MRI parameters at 1.5 T.[5]

Tissue	M_0 (arbitrary units)	T_1/ms	T_2/ms
White matter	80	700	70
Gray matter	85	950	95
Cerebrospinal fluid	100	2500	250

Requiring exogenous contrast agents can be problematic, especially given the relatively large amounts (~10^{-3} M, ~2-3 g for an average adult) of gadolinium-based contrast agents that can be required to provide adequate signal. As with contrast CT, injecting patients with non-biogenic chemicals in appreciable amounts can lead to allergy concerns, as well as GBCAs being linked to insoluble plaques in the brain, and a suspected cause of nephrogenic systemic fibrosis (NSF) for those with chronic kidney disease (CKD).[8–11]

Another consideration is that the acquisition times for MRI can take longer than an hour, far longer than the other techniques, due to the low sensitivity and relatively complex procedure required to provide a signal to the MRI detector.

1.1.1.2 Positron Emission Tomography

Positron emission tomography (PET) uses radioisotopes that emit positrons (β^+) upon decay. These positrons travel a distance through tissue (the positron's initial kinetic energy – an intrinsic property of the radioisotope decay – is proportional to the distance travelled), until they annihilate with an electron (β^-).[12,13] This process converts all of the mass of the positron-electron pair (0.511 MeV each) into two photons with equivalent energy (hence gamma rays), that (if initial momentum was 0) are emitted at

180° to each other.[14] Gamma detectors are arranged in a loop around the patient to detect the gamma rays, and *via* corrections for variation from 180°, attenuation, Compton scattering, and detector penetration.[13] these events are reconstructed to show where the annihilations occurred in 3D space – Figure 1.4. Because PET produces gamma rays, the technique has an essentially limitless depth of penetration in biological tissue.

Clinically, the most used PET tracer is [^{18}F]-2-fluoro-2-deoxy-glucose ([^{18}F]FDG), whose popularity arises from its biodistribution. As an ^{18}F analogue of glucose,[15] tracer uptake is largest in areas of the body whose increased metabolism requires the most glucose – crucial in locating tumorous cells which will be aggressively dividing compared to non-cancerous cells. The ability to tag a targeting agent with a PET tracer – either organic-based such as^{18}F and ^{11}C, or inorganic-based e.g. ^{68}Ga and ^{64}Cu – makes PET a very versatile technique imaging a vast number of diseases and biological processes.

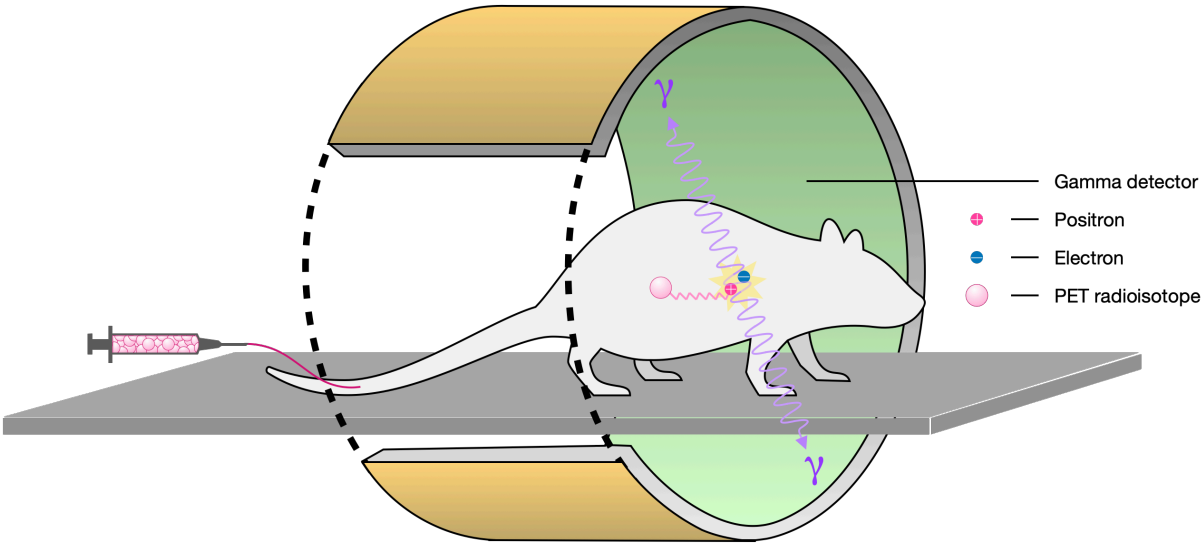


Figure 1.4: Schematic showing the basic make-up of a small animal PET scanner: a PET radioisotope is introduced to the test subject, and after radioactive decay, a positron is produced which travels away from the decayed nucleus until it meets an electron and annihilates. This annihilation produces two 0.511 MeV photons which leave the annihilation at ~180° from each other and are detected by the gamma detector surrounding the test subject.

The main disadvantage with PET is the ionising radiation produced, as beyond the expected two gammas, radioisotopes can often decay *via* alternative routes which will have their own higher energy prompt gammas that can both increase the radioactive dose and can also contaminate the signal received by the detectors.[13,15]

The radiation produced limits the amount of scans a single patient can receive, and also impacts the clinicians producing and delivering the PET isotope. Additionally, the images a PET scan produces only

show where the annihilations occurred in 3D space. As such, PET scans lack reference to the actual body, which means that they often have to be overlayed onto other anatomical imaging techniques like CT or MRI. This increases the cost of the technique, and in the case of PET-CT increases the radiation dose to the patient.[16]

Compounding this, the location of the annihilations (which is what the detectors trace back to) is not where the positron emission occurred. This lowers the resolution as every annihilation event will have a sphere of where the decay *could* have happened. The radius of this sphere will depend on how much kinetic energy the positron has after the radioisotope's decay – for example: ^{19}F (0.64 MeV) travels around 0.5 mm, ^{15}O (1.72 MeV) 2.5 mm, and ^{68}Ga (1.90 MeV) 2.8 mm.[14,15] Furthermore, every PET isotope will always produce gamma rays of 0.511 MeV, whereas other modalities such as optical fluorescence or SPECT can produce photons of different energies allowing multiplexing – imaging multiple biomarkers of interest in concert providing a more complete representation of the tissue.

1.1.1.3 Other imaging techniques

1.1.1.3.1 Single-Photon Emission Computed Tomography

Single-photon emission computed tomography (SPECT) shares a lot of theoretical principles with PET. A SPECT radioisotope, such as ^{99m}Tc or ^{123}I, is administered and undergoes decay to produce a single gamma ray (whose energy is dependent on the radioisotope – and the decay pathway). This gamma ray can then be detected by a rotating gamma camera equipped with a collimator that removes all gammas that are not perpendicular to the camera. These signals are then computationally reconstructed to provide 2D slices.[1] Removing all non-perpendicular gamma rays reduces the sensitivity compared to PET as most decay events will contribute to patient dose but not be recorded.

As mentioned previously, as the gamma rays produced by SPECT isotopes have different energies, multiple SPECT radioisotopes can be simultaneously detected, allowing imaging of different processes/targets at the same time. However, SPECT shares similar limitations to PET as the technique uses ionising radiation that requires special consideration for each patient and has a lower resolution than PET as the radiation from a single decay does not provide information of where the decay occurred.

1.1.1.3.2 X-ray/Computed Tomography

X-ray and computed tomography (CT) scans pass X-ray radiation through tissue and detect what remains, taking advantage of differing levels of attenuation depending on density and chemical composition of the tissues under study. Strongly absorbing tissue like bone – rich in calcium (attenuation is proportional to atomic number) – allow for fewer X-rays to reach the detectors. Classic X-ray scans are 2-dimensional flat images, whereas CT scans capture images in slices which rotate about the body to produce 3-dimensional images of the subject being scanned after computational processing.

To increase the spatial resolution and contrast, CT scans will typically use an iodinated contrast agent which are generally 1,3,5-triiodobenzene derivatives. As per Table 1.1, CT scans are cheaper than PET or MRI, image at all depths, and provide excellent spatial resolution. The main disadvantage, however, is in the nature of how the images are produced. CT scans require ionising X-ray radiation to be absorbed across all of the tissue that is imaged in the slice in order to get any useful information.[16] The intentional irradiation of the patient limits the technique both in terms of who can be scanned – those that cannot (or should not) be subjected to radiation, i.e., children and pregnant people – and also how many times a patient can receive repeat CT scans – which limits functional or disease progression tracking capability. However, as with all of the techniques in Table 1.1, developments are being made to address their limitations – X-ray detectors are being engineered to be more sensitive, allowing for lower doses for the same level of information.[1]

1.1.1.3.3 Ultrasound

Medical ultrasound (US) scanning takes advantage of the different speeds that high-frequency sound waves travel through different tissues, and their reflection/echo on return. As part of the technique, a transducer sends timed high-frequency wave pulses into a body. These waves travel through tissues at different speeds, largely dependent on tissue density, and reflect when they meet tissue interfaces, then these reflections are detected by the transducer and converted into a 2D image.[1] US depth penetration is a balance between frequency and resolution. As the frequency increase, the wavelength shortens and so interacts with tissue more – providing better images at the cost of depth penetration. The technique is *far more* affordable than the previous techniques, and the equipment is smaller and more portable.

1.2 Dynamic Contrast Enhanced MRI

In dynamic contrast enhanced MRI (DCE-MRI), contrast agents are used to shorten the T_1 and T_2 relaxation times for protons in their local environment. This provides a greater contrast between regions that have contrast agents in them, and those that do not – for example, MR angiography uses "blood pool agents" which, due to their increased size and affinity to blood serum albumin,[17] stay in the vascular system far longer than lower molecular weight contrast agents. This allows clinicians to image blood vessels with greater contrast than surrounding tissue.

MRI contrast agents can be pooled according to their mode of action – T_1 shortening, T_2 shortening, and hyperpolarised agents:

- T_1 agents contain paramagnetic elements such as gadolinium or manganese, and the magnetic fields caused by these elements cause the T_1 relaxation times to quicken, returning the longitudinal magnetisation to M_0 faster producing brighter signal in T_1 weighted images.

- T_2 agents such as superparamagnetic iron oxide nanoparticles affect the B_0 in the vicinity of the nanoparticle causing T_2^* (combining T_2 and T_{2i} from field inhomogeneity) to reduce, dephasing the ^{1}H nuclei precessions faster, reducing the transverse magnetisation signal, which produces darker images in T_2 weighted images.

- Hyperpolarised agents work by addressing the polarisation problem in Section 1.1.1.1 which contribute to the low sensitivity of MRI. These agents can increase the signal to noise ratio (SNR) of conventional MRI by a factor of >10,000.[1,18] This allows the incorporation of non-^{1}H nuclei into the agent such as hyperpolarised ^{13}C – imageable endogenous ^{13}C exists in concentrations so low as to produce practically no background for hyperpolarised MRI. After hyperpolarisation the nuclei will rapidly return to thermal equilibrium (within 5 T_1),[1] thus hyperpolarised agents have applications in imaging very fast processes such as glycolysis of ^{13}C-pyruvate, or in vascular and perfusion imaging.

The focus of the work in this thesis is on contrast agents that can reduce T_1. For this discussion these relaxation times will be referred to as T_i (i = 1,2) and specified where necessary.

The concept of relaxivity stems from the ability of a contrast agent (a paramagnetic contrast agent in the case of T_1) to affect the field experienced by ^{1}Hs to make their T_is shorter, and hence relaxation rate faster. This effect can be visualised for GBCAs in Figure 1.5.

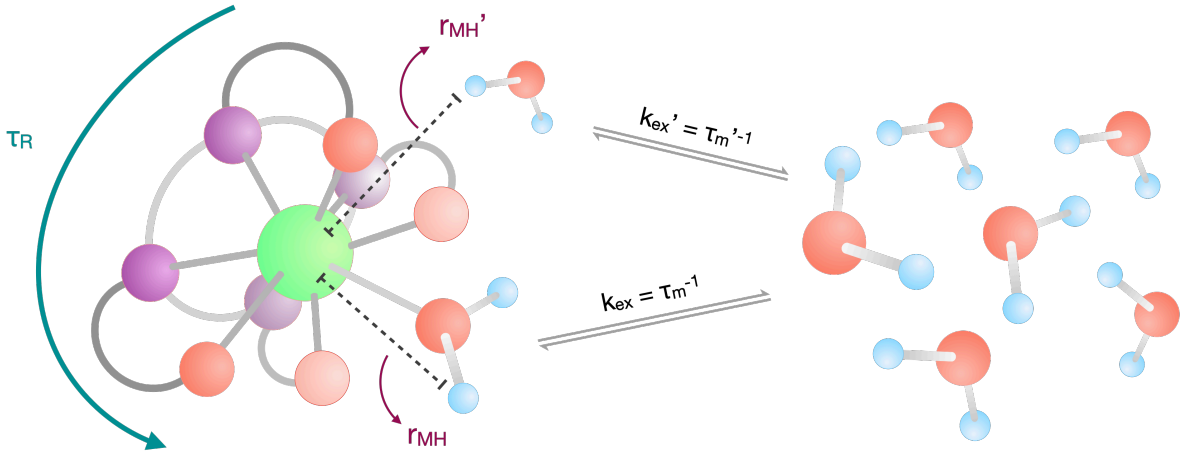


Figure 1.5: Schematic of a GBCA in solution, with factors that affect relaxation rates indicated. Directly bound to the gadolinium core is a water molecule within the inner sphere (IS) at Gd-H length r_{MH}, that can exchange with bulk water at rate τ_m^{-1}. Loosely bound to the GBCA in the second sphere (SS), water here has its hydrogens at Gd-H length r_{MH}', and these water molecules exchange with bulk water at rate $\tau_M'^{-1}$. In addition to this, the rotation time (τ_R) of the GBCA affects the relaxivity, with slower rotation increasing r_1.

Contrast agents take advantage of the massive difference in gyromagnetic ratios (the ratio of magnetic moment over spin) between a proton's nucleus and that of an unpaired electron (γ_H is 4.26×10^1 MHz/T; γ_e is -2.8×10^4 MHz/T; a magnitude 658 times larger).[5] This affects the field experienced by the contrast enhanced proton, drastically reducing T_1 and providing greater signal.

The relaxivity of a contrast enhancement (r_i) is a measure of the change in relaxation rate (R_i which is the inverse of relaxation time (T_i)) normalised by dividing by its concentration, Equation 1.3.

$$r_i = \frac{\Delta \frac{1}{T_i}}{[CA]}$$

Equation 1.3: Contrast enhancement is the change in the relaxation rate normalised by contrast agent concentration.

The observed T_i is a summation of the T_is from the inner sphere, second sphere and the outer sphere, Equation 1.4.

$$r_i^{obs} = r_i^{IS} + r_i^{SS} + r_i^{OS}$$

Equation 1.4: Observed relaxivity is composed of contributions from different water populations.

Relaxation of water bound to a paramagnetic species has been described by equations from McConnell applied to MRI by Swift and Connick, Equation 1.5.[19–21]

$$r_i^{IS} = \frac{q}{[H_2O]} \cdot \frac{1}{T_{im} + \tau_m}$$

Equation 1.5, where T_{im} is the relaxation time of bound water, τ_m is the length of time water is coordinated, q *is the hydration number.*

The second sphere essentially has the same equations but with modified q', T_{im}', and τ_m'. The contribution from these second sphere and outer sphere T_1s are largely independent of contrast agent structure, as seen by the uniform distribution of r_i values seen for $q = 0$ species. Delving deeper into T_{im}, this relaxation is governed by a set of equations by Solomon, Bloembergen, Morgan, and Gueron,[22–24] Equation 1.6-Equation 1.13. For T_{1m} only the dipole-dipole interactions appreciably contribute at fields greater than 1.5 T.[21]

$$\frac{1}{T_{im}} = \frac{1}{T_i^{DD}} + \frac{1}{T_i^{SC}} + \frac{1}{T_i^{CS}}$$

Equation 1.6, where T_{im} is the relaxation time of metal bound water; DD is the dipole-dipole interaction; SC is the scalar relaxation; CS is the Curie Spin relaxation.

$$\frac{1}{T_1^{DD}} = \frac{2}{15} \cdot \left(\frac{\mu_0}{4\pi}\right) \cdot \frac{\gamma_H^2 g_e^2 \mu_B^2 S(S+1)}{r_{MH}^6} \left[\frac{3\tau_C}{1 + \omega_H^2 \tau_C^2}\right]$$

$$\frac{1}{T_2^{DD}} = \frac{1}{15} \cdot \left(\frac{\mu_0}{4\pi}\right) \cdot \frac{\gamma_H^2 g_e^2 \mu_B^2 S(S+1)}{r_{MH}^6} \left[4\tau_C + \frac{3\tau_C}{1 + \omega_H^2 \tau_C^2}\right]$$

$$\frac{1}{T_2^{SC}} = \frac{1}{3} \cdot \left(\frac{A}{\hbar}\right) \cdot S(S+1)[\tau_{SC}]$$

$$\frac{1}{T_2^{CS}} = \frac{1}{5} \cdot \left(\frac{\mu_0}{4\pi}\right)^2 \cdot \frac{\omega_H^2 g_e^4 \mu_B^4 S^2 (S+1)^2}{(3k_b T)^2 r_{MH}^6} [4\tau_{CS}]$$

Equations 1.7, 1.8, 1.9, and 1.10, where μ_0 is the permittivity of vacuum; γ_H is the proton gyromagnetic ratio; g_e is the electronic g-factor; μ_B is the Bohr magneton constant; S is the spin quantum number; r_{MH} is the distance between metal and hydrogen; ω_H is the proton Larmor frequency, $A/\hbar$ is the hyperfine coupling constant between the metal and the hydrogen nucleus, k_b is Boltzmann constant.

$$\frac{1}{\tau_C} = \frac{1}{\tau_R} + \frac{1}{\tau_m} + \frac{1}{T_{1e}}$$

$$\frac{1}{\tau_{SC}} = \frac{1}{\tau_m} + \frac{1}{T_{1e}}$$

$$\frac{1}{\tau_{CS}} = \frac{1}{\tau_R} + \frac{1}{\tau_m}$$

Equations 1.11, 1.12, and 1.13, describing the three relaxation mechanisms that contribute to paramagnetic relaxation, where $1/T_{1e}$ is the relaxation rate of the unpaired electrons.

What can be gleaned from these equations and Figure 1.5, is that from a synthetic chemist point-of-view, we can essentially exert control over only a few parameters – τ_R, τ_m, q, and the number of unpaired electrons (Gd^{3+} vs Mn^{2+} etc.). τ_R has also been defined, Equation 1.14, in which τ_R is a function of molecule volume (approximated by a sphere) and solution viscosity.[25] More intuitively, τ_R can be thought of as a function of molecular weight (the heavier the molecule, the larger the molecule) – affording another parameter to influence relaxivity.

$$\tau_R = \frac{4\pi a^3 \eta}{3k_b T}$$

Equation 1.14, where a is the radius of the molecule approximated by a sphere, η is solution viscosity, k_b is the Boltzmann constant, and T is temperature.

1.3 Gadolinium Based Contrast Agents

Paramagnetic T_1 shortening contrast agents rely on their unpaired electrons at their core to produce magnetic fields that greatly facilitate the relaxation rate (both T_1 and T_2) of water molecules both directly and indirectly attached. Therefore, a good paramagnetic element should be able to both accommodate the loss of denticity to coordinate water *and* have a stable paramagnetic oxidation state. For the first criterion, larger species (naturally having greater coordination numbers) are suitable. Gadolinium – an f-block element with empirical atomic radii of 180 pm[26] – has a stable 3+ oxidation state in which it is [Xe] $4f^7$, with each of these 4f electrons being unpaired.[27] This makes Gd^{3+} very paramagnetic (μ = 7.94 BM) and has cemented its position as the element of choice for MRI contrast agents in the last three and a half decades.[28] Gadolinium-based contrast agents (GBCAs) that are approved for use in Europe are shown in Figure 1.6.

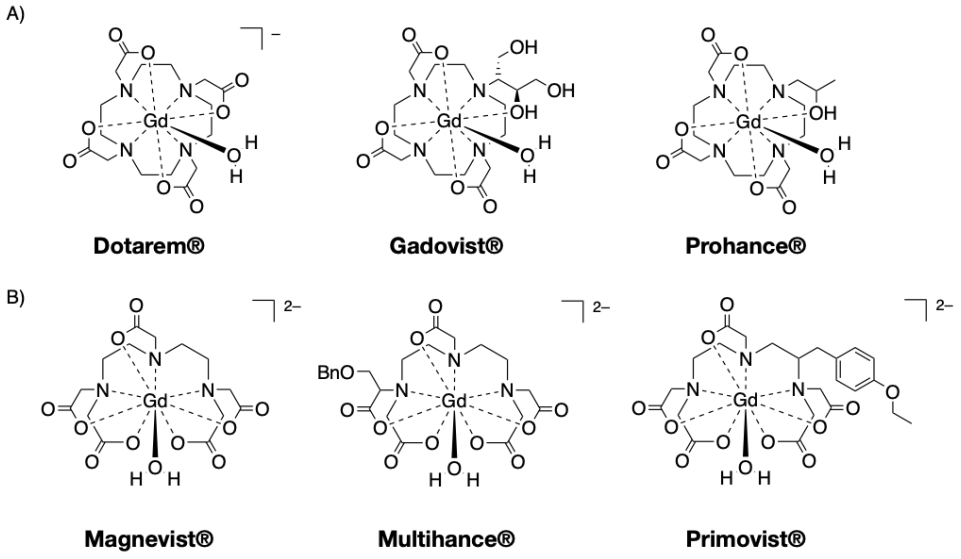


Figure 1.6: A) Macrocyclic GBCAs – left: Dotarem® (gadoteric acid); centre: Gadovist® (gadobutrol); right: Prohance® (gadoteridol); B) Acyclic GBCAs – left: Magnevist® (gadopentetic acid), centre: Multihance® (gadobenic acid), right: Primovist® (gadoxetic acid).

As tissues typically provide enough contrast difference between each other, DCE-MRI is typically used for tumour visualisation, inflammation, and angiography.[1,5] Quantitative MRI techniques were used to judge the health of my heart in March 2022 at London Bridge Hospital after experiencing irregular heartbeats after a long-COVID relapse, Figure 1.7.[29] At the start of the scan, contrast-free MRI was measured – T_1 images looked for heart muscle scarring (brighter lines around the heart), and T_2 for potential fluid build-up from oedema (Figure 1.7, top); Gadovist® was intravenously introduced (Figure 1.7, bottom), which allowed for early- (looking for clot formations and any microvascular obstructions) and late-gadolinium enhancement imaging (leveraging the different washout times for

each tissue to look for scar formation, but at a higher resolution than before). The thin white line around the late-gadolinium enhancement image showed a healthy thin layer of fat surrounding my heart, as opposed to potential scarring from myocarditis, which had been associated with COVID patients at the time, and for young males after mRNA vaccination.[30–33]

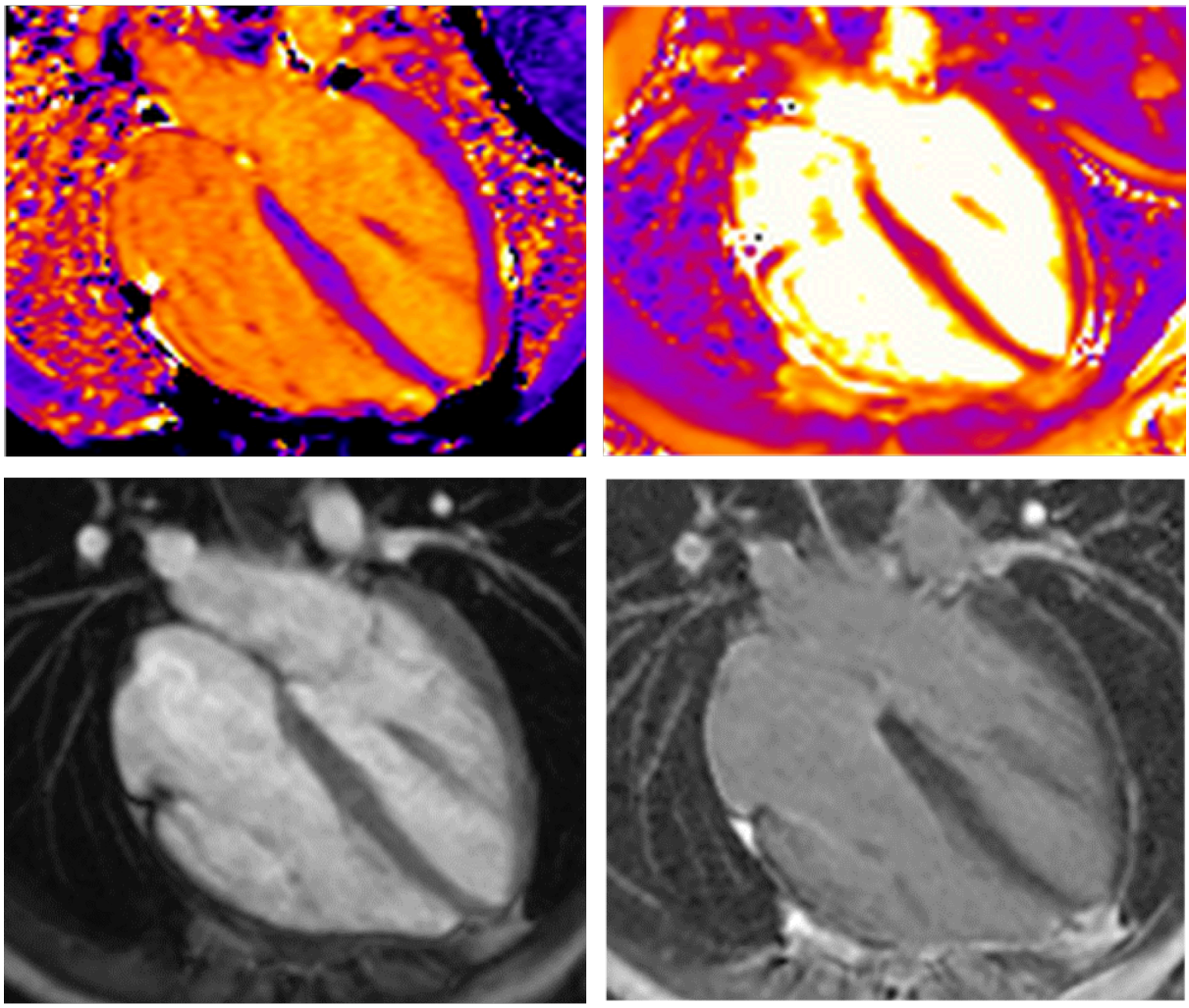


Figure 1.7: Dynamic contrast enhance MRI (DCE-MRI) scan performed on my heart in March 2022 at London Bridge Hospital. Top left: quantitative T_1 *mapping; top right: quantitative* T_2 *mapping; bottom left: early gadolinium enhancement imaging, bottom right: late gadolinium enhancement imaging.*[29]

The over-whelming majority of dynamic contrast-enhanced MRI (DCE-MRI) scans result in total and safe elimination of contrast agent, with over 500 million GBCA enhanced MRI scans performed since 1988, and approximately a third of all MRI scans are performed with GBCAs.[34]

1.3.1 Nephrogenic Systemic Fibrosis

However, from 2006, reports from the clinical imaging community raised concerns over a correlation between GBCAs and nephrogenic systemic fibrosis (NSF) in those with chronic kidney disease (CKD), with free gadolinium being detected in the biopsies of these patients.[35] NSF is a disease characterised by the thickening and scarring of connective tissue (fibrosis) occurring across several organs (systemic) – first noted on the skin, but also seen in joints, eyes, and internal organs – and typically seen in those with reduced kidney function (nephrogenic).[35] CKD is staged by a patient's estimated glomerular filtration rate (eGFR), with values between 30-90 mL/min/1.73 m^2 comprising stages 1-3 (mild to very little reduction in kidney function), and values between 15-30 for stage 4 and < 15 for stage 5 (essentially kidney failure).[36]

In 2017, the Lancet's Global Burden of Disease (GBD) report for CKD estimated that the prevalence of CKD world-wide was at 9.1% and increasing. They estimated that CKD mortality was responsible for 4.6% of global deaths, making it the 12th biggest cause of death in the world.[37] Foreman and co-workers then modelled the 250 causes of death from the GBD Study to predict the leading causes of death until 2040. By 2040, CKD is estimated to be the 5th biggest cause of death in the world, particularly prevalent in Central Latin America (1st) and southeast Asia (3rd).[38] As of 2023, Kidney Research UK estimates that the NHS is spending £1.95 billion on CKD (all stages), and an additional £1.05 billion on end stage kidney disease.[39]

In the late 2000s, evidence suggested that DCE-MRI with gadodiamide (Omniscan®) performed on patients with severe CKD was correlated to the presentation of NSF in these patients.[40] The American College of Radiology concluded that the risk for developing NSF was greatest for those with stage 4 and 5 CKD, and acyclic GBCAs (as opposed to macrocyclic ones).[34] Compounding this, in 2014 Kanda and co-workers reported increased signal intensity located in the dentate nucleus and globus pallidus in unenhanced T_1-weighted MRI images for patients that had multiple GBCA contrast enhanced MRI scans, with the signal being positively correlated to the number of prior GBCA administrations.[41] In that report, only Magnevist® and Omniscan® were used (both acyclic GBCAs), none of the patients had NSF, and only 2 of 381 patients had stage 4 or 5 CKD. Supporting studies have shown that this effect is seen in patients with no intracranial abnormalities, with GBCAs that are not associated with the greatest number of NSF cases, and with gadolinium being deposited in other organs as well.[42–46]
This flurry of GBCA research concluded with the European Medicines Agency (EMA) suspending the authorisation of multiple acyclic contrast agents, and the Food and Drug Administration (FDA) providing warnings for their use.[47,48] The presence of free gadolinium in NSF biopsies indicated that decomplexation was an issue for patients that could not efficiently renally eliminate GBCAs. The evidence of acyclic GBCAs being worse than macrocyclic GBCAs further emphasised that this was an issue around thermodynamic stability.

Table 1.3: EMA restrictions on GBCAs and their stability constants.

GBCA	Type	Log(K_{GdL})	EMA Status[47]
Dotarem®	Macrocyclic	25.6	Maintain
Prohance®	Macrocyclic	23.8	Maintain
Gadovist®	Macrocyclic	21.8	Maintain
Primovist®	Acyclic	23.5	Maintain
Multihance®	Acyclic	22.6	Liver scans only
Magnevist®	Acyclic	22.5	Suspend
Omniscan®	Acyclic	16.9	Suspend
Optimark®	Acyclic	16.8	Suspend

The measure of this stability is defined by log(K_{ML}) values of the equilibrium constant of complexing a ligand to a metal, Equation 1.15.

$$Log(K_{ML}) = \frac{[ML]}{[M][L]}$$

Equation 1.15: Equation for the simplified stability constant, with the concentration of water has been removed for simplicity.

The difference between the stability of macrocyclic and acyclic ligands for gadolinium is as a result of the lower reduction in entropy for a pre-ordered system spatially designed for macrocyclic complexation compared to the larger reduction in entropy of a flexible, linear, acyclic ligand contorting itself to complex a metal. This is seen in Table 1.3. Aime and co-workers posited that linear GBCAs disrupt iron homeostasis, interacting with iron/ferroportin macrophages and releasing the iron which takes up the ligand.[45] The poor sensitivity of MRI compounds this issue, as large quantities of GBCA are required for each scan. Gadovist® is administered at 0.1 mmol/kg scale – therefore 7 mmol boluses are required for the ‘average' 70 kg European adult, which corresponds to 4.22 g of contrast agent per scan or around ~1.09 g of gadolinium.[49] Despite the concern of having insoluble gadolinium deposits in tissues throughout the body, no adverse health events have been found as of yet. Nonetheless, the saga highlights the hazard of introducing exogenous metals into a biological system that has not evolved to naturally remove it and highlights the need to optimise log(K_{ML}) values for complexes to minimise decomplexation risk.

1.4 Alternatives to gadolinium

It would be preferable to remove the hazard of gadolinium deposits and associated NSF risk by finding an alternative to gadolinium-based contrast agents. Manganese as high-spin d^5 Mn^{2+} has 5 unpaired electrons, allowing it to provide paramagnetic enhancement necessary for MRI – $\mu_{eff} = 5.92$.[10] Loosely equivalent as the transition metal equivalent of gadolinium, sitting in the middle of its period, with an unpaired electron in its valence orbitals. High-spin Mn^{3+} is also paramagnetic, giving a weaker relaxivity (per hydration number) compared to Mn^{2+}, as it only has four unpaired electrons – $\mu_{eff} = 4.90$. Additionally, as a biogenic element (serum concentration of 0.4-0.85 μg/L; whole body content 12-20 mg),[50] both Mn^{2+} and Mn^{3+} have a safer risk profile than gadolinium, which is not known to naturally exist in the body.[51–53]

Overexposure to free Mn^{2+} however, has been linked to the development of the neurodegenerative condition "manganism", in which the patient presents with Parkinson's-like symptoms.[54] The majority of cases of manganism stem from chronic occupational exposure at high concentrations of manganese in mining industries, cell battery factories and welding industries.[51] Mn^{2+} is known to almost exclusively (>95% excretion pathway) be cleared *via* the liver with trace amounts cleared *via* the kidneys,[51,54] therefore a manganese-based contrast agent (MnBCA) should be suitable for those with eGFRs that preclude them from gadolinium DCE-MRI. Elimination half-lives for manganese are dependent on where it is deposited. For example, free manganese has a 2-hour half-life in the blood, whereas it has a 8-9 year half-life once deposited in bone.[53] T_1-weighted hyper-intense MRI signal has been used to quantifiably track manganese brain deposition upon chronic exposure (smelters and intravenous ephedrone users – $KMnO_4$ is used in its illicit synthesis to oxidise pseudoephedrine). After 6 months of no exposure, global pallidus signal intensity essentially disappeared, indicating a natural pathway to remove brain deposited manganese. Free $MnCl_2$ has an LD_{50} in rats of 0.70–1.20 mmol kg^{-1} when introduced parenterally,[51] and toxic effects in humans can be seen at 5 μmol kg^{-1} when administered intravenously. These effects are thought to occur due to manganese replacing calcium in calcium channels, affecting the heart, nervous system, and brain. Compared to the typical dosage in MRI of 0.05–0.10 mmol kg^{-1},[49] the need to develop manganese-based complexes resistant to decomplexation with efficient excretion remains.

1.4.1 Manganese-Enhanced Magnetic Resonance Imaging

To date, two manganese-enhanced magnetic resonance imaging (MEMRI) agents have been given clinical approval – Teslascan® (mangafodipir) and LumenHance®, with a third agent developed and patented – EVP1001-1. MEMRI takes advantage of free Mn^{2+} ions to access intracellular imaging as it is able to act as a calcium analogue in L-type calcium channels, allowing transport into cells.[55] This gives it the ability to image cardiomyocyte viability, and an advantage over gadolinium-delayed

enhancement MRI which has a lower resolution as it is an extracellular imaging mode. MEMRI has so far been used in animal studies to image the liver, kidneys, heart, pancreas, brain (cytoarchitecture);[56–60] to image cellular activity (cardiac tissue via calcium channels);[61,62] and carry out neuronal tract tracing in specific areas of the brain.[63,64] Mangafodipir was developed as a MnBCA, however *in vivo* results showed that the complex dissociated – highlighting the main difficulty of low stabilities in MnBCA ligand design; with stronger sales in GBCAs, the makers of mangafodipir voluntarily withdrew its EMA approval in 2012.[65]

1.4.2 Manganese-Based Contrast Agents

The metrics used to assess ligands used in GBCAs can be used for manganese-based contrast agents (MnBCAs), and these are $\log(K_{MnL})$ (to assess thermodynamic stability), r_1 (to assess the strength of paramagnetic contrast enhancement), and q (inner shell water coordination). As kinetic instability can be due to multiple mechanisms, there is not a measure that will guarantee proportionality with stability. However, resistance to transmetalation can be quantified by measuring the pseudo-first-order rate constants when the MnBCA is challenged with Zn^{2+} or Cu^{2+} solution under physiological conditions. As a benchmark, Magnevist® (Gd-DTPA: $\log(K_{GdL}) = 22.46$, $r_1 = 3.6\ mM^{-1}\ s^{-1}$, $q = 1$) and Dotarem (Gd-DOTA: $\log(K_{GdL}) = 24.7$, $r_1 = 3.26\ mM^{-1}\ s^{-1}$, $q = 1$) can be used, two of the most widely used GBCAs.[10] Both GBCAs have larger $\log(K_{ML})$s than typically seen for MnBCAs, as Mn^{2+} has a lower charge/radius ratio compared to Gd^{3+} and high-spin Mn^{2+} lacks any crystal field stabilisation. However, it is important to note that GBCAs require higher $\log(K_{ML})$s due to the larger toxicity of free Gd^{3+}.

The difficulty in stably complexing manganese can be rationalised by a combination of a lack of crystal field stabilisation, with high-spin Mn^{2+} having one electron in each of its 5 d-orbitals (leading to no stabilisation); and the large ionic radius of Mn^{2+} leading to weaker bonds with its ligands. As such, Mn^{2+} is the least stable in the Irving-Williams series. This series looks at the stability constants of top-row transition metals when their aqua ligands are displaced with both L and X type ligands: $Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$.[66] Altogether, these factors make unspecialised ligands very labile when bound to manganese, further increasing the difficulty in creating MnBCAs.

The effect can also be seen in $\log(K_{ML})$ values for DOTA or EDTA complexation with a selection of elements typically used in molecular imaging (MRI and PET) or for biogenic elements, Table 1.4.

Table 1.4: Stability constants for a range of molecular imaging metals with EDTA and DOTA.

Element	EDTA[a]	DOTA
Calcium(II)	10.65	17.23[a]
Manganese(II)	13.88	20.0[b]
Iron(II)	14.30	20.2[d]
Nickel(II)	18.4	20.0[b]
Copper(II)	18.78	22.21[a]
Zinc(II)	16.5	21.05[a]
Gallium(III)	21.0	21.3[c]
Indium(III)	24.9	23.9[c]
Gadolinium(III)	17.32	24.67[a]

[a] Smith and Martell[67], [b] Anderegg and co-workers (25 °C, 0.1 M ionic strength)[68], [c] Baranyai and co-workers[69], [d] Byegård and co-workers[70].

Table 1.4 shows the poor performance of manganese (Mn^{2+}) in chelating EDTA compared to other elements of interest, which also has implications for kinetic stability – in the body the competing ions (such as Zn^{2+}, Ca^{2+}, $Fe^{2+/3+}$) will form more stable complexes with most ligands. It perhaps also indicates the ability of Mn^{2+} to outcompete Ca^{2+} in biological processes and hence its toxicity in overexposure.

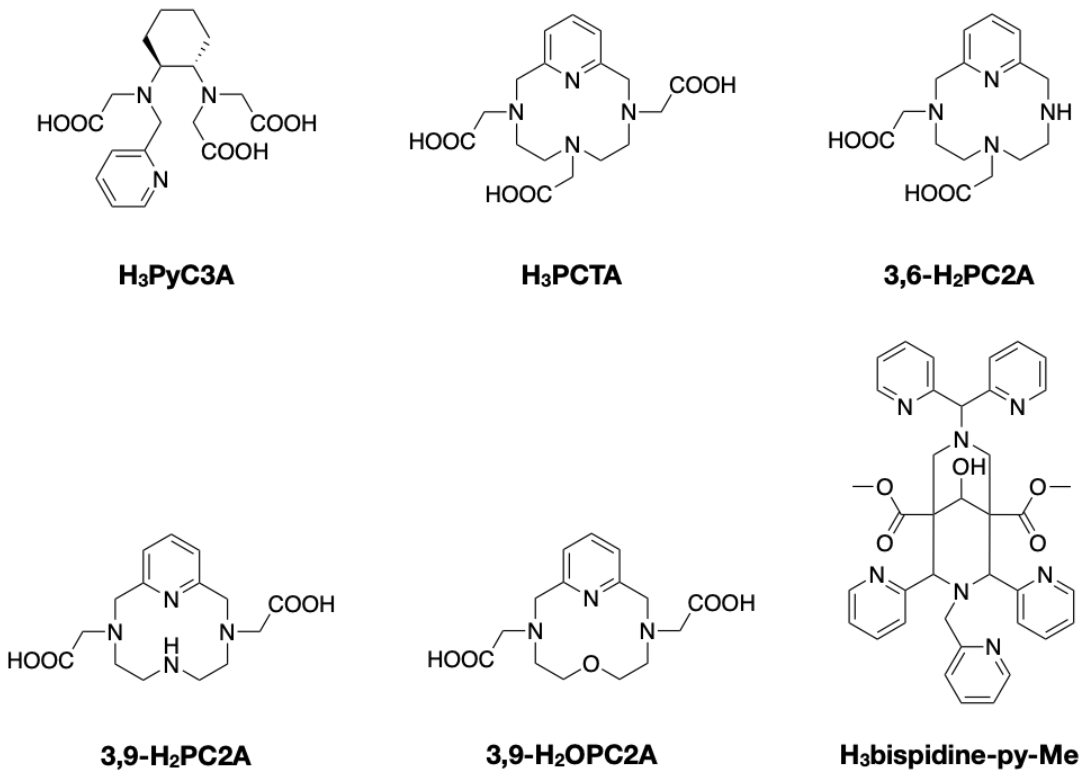


Figure 1.8: Selection of ligands in Mn^{2+} MnBCAs

Examples of MnBCAs in the literature are split into macrocyclic and acyclic ligand families. Of these, acyclic or linear ligands have been most investigated with the greatest interest applied to functionalising ethylenediaminetetraacetic acid (EDTA). However, as with GBCA research macrocyclic ligands have taken over with the pyclen motif particularly fashionable.[10,71–75]

Caravan and co-workers synthesised the PyC3A ligand based on *trans*-cyclohexylene-diamine-tetraacetic acid (CDTA, a cyclohexylene analogue of EDTA), Figure 1.8, top left. This was rationally designed to increase thermodynamic and kinetic stability with a *trans*-cyclohexylene backbone – shown to decrease Mn^{2+} dissociation compared to *trans*-phenylene and ethylene[73] – and a pyridyl ring added to rigidify the chelate, and add lipophilic character – important to drive hepatobiliary clearance, with a view for CKD applications.

Tircsó's research group have published a number of pyclen-based ligands with variations on the substituents on the nitrogen atoms, and swapping nitrogen for oxygen. Tircsó found that fully saturating the nitrogen atoms with acetate arms provided a very stable complex ($\log(K_{MnL})$ of 16.83) but at the cost of relaxivity (25 °C, 20 MHz, r_1: 1.50 mM^{-1} s^{-1}) – equivalent to a $q = 0$ complex, Figure 1.8, top centre.[76] This led to removing an arm (Figure 1.8, top right and bottom left). Tircsó found that 3,9-PC2A was more thermodynamically stable than 3,6-PC2A (17.09 vs. 15.53) and had a higher relaxivity (25 °C, r_1: 2.91 vs 2.71 mM^{-1} s^{-1}), however 3,9-PC2A had a lower half life when faced with a large Cu^{2+} excess (21.0 vs 63.2 hours).[71] In 2022, the group replaced the free amine in 3,9-PC2A with an oxygen instead. Unfortunately, this proved to be a less stable complex ($\log(K_{MnL})$ of 13.03), but provided better contrast (25 °C, r_1: 3.13 mM^{-1} s^{-1}).[77]

Finally, bispidine ligands have recently been shown to form highly kinetically and thermodynamically inert complexes, strongly outperforming the rest of the MnBCA field.[78,79] This is thought to be due to the rigidity of the diazaadamantane-derived structure, as well as the large cavity size being ideal for the large Mn^{2+} ion (effective ionic radius compared to competing endogenous ions - Mn^{2+}: 83 pm, Cu^{2+}: 73 pm, Zn^{2+}: 74 pm; ~13% increase in size).[80] Furthermore, the relaxivities (25 °C, 20 MHz) of these complexes were exceptionally high, 5.04 mM^{-1} s^{-1} and 3.62 mM^{-1} s^{-1} – very large for an octadentate MnBCA such as $[Mn(\text{bispidine-py-Me})(H_2O)]^{2+}$.

Compared to most GBCAs that are nonadentate, Mn^{2+} is usually limited to hexa- or hepta-coordinated complexes – which leads to a smaller chelate effect and reduction in stability. This lower stability is then exacerbated as MRI MnBCAs require intentionally leaving a potential coordination site free for water exchange with surrounding water, essentially tying one arm behind your back when developing stable MnBCAs. Table 1.4 shows that when the coordination sphere is completely filled with donors

(as with DOTA), Mn^{2+} is comparable to other elements like Fe^{2+} and Ni^{2+}, and is highly stable – for comparison, in Table 1.3 the cut-off for EMA GBCA approval is around 20.0.

Additionally, using harder Lewis base donors – a strategy that would provide success with other transition metals and elements – will favour oxidation of the Mn^{2+} to Mn^{3+} which will reduce the effective magnetic moment, μ_{eff}, and reduce the capability of the MnBCA. This has been exploited by Caravan and co-workers to produce a "Janus chelator", capable of switching between ligand donor modes (Figure 1.9); oxidative conditions (*via* peroxidase) convert the Mn^{2+} form into the Mn^{3+} and it can be reduced back *via* L-cysteine.

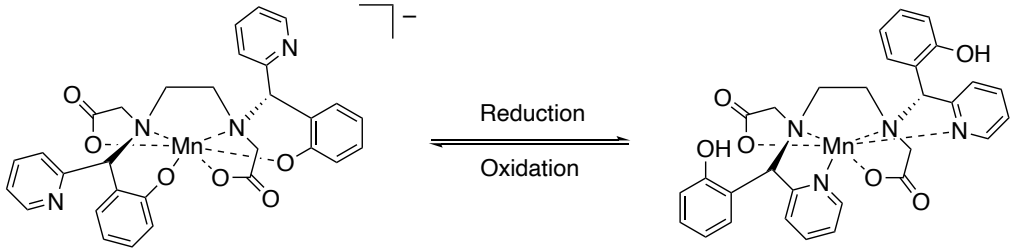


Figure 1.9: Janus chelator designed by Caravan and co-workers.[81]

1.4.2.1 MnBCA design considerations

One major issue when developing MnBCAs is translatability. Its set of design considerations mean that the best ligands for Mn^{2+} complexation will be specialised to meet the challenging stability problem, but not necessarily be good ligands for other systems – reducing the attractiveness of designing MnBCAs.

The primary consideration with respect to safety is to maximise thermodynamic and kinetic stability. This is typically achieved *via* increasing rigidity in the ligand by restricting rotation by using a scaffold that is macrocyclic or contains aromatic portions in its skeleton. In order to provide acceptable contrast ($> 2\ mM^{-1}\ s^{-1}$), the coordination shell cannot be saturated – there needs to be one coordination site left vacant for water exchange. Therefore, a hexadentate ligand is typically required, or in very specialised systems (like bispidine) the ligand can be heptadentate. Additionally, the complex needs to be stable with respect to oxidation, often achieved by including softer donors without compromising thermodynamic stability.

In the context of an *MRI* MnBCA – i.e., to replace GBCAs, the primary focus should be on driving hepatobiliary clearance for patients affected by CKD. To favour this, the ligand should be designed to be more hydrophobic and the complex should be overall be monoanionic.[73] And as with GBCAs, the

larger the ligand, the longer the τ_R. This in turn increases r_1 – a trend that has also been reported by the Caravan group for MnBCAs.[9]

1.4.2.2 Mn^{3+} MnBCAs

Mn^{3+} complexes are not very popular choices for MnBCAs due to the lower effective magnetic moment (5.92 vs 4.90 BM). However, due to the higher thermodynamic stabilities that can be achieved once CFSE is restored, coupled with a lack of threat from oxidation with harder donors, $q = 2$ complexes are viable MnBCAs.

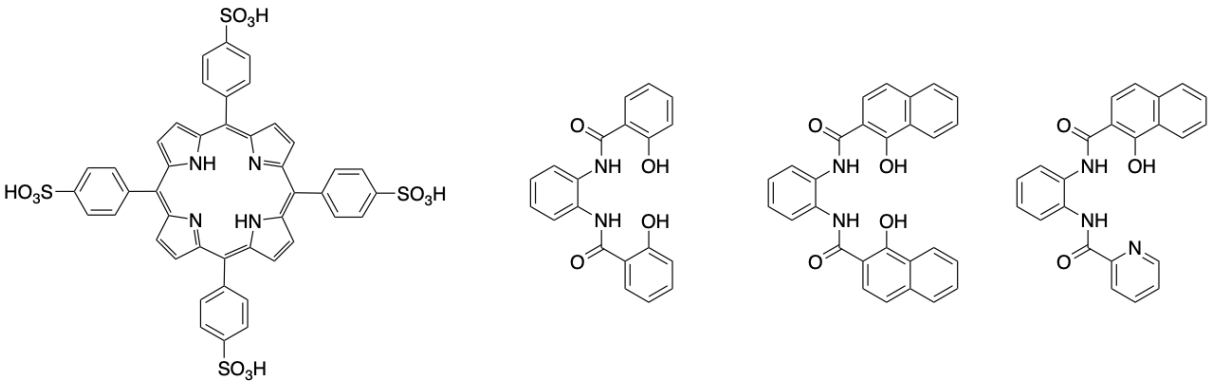


Figure 1.10: Selection of ligands in Mn^{3+} MnBCAs.

Porphyrins have been shown to provide large relaxivity values ($[Mn(TPPS_4)(H_2O)_2]^{3-}$ r_1: 10.4 $mM^{-1}s^{-1}$), atypical of Mn^{3+} based contrast agents, Figure 1.10.[82] Koenig and co-workers have attributed this 'anomalously high' r_1 to labile water exchange, $q = 2$, slow rotational tumbling, and anisotropy from the high-spin d^4 configuration leading to the elongation of the $3d_{z^2}$ orbital which causes better wavefunction overlap with the ^1H from the coordinated water (shortening the r_{MH} – which has a r^{-6} relationship with contrast enhancement).[83] Log(K_{ML}) values are difficult to achieve in Mn^{3+} complexes *via* standard pH potentiometry due to a willingness to disproportionate to Mn^{2+} and Mn^{4+}, however challenge with competing ligands can indicate thermodynamic stability.[9] Due to a favourable size match between Mn^{3+} and the porphyrin pocket, complexes like $[Mn(TPPS_4)]^{3-}$ are both thermodynamically and kinetically stable.[9] This is due to the strong overlap in orbital wavefunctions from the porphyrin to the Mn^{3+} centre.[11] Ligands based on the phenylenediamino backbone have been developed by the Jasanoff group, Figure 1.10. These Mn^{3+} MnBCAs show large r_1 values (r_1: 4.1, 5.4, and 5.1 $mM^{-1}s^{-1}$ for $[Mn(PDA\text{-}1)(H_2O)_2]^-$, $[Mn(PDA\text{-}2)(H_2O)_2]^-$, $[Mn(PDA\text{-}3)(H_2O)_2]$ respectively) and resistance to challenge with EDTA over 48 hours – indicating thermodynamic and kinetic stability.[84]

The difference between Mn^{2+} and Mn^{3+} thermodynamic stability is clear when you compare the ligands that must be used to bind them. Stable Mn^{3+} ligands can be tetradentate with hard donors such as phenolates – tetradentate ligands for Mn^{2+} complexes have log(K_{MnL}) values below 10 log units (rendering them useless), and hard donors will cause oxidation – seen in the experimental for the synthesis of Jasanoff's PDA ligands.[75,84] The challenge in designing Mn^{2+} ligands is to make a highly specific ligand that requires intentionally reducing its already inherently low thermodynamic stability to allow water to freely exchange with bulk solution.

1.5 PET and dual-modal imaging

1.5.1 PET imaging

Clinically, the most used PET tracer is [^{18}F]-2-fluoro-2-deoxy-glucose ([^{18}F]FDG), which is an ^{18}F analogue of glucose. This means that the tracer will be present in areas of the body whose increased metabolism requires the most glucose to run – crucial in locating tumorous cells which will be aggressively dividing compared to non-cancerous cells. The most common PET radiometal used is ^{68}Ga which is able to be produced *via* generator from ^{68}Ge, this ease in supply likely leads to its popularity.

The ability to tag a targeting agent with a PET tracer – either organic-based, such as ^{18}F and ^{11}C, or inorganic based such as ^{68}Ga and ^{64}Cu – makes PET a very versatile technique imaging a vast number of diseases and biological processes, with application dependent on the half-life of the radionuclide, Table 1.5.[1]

Table 1.5: Properties of the most widely used clinically relevant PET radioisotopes. Data from the Lund/LBNL Nuclear Data Search.[85]

Radionuclide	β_{max} energy/ MeV (%)[a]	Mean range in water/mm[b]	Prompt γ energies/ MeV and (%)	Half-life
Fluorine-18	0.634 (96.9%)	0.6	-	110 minutes
Carbon-11	0.960 (99.8%)	1.2	-	20.4 minutes
Manganese-51	2.186 (96.9%)	4.3	-	46.2 minutes
Manganese-52	0.576 (29.6%)	0.6	0.744 (90%), 0.936 (94.5%), 1.434 (100%)	5.591 days
Copper-64	0.653 (17.4%)	0.7	-	12.7 hours
Gallium-68	1.899 (88.0%)	3.5	1.077 (3%)	67.8 minutes
Zirconium-89	0.902 (22.7%)	1.3	0.909 (100%)	78.4 minutes

[a] *Conti and co-workers*[15], [b] *Jødal and co-workers*[86].

^{51}Mn decays *via* a 96.9% positron emission pathway, with a half-life of 46.2 minutes, which positions it between ^{68}Ga and ^{11}C for an application to track fast biological processes and has negligible amounts of prompt gamma rays. However, ^{51}Mn produces quite a high energy positron (similar to ^{68}Ga) leading to a 4.3 mm mean range in water, which contributes to much poorer resolution than industry leading ^{18}F.[86]

^{52g}Mn (herein referred to as ^{52}Mn) decays with only a 29.6% positron emitting pathway, with a half-life of 5.591 days, making it suitable for tracking slow biological processes such as pharmacokinetics. ^{52}Mn offers very good resolution compared to other radionuclides, with a mean range in water of only 0.6 mm, however prompt gammas are an issue both from a gamma detection point-of-view and patient radioactivity dose.[15,86]

1.5.2 Dual-modal imaging

Recently, molecular imaging efforts have been focused on combining PET and MRI modalities to create techniques that can bring the advantages of individual modalities together, providing more useful, and more convenient images. This follows the success of PET-CT which, since its introduction in 1997, has become widely available across the world. PET-CT combines the functional imaging acquired by PET with the anatomical images of CT/X-ray, allowing clinicians to precisely locate biological processes – crucial in oncology.

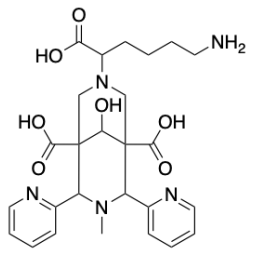
PET-MRI can offer a series of benefits over PET-CT due to the inherent advantages of MRI over CT. Firstly, MRI has far better soft-tissue contrast – MRI is the technique of choice in brain imaging, and PET-MRI has been shown to be useful in staging tumours in the head, neck, and liver, being able to match the detection rate of PET-CT, and in some cases providing better tumour assessment.[87] The inherent longer acquisition times of MR allows for increased time to capture more PET data, and hence provide better image quality with increased detector counts. MRI images are typically acquired ‘motion-free' due to respiratory gating with images taken at full breath holds, which in turn helps to sharpen PET resolution by decreasing motion artefacts.[16] PET-CT combines two techniques which both rely on ionising radiation, precluding paediatric scanning and long-term monitoring of therapy and staging in adult populations. >50% of the ionising radiation dose from a PET-CT scan can be attributed to the anatomical CT scan, replacing this component with MR imaging allows clinicians to procure functional images without losing image quality, and improves diagnosis for people previously unable to undergo PET-CT imaging.[16]

The challenge in developing combined contrast agents for both PET and MRI lies in the huge difference in sensitivity. DCE-MRI typically requires mM amounts of GBCA for adequate signal compared to nM (or even pM) amounts of contrast for PET imaging. This immediately rules out simply [^{18}F]-fluorinating

a Dotarem® derivative, as the radiation dose from having mM amounts of ^{18}F would be too high, and the detector would be saturated far beyond useful imaging limits. Alternatively, GBCAs could be co-administered with [^{18}F]-FDG however this would lead to tracers with differing pharmacokinetics.

In Table 1.5, of the typically used PET radionuclides, only manganese has stable paramagnetic oxidation states required for MRI. Doping nM amounts of either ^{51}Mn or ^{52}Mn into mM amounts of ^{55}Mn, and then complexing with a suitable ligand for MR imaging, would provide a bi-modal imaging agent with identical pharmacokinetics between the PET and MR component.

Vanasschen and co-workers used CDTA as a ligand for their studies into radiolabelling MnBCAs.[88] They combined ~8 MBq [^{52}Mn]$MnCl_2$ with 64 mM of CDTA for 30 minutes. A radiochemical yield (RCY) of >99% was achieved, with no uncomplexed ^{52}Mn present after stirring; excess CDTA was able to be removed via HPLC – necessary for clinical translation. An isotopic mixture was able to be prepared in the same way, but with a 0.1 equivalent (relative to CDTA) of $^{55}MnCl_2.4H_2O$ added with ~2 MBq of $^{52}MnCl_2$, leading to the development of a chemically equivalent PET-MR imaging probe.



H_3bispidine-BuNH_2

Figure 1.11: H_3bispidine-BuNH_2 synthesised by Ndiaye and co-workers.[78]

In 2020, H_3bispidine-BuNH_2 (Figure 1.11) was radiolabelled with ^{52}Mn.[78] Ndiaye and co-workers achieved a 99% RCY at 70 °C, 1 hour, pH 7, in the presence of ascorbic acid (to prevent oxidation). The complex was found to have a log P of –1.96, indicating that [^{52}Mn][Mn(bispidine-BuNH_3)(H_2O)]$^{2+}$ is very hydrophilic. This was seen in its biodistribution with *ex vivo* ICP-OES quantification after 1.5 hours of injection showing a > 5× proportion in the kidneys than in the liver.

For an effective PET agent, tight binding is required to ensure the radioactive tracer is not going to undesired places, which typically requires saturation of the coordination sphere. MRI contrast agents on the other hand require thermodynamic stability to be sacrificed for efficient water exchange, requiring an unsaturated coordination sphere. Therefore, the design of a PET-MRI agent for manganese understandably prioritises MRI contrast enhancement at the cost of stability.

1.6 DFT and analytics

Density functional theory (DFT) is a computational method of optimising the positions of nuclei within a desired input molecule to give a structure that is energetically stable, and then with this structure, working out properties of interest for this optimised system. It relies on two Hohenberg-Kohn theorems that showed that the ground state electron density of a many-electron system could predict that system's ground state properties.[89]

In the simplest system (a single hydrogen atom – one proton, one electron) Schrödinger's equation can be exactly solved to know the electron's wavefunction (and hence electron density throughout space). As complexity builds, solving this equation for many-electron systems becomes massively more complex and prohibitively computationally expensive. Therefore, density functional approximations (DFAs) have been developed which address the complexity problem by applying 'simplifications' to reduce the amount of computation required. Each DFA can be thought of as a different approach to optimise the ground-state electron density. They are judged by their ability to accurately model the system under study, and typically with every incremental increase in accuracy (by considering more and more complex electronic effects – such as dispersion effects or long range e-e interactions), there is a trade-off with increased computational cost.[90]

Alongside functionals, DFT requires basis sets. These help to build the wavefunctions to describe the orbitals that the functional will use to optimise the structure. They too range by theory – from the level of theory seen at A-level to describe orbitals, up to functionals that take into consideration more and more quantum mechanical theory, once again at the trade-off with increased computational cost. For this work the Gaussian 16 (C.01)[91] software package was used, with inputs typically supplied after a pre-optimisation with Avogadro 1.2.0.[92]

At its core, any optimisation *via* DFT is a model of a real-world system, the same way a skeletal diagram is not a true reflection of what a molecule looks like but is a useful model for us to communicate and understand the chemical information. In this way, DFT calculations can be likened to digital photography – we are trying to capture what the world looks like with a tool that will greatly impact how accurately our picture matches reality. If the Gaussian software is our digital camera, the functionals are the processor-detector – manipulating the received data to create the digital image; and the basis sets are the lens, mirrors, and shutters – they alter the light (information) that the processor receives before it creates an image. Detectors and processors with better technology can more accurately turn the light they receive into an accurate picture, but if the light they receive is poor – i.e., using a poor quality, dirty lens – then the image a high-end camera takes will not be accurate. Similarly,

computationally expensive methods will not provide accurate results if the basis set used is too simplistic (for the system being studied).
In addition to functional and basis sets, to accurately model a real-world system we need to take solvation into consideration. This substantially complicates the real-world problem we are modelling, from a molecule as an ideal gas to a molecule subject to the combined forces of many non-fixed solvent molecules.[93]

As such, multiple approaches have been developed to tackle (and simplify) the problem. These range from molecular dynamics (studying the interaction between solute and solvent on a molecule by molecule basis), to an implicit solvation model (which treats the solvent as a continuous isotropic medium), and mixtures of the two.[94] Gaussian approaches this problem with a selection of Self-Consistent Reaction Field (SCRF) methods, where the solvent is implicit (captured by being a continuous, uniform dielectric medium with dielectric constant ε). SCRF methods treat the molecules under study as being dilute enough so as to not interact with each other (some methods may require explicit solvent models to be added to the calculation).

Practically, this means we have a cavity inside a 'sea of solvent' represented only by its properties – for the simplest SCRF method, that will just be ε. Within this cavity we have the molecule under study which will interact with the 'sea of solvent' *via* its own dipole(s) to mutually polarise each other. This in turn provides Gaussian with a set of parameters to optimise with, and then iteratively this optimisation occurs with mutual polarisation of the solute surface/solvent, until the system finds an equilibrium – hence 'self-consistent'.[93] This is the basis for the polarisable continuum models (PCM) typically used in Gaussian calculations, the default of these is the integral equation formalism model known as IEFPCM.[95] By using this approximation of the situation, calculations with SCRF methods can separate the problem of optimising a solute molecule from the much more complex problem of optimising the interactions of 10s-100s of solvent molecules around the solute, to produce a compromise between accuracy and computational cost. For most methods the cavity that the solvent sits in is determined by coinciding spheres from their van der Waals radii (from radii calculated with the UFF force field [96]), as can be seen in the final solvation cavity of a HF/3-21G/IEFPCM optimisation on vinyl acetate, Figure 1.12.

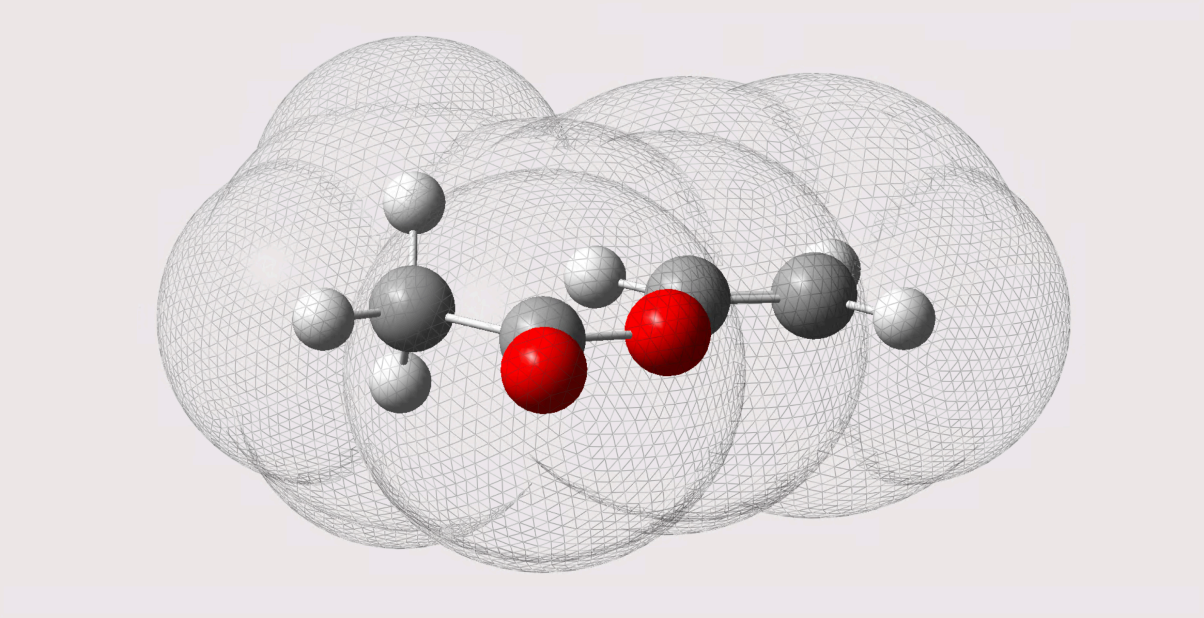


Figure 1.12: Optimised geometry of vinyl acetate at HF/3-21G/IEFPCM, with the final solvation cavity shown.

DFT analysis has not been widely applied to studying MnBCAs, and generally only optimisations are carried out to find a minimum energy, which is then used to analyse steric properties such as bond-length and bond angle. The Platas-Iglesias group have largely been the only group carrying out DFT analysis for geometry optimisation. They prefer to use the hybrid meta-GGA M06-2X functional along with the TZVP basis set.[75,76] This has been shown by them to better match experimental data than the extremely well-used (yet older) hybrid GGA B3LYP functional, which they found to over stabilise systems with low coordination numbers.[97]

In terms of analytics for MnBCAs, there have been a number of publications detailing useful trends for DFT and ligand design. DFT optimisations by Esteban-Goméz and co-workers on GBCAs have shown that the explicit inclusion of second sphere H_2O molecules in the optimisation (in addition to employing an aqueous IEFPCM) was required to accurately predict H_2O-Gd bond length.[27] With every inclusion of an explicit water molecule, the predicted length closer matched the experimental value, Table 1.6. The number of H_2Os included follows the law of diminishing returns, as the increase in accuracy drops off whilst every addition becomes more and more computationally expensive. This GBCA conclusion has also been shown to apply to MnBCAs.[98]

Table 1.6: Averaged bond distances (in Å) of Gd-O_{water}, by number of explicit second sphere H_2O molecules for $[Gd(L)(H_2O)]^{n+/-}$ complexes calculated at the TPSSh/LCRECP/6-31G(d)/IEFPCM level.

Ligand	0	1	2	Exp.
DOTA (SAP)	2.606	2.523	2.494	2.456
HP-DO3A	2.568	2.544	2.514	2.507
DTPA	2.639	2.542	2.501	2.490
BPEDA	2.669	2.551	2.516	2.52

Peters and Geraldes published an analysis on r_1 and q data showing that increased literature q value corresponded to higher r_1, however there was a large spread in r_1 at each q value. They inferred that formula weight as an additional contribution would be better able to model r_1 – Equation 1.14 implied that molecular volume would affect τ_R, and hence r_1. They then plotted r_1/q_{lit} against formula weight and ran through all best-fit models that *CurveExpert 1.4* provided to produce a 'semi-empirical' formula to estimate hydration number from formula weight, Equation 1.16.[99]

$$q_{calc} = \frac{r_1}{9.16(1 - e^{(-2.97 \times FW \times 10^{-3})}}$$

Equation 1.16: Equation to estimate hydration number published by Peters and Geraldes.[99]

The authors do not suggest any reasoning behind the numbers, hence the 'semi-empirical' nature of their method. Additionally, in order to avoid complications from MnBCAs with double dispersions in their ^{1}H NMRD profiles, they chose to take r_1 data at 0.01 MHz (most studies report r_1 at 20, 64, or 128 MHz). Due to the nature of the data that went into the calculation (the majority of the r_1 values reported are to 0 decimal places), it is improper to use Equation 1.16 beyond an estimation at 0 decimal places.

Uzal-Varela and co-workers have developed an empirical formula which can predict $\log(K_{MnL})$ values based on the summation of the contributions to $\log(K_{MnL})$ from individual parts of the ligand.[75] This work follows their paper on predicting the $\log(K_{GdL})$ for gadolinium complexes, in which they postulated that the individual components of a ligand would provide mutually exclusive contributions to the stability constant.[100] The manganese formula was developed by performing a least squares fit on 150 manganese complexes (practically the entire MnBCA field), with each structural component as an independent variable affecting $\log(K_{MnL})$. This fit was able to reproduce the $\log(K_{MnL})$ data it was trained on with a mean absolute error of 0.99 log units, implying strong predictive power for new MnBCAs so long as they only contain the structural components covered in their model.

Because the model is so dependent on the training set (the 150 MnBCAs) for its prediction it cannot handle new ligand types. Additionally, new ligand families (or even current ligand families modified to

include new coordinating groups) would require multiple examples in order to give the least squares fitting enough data to provide confidence in its prediction. The model is also blind to structural or geometric isomers, for example, it predicts that *cis*-CDTA and *trans*-CDTA would have equal log(K_{MnL}) values. Nevertheless, this work represents a massive leap forward for the use of data and predictive analytics in the MnBCA space.

1.7 Aims of this work

The focus of this PhD project has been the development of contrast agents primarily focused on manganese-based alternatives to GBCAs for MRI, whilst simultaneously developing manganese complexes capable of delivering PET active ^{52}Mn. Both synthetic and computational studies were undertaken to aid these efforts and to provide tools to predict their physical properties. Alongside this, computational studies were used to further understand the chemistry of two PET tracers (^{89}Zr and ^{68}Ga).

The first aim of the project was to develop novel ligand families for MnBCAs using design motifs that are under- or un-used in the field. This aim was supported by DFT geometry optimisations to aid in ligand design, with the overall effort directed towards providing a MnBCA that is capable of replacing GBCAs in renally compromised patients. This required a consideration of increased hydrophobicity, overall negative charge, and suitable mix of hard and soft donors.

Complementary to these efforts, an analytical project was designed to discover correlation between DFT optimised geometries and the key performance indicators (KPIs) for a successful MnBCA – namely log(K_{MnL}), r_1, k_{ex}, q; this project was designed to use both simple supervised machine learning techniques such as linear regression, and also more complex techniques such as neural networks and support vector machine (SVM) learning.

The second aim of the project was to use computational chemistry to provide insight into two different applications of PET radiometals. The first of these was to investigate the possible species formed when [^{89}Zr][Zr(oxinate)$_4$] is introduced into a doxorubicin loaded liposome and predict the dominant species under different simulated pH environments. Finally, to aid other research efforts in the Long group, a predictive tool was developed to estimate the log P of ^{68}Ga complexes in an attempt to provide a computational prototype that could evaluate the field's KPI before lengthy synthesis is undertaken.

1.8 References

1 M. L. James and S. S. Gambhir, *Physiol. Rev.*, 2012, **92**, 897–965, DOI: 10.1152/physrev.00049.2010.

2 T. Tang, A. Y. Louie, E. M. Gale, P. Caravan, C. T. Farrar and B. Keil, in *Contrast Agents for MRI: Experimental Methods*, 2017, pp. 499–584, DOI: 10.1039/9781788010146-00499.

3 M. Brandt, J. Cardinale, I. Rausch and T. L. Mindt, *J. Label. Compd. Radiopharm.*, 2019, **62**, 541–551, DOI: 10.1002/jlcr.3754.

4 A. Gabizon, H. Shmeeda and Y. Barenholz, *Clin. Pharmacokinet.*, 2003, **42**, 419–436, DOI: 10.2165/00003088-200342050-00002.

5 R. B. Buxton, *Introduction to Functional Magnetic Resonance Imaging*, Cambridge University Press, New York, 2nd edn., 2009, DOI: 10.1017/CBO9780511549854.

6 G. Saar, C. M. Millo, L. P. Szajek, J. Bacon, P. Herscovitch and A. P. Koretsky, *Mol. Imaging Biol.*, 2018, **20**, 562–574, DOI: 10.1007/s11307-018-1162-6.

7 Q. N. Do, J. S. Ratnakar, Z. Kovács, G. Tircsó, F. Krisztián Kálmán, Z. Baranyai, E. Brücher and I. Tóth, in *Contrast Agents for MRI: Experimental Methods*, eds. V. C. Pierre and M. J. Allen, Cambridge, 2017, pp. 1–120, DOI: 10.1039/9781788010146-00001.

8 A. Merbach, L. Helm and É. Tóth, *The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging*, John Wiley & Sons, Ltd, Chichester, UK, 2013, DOI: 10.1002/9781118503652.

9 A. Gupta, P. Caravan, W. S. Price, C. Platas-Iglesias and E. M. Gale, *Inorg. Chem.*, 2020, **59**, 6648–6678, DOI: 10.1021/acs.inorgchem.0c00510.

10 J. Wahsner, E. M. Gale, A. Rodríguez-Rodríguez and P. Caravan, *Chem. Rev.*, 2019, **119**, 957–1057, DOI: 10.1021/acs.chemrev.8b00363.

11 H. Liu and X.-A. Zhang, in *Encyclopedia of Inorganic and Bioinorganic Chemistry*, John Wiley & Sons, Ltd, Chichester, UK, 2018, pp. 1–16, DOI: 10.1002/9781119951438.eibc2626.

12 Z. Li and P. S. Conti, *Adv. Drug Deliv. Rev.*, 2010, **62**, 1031–1051, DOI: 10.1016/j.addr.2010.09.007.

13 S. M. Sephton and S. M. Ametamey, in *Molecular Imaging Techniques: New Frontiers*, Future Science Ltd, Unitec House, 2 Albert Place, London N3 1QB, UK, 2013, pp. 6–20, DOI: 10.4155/ebo.12.504.

14 W. W. Moses, *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, 2011, **648**, S236–S240, DOI: 10.1016/j.nima.2010.11.092.

15 M. Conti and L. Eriksson, *EJNMMI Phys.*, 2016, **3**, 8, DOI: 10.1186/s40658-016-0144-5.

16 E. C. Ehman, G. B. Johnson, J. E. Villanueva-Meyer, S. Cha, A. P. Leynes, P. E. Z. Larson and T. A. Hope, *J. Magn. Reson. Imaging*, 2017, **46**, 1247–1262, DOI: 10.1002/jmri.25711.

17 F. K. Kálmán, F. K. Kálmán, V. Nagy, B. Váradi, Z. Garda, E. Molnár, G. Trencsényi, J. Kiss,

S. Même, W. Même, É. Tóth and G. Tircsó, *J. Med. Chem.*, 2020, **63**, 6057–6065, DOI: 10.1021/acs.jmedchem.0c00197.

18 J. H. Ardenkjær-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M. H. Lerche, R. Servin, M. Thaning and K. Golman, *Proc. Natl. Acad. Sci.*, 2003, **100**, 10158–10163, DOI: 10.1073/pnas.1733835100.

19 T. J. Swift and R. E. Connick, *J. Chem. Phys.*, 1962, **37**, 307–320, DOI: 10.1063/1.1701321.

20 E. J. Werner, A. Datta, C. J. Jocher and K. N. Raymond, *Angew. Chem. Int. Ed.*, 2008, **47**, 8568–8580, DOI: 10.1002/anie.200800212.

21 P. Caravan, C. T. Farrar, L. Frullano and R. Uppal, *Contrast Media Mol. Imaging*, 2009, **4**, 89–100, DOI: 10.1002/cmmi.267.

22 I. Solomon, *Phys. Rev.*, 1955, **99**, 559–565, DOI: 10.1103/PhysRev.99.559.

23 N. Bloembergen and L. O. Morgan, *J. Chem. Phys.*, 1957, **27**, 572–573, DOI: 10.1063/1.1743771.

24 M. Gueron, *J. Magn. Reson.*, 1975, **19**, 58–66, DOI: 10.1016/0022-2364(75)90029-3.

25 L. Vander Elst, S. Laurent, H. Masaka Bintoma and R. N. Muller, *Magn. Reson. Mater. Physics, Biol. Med.*, 2001, **12**, 135–140, DOI: 10.1016/S1352-8661(01)00110-7.

26 J. C. Slater, *J. Chem. Phys.*, 1964, **41**, 3199–3204, DOI: 10.1063/1.1725697.

27 D. Esteban-Gómez, A. de Blas, T. Rodríguez-Blas, L. Helm and C. Platas-Iglesias, *ChemPhysChem*, 2012, **13**, 3640–3650, DOI: 10.1002/cphc.201200417.

28 T. J. Clough, L. Jiang, K.-L. Wong and N. J. Long, *Nat. Commun.*, 2019, **10**, 1420, DOI: 10.1038/s41467-019-09342-3.

29 D. M. Sado, *Shane Angoh Cardiac MRI report*, London, 2022.

30 V. M. Ferreira, S. Plein, T. C. Wong, Q. Tao, Z. Raisi-Estabragh, S. S. Jain, Y. Han, V. Ojha, D. A. Bluemke, K. Hanneman, J. Weinsaft, M. K. Vidula, N. A. B. Ntusi, J. Schulz-Menger and J. Kim, *J. Cardiovasc. Magn. Reson.*, 2023, **25**, 21, DOI: 10.1186/s12968-023-00933-0.

31 L. Chelala, J. Jeudy, R. Hossain, G. Rosenthal, N. Pietris and C. S. White, *Am. J. Roentgenol.*, 2022, **218**, 651–658, DOI: 10.2214/AJR.21.26853.

32 R. M. Inciardi, L. Lupi, G. Zaccone, L. Italia, M. Raffo, D. Tomasoni, D. S. Cani, M. Cerini, D. Farina, E. Gavazzi, R. Maroldi, M. Adamo, E. Ammirati, G. Sinagra, C. M. Lombardi and M. Metra, *JAMA Cardiol.*, 2020, **5**, 819–824, DOI: 10.1001/jamacardio.2020.1096.

33 N. P. Neupane, K. Rajlawot, C. Adhikari, D. Kandel and I. Mustafa, *IDCases*, 2022, **29**, e01579, DOI: 10.1016/j.idcr.2022.e01579.

34 J. C. Weinreb, R. A. Rodby, J. Yee, C. L. Wang, D. Fine, R. J. McDonald, M. A. Perazella, J. R. Dillman and M. S. Davenport, *Radiology*, 2021, **298**, 28–35, DOI: 10.1148/radiol.2020202903.

35 T. Grobner and F. C. Prischl, *Kidney Int.*, 2007, **72**, 260–264, DOI: 10.1038/sj.ki.5002338.

36 T. K. Chen, D. H. Knicely and M. E. Grams, *JAMA - J. Am. Med. Assoc.*, 2019, **322**, 1294–

1304, DOI: 10.1001/jama.2019.14745.

37 P. Cockwell and L. A. Fisher, *Lancet*, 2020, **395**, 662–664, DOI: 10.1016/S0140-6736(19)32977-0.

38 K. J. Foreman, N. Marquez, A. Dolgert, K. Fukutaki, N. Fullman, M. McGaughey, M. A. Pletcher, A. E. Smith, K. Tang, C.-W. Yuan, J. C. Brown, J. Friedman, J. He, K. R. Heuton, M. Holmberg, D. J. Patel, P. Reidy, A. Carter, K. Cercy, A. Chapin, D. Douwes-Schultz, T. Frank, F. Goettsch, P. Y. Liu, V. Nandakumar, M. B. Reitsma, V. Reuter, N. Sadat, R. J. D. Sorensen, V. Srinivasan, R. L. Updike, H. York, A. D. Lopez, R. Lozano, S. S. Lim, A. H. Mokdad, S. E. Vollset and C. J. L. Murray, *Lancet*, 2018, **392**, 2052–2090, DOI: 10.1016/S0140-6736(18)31694-5.

39 Kidney Research UK, *Kidney disease: A UK public health emergency The health economics of kidney disease to 2033*, 2023.

40 P. Marckmann, *Eur. J. Radiol.*, 2008, **66**, 187–190, DOI: 10.1016/j.ejrad.2008.01.032.

41 T. Kanda, K. Ishii, H. Kawaguchi, K. Kitajima and D. Takenaka, *Radiology*, 2014, **270**, 834–841, DOI: 10.1148/radiol.13131669.

42 T. Kanda, T. Fukusato, M. Matsuda, K. Toyoda, H. Oba, J. Kotoku, T. Haruyama, K. Kitajima and S. Furui, *Radiology*, 2015, **276**, 228–232, DOI: 10.1148/radiol.2015142690.

43 R. J. McDonald, J. S. McDonald, D. F. Kallmes, M. E. Jentoft, D. L. Murray, K. R. Thielen, E. E. Williamson and L. J. Eckel, *Radiology*, 2015, **275**, 772–782, DOI: 10.1148/radiol.15150025.

44 R. J. McDonald, J. S. McDonald, D. F. Kallmes, M. E. Jentoft, M. A. Paolini, D. L. Murray, E. E. Williamson and L. J. Eckel, *Radiology*, 2017, **285**, 546–554, DOI: 10.1148/radiol.2017161595.

45 E. Gianolio, E. Di Gregorio and S. Aime, *Eur. J. Inorg. Chem.*, 2019, **2019**, 137–151, DOI: 10.1002/ejic.201801220.

46 N. Murata, L. F. Gonzalez-Cuyar, K. Murata, C. Fligner, R. Dills, D. Hippe and K. R. Maravilla, *Invest. Radiol.*, 2016, **51**, 447–453, DOI: 10.1097/RLI.0000000000000252.

47 European Medicines Agency, *EMA's final opinion confirms restrictions on use of linear gadolinium agents in body scans*, 2017, vol. November.

48 FDA, FDA Drug Safety Communication: FDA warns that gadolinium-based contrast agents (GBCAs) are retained in the body; requires new class warnings, https://www.fda.gov/drugs/drug-safety-and-availability/fda-drug-safety-communication-fda-warns-gadolinium-based-contrast-agents-gbcas-are-retained-body, (accessed 4 September 2023).

49 Bayer, *Gadovist® 1.0 mmol/ml solution for injection*, 2022.

50 CDC, Toxicological Profile for Manganese, https://www.atsdr.cdc.gov/toxprofiles/tp151-c2.pdf, (accessed 4 September 2023).

51 B. Misselwitz, A. Mühler and H.-J. Weinmann, *Invest. Radiol.*, 1995, **30**, 611–620, DOI:

10.1097/00004424-199510000-00007.

52 B. Drahoš, I. Lukeš and É. Tóth, *Eur. J. Inorg. Chem.*, 2012, **2012**, 1975–1986, DOI: 10.1002/ejic.201101336.

53 S. L. O'Neal and W. Zheng, *Curr. Environ. Heal. Reports*, 2015, **2**, 315–328, DOI: 10.1007/s40572-015-0056-x.

54 J. Crossgrove and W. Zheng, *NMR Biomed.*, 2004, **17**, 544–553, DOI: 10.1002/nbm.931.

55 N. B. Spath, G. Thompson, A. H. Baker, M. R. Dweck, D. E. Newby and S. I. K. Semple, *Heart*, 2019, heartjnl-2019-315227, DOI: 10.1136/heartjnl-2019-315227.

56 S. Serai, L. Leoni, M. Haque and B. B. Roman, *J. Diabetes Res. Clin. Metab.*, 2012, **1**, 21, DOI: 10.7243/2050-0866-1-21.

57 Y. S. Kang and J. C. Gore, *Invest. Radiol.*, 1984, **19**, 399–407, DOI: 10.1097/00004424-198409000-00012.

58 T. C.-C. Hu, T. F. Christian, A. H. Aletras, J. L. Taylor, A. P. Koretsky and A. E. Arai, *Magn. Reson. Med.*, 2005, **54**, 196–200, DOI: 10.1002/mrm.20516.

59 P. F. Antkowiak, S. A. Tersey, J. D. Carter, M. H. Vandsburger, J. L. Nadler, F. H. Epstein and R. G. Mirmira, *Am. J. Physiol. Metab.*, 2009, **296**, E573–E578, DOI: 10.1152/ajpendo.90336.2008.

60 I. Aoki, Y.-J. L. Wu, A. C. Silva, R. M. Lynch and A. P. Koretsky, *Neuroimage*, 2004, **22**, 1046–1059, DOI: 10.1016/j.neuroimage.2004.03.031.

61 N. B. Spath, D. M. L. Lilburn, G. A. Gray, L. M. Le Page, G. Papanastasiou, R. J. Lennen, R. L. Janiczek, M. R. Dweck, D. E. Newby, P. C. Yang, M. A. Jansen and S. I. Semple, *Contrast Media Mol. Imaging*, 2018, **2018**, 1–13, DOI: 10.1155/2018/9641527.

62 T. C.-C. Hu, R. G. Pautler, G. A. MacGowan and A. P. Koretsky, *Magn. Reson. Med.*, 2001, **46**, 884–890, DOI: 10.1002/mrm.1273.

63 R. G. Pautler, A. C. Silva and A. P. Koretsky, *Magn. Reson. Med.*, 1998, **40**, 740–748, DOI: 10.1002/mrm.1910400515.

64 K.-H. Chuang and A. P. Koretsky, *Magn. Reson. Imaging*, 2009, **27**, 594–600, DOI: 10.1016/j.mri.2008.10.006.

65 EMA, *Teslascan (mangafodipir) Withdrawal of the marketing authorisation in the European Union*, 2012.

66 H. Irving and R. J. P. Williams, *J. Chem. Soc.*, 1953, 3192–3210, DOI: 10.1039/jr9530003192.

67 A. E. Martell and R. M. S. Editors, *Critical Stability Constants Volume 5*, 1983, vol. 11, DOI: 10.1016/0307-4412(83)90054-7.

68 G. Anderegg, F. Arnaud-Neu, R. Delgado, J. Felcman and K. Popov, *Pure Appl. Chem.*, 2005, **77**, 1445–1495, DOI: 10.1351/pac200577081445.

69 Z. Baranyai, G. Tircsó and F. Rösch, *Eur. J. Inorg. Chem.*, 2020, **2020**, 36–56, DOI: 10.1002/ejic.201900706.

70 J. Byegård, G. Skarnemark and M. Skålberg, *J. Radioanal. Nucl. Chem.*, 1999, **241**, 281–290, DOI: 10.1007/BF02347463.

71 Z. Garda, E. Molnár, N. Hamon, J. L. Barriada, D. Esteban-Gómez, B. Váradi, V. Nagy, K. Pota, F. K. Kálmán, I. Tóth, N. Lihi, C. Platas-Iglesias, É. Tóth, R. Tripier and G. Tircsó, *Inorg. Chem.*, 2021, **60**, 1133–1148, DOI: 10.1021/acs.inorgchem.0c03276.

72 M. R. Brandt, C. Vanasschen, J. Ermert, H. H. Coenen and B. Neumaier, *Dalton Trans.*, 2019, **48**, 3003–3008, DOI: 10.1039/C8DT04996C.

73 E. M. Gale, I. P. Atanasova, F. Blasi, I. Ay and P. Caravan, *J. Am. Chem. Soc.*, 2015, **137**, 15548–15557, DOI: 10.1021/jacs.5b10748.

74 F. K. Kálmán and G. Tircsó, *Inorg. Chem.*, 2012, **51**, 10065–10067, DOI: 10.1021/ic300832e.

75 R. Uzal-Varela, F. Pérez-Fernández, L. Valencia, A. Rodríguez-Rodríguez, C. Platas-Iglesias, P. Caravan and D. Esteban-Gómez, *Inorg. Chem.*, 2022, **61**, 14173–14186, DOI: 10.1021/acs.inorgchem.2c02364.

76 Z. Garda, E. Molnár, F. K. Kálmán, R. Botár, V. Nagy, Z. Baranyai, E. Brücher, Z. Kovács, I. Tóth and G. Tircsó, *Front. Chem.*, 2018, **6**, 1–14, DOI: 10.3389/fchem.2018.00232.

77 T. Csupász, D. Szücs, F. K. Kálmán, O. Hollóczki, A. Fekete, D. Szikra, É. Tóth, I. Tóth and G. Tircsó, *Molecules*, 2022, **27**, 371, DOI: 10.3390/molecules27020371.

78 D. Ndiaye, M. Sy, A. Pallier, S. Même, I. de Silva, S. Lacerda, A. M. Nonat, L. J. Charbonnière and É. Tóth, *Angew. Chem. Int. Ed.*, 2020, **59**, 11958–11963, DOI: 10.1002/anie.202003685.

79 P. Cieslik, P. Comba, B. Dittmar, D. Ndiaye, É. Tóth, G. Velmurugan and H. Wadepohl, *Angew. Chem. Int. Ed.*, 2022, **61**, 1–8, DOI: 10.1002/anie.202115580.

80 R. D. Shannon, *Acta Crystallogr. Sect. A*, 1976, **32**, 751–767, DOI: 10.1107/S0567739476001551.

81 E. M. Gale, C. M. Jones, I. Ramsay, C. T. Farrar and P. Caravan, *J. Am. Chem. Soc.*, 2016, **138**, 15861–15864, DOI: 10.1021/jacs.6b10898.

82 C. wan Chen, J. S. Cohen, C. E. Myers and M. Sohn, *FEBS Lett.*, 1984, **168**, 70–74, DOI: 10.1016/0014-5793(84)80208-2.

83 S. H. Koenig, R. D. Brown and M. Spiller, *Magn. Reson. Med.*, 1987, **4**, 252–260, DOI: 10.1002/mrm.1910040306.

84 A. Barandov, B. B. Bartelle, B. A. Gonzalez, W. L. White, S. J. Lippard and A. Jasanoff, *J. Am. Chem. Soc.*, 2016, **138**, 5483–5486, DOI: 10.1021/jacs.5b13337.

85 S. Y. F. Chu, L. P. Ekström and R. B. Firestone, The Lund/LBNL Nuclear Data Search, http://nucleardata.nuclear.lu.se/toi/, (accessed 4 September 2023).

86 L. Jødal, C. Le Loirec and C. Champion, *Phys. Med. Biol.*, 2014, **59**, 7419–7434, DOI: 10.1088/0031-9155/59/23/7419.

87 V. Ruhlmann, M. Ruhlmann, A. Bellendorf, J. Grueneisen, L. M. Sawicki, H. Grafe, M. Forsting, A. Bockisch and L. Umutlu, *Eur. J. Radiol.*, 2016, **85**, 1941–1947, DOI:

10.1016/j.ejrad.2016.08.020.

88 C. Vanasschen, M. Brandt, J. Ermert and H. H. Coenen, *Dalton Trans.*, 2016, **45**, 1315–1321, DOI: 10.1039/C5DT04270D.

89 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138, DOI: 10.1103/PhysRev.140.A1133.

90 M. Witwicki, P. K. Walencik and J. Jezierska, *J. Mol. Model.*, 2020, **26**, 10, DOI: 10.1007/s00894-019-4268-0.

91 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian, Inc., Wallingford, CT, 2016.

92 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeerschd, E. Zurek and G. R. Hutchison, *J. Cheminform.*, 2012, 4, 1–17, Version 1.2.0, DOI: 10.1186/1758-2946-4-17.

93 J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc., Wallingford, CT, 3rd edn., 2015.

94 R. E. Skyner, J. L. McDonagh, C. R. Groom, T. Van Mourik and J. B. O. Mitchell, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6174–6191, DOI: 10.1039/c5cp00288e.

95 E. Cancès, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 1997, **107**, 3032–3041, DOI: 10.1063/1.474659.

96 A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035, DOI: 10.1021/ja00051a040.

97 M. Regueiro-Figueroa, D. Esteban-Gómez, R. Pujales-Paradela, L. Caneda-Martínez, A. de Blas and C. Platas-Iglesias, *Int. J. Quantum Chem.*, 2016, **116**, 1388–1396, DOI: 10.1002/qua.25191.

98 G. A. Rolla, C. Platas-Iglesias, M. Botta, L. Tei and L. Helm, *Inorg. Chem.*, 2013, **52**, 3268–3279, DOI: 10.1021/ic302785m.

99 J. A. Peters and C. F. G. C. Geraldes, *Inorganics*, 2018, **6**, 116, DOI: 10.3390/inorganics6040116.

100 R. Uzal-Varela, A. Rodríguez-Rodríguez, H. Wang, D. Esteban-Gómez, I. Brandariz, E. M. Gale, P. Caravan and C. Platas-Iglesias, *Coord. Chem. Rev.*, 2022, **467**, 214606, DOI:

10.1016/j.ccr.2022.214606.

Chapter 2

Synthesis of Ligands for Manganese-Based Contrast Agents

2.1 Introduction

The field of manganese-based contrast agents for MRI has been spurred on by the discovery of gadolinium deposition in patients with repeated DCE-MRI scanning with GBCAs, and a causal link between acyclic GBCA use and nephrogenic systemic fibrosis for those with stage 4 and 5 chronic kidney disease.[1] To meet this clinical need, there have been twenty years of research into providing a safer alternative to GBCAs with manganese as a replacement for their paramagnetic gadolinium centre. This is centred both on the body's natural ability to remove excess manganese (as a biogenic element) protecting against deposition, as well as the favourable properties of manganese as an MR-active metal – five unpaired d electrons in high-spin manganese complexes (d^5), fast exchange rates, and smaller metal-water bond lengths than for GBCAs.[2] Despite this, only one MnBCA has ever been clinically approved (Mangafodipir®), a $q = 0$ complex, whose mode-of-action relies on fully dechelating the manganese ion so that it can accumulate in the liver during its elimination to provide MR contrast for the liver.[3]

Most MnBCAs are based around a polyaminocarboxylate scaffold (acyclic or macrocyclic) chelating a Mn^{2+} centre. The design of these is typically optimised to ensure thermodynamic stability above contrast enhancement – Mn^{2+} complexes generally have stability constants in the range of 10^8-10^{11} times smaller than GBCAs primarily due to the lack of ligand field stabilisation energy (for high-spin d^5 Mn^{2+}) and lower charge of the ion. Whilst biogenic, free manganese at the quantities required for MR contrast enhancement (300-600 mg per dose, assuming 0.1 mmol/kg as for Gadovist®), could still present possible neural and cardiovascular toxicity.[2] Transmetalation reactions with endogenous ions (Zn^{2+}, Ca^{2+}, Cu^{2+}) can also reveal kinetic instability of MnBCAs when challenged with large concentrations of competitor ions. MnBCAs can be split into acyclic and macrocyclic complexes, with macrocyclic complexes generally showing higher stability constants than acyclic ones due to the macrocycle effect (a specific application of the chelate effect), in which the solution conformation of the macrocycle (known as being 'pre-organised') is closer to its bound form than that of the acyclic ligand, resulting in a lower reduction in entropy for chelation.

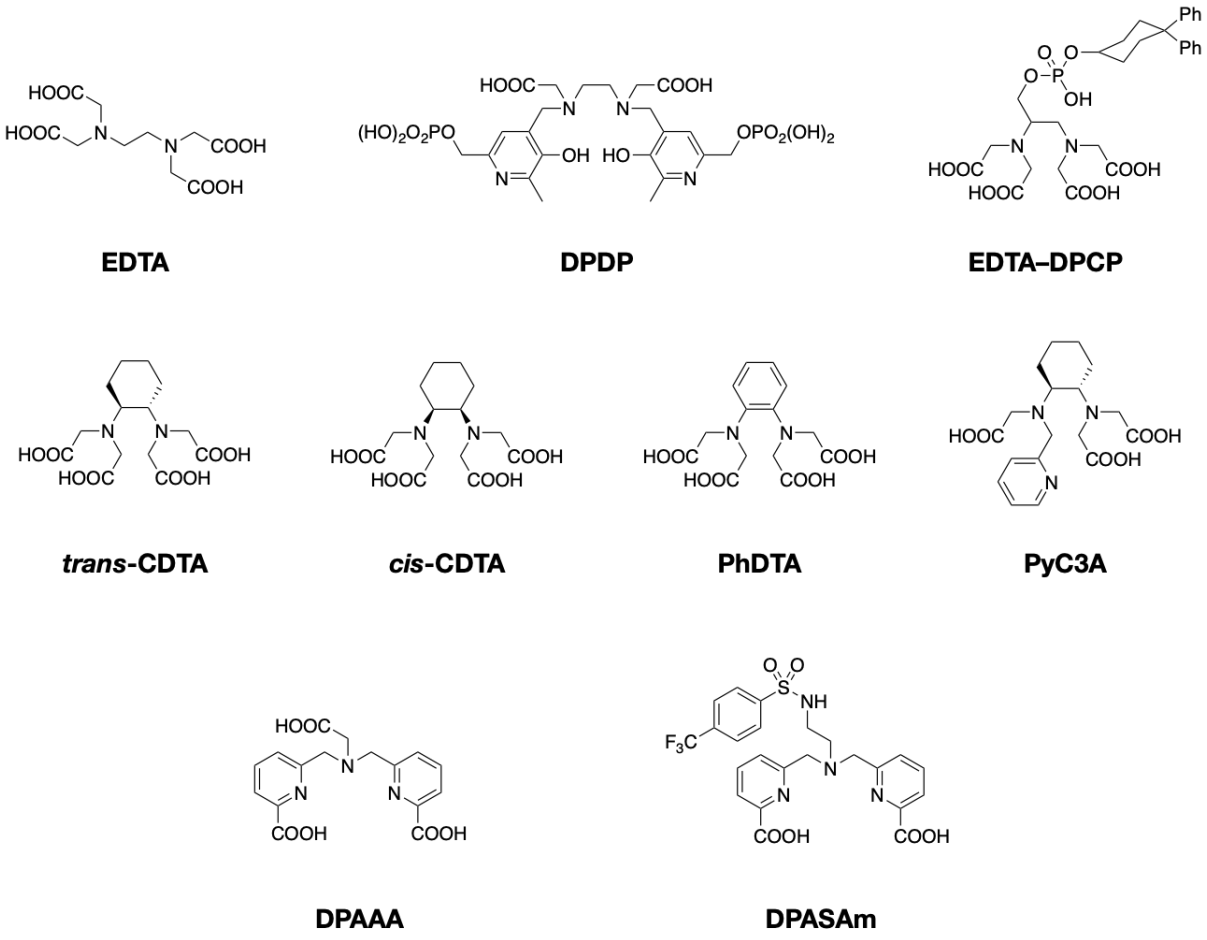


Figure 2.1: Selected acyclic ligands for Mn^{2+} discussed in this work.

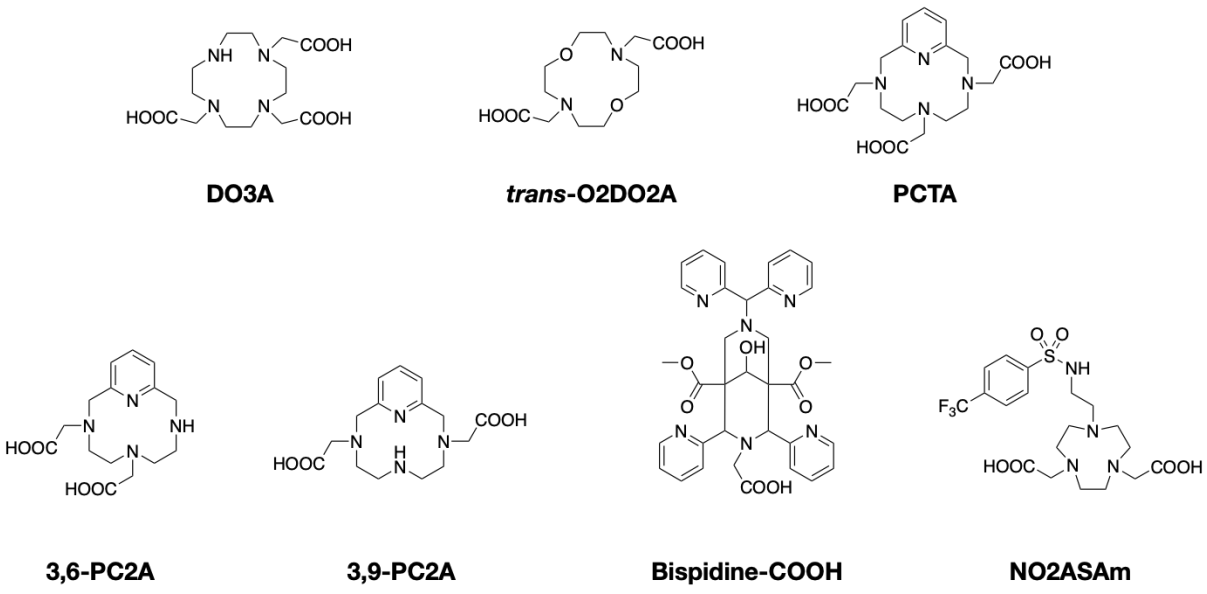


Figure 2.2: Selected macrocyclic ligands for Mn^{2+} discussed in this work.

2.1.1 Acyclic ligands

Of the acyclic ligands published, those based on EDTA (ethylenediaminetetraacetate) are the most studied and derivatised. $[Mn(EDTA)(H_2O)]^{2-}$ itself is heptacoordinate, with a free space for one bound water, and produces an r_1 value of 3.2 mM^{-1} s^{-1} and log(K_{MnL}) of 12.46, Figure 2.1.[4] These are typical values for q = 1 MnBCAs indicating the much lower stability constants of MnBCAs compared to GBCAs and target r_1 values to expect at 20 MHz and 298 K. Mangafodipir® (also known as MnDPDP, Figure 2.1) also uses an diaminoethylene backbone with two acetate arms, and the other two acetate arms replaced with two 3-hydroxy-2-methyl-5-((phosphonooxy)methyl)pyridin-4-yl arms. A manganese analogue of Gadofosveset® (a GBCA designed to act as a blood-pool agent *via* binding to HSA, based on DTPA (diethylenetriaminepentaacetate)) has been designed which also shows an increased r_1 from 5.8 to 48 mM^{-1} s^{-1} (20 MHz, 37 °C) once bound to HSA due to reduced rotational tumbling.[5] Replacement of the ethylene portion of EDTA, with either (±)-*cis*- or (±)-*trans*-1,2-diaminocyclohexene (CDTA) resulted in better thermodynamic stability over EDTA – 14.19 and 14.69 respectively, and a similar r_1 of 3.62 and 3.79 mM^{-1} s^{-1}.[4] The increase in thermodynamic stability is likely due to pre-forming an ideal coordination site for manganese coordination, as opposed to the often-cited increase in rigidity – replacement with a phenyl backbone actually reduces the stability constant to 11.79.[6] This highlights an oversimplification/lack-of-data-before-conclusion problem in ligand design discourse in the MnBCA space – generalisations can be made from the difference between only two results (and applied to all MnBCAs) which may have more nuanced explanations than the obvious. Perhaps the most celebrated of the acyclic ligands in this space is PyC3A, Figure 2.1.[7] This ligand incorporates the (±)-*trans*-1,2-diaminocyclohexenyl backbone, and replaces an acetate arm for a methylpyridyl group – which as a more lipophilic L donor both decreases overall charge to -1 and increases liver clearance. $[Mn(PyC3A)(H_2O)]^{-}$ produced a stability constant of 14.14, and a marginal increase in relaxivity to 3.3 mM^{-1} s^{-1} compared to $[Mn(EDTA)]^{2-}$. The EDTA type complexes discussed above all have a capped-trigonal prism geometry when bound to water. Ligands with picolinate arms have shown promise, DPAAA and DPASAm (Figure 2.1) centre around a tertiary amine which takes a pentagonal bipyramidal geometry when bound to water. DPAAA and DPASAm have stability constants of 13.53 and 13.19, and relaxivities of 3.8 and 3.9 mM^{-1} s^{-1}. Surveying the 150 MnBCA ligands collated by Uzal-Varela and co-workers,[8] of the acyclic, q = 1 complexes (n = 18), the average stability constant is 13.12 (σ = 1.31) and r_1 is 3.56 mM^{-1} s^{-1} (σ = 0.42 mM^{-1} s^{-1}). These statistics indicate that outside of introducing completely novel designs, there is not much potential in producing acyclic MnBCA for KPIs that are larger than those measured from $[Mn(CDTA)(H_2O)]^{2-}$ or $[Mn(PyC3A)(H_2O)]^{-}$.

2.1.2 Macrocyclic ligands

Macrocyclic ligands tend to have larger stability constant values. Repeating the above analysis for q = 1, macrocyclic complexes (n = 16) – stability constant: μ = 13.55, σ = 3.12; r_1: μ = 2.93 mM^{-1} s^{-1}, σ = 0.51 mM^{-1} s^{-1}. The average stability constant is essentially the same (p = 0.60) but has much larger

variance – indicating greater potential for synthesising complexes that maximise this metric. Relaxivity on average for the macrocyclic complexes is lower ($p < 0.001$) – this may be indicating that the cage-like structures for macrocycles reduce the non-integer value of q – though as seen in Chapter 1, relaxivity is a function of many variables.

In general, structures based on DO2A, DO3A, or DOTA (DO*X*A complexes) tend to saturate the manganese coordination sphere producing $q = 0$ complexes. Whilst these complexes have larger stability constants, relaxivity suffers. For example, $[Mn(DO3A)]^-$ (Figure 2.2) has a stability constant of 19.43, but relaxivity of 1.3 mM^{-1} s^{-1} (indicating $q = 0$).[9] Of the DO*X*A based ligands, the $q = 1$ complexes tend to suffer from low stability. $[Mn(cis\text{-}BnDO2Am)(H_2O)]^{2+}$ has an r_1 of 3.8 mM^{-1} s^{-1} but a stability constant of 11.54;[10] and Mn(*trans*-O2DO2A), designed not to saturate and have X donors, has an r_1 of 2.86 mM^{-1} s^{-1} and stability constant of 9.38.[11] In 2018, Garda and co-workers (Tircsó/I. Tóth groups) published a Mn(PCTA) complex which, with its heptadentate ligand, was $q = 0$, $r_1 = 1.5$ mM^{-1} s^{-1}, and stability constant of 16.83.[12] This was improved by removing an acetate arm to better suit the ligand to manganese for 3,6- and 3,9-PC2A, Figure 2.2.[13] These complexes were able to provide much higher stability constants for $q = 1$ complexes than seen for the acyclic complexes seen in earlier published MnBCA work: 15.53 and 17.09 respectively. These complexes showed, however, a reduced r_1 (2.72 and 2.91 mM^{-1} s^{-1}) suggesting a compromise between stability and relaxivity for this family of ligands. Most impressive amongst the macrocyclic ligands is the bispidine family that was recently published. In 2022, the É. Tóth group reported the first complex to combine both exceptional stability (19.47), relaxivity (4.44 mM^{-1} s^{-1}), and kinetic stability against Zn^{2+}.[14] This is in large part due to the pre-organised shape of the bispidine matching the (relatively) large size of the Mn^{2+} ion extremely well, anchoring it in a position that leaves a space for a bound water molecule despite being a heptadentate donor.

2.1.3 Phosphasalens

In general, published MnBCAs rely on tertiary amines, carboxylates, or pyridines for donors. Apart from these, there are examples of phosphonates, sulfonamides, phenols (likely not deprotonated), and ethers being used as donors. The aim of this work was to investigate under-explored and novel ligand types/structures for Mn^{2+}. The first of these ligand families were phosphasalens, the phosphorus analogues of salens, where the imine bond is replaced *via* iminophosphoranes ($R_3P{=}N$ vs. $R_2C{=}N$). The iminophosphorane bond (and substituents on the phosphorus) allow for different, if not better, coordination chemistry characteristics than the more well-known imine bond due to larger electron-donation, more steric-hindrance (due to the often aryl substituents on the phosphorus), greater flexibility for the ligand skeleton (due to the intermediate nature between zwitterionic P^+-N^- or P=N forms), and theoretical tunability of electron donation *via* affecting the aryl substituents.[15] These factors are used to

influence catalysis reactions, with the majority of applications of phosphasalen research in the ring opening polymerisation of lactide, Figure 2.3.

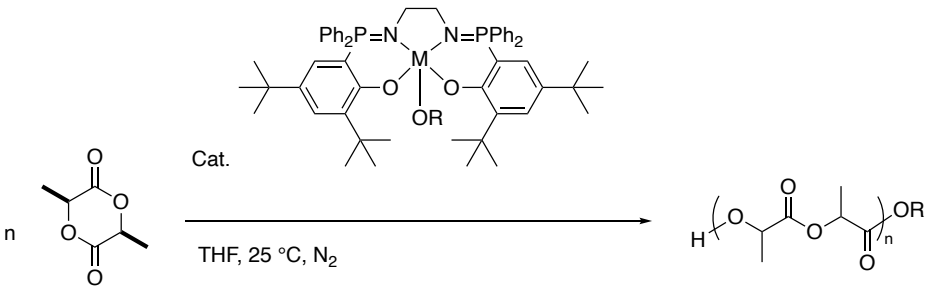


Figure 2.3: Representative lactide ring-opening polymerisation reaction catalysed by phosphasalen complexes, where M = Y, Sc, Zr.

As written above, the iminophosphorane bond, although shown as P=N, is better described as a P^{+}-N^{-} bond with two lone pairs on the nitrogen atom (one hybridised most closely to sp^2 (σ-donating); and the second as a pure p-orbital (π-donating)).[15] These two orbitals are able to donate into the phosphorus' σ* (P-R) orbitals (which are higher in energy), resulting in a net stabilisation and shorter P-N bond than would be seen from a P-N single bond (negative hyperconjugation).[16] Due to the lack of a 'double bond' between phosphorus and nitrogen, unlike the salen bond, iminophosphoranes lack π-accepting ability. Their donor ability is able to be influenced *via* modification of the substituents on the phosphine or the nitrogen. In addition, the steric bulk afforded to the iminophosphorane due to the phosphorus substituents should aid in preventing transmetalation.

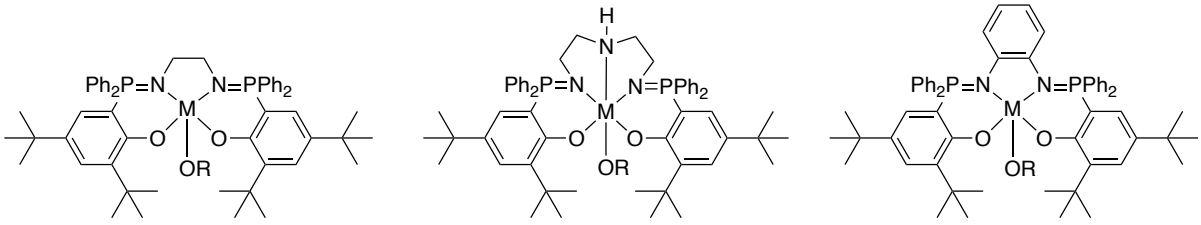


Figure 2.4: Representative phosphasalen complexes in literature, where M = Y, Sc, Zr, In, Ti.

Phosphasalens have typically been used as initiators for racemic lactide polymerisation, affording excellent reaction rates and *iso*-selectivities. In 2012 and 2013, the Williams, Long, and Auffrant groups published three articles focusing on yttrium phosphasalens that showed high activity and stereoselectivity in polymerising racemic lactide (plant-derived) to produce the biodegradable polymer polylactide – a feasible alternative to oil-derived plastics for some applications. The design of these phosphasalen ligands followed the general design on salen ligands, with a phenol substituent on the

phosphorus and an amine bridge between the two iminophosphoranes, Figure 2.4.[17–19] In 2015, Bakewell (of the Long and Williams groups) also used scandium as the metal in the complex in Figure 2.4 (centre) to initiate the ring-opening polymerisation of caprolactone; alongside this indium, zirconium, and titanium have been shown to form complexes with phosphasalens that catalyse ring-opening polymerisation reactions.[20,21] In each of these cases, mild coordination conditions (THF or toluene, N_2, 25 °C, 16 hours, metal halide) were able to provide the desired metal complexes – fast, room-temperature complexations are desirable for clinical translation (especially for a PET contrast agent).

2.1.4 Sulfonamides and amides

As emphasised with Caravan's Janus ligand,[22] there is a balance that must be struck between hard and soft donors, and stability. From a stability viewpoint, hard donors will chelate most strongly to the Mn^{2+} centre, disfavouring dechelation and optimising stability. However, pushing this balance too far will promote oxidation to the smaller, more charge dense Mn^{3+} centre, which could sacrifice contrast enhancement. The difficulty in MnBCA synthesis is to find donors that have more of an intermediate character, or a ligand containing mixed hard/soft donors, whilst also providing X donors to overall provide a monoanionic species for hepatobiliary clearance.[7]

Sulfonamides are able to provide a softer X donor than, for example, phenolates. These types of donors are under-explored in the MnBCA and GBCA space. So far, they have been typically used as pH sensitive handles on Gd-DO3A derivatives, to sense hypoxia in tumours. This has been inspired by work from Lowe and Parker, which showed that lanthanide luminescence could be modulated by the *para*-substituent on arylsulfonamide DO3A derivatives, Figure 2.5.[23] At a physiological pH of 7.4, these complexes largely existed as deprotonated, sulfonamide-bound species, which subsequently protonate as the pH drops. The more electron-withdrawing the substituent, the lower the pH required for sulfonamide protonation, and the greater the proportion of the ligand is existing as N-bound to Eu.

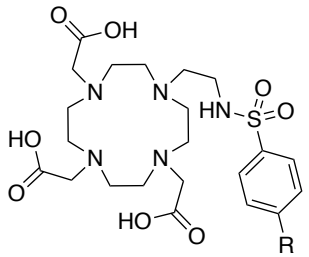


Figure 2.5: Sulfonamide ligand published by Lowe and Parker, where R = CF_3, CH_3, or OCH_3.[23]

In 2012, Iwaki and co-workers investigated the effect of *para*-substituents on relaxivity for sulfonamide GBCAs. Under hypoxic conditions, bio-reduction *via* nitro-reductases can transform nitro-substituted aryls to amino-substituted aryls, and as such, their study examined the change in contrast enhancement, r_1, between nitro- and amino-DO3A-sulfonamide derivatives. The study showed that the difference in r_1 at physiological pH changed depending on the position of the nitro (and therefore amino) group. However, more importantly the study showed that the Gd-sulfonamide bond was stable at pH 7.4 for all but one of their nitro-complexes (2,4,6-trimethyl-3-nitro-). In 2020, Platas-Iglesias and co-workers showed that a pendant sulfonamide could be used to create a pH-responsive MnBCA which changes from a q = 1 agent to a q = 2 agent at low pH, taking advantage of the increase in relaxivity in acidic conditions, Figure 2.1 and Figure 2.2.[24] These complexes showed moderate to good thermodynamic stabilities, log(K_{MnL}) values: NO2ASAm 13.53, DPASAm 15.29.

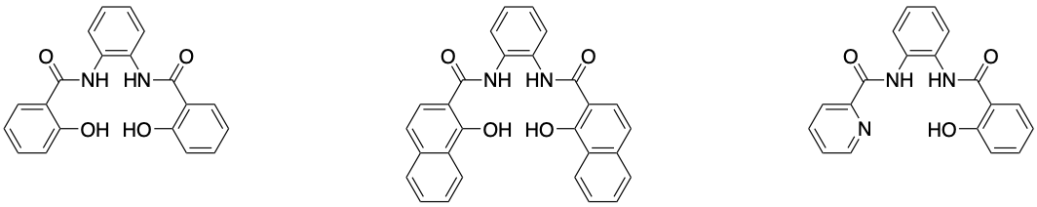


Figure 2.6: Phenylenediamide ligands for Mn^{3+} MnBCAs published by the Jasanoff group.[25]

Aside from q = 1 Mn^{2+} MnBCAs, q = 2 Mn^{3+} complexes have comparable, or better, relaxivity values. The Jasanoff group in 2016, published a set of phenylenediamide Mn^{3+} complexes, Figure 2.6, which provided large relaxivity values of 4.1, 5.4, and 5.1 mM^{-1} s^{-1} for the three complexes measured at 7 T at 22 °C.[25] These values show the potential utility of Mn^{3+} complexes for MnBCAs despite the lower number of unpaired electrons. In these complexes, the ligand design of two amide and two phenol donors provided hard enough X donors to ensure the Mn^{3+} oxidation state. The complexes synthesised by Jasanoff's group were then further functionalised to be both intracellular calcium ion sensors,[26] and nitric oxide sensors.[27]

2.2 Results and Discussion

2.2.1 Phosphasalens

Initial ligand families investigated were chosen to take advantage of the Long group's experience with phosphasalens, with the aim of creating rationally designed ligands based on the phosphasalen scaffold. Phosphasalens were chosen due to their tuneable electronic and geometric properties – which should provide control over thermodynamic and kinetic stability; and due to their mild transition metal complexation conditions – which should help in translating to a clinical application. Previous examples of phosphasalens in the literature have focused on ring-opening polymerisation,[17–20] however there are

no reported manganese phosphasalen complexes. Conclusions from studies on r_1 and the stability constant for MnBCAs were used to instruct the design of three initial phosphasalen motifs. These were designed to maximise kinetic and thermodynamic stability, and hydration number, whilst maintaining the general design of phosphasalens used in the group. This final consideration was key for being able to use the protocols developed in the Long group to leverage the expertise of fellow researchers. Once these 'prototype' designs were validated the intention was to either substantially increase ligand hydrophilicity and modify the protocols to accommodate the new ligands or to include handles for further functionalisation with PEG chains or sulfonates to improve solubility.

2.2.1.1 Initial phosphasalen ligands and DFT evaluation

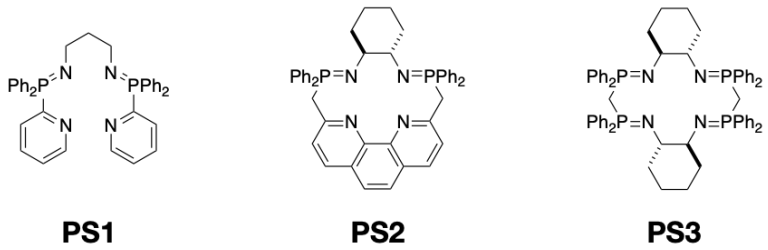


Figure 2.7: Initial phosphasalen designs.

Three phosphasalens were initially designed, Figure 2.7. These were designed as tetradentate ligands with phenyl groups to sterically prevent attack from endogenous metal ions and increase hydrophobicity for improved liver clearance. Pyridyl donors were chosen in **PS1** to accompany the iminophosphorane N donors to add to the soft donor nature of the ligand, thus avoiding spontaneous oxidation of Mn^{2+} to Mn^{3+}. The low coordination number (four) maximises the manganese hydration number, which was expected to be around 2. The desired Mn^{2+} complex (without any coordinated water) would have a molecular weight of 651.59 Da, which would position it as one of the largest molecular weight small molecular MnBCAs. Hydrated Mn(bispidine) complexes boast some of the highest r_1 values (4.3-5.0 mM^{-1} s^{-1}) for q = 1 complexes – their large r_1 values are likely a function of their molecular weight (the most effective are above 700 Da with Mn). Designing large molecular weight complexes follows a trend seen in GBCA work, where r_1 positively and linearly correlates with molecular weight (R^2 of 0.991) as investigated by Carniato and Botta.[28] This is expected to slow the rotational correlation time, (τ_R), as modelled in the Debye-Stokes equation (Chapter 1) where τ_R is proportional to 'effective radius' cubed i.e. larger molecular weight complexes will have larger effective radii.

PS2 incorporated a *trans*-1,2-diaminocyclohexane linker between iminophosphorane bonds – as they are expected to provide greater kinetic stability compared to a simple ethylene or *cis*-1,2-diaminocyclohexane linker.[3,29] This was linked to a phenanthroline unit connecting to give a 6-membered ring between the iminophosphorane, the manganese, and the phenanthroline nitrogen. The phenanthroline unit was chosen as a successor to the **PS1** design, linking the pyridyls, increasing rigidity in the ligand, enforcing planarity between the two N donors, and providing a macrocyclic unit which would increase thermodynamic stability *via* the chelate effect. As with **PS1**, **PS2** was designed to complex to two water molecules in solution, this (in addition to the large molecular weight (Mn(**PS2**) MW: 741.72) due to the four phenyl units) was expected to result in large r_1 values.

PS3 was designed to be one of the heaviest small molecule MnBCAs, with a molecular weight of 1044.06 Da. It comprises of two commercially available parts – the *trans*-1,2-diaminocyclohexane discussed before for greater kinetic stability, and bis(diphenylphosphino)methane – undergoing a [1+1+1+1] addition to create a more flexible macrocycle than **PS2**.

These designs were validated *via* DFT geometry optimisations, Figure 2.8–Figure 2.11, at UB3LYP/6-31G(d,p)/IEFPCM level. As expected, [Mn(**PS1**)$(H_2O)_2$].4H_2O optimised in an octahedral geometry with the ligand taking axial positions and the water molecules coordinating *trans* to each other. The size and steric bulk of the phenyl groups are captured in Figure 2.8. The flexibility in the ligand allows for both water molecules to comfortably coordinate to the manganese centre – Mn-O_{water} bond lengths of 2.22 and 2.19 Å. The bond angles about the Mn centre suggested an octahedral geometry – pyridyl-Mn-pyridyl had a bond angle of 110.5°, the deviation from a perfect 90° is likely due to the steric interactions between the C-Hs on the pyridyls preventing the pyridyls closing in around the Mn, or perhaps the large size of the manganese ion itself is the cause. The effect of this can also be seen in the slight deviation from being on the equatorial plane of the ligand. In the optimised geometry the axial water molecules had a O-Mn-O bond angle of 152°, which appeared to be a consequence of the phenyls optimising to squeeze the water molecules away from them. In solution, this may not be so pronounced as these optimisations required explicitly modelling the second sphere water molecules to capture a more accurate Mn-O_{water} bond length, which themselves are perhaps overexerting influence on the O-Mn-O bond angle.

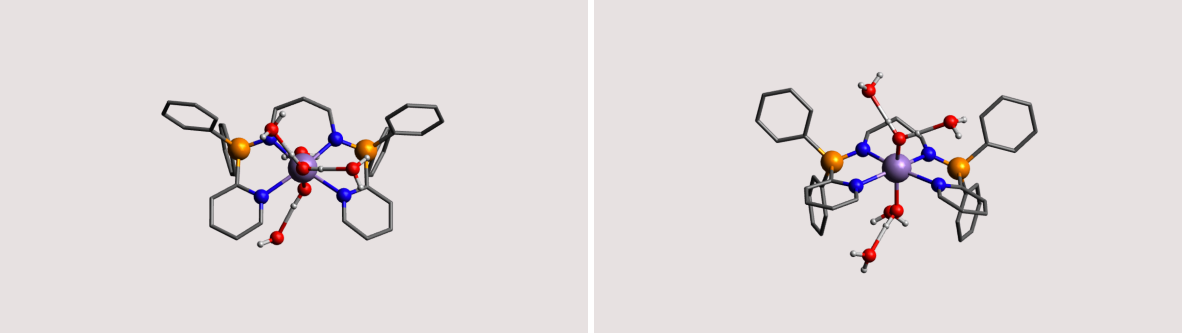


*Figure 2.8: UB3LYP/6-31G(d,p)/IEFPCM optimised geometry of [Mn(**PS1**)(H_2O)$_2$].4H_2O, two explicit water molecules per coordinated water molecule were included to aid Mn-O_{water} bond optimisation. C-H bonds removed, and C atoms represented with wireframe for clarity.*

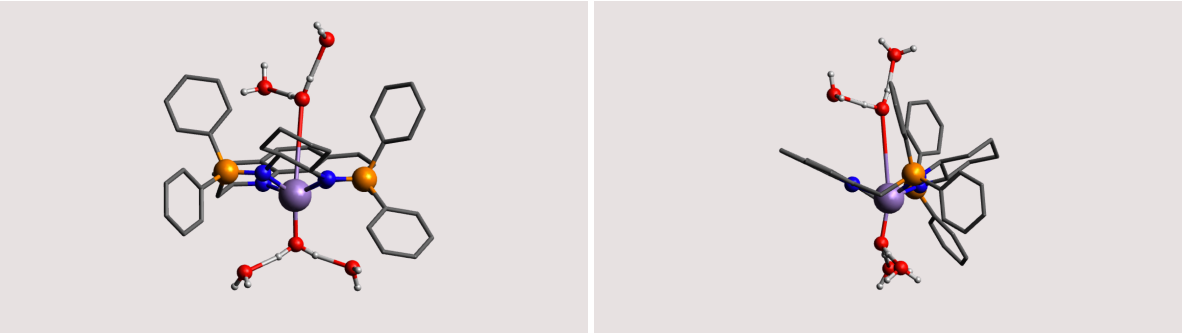


*Figure 2.9: UB3LYP/6-31G(d,p)/IEFPCM optimised geometry of [Mn(**trans-PS2**)(H_2O)].5H_2O, two explicit water molecules per coordinated water molecule were included to aid Mn-O_{water} bond optimisation. C-H bonds removed, and C atoms represented with wireframe for clarity. Right image shows the abnormally large Mn-O_{water} bond for one of the water molecules suggesting the complex would be* q = *1.*



*Figure 2.10: UB3LYP/6-31G(d,p)/IEFPCM optimised geometry of [Mn(**cis-PS2**)(H_2O)$_2$].4H_2O, two explicit water molecules per coordinated water molecule were included to aid Mn-O_{water} bond optimisation. C-H bonds removed, and C atoms represented with wireframe for clarity.*

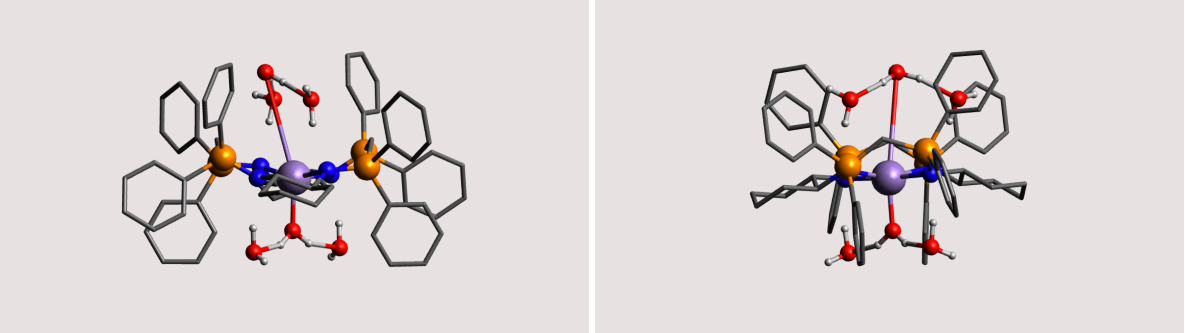


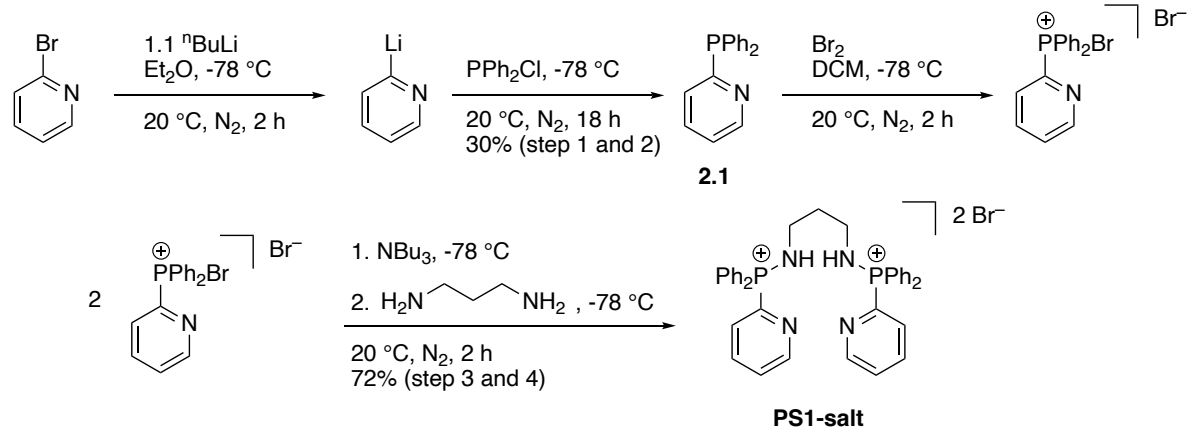
*Figure 2.11: UB3LYP/6-31G(d,p)/IEFPCM optimised geometry of [Mn(**PS3**)(H_2O)].5H_2O, two explicit water molecules per coordinated water molecule were included to aid Mn-O_{water} bond optimisation. C-H bonds removed, and C atoms represented with wireframe for clarity. Both images show the abnormally large Mn-O_{water} bond for one of the water molecules suggesting the complex would be* q = *1.*

The optimisation for [Mn(**PS2**)(H_2O)$_2$].4H_2O, Figure 2.9, suggested that the complex would actually be q = 1 in solution, as the second water molecule optimised with a Mn-O_{water} bond length of 4.21 Å – indicating that the bond would not form in solution. The Mn centre optimised towards a square pyramidal molecular geometry, akin to $MnCl_5^{2+}$, with the Mn^{2+} sitting above the pocket, indicating a problem with the design (where the expected behaviour was for the manganese ion to sit inside the pocket and coordinate to two water molecules). Interestingly, the ***cis*-** variant of [Mn(**PS2**)(H_2O)$_2$].4H_2O optimised with both water molecules with regular Mn-O_{water} bond lengths, suggesting that the rigidity of the *trans*-1,2,-diaminocyclohexane linker may not always be more beneficial to the overall MnBCA characteristics, Figure 2.10. The Mn ion sitting outside of the *trans*-complex's pocket potentially indicates competition with endogenous divalent donors, which may reduce the hydration number and r_1 value.

Similar to the ***trans*-PS2** complex, [Mn(**PS3**)(H_2O)$_2$].4H_2O optimised with one of the bound water molecules being ejected into the second hydration sphere – this is likely because of the large hydrophobic phenyls. The manganese ion however sits in the pocket a little better than **PS2**, suggesting the increased flexibility by removing the planar, aromatic phenanthroline resulted in a more regular geometry.

2.2.1.2 Synthesis

2.2.1.2.1 PS1



*Scheme 2.1: Synthesis of **PS1-salt**.*

The general method for phosphasalen synthesis follows a Kirsanov reaction to first react the phosphine with bromine, forming a reactive bromine-phosphorus intermediate adduct which is then subjected to nucleophilic attack by a primary amine to form a phosphasalen salt. This salt is deprotonated to give the iminophosphorane bond. For **PS1**, the required phosphine was generated *via* the synthesis reported by Brück and co-workers,[30] providing **2.1** in 30% yield. Phosphine **2.1** was then moved on to undergo a Kirsanov reaction to form **PS1-salt** in 22% overall yield, Scheme 2.1.

This synthesis followed that of Cao and co-workers,[18] and a light brown/off-white powder (72%) was achieved. Single crystals were grown *via* vapour diffusion of hexane in DCM to provide orange needles. X-ray structure elucidation was carried out by Dr Andrew White, Imperial College, showing the bromide ions H-bonding to the amine hydrogens of the phosphasalen salt, Figure 2.12. The P-N bond length was 1.623 Å, with the C-N-P bond angle at 124°, and P-N-H angle at 117°, the dihedral between C-H-N-P was 150.5°. These values indicate that the character on the nitrogen is between sp^2 and sp^3, being closer represented as a sp^2 N – the N-P bond length is not too far from that of N=P seen in literature, 1.591 Å.[31]

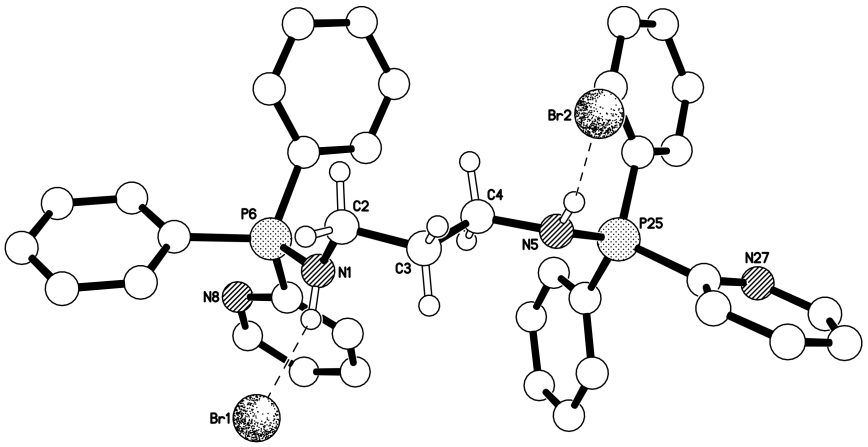
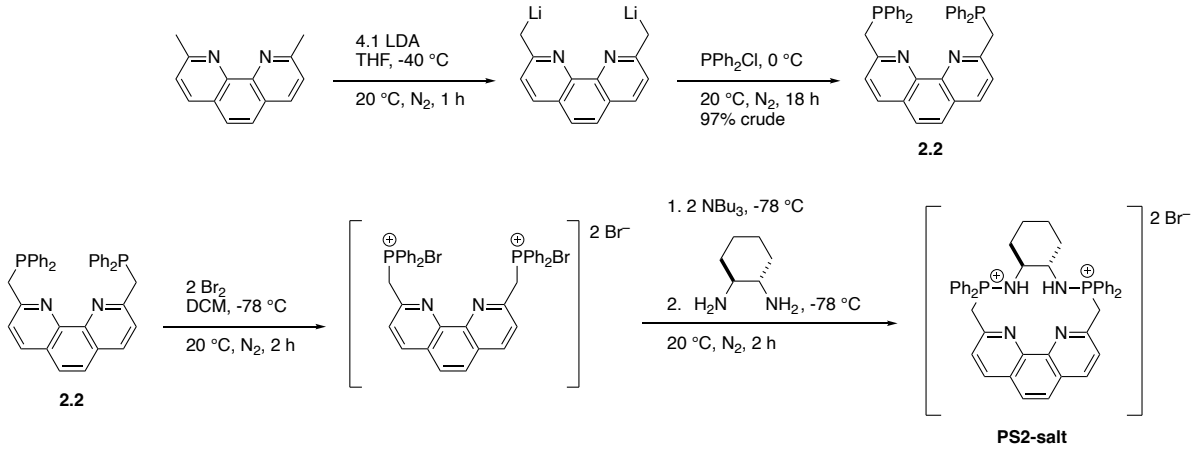


*Figure 2.12: X-ray crystal structure diagram of **PS1-salt**. The ellipsoids are drawn at the 50% probability level.*

Further synthetic efforts were abandoned after the first COVID-19 lockdown and subsequent project redesign due to phosphasalen limitations which is further elaborated on in Section 2.2.1.3.

2.2.1.2.2 PS2



*Scheme 2.2: Synthesis of phosphine **2.2**,[32] and **PS2-salt** via established Kirsanov route.*

The phosphine for **PS2** was synthesised *via* an adapted protocol from Takeshita and co-workers,[32] Scheme 2.2. **2.2** was synthesised in 97% crude yield. Attempts to purify consistently led to oxidation to the phosphine oxide. A dry silica plug under a nitrogen atmosphere was used which was also unable to cleanly separate the phosphine from the phosphine oxide, Figure 2.13. LC-MS analysis showed that the major compound was the expected product found at *m/z* 577.2 $[M+H]^+$ (calc'd 577.19). Additional peaks were attributed to mixed single/double phosphines and their respective oxides. The crude mixture was taken forward, with the intention of purifying the final ligands *via* flash column chromatography. The phosphine **2.2** was reacted to form the **PS2-salt**, and the phosphasalen salt from the

monophosphinated neocuproine, as well as the deprotonated, final phosphasalen forms of both of the salts. Multiple flash column chromatography separations with conditions suggested by the Williams group were unsuccessful,[17,18] suggesting that the purification of the phosphine would need to prioritised.

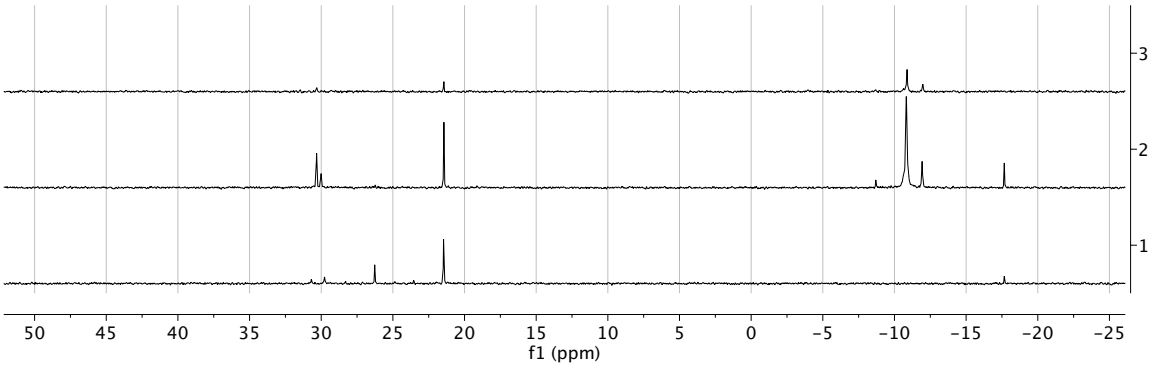
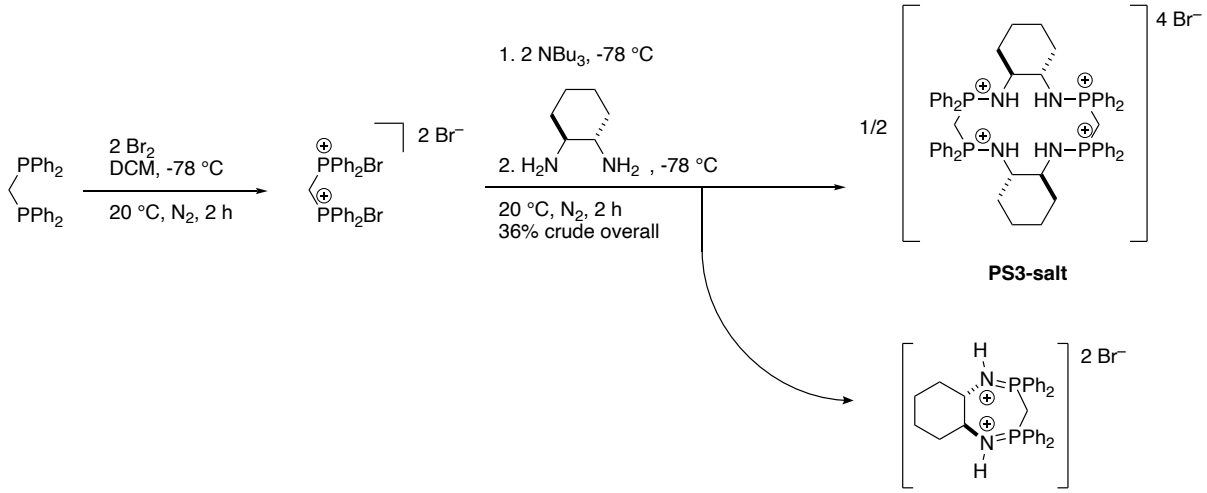


*Figure 2.13: $^{31}P\{^{1}H\}$ NMR (121.5 MHz, $CDCl_3$, δ) spectra of the fractions collected from the silica plug to purify **2.2**. Product expected at -10.7 ppm. Fraction 1 (bottom) only contained expected oxidised products. Fraction 2 (middle) contains both oxides and phosphines. Fraction 3 (top) has a greatly reduced proportion of oxides, however an increased proportion of (expected) phosphine by-product at -11.5 ppm.*

As above, further synthetic efforts were abandoned after the first COVID-19 lockdown and subsequent project redesign due to phosphasalen limitations which is further elaborated on in Section 2.2.1.3.

2.2.1.2.3 PS3



*Scheme 2.3: Synthesis of **PS3-salt** via established Kirsanov route, possible by-product also shown.*

The Kirsanov reaction using 1,1-bis(diphenylphosphino)methane (DPPM) was carried out to produce an off-white powder (0.18 g, 36% crude), Scheme 2.3. $^{31}P\{^{1}H\}$ NMR analysis of this solid showed the presence of four environments, three of which were unexpected, Figure 2.14. ES^{+} MS results showed evidence that the expected product could have formed – $[M]^{4+}$ *m/z* calc'd: 248.1098, found: 248.1093; the main peak was at *m/z* 495.2109 – which implies a $[M]^{2+}$ species (calc'd: 495.2114). Perhaps more thermodynamically feasible than a macrocycle formation from four individual pieces, is for one amine bridge to fully react with the DPPM. This would provide an identical MS signature to the desired macrocycle but would be entropically more favourable.

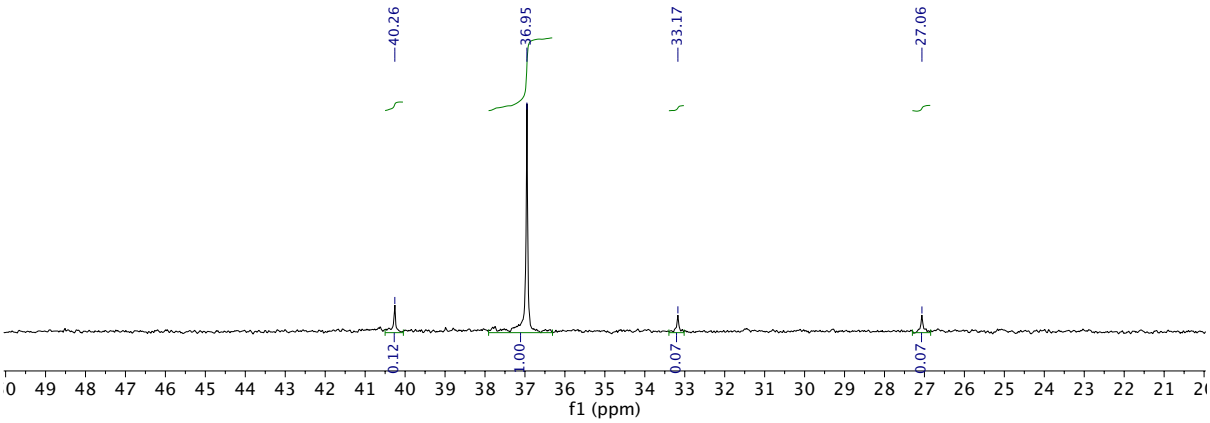
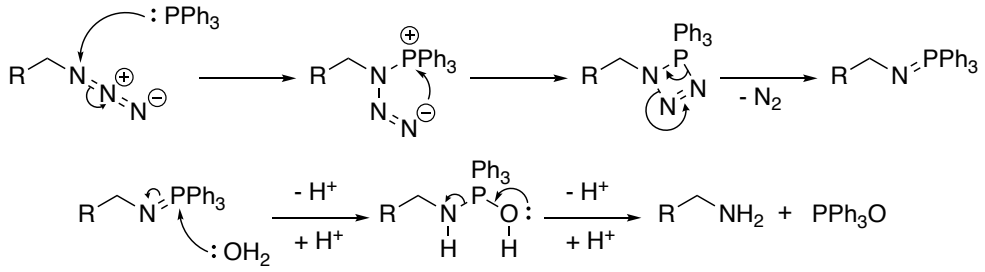


*Figure 2.14: $^{31}P\{^{1}H\}$ NMR (121.5 MHz, $CDCl_3$, δ) spectra of collected solid in the attempted synthesis of **PS3-salt**.*

As above, further synthetic efforts were abandoned after the first COVID-19 lockdown and subsequent project redesign due to phosphasalen limitations which is further elaborated on in Section 2.2.1.3.

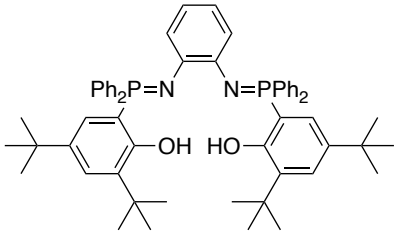
2.2.1.3 Limitations



Scheme 2.4: Reaction mechanism of the Staudinger reaction for the conversion of alkyl azide to alkyl amine.

Synthetic efforts were diverted away from the phosphasalen family after a review of different routes of formation revealed that the critical step in the alternative route (the Staudinger reaction, the reduction of an alkyl azide to alkyl amine, Scheme 2.4) consumes the formed phosphasalen by nucleophilic attack with water. As the driver for the reaction is also the intended final solvent of our MnBCAs, this likely invalidates any potential route with phosphasalens. It is also expected that any phosphasalen MnBCAs will exist in a dynamic equilibrium between complex, and metal + ligand; hydrolysis of the free ligand to thermodynamically stable products will eventually consume the complex leaving free manganese, phosphine oxides, and bridging amines throughout the body.

To confirm these suspicions, a previously made phosphasalen from another researcher in the Long group, **CB134** (Figure 2.15), was stirred in water for 24 hours, and showed 93% conversion of the phosphasalen peak to a broad environment at 25.5 ppm (Figure 2.16, expected phosphine oxide range is between 25-30 ppm). The instability to water, difficulty in phosphine synthesis (synthesising tuneable, water-soluble phosphines is likely to require very harsh conditions), and orthogonality of literature's mild coordination conditions to the desired final water-soluble complexes meant that whilst the phosphasalen family are theoretically very interesting with a lot of potential, in practice they would not be worthwhile ligands for a MnBCA.



CB134

*Figure 2.15: Phosphasalen **CB134** used for hydrolysis testing.*

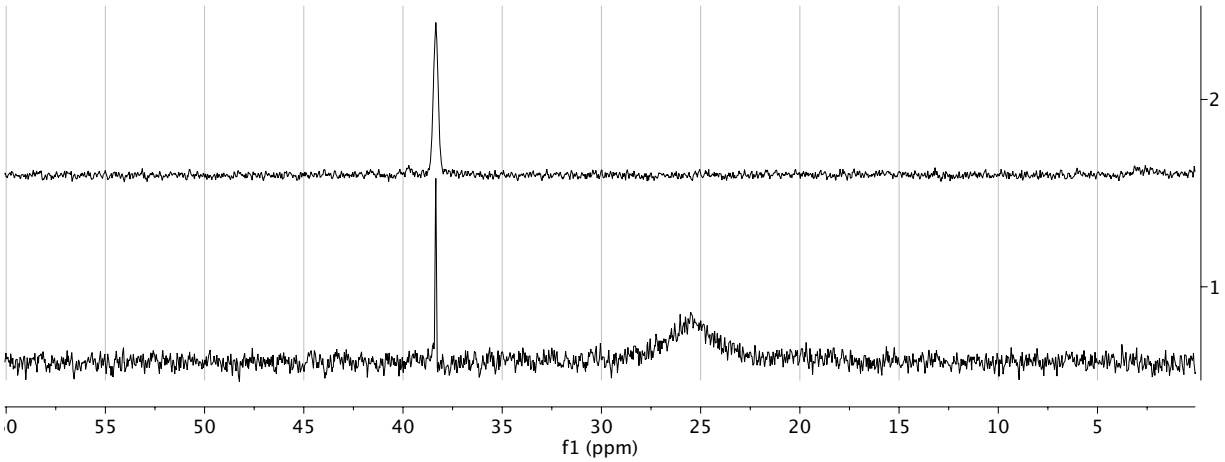


*Figure 2.16: $^{31}P\{^{1}H\}$ NMR (121.5 MHz, $CDCl_3$, δ) spectrum of **CB134** before hydrolysis (top), and after (bottom). Typical range of phosphine oxides is 25-30 ppm.*

2.2.2 Sulfonamides

The designs of the sulfonamide family were created to exactly replace the ill-fated iminophosphorane bond in the phosphasalen designs, whilst improving on their design to aim for maximising the design considerations from Section 1.4.2.1. The sulfonamide bonds would provide two X donors, countering the Mn^{2+} charge, providing an opportunity to find suitable amine bridges which could both provide another X donor and make the overall complex monoanionic.

The original design of these sulfonamide complexes followed the phosphasalen design in which a neocuproine unit would make up the lower half of the ligand, and then sulfonamides and an amine bridge would complete it – Figure 2.17. This leads to an equatorial arrangement of one 5-membered ring, and three 6-membered rings.

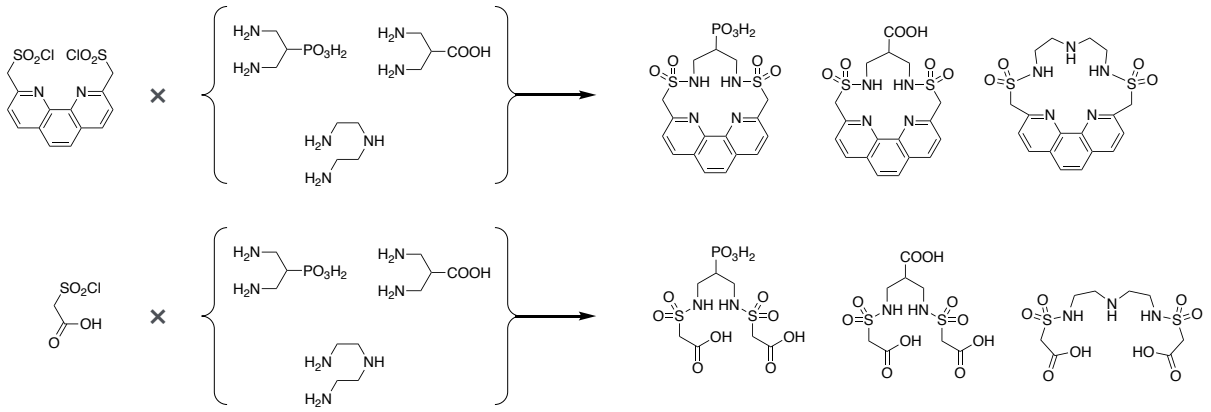


Figure 2.17: Initial designs for the sulfonamide family.

However, DFT geometry optimisations showed that the methylene bridges were leading to sub-optimal coordination, Figure 2.18, and that only having four equatorial donors in a 15-membered binding pocket led to the phenanthroline unit residing off the equatorial plane, with non-ideal binding. Removing these methylenes led to the manganese ion sitting above the pocket, which appeared to be a combination of a short acetate arm, and strain in the equatorial ring still tilting the phenanthroline. This was remedied by moving towards a pentagonal bipyramidal structure in which the phenanthroline unit sits on the equatorial plane and the acetate arm has enough length to freely bind axially.

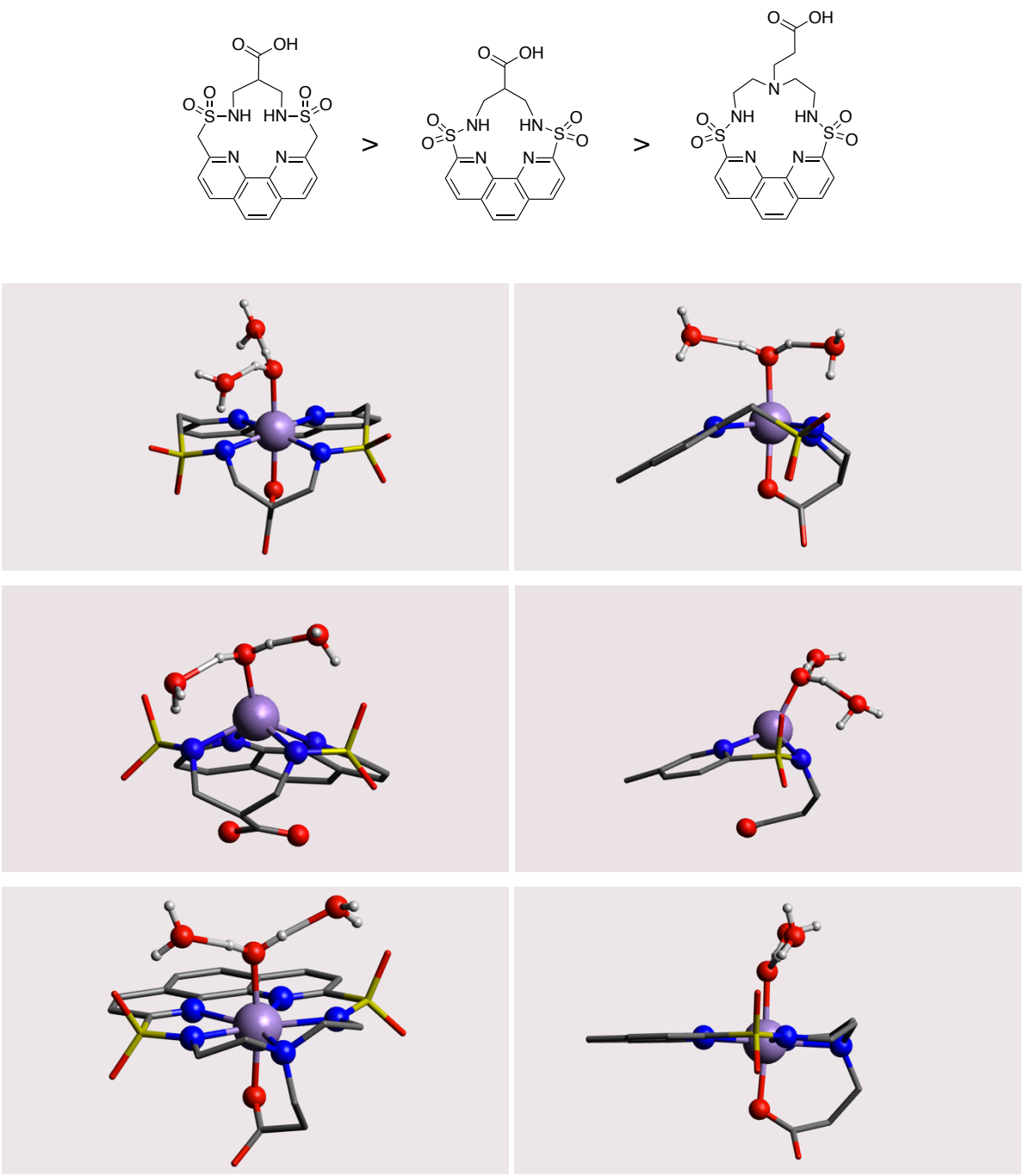


Figure 2.18: Evolution of the phenanthroline designs from octahedral to pentagonal bipyramidal. UM06-2X/TZVP/IEFPCM optimised geometries of sulfonamide phenanthroline designs, with two explicit water molecules per coordinated water molecule were included to aid Mn-O_{water} bond optimisation. C-H bonds removed, and C atoms represented with wireframe for clarity. First design shows imperfect phenanthroline binding to Mn^{2+} (upper), then out of pocket binding and lack of binding from carboxylate to Mn^{2+} (middle), final design aims for pentagonal bipyramidal (lower).

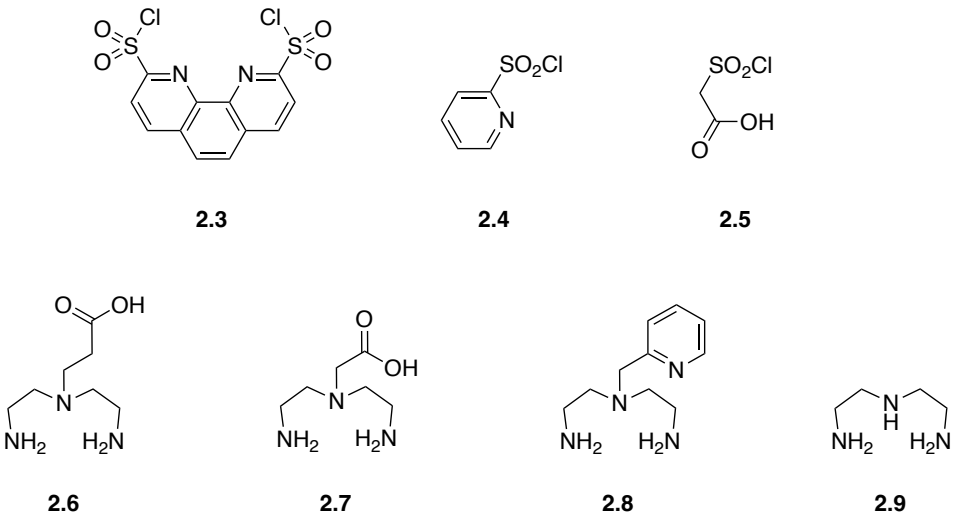
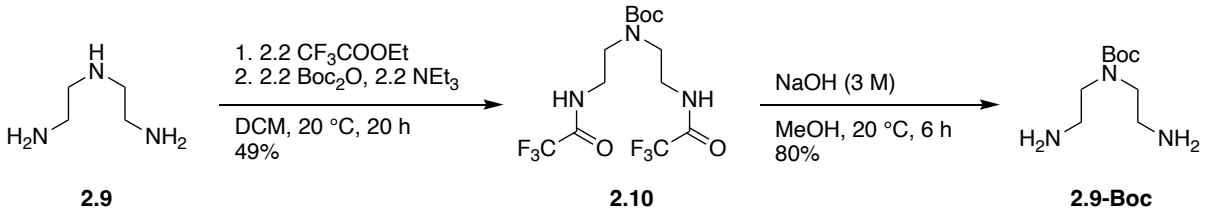


*Figure 2.19: Redesigned sulfonamide components to produce both acyclic and macrocyclic sulfonamide ligands; planned sulfonyl chlorides (**2.3-2.5**); redesigned set of amine bridges (**2.6-2.9**).*

This led to the development of seven components for sulfonamides, Figure 2.19. Those based around **2.3** would be macrocyclic with a phenanthroline base joined *via* sulfonamide bonds to the set of amine bridges, the other ligands would be acyclic with acetate or pyridyl arms. The amine bridges (Figure 2.19, lower) were designed to provide an extra X donor to aid in producing an overall monoanionic complex (**2.6** and **2.7**), or in the case of the acyclic sulfonamides from **2.5**, provide a link between arms and a final L donor to leave one coordination site free for water coordination.

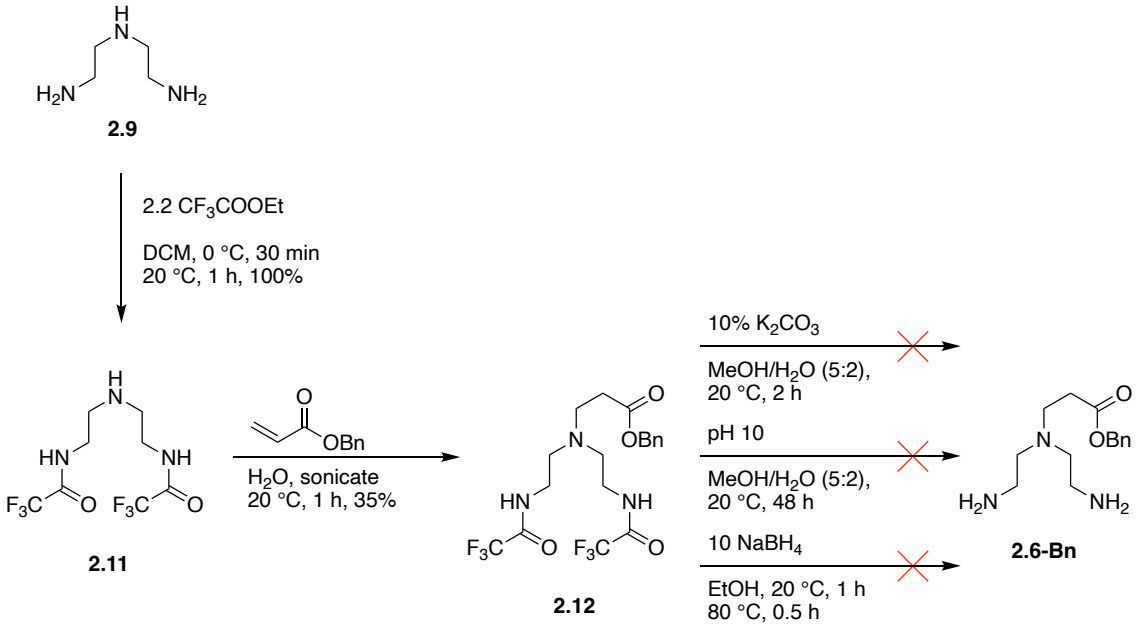
2.2.2.1 Development of amine bridges

Amine bridge development for all of the functionalised amine bridges centred around diethylenetriamine (Figure 2.19, **2.9**). The main synthetic challenge was finding orthogonal chemistry for primary and secondary amine (de)protection and functionalisation. In its final state before sulfonamide linking, an amine bridge needs to have free primary amines, whilst also potentially protecting any functional group on the secondary amine.

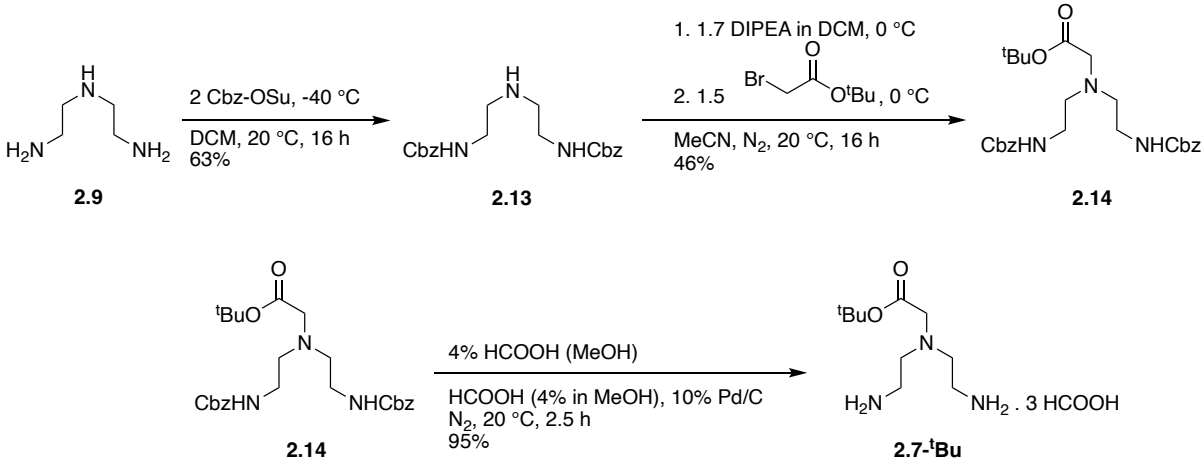


*Scheme 2.5: Synthesis of **2.9-Boc**.*

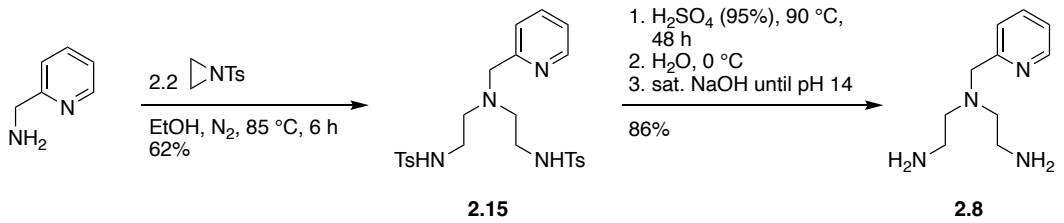
The simplest of the amine bridges, **2.9-Boc** (Boc protected **2.9**), was achieved *via* orthogonal protecting strategies, preferentially protecting the primary amines before Boc protection of the secondary amine, and finally deprotecting the primary amines in 39% overall yield, Scheme 2.5.



*Scheme 2.6: Synthetic attempts for **2.6-Bn**.*



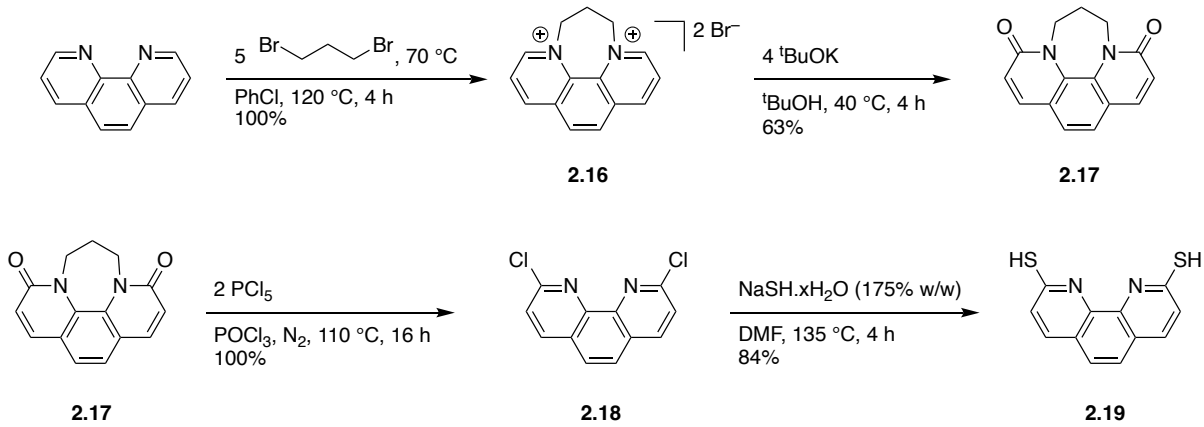
*Scheme 2.7: Synthesis of **2.7-^{t}Bu**.*



*Scheme 2.8: Synthesis of **2.8**.*

The synthetic attempts for **2.6-Bn** are shown in Scheme 2.6. Difficulty in synthesising **2.6-Bn** *via* standard N-alkylation routes led to an aza-Michael addition aided by sonication with benzyl acrylate, which was able to provide the completely protected product at 35% yield. However, attempts to selectively deprotect the primary amines failed. This led to synthesising a version of **2.6-Bn** with a shorter acetate arm, **2.7-^{t}Bu**, Scheme 2.7. **2.8** was synthesised from 2-(aminomethyl)pyridine in an overall yield of 53%, Scheme 2.8.

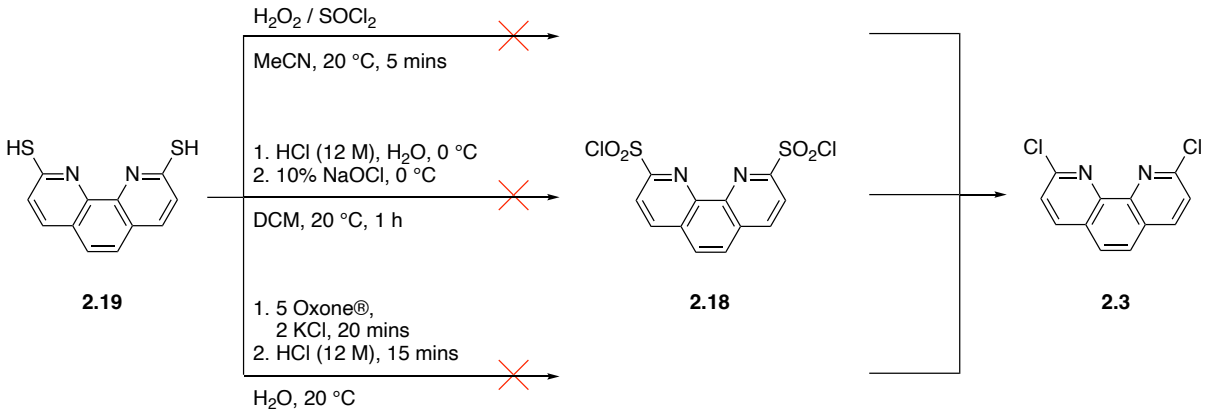
2.2.2.2 Attempted syntheses of 1,10-phenanthroline-2,9-disulfonyl dichloride



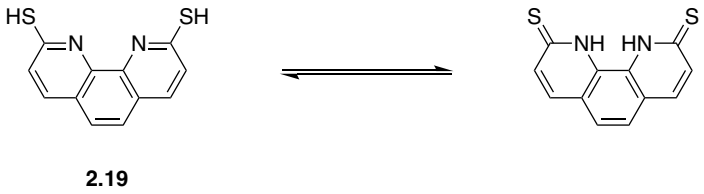
*Scheme 2.9: Synthesis of **2.19** from phenanthroline.*

The typical route towards a sulfonyl chloride is *via* oxidative chlorination of the thiol. As such, the precursor to **2.3**, **2.19** was synthesised in 53% overall yield, Scheme 2.9, *via* a combination of routes published by Guo and co-workers, and Krapcho and co-workers.[33,34] Attempts to convert **2.19** to **2.3** are shown in Scheme 2.10. These were based on adaptations of methods to produce 2-pyridinesulfonyl chloride and are the standard or general methods to synthesise sulfonyl chlorides. All three oxidative chlorination methods produced 2,9-dichloro-1,10-phenanthroline (**2.18**) instead. These methods vary in chloride source ($SOCl_2$, HCl, KCl/HCl) and oxidation agent (H_2O_2, NaOCl, Oxone®), which points to the starting material itself, **2.19**, being the issue. This is thought to be due to **2.19** tautomerising to the thioketone due to the nitrogen atoms on the phenanthroline readily accepting the thiol hydrogen,

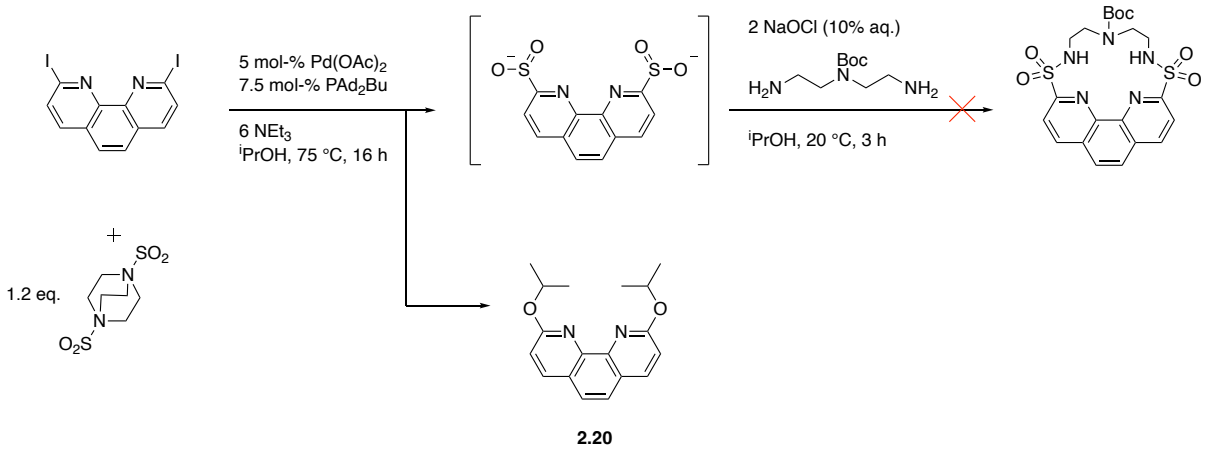
Scheme 2.11. This leaves a highly electron deficient centre to attack, with multiple possible routes to oxidatively remove the sulfur and leave a chloride in its place.



*Scheme 2.10: Synthetic attempts to form **2.3** via typical sulfonyl chloride methods resulted in **2.18**.*



*Scheme 2.11: Possible tautomerism of **2.19** preventing oxidative chlorination.*

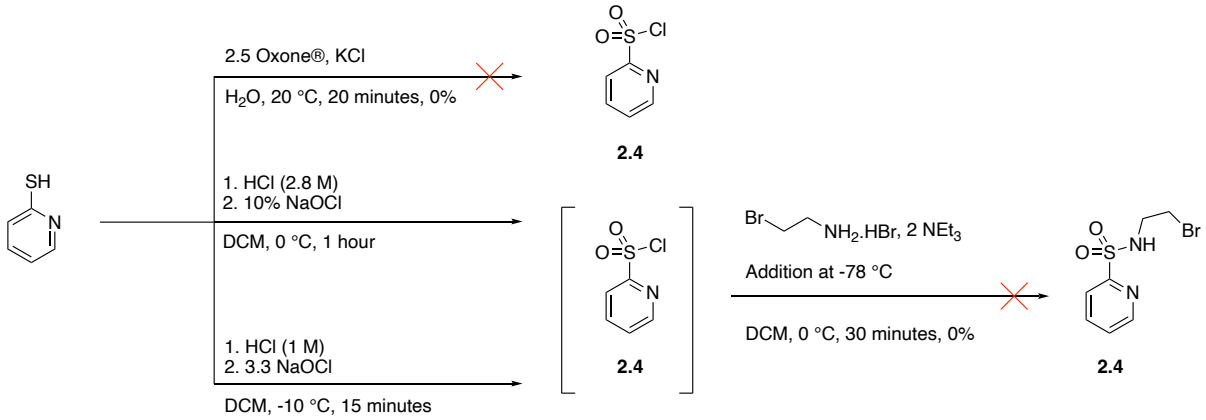


Scheme 2.12: Final synthetic attempt at forming phenanthroline sulfonamides resulted in the trapping of the phenanthroline species as a di-isopropyl ether.

With a potential inability of **2.19** to be a precursor for **2.3** *via* typical methods, a final method *via* palladium coupling of 2,9-iodo-1,10-phenanthroline was attempted, Scheme 2.12. This method was adapted from syntheses published by researchers in the Willis lab,[35] producing an aryl sulfinate *in-situ*

that is trapped by an amine (in this case readily synthesised **2.9-Boc**). Instead of any sulfonamide coupling products, 2,9-diisopropoxy-1,10-phenanthroline (**2.20**) was produced, most likely as a result of reductive elimination of phenanthroline with the iPrOH solvent from the palladium catalyst. This behaviour is likely due to the coordinating nature of the starting material as a ligand for the palladium catalyst, blocking the sulfinate formation pathway. With this unsuccessful attempt, all known routes towards **2.3** were exhausted.

2.2.2.3 Attempted syntheses of pyridine-2-sulfonyl chloride



Scheme 2.13: Synthetic attempts of oxidative chlorination of 2-mercaptopyridine.

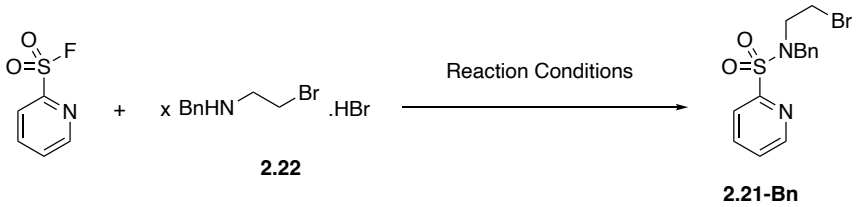
As with the attempted synthesis of **2.3**, following published methodologies to synthesise **2.4** also ended without success, Scheme 2.13. As above, oxidative chlorination was attempted *via* Oxone®/KCl with no evidence of the product forming *via* ^{1}H NMR analysis. Literature review of syntheses of chemicals similar to **2.4**, revealed that these types of aza-aromatic chemicals can be unstable once isolated, or can decompose to the chloride (as repeatedly seen above with **2.3**) without cautious temperature control. With this in mind, conditions were found which carried out *in-situ* oxidative chlorination and trapping of the sulfonyl chloride with a test amine at reduced temperature, also without success.

2.2.2.4 Sulfonamide formation reactions

Experimental attempts to synthesise any of the sulfonyl chlorides consistently failed, including attempts to convert methyl thioglycolate to the methyl protected analogue of **2.5**. Therefore, both methyl 2-(chlorosulfonyl)acetate and PyFluor (pyridine-2-sulfonyl fluoride) were purchased. Sulfonyl fluorides (such as PyFluor) are able to participate in sulfur(VI) fluoride exchange (SuFEx) reactions, which as a subset of 'click' chemistry reactions, are high-yielding, fast, water tolerant reactions that lead to thermodynamically stable products – in this case sulfonamide formation.[36]

2.2.2.4.1 SuFEx reactions

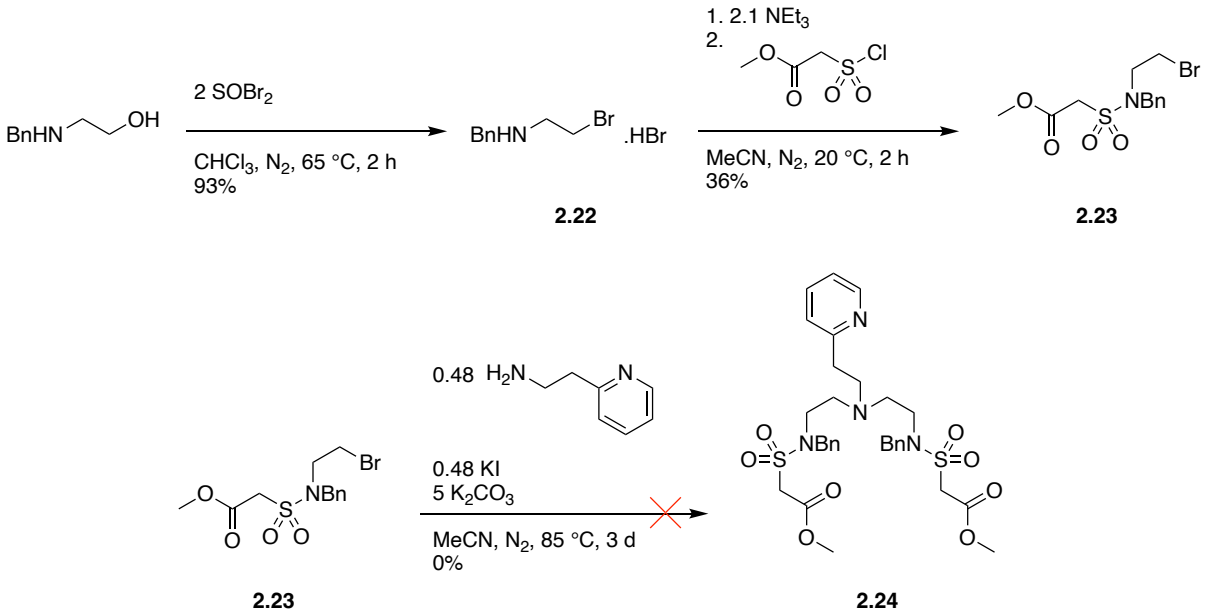
Table 2.1: Reaction conditions and yields for the SuFEx reaction between PyFluor and N-benzyl 2-bromoethan-1-amine hydrobromide.



Amine eq. (x)	Solvent	Base	Time	Temp/°C	Yield/%	Other conditions	Reference
2	H_2O	-	3 d	20	0 (all PyFluor)	-	-
2	THF	-	3 d	20	0 (all PyFluor)	-	[37]
1	DCM	-	3 d	20	0 (all PyFluor)	-	[38]
1	MeCN	NEt_3 (2.1 eq.)	3 d	20	0 (all PyFluor)	-	[39]
0.5	DMF[a]	DBU (1.75 eq.)	6 h	80	0	-	[40]
1.1	MeCN	Cs_2CO_3 (3.1 eq.)	2 h	20	0 (all amine)	-	-
1.2	MeCN (anhydrous)	NEt_3 (2.5 eq.)	1 h	20	0 (all PyFluor)	N_2	[41]
1.2	MeCN-d_3	NEt_3 (2.5 eq.)	1 h	60	0 (all PyFluor)	N_2	-
1.2	MeCN (anhydrous)	NEt_3 (2.5 eq.)	10 mins	150	0 (PyFluor and degradation)	Microwave reaction	-

Reaction conditions were tested to form a sulfonamide with **2.22**, with the view of using **2.21-Bn** to doubly N-alkylate an ethylamine donor to build the amine bridge piece by piece (analogous to Scheme 2.14). Nine attempts were made without appreciable success, Table 2.1. Ultimately, it may have been the steric bulk of the benzyl protecting group that stopped this reaction working. With both the sulfonyl chloride synthesis failure and the PyFluor inability to couple to form desired intermediates, the pyridyl sulfonamides were abandoned.

2.2.2.4.2 Methyl acetate sulfonamides



*Scheme 2.14: Attempted synthesis of **2.24** via N-alkylation of 2-pyridylethylamine.*

Unlike in Section 2.2.2.4.1, methyl 2-(chlorosulfonyl)acetate was much more pliable to forming sulfonamides. The equivalent planned reaction from Section 2.2.2.4.1, was carried out in Scheme 2.14, with a yield of 36%. Attempts to couple with 2-pyridylethylamine were unsuccessful, both with and without KI. Unfortunately, this reaction avenue could not be fully explored due to ill health, further elaborated in Section 2.3.2.

2.2.3 Amides

As with the sulfonamide family, the amide family neatly replaced iminophosphorane bonds whilst also being X donors themselves. The fear with producing amido-based MnBCAs is the tendency to spontaneously convert to less effective Mn^{3+} MnBCAs (as well as the drop in the number of unpaired electrons, the ligand designed for the desired Mn^{2+} ion would be less than ideal for a Mn^{3+} ion). Jasanoff and co-workers synthesised planar tetracoordinate ligands with hard donors such as phenolates favouring the oxidised Mn^{3+} product.[25] A similar effect can be seen with Jacobsen's catalyst (a salen with two phenolates),[42] where $Mn(OAc)_2.4H_2O$ can be used as the precursor metal and bubbling with air can cause spontaneous oxidation to a Mn^{3+} complex. However, the Janus chelator designed by Caravan and co-workers shows that controlling the donor hardness and softness can lead to one form over the other.[22] With this in mind, the amide family either modifies the structures of the phenanthroline sulfonamide ligands above, or takes the diamide structure from Jasanoff's ligands and includes softer donors (carboxylates and pyridyls) in an attempt to favour Mn^{2+} over Mn^{3+}.

2.2.3.1 Attempted syntheses of macrocyclic amides

The previous design for a macrocyclic phenanthroline-based sulfonamide complex optimised very neatly *via* DFT despite its inability to be synthesised. This ligand was modified to contain amides instead and produced a sensible optimised geometry *via* DFT at M06-2X/TZVP/IEFPCM level, Figure 2.20. The complex largely adheres to the designed pentagonal bipyramidal geometry, and when viewed through the equatorial plane the phenanthroline unit is binding at the expected angle, as well as the other three nitrogen atoms in the ring. The axial donors had a O-Mn-O_{water} angle of 169°, and the O_{water}-Mn-N_{axial} average bond angle was 88.5° suggesting the desired complex would have a regular pentagonal bipyramidal geometry around the manganese.

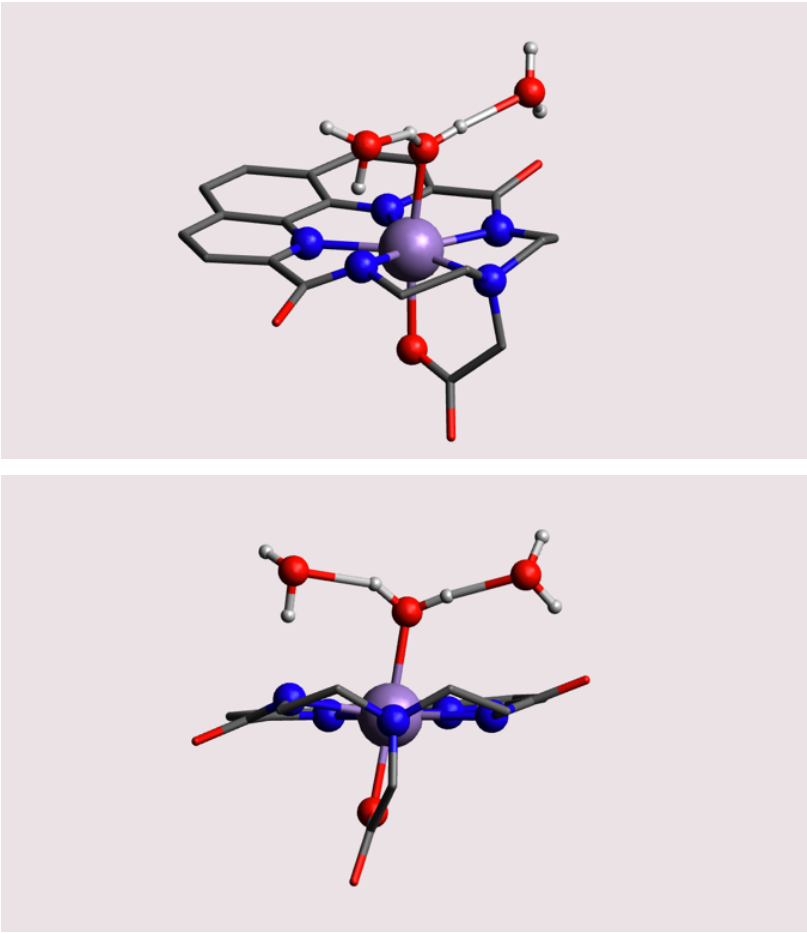
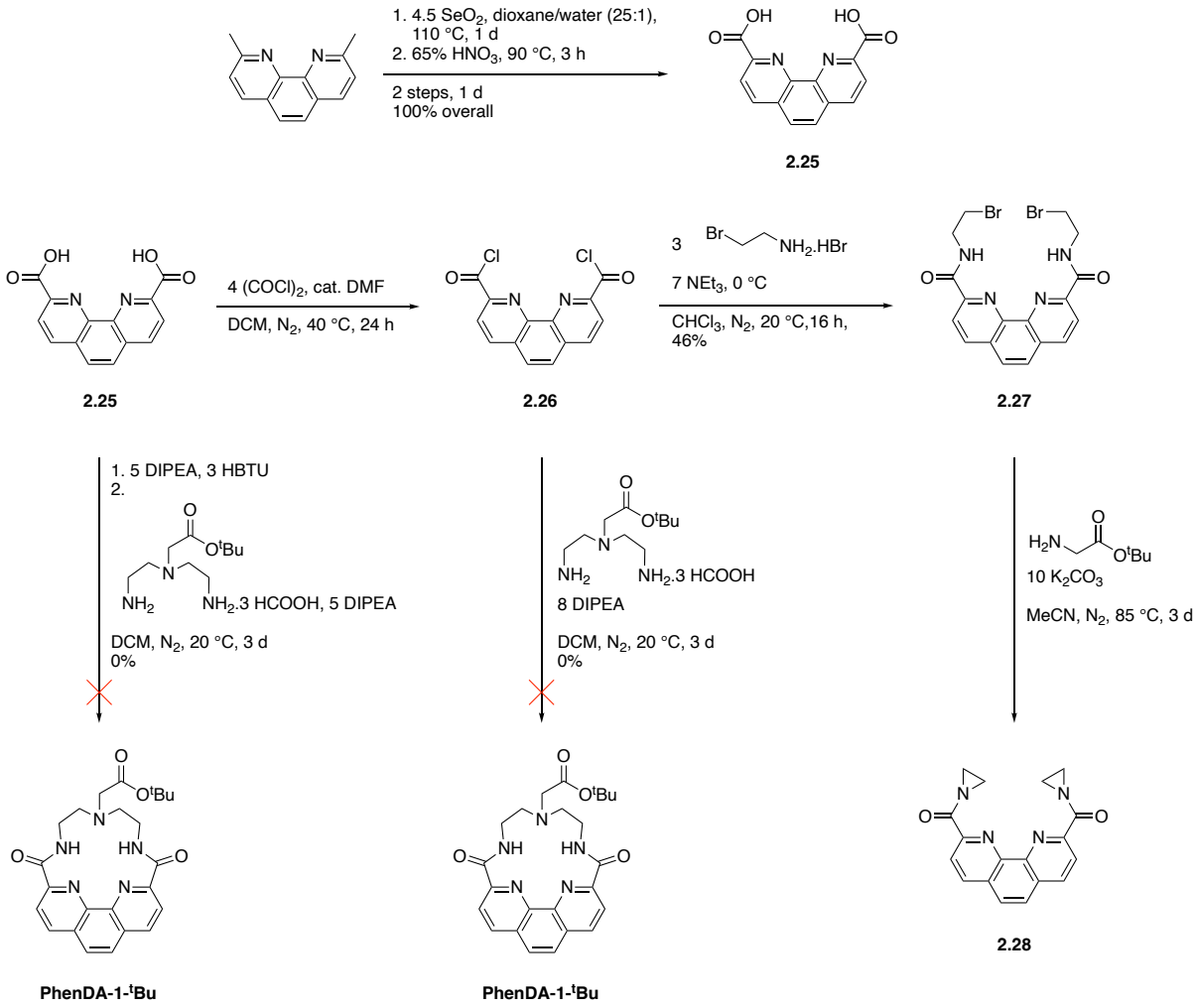


*Figure 2.20: UM06-2X/TZVP/IEFPCM optimised geometry of [Mn(**PhenDA-1**)(H_2O)].2H_2O, two explicit water molecules per coordinated water molecule were included to aid Mn-O_{water} bond optimisation. C-H bonds removed, and non-donor atoms represented with wireframe for clarity.*



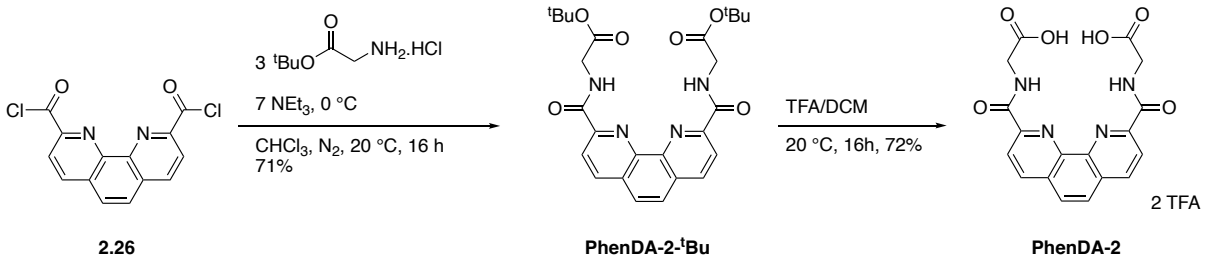
*Scheme 2.15: Attempted synthesis of **PhenDA-1-^{t}Bu**, and eventual synthesis of the aziridine amide instead.*

Synthetically, producing the complex was as challenging and as successful as its sulfonamide analogues, Scheme 2.15. The precursor carboxylic acid was produced from neocuproine in 100% yield, and acyl chloride was produced with oxalyl chloride and reacted without purification. From these, three strategies were attempted to either couple **2.7-^{t}Bu** or form the macrocycle *via* N-alkylation with *tert*-butyl glycinate. Both of the direct routes (and their repeat attempts) did not produce evidence of the macrocycle *via* LC-MS of the reaction mixture, nor in collected fractions from flash column chromatography. With these two direct approaches failing, a different route was designed to form the macrocycle *via* ring-closure at the top of the structure, Scheme 2.15. The synthesis of **2.27** was completed in 46% yield (additionally confirming the formation of the acyl chloride). However, the N-alkylation was not as successful, with LC-MS monitoring during the reaction showing that **2.27** was preferentially reacting with itself, deprotonating the amide, and forming aziridines. Unfortunately, these aziridines did not go on to ring-open with the *tert*-butyl glycinate, finishing the reaction with no

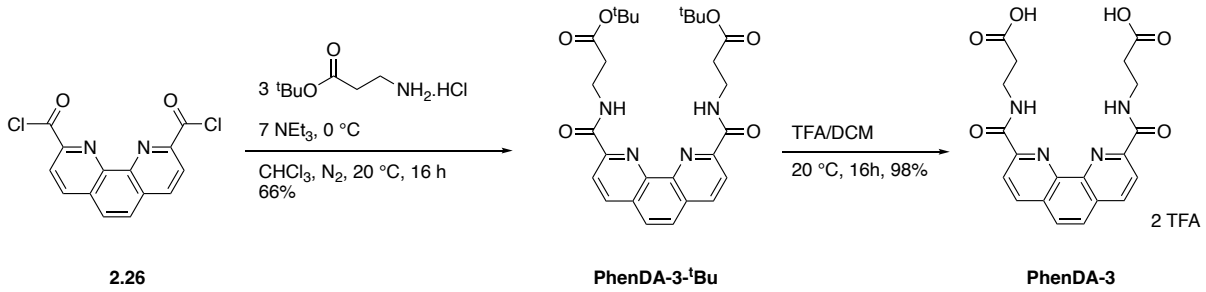
evidence of product formation – or even N-alkylation of just one side – occurring. The repeated failed attempts at synthesising this novel macrocycle drew a line under the many attempts at producing sulfonamide or amide macrocycles for MnBCAs.

2.2.3.2 Acyclic amides

With the difficulty in producing macrocyclic ligands, as well as a desire to test if the acyl chloride was being formed, acyclic diamido-phenanthroline based ligands were synthesised in parallel to the macrocyclic work. These were able to produce suitable ligands, with **PhenDA-2** being formed in an overall yield of 51% from **2.26**, and **PhenDA-3** formed in an overall yield of 65% from **2.26**, Scheme 2.16 and Scheme 2.17. DFT geometry optimisations also showed sensible complexes after optimisation, with the phenanthroline units for both **PhenDA-2** and **PhenDA-3** coordinating manganese at a reasonable angle. Unlike with the other designed 6-coordinate ligands, the geometry does not neatly become pentagonal bipyramidal, even with the longer acetate arms of **PhenDA-3**.



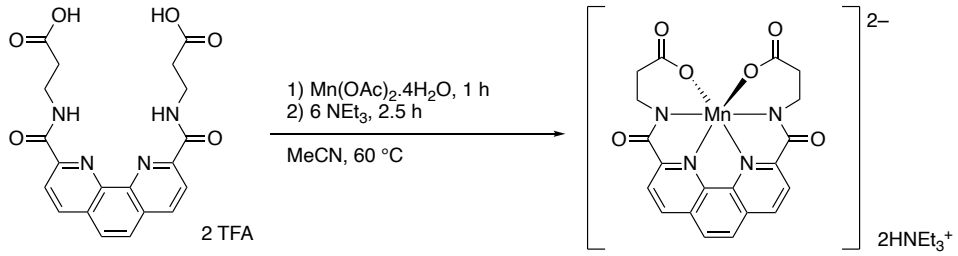
*Scheme 2.16: Synthesis of **PhenDA-2**.*



*Scheme 2.17: Synthesis of **PhenDA-3**.*

Complexation conditions for **PhenDA-3** were sought. The first attempt at complexation followed the synthesis of Mn(oxinate)$_2$ – a Mn^{2+} complex with a high pK_a ligand – as reported by Gawne and co-workers.[43] These conditions (6 KOH, 55 °C, EtOH, 4 h) led to a brown suspension (indicating oxidation to Mn^{3+}), which produced a brown solid insoluble in all of the common laboratory solvents. HR-MS

(ES^+) produced a peak corresponding to $[Mn+L+3H]^+$ suggesting a Mn^{2+} ion, surrounded by L^{4-}, and made monocationic by 3 H^+s. However, the lack of solubility meant that the brown solid could not be tested further. Alternative complexation conditions were sought, altering the base, solvent, manganese source, and protocol. Of all of these, careful pH adjustment (to avoid oxidation) with varying manganese sources in aqueous solutions was not able to show the presence of the complex. The final attempt was able to provide a yellow solid (typical for Mn^{2+} complexes) that could be isolated, Scheme 2.18. Interestingly, whilst the complex could not be found on ES^+ HRMS analysis, MALDI (Matrix-assisted laser desorption/ionisation; a softer MS technique) was able to show $[Mn+L+3H]^+$, once again suggesting a Mn^{2+} complex. The complexation attempts revealed that the ligand's previous solubility appeared to be dependent on being a TFA salt, as once the pH was adjusted to neutral or slightly basic a suspension occurred. Additionally, the fragility of the Mn^{2+} oxidation state (and difficulty in isolating the complex in aqueous solvents) suggests that explicitly mentioning ligands with lower pK_as than these amide complexes should be added to the design considerations for Mn^{2+} ligands.

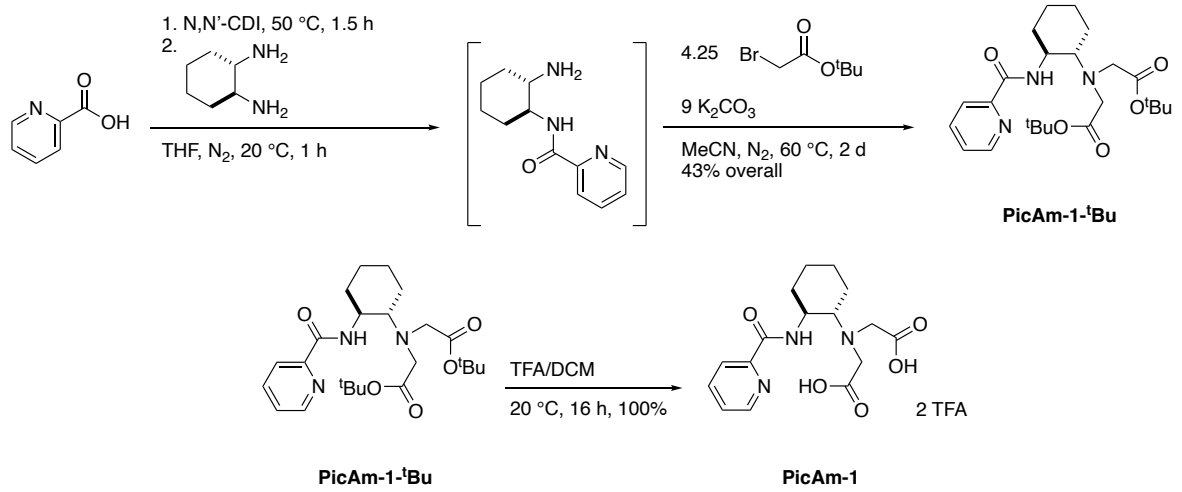


*Scheme 2.18: Complexation conditions for **PhenDA-3**.*

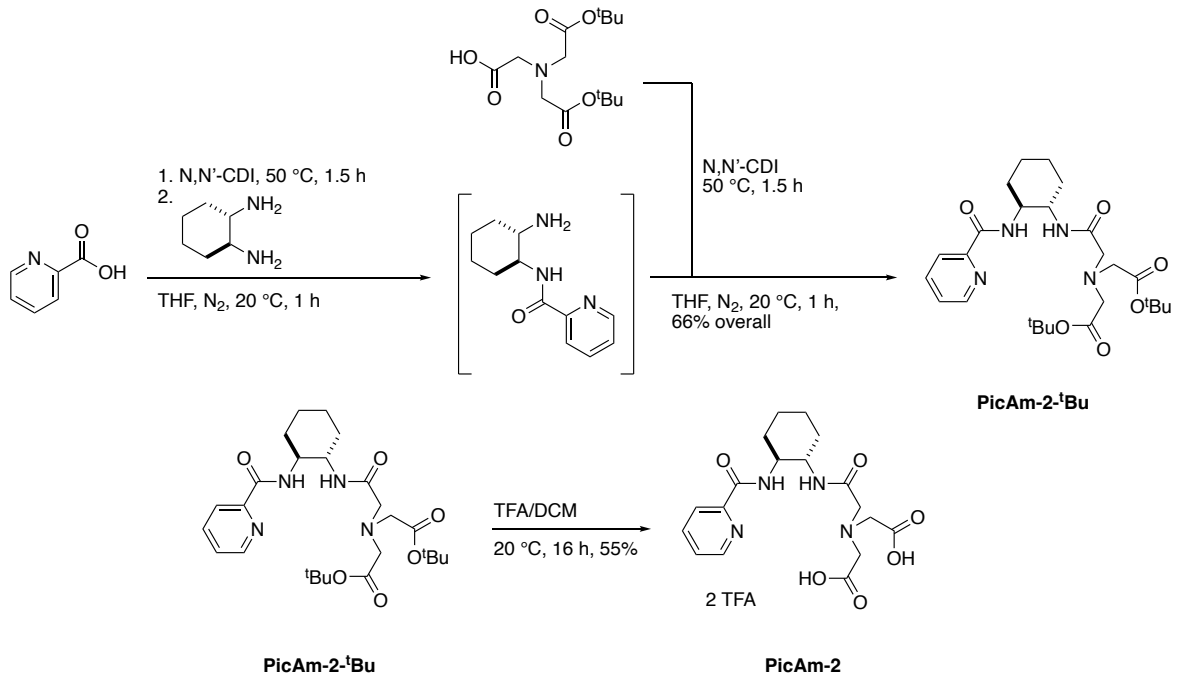
Unfortunately, further testing and optimisation for **PhenDA-3** and complexation studies on **PhenDA-2** could not be fully explored due to ill health, further elaborated in Section 2.3.2.

2.2.3.3 Picolyl amides

As a result of the solubility issues seen when deprotonating the **PhenDA-2** and **PhenDA-3** ligands, the phenanthroline unit was replaced with (±)-*trans*-1,2-diaminocyclohexane, and picolinic acid was used as one half of the structure to provide lipophilic character and a soft pyridyl donor. From this structure, two ligands were designed – one where the remaining amine is N-alkylated with *tert*-butyl acetate twice (**PicAm-1**), and the other forms another amide with 2-(bis(2-(*tert*-butoxy)-2-oxoethyl)amino)acetic acid (**PicAm-2**). The first ligand was designed to only have one amido donor in case two amido donors would favour oxidation of the MnBCA to a Mn^{3+} complex.



*Scheme 2.19: Synthesis of **PicAm-1**.*

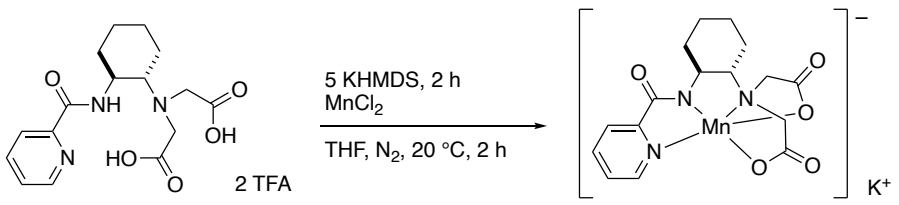


*Scheme 2.20: Synthesis of **PicAm-2.***

The synthesis of the picolyl amide ligands was aided with N,N'-CDI (1,1'-carbonyldiimidazole). In the synthesis of **PicAm-1**, picolinic acid was coupled to (±)-*trans*-1,2-diaminocyclohexane, with the resultant amide alkylated *in-situ* as the N-alkylated products were easier to separate than imidazole and the intermediate. After this, the *tert*-butyl esters were deprotected providing **PicAm-1** at 43% overall yield, Scheme 2.19. **PicAm-2** was synthesised in a similar fashion, but the intermediate amide was then coupled a second time with 2-(bis(2-(*tert*-butoxy)-2-oxoethyl)amino)acetic acid, which after deprotection formed **PicAm-2** in 36% overall yield, Scheme 2.20.

2.2.3.3.1 PicAm-1 complexation

As the ligand synthesised without trouble, complexation conditions were investigated. Due to the higher than usual pK_a of the amido bond, the first attempt used an ethanolic solution of NaOH (with stoichiometric equivalence of NaOH to fully deprotonate). However, the high pH was likely to be responsible for the brown solid produced (most likely Mn^{2+} oxidised to Mn^{3+}), and despite an aqueous solution of the brown solid reporting negative with xylenol orange, and showing an Evans shift, mass spectrometry techniques (LC-MS, ES^+, $MALDI^+$) could not find either the Mn^{2+} or Mn^{3+} complexes. This may indicate weak binding to the metal, with decomplexation occurring within the mass spectrometer. Buffered solutions were tried but were likely too weakly deprotonating to provide the fully deprotonated ligand required for complexation (pHs between 6 and 9). The high pK_a of the amide would suggest that carrying out the reaction in aqueous conditions would simply always reprotonate the amide. Therefore, the complexation conditions for phosphasalens – using KHMDS and anhydrous THF, were tried, Scheme 2.21. This provided a manganese complex which was purified *via* reversed-phase column chromatography, passing both a xylenol orange test, and providing a paramagnetic shift in the Evans method.



PicAm-1 **MnPicAm-1**

*Scheme 2.21: Complexation conditions for **PicAm-1**.*

2.2.3.3.2 Relaxivity measurement on MnPicAm-1

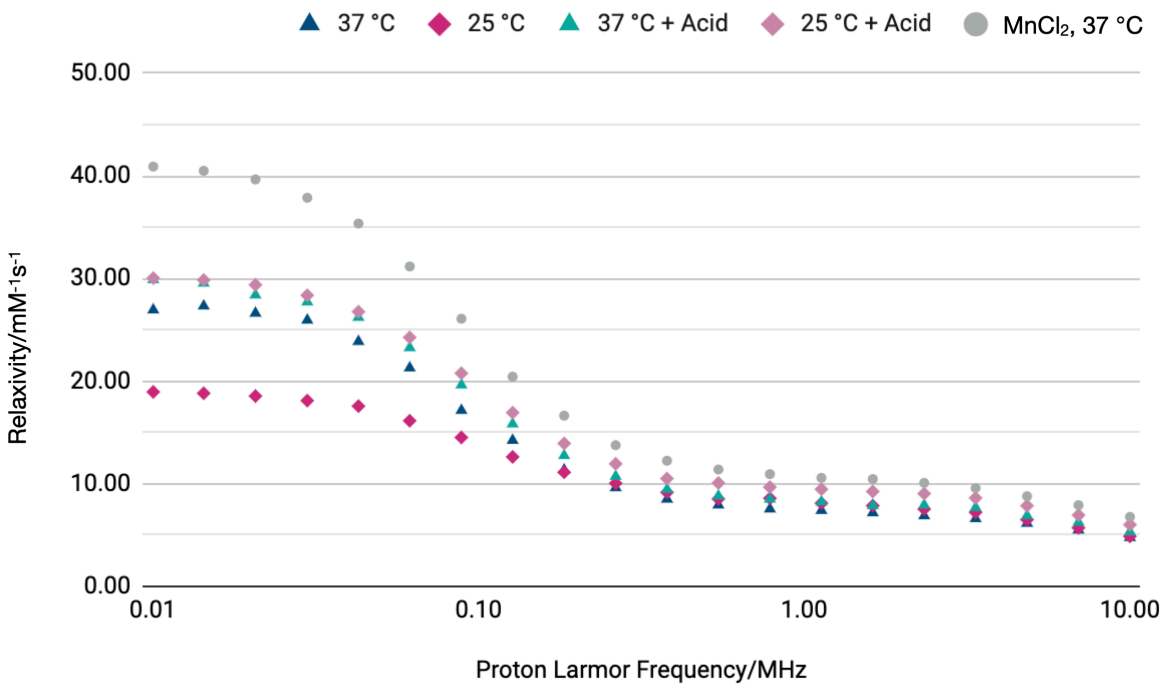


*Figure 2.21: ^{1}H nuclear magnetic relaxation dispersion (NMRD) profile of K_2[Mn(**PicAm-1**)] at different temperatures (diamonds: 25 °C, triangles: 37 °C). Sample also measured at acidic pH at 25 °C and 37 °C. $MnCl_2$ at 37 °C also measured for reference (circles).*

With a potential MnBCA, the ^{1}H nuclear magnetic relaxation dispersion (NMRD) profile for the manganese complex of **PicAm-1** was recorded, Figure 2.21. The profile was measured at both 25 °C and 37 °C (diamonds and triangles respectively). The shape of the profile is atypical of MnBCAs, and more closely resembles that of free Mn^{2+}, as such, the ^{1}H NMRD profiles of both free $MnCl_2.4H_2O$ (at 37 °C, circles) and the sample at pH 1 (to induce complete decomplexation) were also recorded (green triangles at 37 °C, and light pink diamonds at 25 °C). For the non-pH adjusted measurements, the r_1 values increased with increasing temperature, a trend not usually seen with MnBCAs as small Mn^{2+} complexes' relaxivity should be limited by fast rotation. In this case, the expected behaviour is only seen once the Larmor frequency exceeds 0.2 MHz.

Unlike for GBCAs (whose relaxivity is dominated by dipole-dipole (DD) interactions), the relaxivity for MnBCAs is made up of significant contributions from both dipole-dipole interactions and scalar mechanisms. In Figure 2.21, the typical DD dispersion seen between 1 and 10 MHz can be observed – for most MnBCAs (and GBCAs) the relaxivity would be flat between 0.01 and 1 MHz as the scalar contributions are negligible. However, for **MnPicAm-1** another dispersion can be seen which is typically seen for manganese complexes with scalar contributions – usually only seen in $Mn(H_2O)_6^{2+}$.[44]

The Tóth group reported a double dispersion ^{1}H NMRD profile for [Mn_2(ENOTA)(H_2O)$_2$], Figure 2.22. The scalar contribution arises from contact interaction between unpaired electrons and the ^{1}H nucleus, with the low-field dispersion deemed to occur in [Mn_2(ENOTA)(H_2O)$_2$] due to long electronic relaxation times and shorter Mn-O_{water} distance compared to MnBCAs with only one dispersion – which itself was a result of being hexacoordinate.[45]

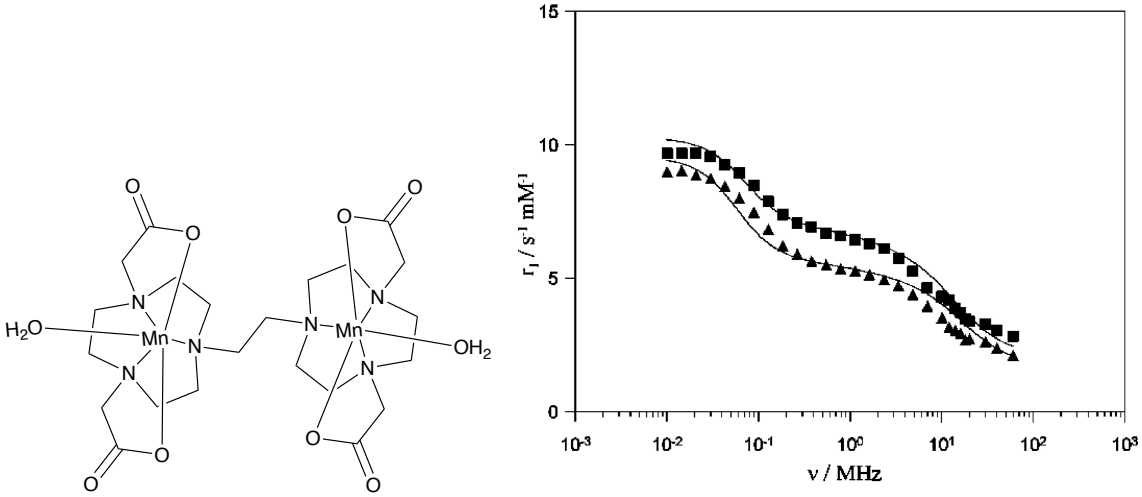


Figure 2.22: [Mn_2(ENOTA)(H_2O)$_2$] published by the Tóth group,(left) which showed a double dispersion in its ^{1}H NMRD profile (right; ^{1}H NMRD profiles at 25 (squares) and 37 °C (triangles)). Data reproduced with permission from Tóth and co-workers.[45]

The shape of the ^{1}H NMRD profile may reveal the nature of **MnPicAm-1** in solution. Free manganese provided much larger relaxivity values, indicating (along with the xylenol orange test) that the complex had not been fully decomplexed. Additionally, the profiles were rerecorded at pH 1 *via* the addition of 1 M HCl, and these did not match the high relaxivities of free Mn^{2+}, suggesting a bound complex at low pH. The relaxivity values at 0.01 MHz for 25 °C and 37 °C (18.9 and 26.9 $mM^{-1}s^{-1}$ respectively) are between half and a third of the values seen for free $Mn(H_2O)_6$ at 25 °C and 37 °C (free Mn^{2+} at 25 °C and 0.01 MHz has an r_1 of approximately 63 $mM^{-1}s^{-1}$.[3] Typical q = 1 complexes have an r_1 at 0.01 MHz of around 5-8 $mM^{-1}s^{-1}$, and q = 2 complexes are around 11-14 $mM^{-1}s^{-1}$.[46] An r_1 of 18.9 $mM^{-1}s^{-1}$ likely suggests a q = 3 complex, Figure 2.23. Whilst an amido-bound Mn^{2+} complex may have been made in the anhydrous complexation conditions, the high pK_a of the ligand may lead to irreversible protonation, leading to a complex that is only tricoordinated from the ligand. This would result in three water molecules inside the inner sphere, and large relaxivity values.

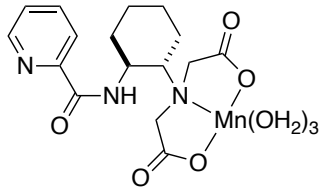
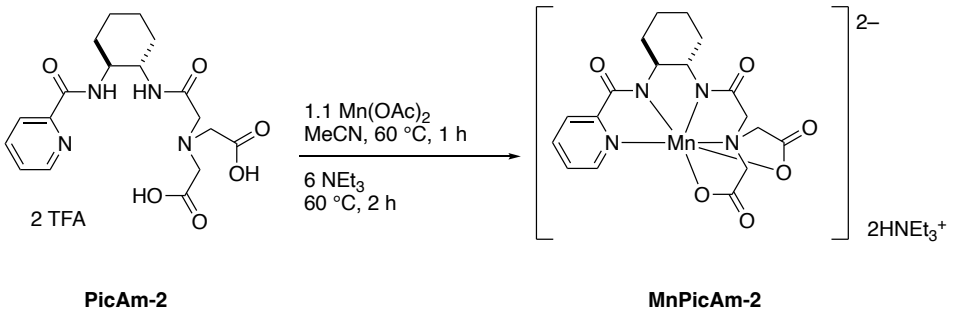


*Figure 2.23: Possible identity of **MnPicAm-1** in solution.*

2.2.3.3.3 PicAm-2 complexation



PicAm-2 **MnPicAm-2**

*Scheme 2.22: Complexation conditions for **PicAm-2**.*

PicAm-2 complexation followed the conditions for **PhenDA-3** complexation and produced a yellow solid which showed a $[Mn+L+3H]^+$ peak when analysed *via* HRMS ES^+, Scheme 2.22. However, further testing and optimisation for **MnPicAm-1** and relaxivity studies on **MnPicAm-2** could not be fully explored due to ill health, further elaborated in Section 2.3.2.

2.3 Conclusion

In the synthetic portion of this work, multiple ligand families were investigated as potential MnBCAs. The first of these, the phosphasalen family, was ultimately unsuitable due to hydrolysis of the iminophosphorane bond. Whilst salen complexes are able to survive in an aqueous environment due to complexation to a metal centre, testing this for phosphasalens would require synthesising a complex hydrophilic enough to be water soluble. This is an example where prototyping that would be used in other industries to validate an idea is very difficult in synthetic chemistry. Commercially available phosphines are typically tri-aryl phosphines which reduce the hydrophilicity, and more broadly, the most successful MnBCA ligands are hexadentate, which is difficult to achieve when the aryl phosphine is usually a dead-end of the ligand rather than a point of functionalisation. The synthesis of three initial phosphasalens was attempted, with **PS1-salt** being synthesised, and difficulty in producing **PS2** and **PS3**, due to the challenging air-sensitive phosphine purification and macrocycle formation.

Sulfonamides and amides were chosen as a way to replace the iminophosphorane bonds, whilst also providing X donors to aid in synthesising an overall monocationic complex. The ligands were split along the sulfonamide/amide bond, leaving amine bridges and the sulfonyl chlorides/acyl chlorides to be synthesised. Synthesis of the amine bridges required detailed understanding of orthogonal protecting group strategies, which after many failed attempts yielded three useful amine bridges for ligand synthesis. Producing the sulfonyl chlorides presented greater challenges than first thought, as obvious methods to produce the sulfonyl chloride (multiple routes of oxidative chlorination of the thiol analogue, and *in-situ* trapping of a sulfinate) ended without success for **2.3**. This is likely due to the nature of the molecule favouring tautomerism to the thioketone. SuFEx chemistry was also investigated as a potential method of forming sulfonamides, *via* PyFluor. This also encountered difficulty, with all attempted sulfonamide preparations from sulfonyl fluorides ending without success.

Amides proved to be easier to work with, with their synthesis being more explored in literature. As with the sulfonamide analogue, the phenanthroline macrocycle was unable to be synthesised. Coupling between carboxylic acid and amine, as well as between acyl chloride and amine ended without evidence of product formation. Attempting to construct the macrocycle by N-alkylation and simultaneously closing at the top ended with aziridine formation which did not go on to further react. Acyclic amide ligands were then synthesised to replace the unachievable macrocyclic ligand. Complexation conditions were sought, eventually finding organic solvents with organic bases able to deprotonate the amide and complex to the manganese. Alongside these, picolyl amides were investigated to overcome solubility issues seen in the phenanthroline amide complexations. These picolyl amide ligands were synthesised without trouble with N,N'-CDI. Complexation was attempted with **PicAm-1**, with conditions borrowed from phosphasalen chemistry aiding to deprotonate the amide in anhydrous THF. ^{1}H NMRD profiles at

25 °C and 37 °C were recorded, showing two dispersions (at both high and low fields) rarely seen in small molecule MnBCAs. The large r_1 at 0.01 MHz suggested a hydration number of 3, which when cross referenced with the ligand structure indicated that the ligand was not coordinating as expected, most likely with the picolyl amide free. Further investigations of the potentially interesting ligands and complexes were unable to be completed due to ill health.

One of the most frustrating parts of MnBCA research is how many techniques are unavailable once a complex is made. For almost all of a chemist's training, ^{1}H NMR is relied upon to provide the majority of the information on the success, yield, or purity of a reaction – but as MnBCAs are paramagnetic characterisation is more challenging. Of the 150 MnBCAs (essentially the entire field) surveyed in Chapter 3, only three have published crystal structures, indicating the difficulty in recrystallising these complexes. The fragile nature of the Mn^{2+} oxidation state means techniques that Gd^{3+} chemists might use – such as increasing the pH and removing free $Gd(OH)_3$ – cannot be used. This pH window also limits the possible ligands and donors available for coordination as realised in the amide ligand work. Realistically, the only technique for confirming complexation success is mass spectrometry. Post- and during the COVID disruption, this service was almost non-existent at Imperial during the synthetic research part of this work.

2.3.1 Future work

The disruption to the third year of this work and eventual end of synthesis work came at an unfortunate time. At that point in the research, a lot of the hard exploratory work and synthetic understanding of the research area had already been carried out. Having the knowledge of what would and would not work, and a better grasp of which ligand types would be more successful, would have certainly improved the output of a final year in the laboratory.

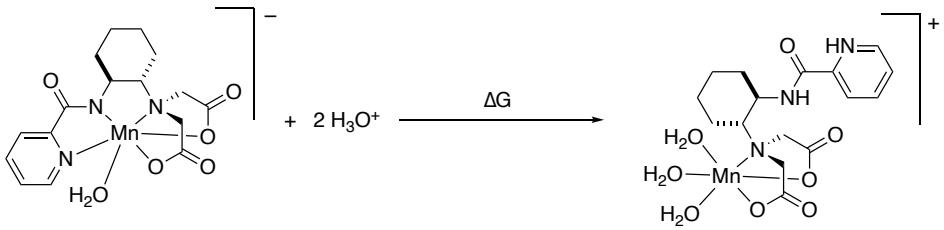
Whilst phosphasalens are not likely to be useful for molecular imaging, they are interesting ligands in other inorganic fields. To fully explore this area, being able to easily produce phosphines with different electron donating properties would be beneficial. This is quite difficult as the harsh conditions required to make different phosphines are usually orthogonal to the functional groups required for electron tunability, or the protecting groups required to shield Mn^{2+} donors.

The most general method of sulfonyl chloride synthesis is *via* $SOCl_2/H_2O_2$ as reported by Bahrami and co-workers. They propose a synthetic route that starts with the thiol converting to a sulfanol, then disulfide, then oxidatively chlorinating to a sulfonyl chloride. If thioketone formation was responsible for a lack of reaction, perhaps formation of the disulfide (Reaxys® reports >100 transformations from 2-mercaptopyridine to 2,2'-dipyridyldisulphide with > 90% yield), and then reaction would have formed the more reactive sulfonyl chloride with more success. This may not have worked for **2.19** – its

neocuproine analogue formed an insoluble red gum with the Bahrami protocol (perhaps suggesting polymerisation). Instead, a SuFEx route may have been more successful – the synthesis of PyFluor starts from 2-mercaptopyridine before oxidation with NaOCl and trapping with fluoride (sat. KHF_2 in MeCN).[47]

The difficulty in complexation and stability of the amide ligands with Mn^{2+} (but not with Mn^{3+} for Jasanoff) suggests a mismatch between Mn^{2+} and the amido donor. Nonetheless, it would be pleasing to have made the macrocyclic amide **PhenDA-1**, and attempt complexation with other metals. The N,N'-CDI reactions worked extremely well in the preparation of **PhenDA-2** and **PhenDA-3**, and it would be interesting to try coupling **2.25** with the prepared amine bridges. Overall, great difficulty was found in purifying and characterising the Mn^{2+} complexes; moving the research towards Mn^{3+} octahedral complexes is likely to be simpler, with fewer difficulties in preparation and stability (and therefore ligand design). All things being equal, moving from high-spin Mn^{2+} to high-spin Mn^{3+} results in a 33% reduction in relaxivity due to the S(S+1) terms in Equations 1.7, 1.8, 1.9, and 1.10. However, this can be mitigated by designing $q = 2$ complexes and by gaining from the higher thermodynamic stability.

Whilst the DFT geometry optimisations show that kinetically the complexes are feasible, a computational pK_a study would be able to show if they are thermodynamically feasible. This would require modelling the second sphere of water as well – geometry optimising $Mn(OH_2)_6^{2+}$ often results in an incorrect geometry where H-bonding between water molecules causes a distortion of the Mn-O_{water} bond angles and deviation from octahedral. This is typically overcome by optimising $[Mn(H_2O)_6.(H_2O)_{12}]^{2+}$ with each bound water having two second sphere water molecules to better sample the actual space around manganese centre in solution.[48] Finding the Gibbs energy change between two states (and likely correlating these changes to actual $\log(K_{MnHL})$ values) would have found the problem with **MnPicAm-1** before synthesis, Scheme 2.23. If viable, this could potentially allow for better predictive tools for all bioinorganic chemists – the real question then is whether computational chemistry is developed enough to provide accurate predictions.



Scheme 2.23: Example reaction proposed to investigate $\log(K_{MnH2L})$ via *DFT on **MnPicAm-1**(H_2O).*

Surprisingly, to our knowledge, there has only been one reported log P measurement of a MnBCA.[49] As driving hepatobiliary clearance *via* lipophilicity is a cornerstone of assumed knowledge and ligand design, it would be beneficial to test for this rather than rely on intuition – as explored in Chapter 5 for gallium complexes. Currently, understanding biodistribution for MnBCAs only happens with *ex* or *in vivo* testing. Computationally modelling this with DFT provides a route to predicting a KPI before synthesis, increasing the efficiency of the MnBCA design process.

Out of necessity, the remainder of this thesis was changed from the project's original synthetic-based aims due to ill health. The project has been transformed to be completely remote-working, and instead focuses on the applications of computational chemistry in the field of medical imaging. This has required extensive re-skilling to transform what was a synthetic chemistry researcher into a computational chemist and data analyst. The following chapters detail the efforts in producing analyses which help to provide tools for synthetic chemists and predict the behaviour of certain molecular imaging agents.

2.3.2 COVID-19 disruption

Oct 19	Nov 19	Dec 19	Jan 20	Feb 20	Mar 20	Apr 20	May 20	Jun 20	Jul 20	Aug 20	Sep 20
100%	100%	Annual Leave	100%	100%	Annual Leave	0%	0%	0%	40-50%	50%	50%

Oct 20	Nov 20	Dec 20	Jan 21	Feb 21	Mar 21	Apr 21	May 21	Jun 21	Jul 21	Aug 21	Sep 21
50%	50%	50%	50%	50%	50%	50%	60%	80%	100%	0%	0%

Oct 21	Nov 21	Dec 21	Jan 22	Feb 22	Mar 22	Apr 22	May 22	Jun 22	Jul 22	Aug 22	Sep 22
Annual Leave	75%	85%	85%	85%	85%	Annual Leave	85%	85%	75%	50%	0%

Oct 22	Nov 22	Dec 22	Jan 23	Feb 23	Mar 23	Apr 23	May 23	Jun 23	Jul 23	Aug 23	Sep 23
0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	Writing

Figure 2.24: Project timeline Oct 2019 - Sep 2023, split by month. Percentages indicate lab occupancy or ability to work on synthesis due to COVID-19 (red: 0%, orange: ≤50%, yellow: <100%, green: 100%). Dec 2022, project moved to remote-working, computational chemistry, and data analytics.

Disruption due to COVID-19 totalled around 18 months of possible synthesis time in the project, Figure 2.24. Between Mar 2020 - Jun 2021, average lab occupancy was below 40% due to imposed Chemistry Department working rotas and regulations in accordance with UK Government rules and regulations. In August 2021, I developed long COVID after two months of severe musculoskeletal and neurological COVID-19 symptoms. A phased re-entry to working was conducted from Nov 2021 - Dec 2021. However, Jan 2022 - Aug 2022 was spent managing chronic pain, as both the physical adjustments and work adjustments made towards making synthetic chemistry possible whilst suffering from fibromyalgia/long COVID failed, with a very large relapse of symptoms in Aug 2022. Ultimately, fibromyalgia is incompatible with synthetic chemistry. This has left a large portion of the synthetic efforts in this work unable to be completed.

2.4 References

1 A. Gupta, P. Caravan, W. S. Price, C. Platas-Iglesias and E. M. Gale, *Inorg. Chem.*, 2020, **59**, 6648–6678, DOI: 10.1021/acs.inorgchem.0c00510.

2 H. Liu and X.-A. Zhang, in *Encyclopedia of Inorganic and Bioinorganic Chemistry*, John Wiley & Sons, Ltd, Chichester, UK, 2018, pp. 1–16, DOI: 10.1002/9781119951438.eibc2626.

3 M. Botta, F. Carniato, D. Esteban-Gómez, C. Platas-Iglesias and L. Tei, *Future Med. Chem.*, 2019, **11**, 1461–1483, DOI: 10.4155/fmc-2018-0608.

4 F. K. Kálmán and G. Tircsó, *Inorg. Chem.*, 2012, **51**, 10065–10067, DOI: 10.1021/ic300832e.

5 J. S. Troughton, M. T. Greenfield, J. M. Greenwood, S. Dumas, A. J. Wiethoff, J. Wang, M. Spiller, T. J. McMurry and P. Caravan, *Inorg. Chem.*, 2004, **43**, 6313–6323, DOI: 10.1021/ic049559g.

6 K. Pota, Z. Garda, F. K. Kálmán, J. L. Barriada, D. Esteban-Gómez, C. Platas-Iglesias, I. Tóth, E. Brücher and G. Tircsó, *New J. Chem.*, 2018, **42**, 8001–8011, DOI: 10.1039/C8NJ00121A.

7 E. M. Gale, I. P. Atanasova, F. Blasi, I. Ay and P. Caravan, *J. Am. Chem. Soc.*, 2015, **137**, 15548–15557, DOI: 10.1021/jacs.5b10748.

8 R. Uzal-Varela, F. Pérez-Fernández, L. Valencia, A. Rodríguez-Rodríguez, C. Platas-Iglesias, P. Caravan and D. Esteban-Gómez, *Inorg. Chem.*, 2022, **61**, 14173–14186, DOI: 10.1021/acs.inorgchem.2c02364.

9 A. Takács, R. Napolitano, M. Purgel, A. C. Bényei, L. Zékány, E. Brücher, I. Tóth, Z. Baranyai and S. Aime, *Inorg. Chem.*, 2014, **53**, 2858–2872, DOI: 10.1021/ic4025958.

10 A. Forgács, L. Tei, Z. Baranyai, D. Esteban-Gómez, C. Platas-Iglesias and M. Botta, *Dalton Trans.*, 2017, **46**, 8494–8504, DOI: 10.1039/C7DT01508A.

11 F. K. Kálmán, V. Nagy, R. Uzal-Varela, P. Pérez-Lourido, D. Esteban-Gómez, Z. Garda, K. Pota, R. Mezei, A. Pallier, É. Tóth, C. Platas-Iglesias and G. Tircsó, *Molecules*, 2021, **26**, 1524, DOI: 10.3390/molecules26061524.

12 Z. Garda, E. Molnár, F. K. Kálmán, R. Botár, V. Nagy, Z. Baranyai, E. Brücher, Z. Kovács, I. Tóth and G. Tircsó, *Front. Chem.*, 2018, **6**, 1–14, DOI: 10.3389/fchem.2018.00232.

13 Z. Garda, E. Molnár, N. Hamon, J. L. Barriada, D. Esteban-Gómez, B. Váradi, V. Nagy, K. Pota, F. K. Kálmán, I. Tóth, N. Lihi, C. Platas-Iglesias, É. Tóth, R. Tripier and G. Tircsó, *Inorg. Chem.*, 2021, **60**, 1133–1148, DOI: 10.1021/acs.inorgchem.0c03276.

14 D. Ndiaye, P. Cieslik, H. Wadepohl, A. Pallier, S. Même, P. Comba and É. Tóth, *J. Am. Chem. Soc.*, 2022, **144**, 22212–22220, DOI: 10.1021/jacs.2c10108.

15 I. M. Marín and A. Auffrant, *Eur. J. Inorg. Chem.*, 2018, **2018**, 1634–1644, DOI: 10.1002/ejic.201701210.

16 T.-P.-A. Cao, PhD Thesis, Ecole Polytechnique, 2012.

17 C. Bakewell, T.-P.-A. Cao, N. Long, X. F. Le Goff, A. Auffrant and C. K. Williams, *J. Am. Chem. Soc.*, 2012, **134**, 20577–20580, DOI: 10.1021/ja310003v.

18 T.-P.-A. Cao, A. Buchard, X. F. Le Goff, A. Auffrant and C. K. Williams, *Inorg. Chem.*, 2012, **51**, 2157–2169, DOI: 10.1021/ic202015z.

19 C. Bakewell, T.-P.-A. Cao, X. F. Le Goff, N. J. Long, A. Auffrant and C. K. Williams, *Organometallics*, 2013, **32**, 1475–1483, DOI: 10.1021/om301129k.

20 A. Thevenon, A. Cyriac, D. Myers, A. J. P. White, C. B. Durr and C. K. Williams, *J. Am. Chem. Soc.*, 2018, **140**, 6893–6903, DOI: 10.1021/jacs.8b01920.

21 A. T. Normand, R. Malacea-Kabbara, R. Lapenta, A. Dajnak, P. Richard, H. Cattey, A. Bolley, A. Grassi, S. Milione, A. Auffrant, S. Dagorne and P. Le Gendre, *Dalton Trans.*, 2020, **49**, 6989–7004, DOI: 10.1039/D0DT00972E.

22 E. M. Gale, C. M. Jones, I. Ramsay, C. T. Farrar and P. Caravan, *J. Am. Chem. Soc.*, 2016, **138**, 15861–15864, DOI: 10.1021/jacs.6b10898.

23 M. P. Lowe and D. Parker, *Chem. Commun.*, 2000, 707–708, DOI: 10.1039/b001629m.

24 R. Uzal-Varela, L. Valencia, D. Lalli, M. Maneiro, D. Esteban-Gómez, C. Platas-Iglesias, M. Botta and A. Rodríguez-Rodríguez, *Inorg. Chem.*, 2021, **60**, 15055–15068, DOI: 10.1021/acs.inorgchem.1c02057.

25 A. Barandov, B. B. Bartelle, B. A. Gonzalez, W. L. White, S. J. Lippard and A. Jasanoff, *J. Am. Chem. Soc.*, 2016, **138**, 5483–5486, DOI: 10.1021/jacs.5b13337.

26 A. Barandov, B. B. Bartelle, C. G. Williamson, E. S. Loucks, S. J. Lippard and A. Jasanoff, *Nat. Commun.*, 2019, **10**, 1–9, DOI: 10.1038/s41467-019-08558-7.

27 A. Barandov, S. Ghosh, N. Li, B. B. Bartelle, J. I. Daher, M. L. Pegis, H. Collins and A. Jasanoff, *ACS Sensors*, 2020, **5**, 1674–1682, DOI: 10.1021/acssensors.0c00322.

28 L. Helm, J. R. Morrow, C. J. Bond, F. Carniato, M. Botta, M. Braun, Z. Baranyai, R. Pujales-Paradela, M. Regueiro-Figueroa, D. Esteban-Gómez, C. Platas-Iglesias and T. J. Scholl, in *Contrast Agents for MRI: Experimental Methods*, 2017, pp. 121–242, DOI: 10.1039/9781788010146-00121.

29 E. Molnár, B. Váradi, Z. Garda, R. Botár, F. K. Kálmán, É. Tóth, C. Platas-Iglesias, I. Tóth, E. Brücher and G. Tircsó, *Inorganica Chim. Acta*, 2018, **472**, 254–263, DOI: 10.1016/j.ica.2017.07.071.

30 A. Brück and K. Ruhland, *Organometallics*, 2009, **28**, 6383–6401, DOI: 10.1021/om900324a.

31 L. Morello, P. Yu, C. D. Carmichael, B. O. Patrick and M. D. Fryzuk, *J. Am. Chem. Soc.*, 2005, **127**, 12796–12797, DOI: 10.1021/ja054467r.

32 T. Takeshita, K. Sato and Y. Nakajima, *Dalton Trans.*, 2018, **47**, 17004–17010, DOI: 10.1039/C8DT04168G.

33 H. C. Guo, R. H. Zheng and H. J. Jiang, *Org. Prep. Proced. Int.*, 2012, **44**, 392–396, DOI: 10.1080/00304948.2012.697745.

34 A. P. Krapcho, S. Sparapani, A. Leenstra and J. D. Seitz, *Tetrahedron Lett.*, 2009, **50**, 3195–3197, DOI: 10.1016/j.tetlet.2009.01.138.

35 E. Flegeau, J. Harrison and M. Willis, *Synlett*, 2015, **27**, 101–105, DOI: 10.1055/s-0035-1560578.

36 J. Dong, L. Krasnova, M. G. Finn and K. Barry Sharpless, *Angew. Chem. Int. Ed.*, 2014, **53**, 9430–9448, DOI: 10.1002/anie.201309399.

37 S. Berg, M. Bergh, S. Hellberg, K. Högdin, Y. Lo-Alfredsson, P. Söderman, S. Von Berg, T. Weigelt, M. Ormö, Y. Xue, J. Tucker, J. Neelissen, E. Jerning, Y. Nilsson and R. Bhat, *J. Med. Chem.*, 2012, **55**, 9107–9119, DOI: 10.1021/jm201724m.

38 A. J. Brouwer, T. Ceylan, T. van der Linden and R. M. J. Liskamp, *Tetrahedron Lett.*, 2009, **50**, 3391–3393, DOI: 10.1016/j.tetlet.2009.02.130.

39 A. V. Bogolubsky, Y. S. Moroz, P. K. Mykhailiuk, S. E. Pipko, A. I. Konovets, I. V. Sadkova and A. Tolmachev, *ACS Comb. Sci.*, 2014, **16**, 192–197, DOI: 10.1021/co400164z.

40 R. Zelli, S. Tommasone, P. Dumy, A. Marra and A. Dondoni, *European J. Org. Chem.*, 2016, **2016**, 5102–5116, DOI: 10.1002/ejoc.201600732.

41 J. J. Rojas, R. A. Croft, A. J. Sterling, E. L. Briggs, D. Antermite, D. C. Schmitt, L. Blagojevic, P. Haycock, A. J. P. White, F. Duarte, C. Choi, J. J. Mousseau and J. A. Bull, *Nat. Chem.*, 2022, **14**, 160–169, DOI: 10.1038/s41557-021-00856-2.

42 J. Hanson, *J. Chem. Educ.*, 2001, **78**, 1266, DOI: 10.1021/ed078p1266.

43 P. Gawne, F. Man, J. Fonslet, R. Radia, J. Bordoloi, M. Cleveland, P. Jimenez-Royo, A. Gabizon, P. J. Blower, N. Long and R. T. M. de Rosales, *Dalton Trans.*, 2018, **47**, 9283–9293, DOI: 10.1039/C8DT00100F.

44 O. Porcar-Tost, A. Pallier, D. Esteban-Gómez, O. Illa, C. Platas-Iglesias, É. Tóth and R. M. Ortuño, *Dalton Trans.*, 2021, **50**, 1076–1085, DOI: 10.1039/d0dt03402a.

45 E. Balogh, Z. He, W. Hsieh, S. Liu and É. Tóth, *Inorg. Chem.*, 2007, **46**, 238–250, DOI: 10.1021/ic0616582.

46 J. A. Peters and C. F. G. C. Geraldes, *Inorganics*, 2018, **6**, 116, DOI: 10.3390/inorganics6040116.

47 M. K. Nielsen, C. R. Ugaz, W. Li and A. G. Doyle, *J. Am. Chem. Soc.*, 2015, **137**, 9571–9574, DOI: 10.1021/jacs.5b06307.

48 W. W. Rudolph and G. Irmer, *Dalton Trans.*, 2013, **42**, 14460, DOI: 10.1039/c3dt51493e.

49 D. Ndiaye, M. Sy, A. Pallier, S. Même, I. de Silva, S. Lacerda, A. M. Nonat, L. J. Charbonnière and É. Tóth, *Angew. Chem. Int. Ed.*, 2020, **59**, 11958–11963, DOI: 10.1002/anie.202003685.

Chapter 3

Can a DFT optimisation predict the key metrics for MnBCAs before they have been synthesised?

3.1 Introduction

The safety of MRI contrast agents is primarily judged by their thermodynamic stability constant, log(K_{ML}). Understanding the speciation of the metal once inside the decomplexing conditions of the body is crucial to prevent unforeseen reactions, and concerns over the toxicity of gadolinium release from GBCAs have intensified focus on producing contrast agents that are thermodynamically stable.[1–3]

Manganese has a safer toxicological profile than gadolinium – as an essential element for life elimination pathways have evolved to manage levels in the body. However, manganese is innately less stable – high spin Mn^{2+} has no CFSE and has the second largest ionic radius in the first row of the transition metals. Coupled with this, MnBCAs are further hampered by the requirement to increase contrast enhancement – most intuitively increased by increasing hydration number, and consequently, ligands for MnBCAs typically have one fewer donor than the maximum coordination number of manganese (7 or 8). Furthermore, assuming dosage follows GBCAs, MnBCAs will require 1 M solutions for contrast enhancement, which for a 70 kg adult would require around 370 mg of manganese in a scan, and 3.15 g of a 450 g mol^{-1} complex. For context, the upper limit of oral supplemental manganese ranges from 5-11 mg/day,[4,5] and manganese toxicity has been shown to occur in drinking water with levels of 23 mg/L manganese.[6]

3.1.1 Thermodynamic stability prediction

Current methods of measuring thermodynamic stability require non-trivial pH-potentiometric titrations, which requires a level of precision that can require expensive automated systems. Additionally, this requires synthesis of the complex in a sufficient quantity for characterisation, which in itself can require months to years of challenging synthetic work before testing for its key performance indicator can be conducted. In 2022, Uzal-Varela applied multivariate regression on 150 manganese complexes between structural descriptors of the ligand (donors and ligand structure) and log(K_{MnL}), to produce a single equation to predict thermodynamic stability from its component parts, Equation 3.1.[7]

$$log(K_{MnL}) = \Sigma\,(n_i \cdot \Delta\, log(K_{MnL})) + A_{MnL}$$

Equation 3.1: Equation for the prediction of log(K_{MnL}) by Uzal-Varela and co-workers.[7] Where n_i is the number of structural descriptors of type i. $\Delta log(K_{MnL})$ is the effect on log(K_{MnL}) due to the structural descriptor, A_{MnL} is the effect on log(K_{MnL}) due to the type of macrocyclic ligand.

This method was able to predict the log(K_{MnL}) of its training dataset with a R^2 of 0.873, MAE of 0.99 log units and RMSE of 1.48, Figure 3.1. Its high R^2 value indicates that the method has high predictive power, and error values are well within a useful range for a synthetic chemist to understand how their complex will perform, before any attempt at synthesis is made. Due to being a solely phenomenological model, there are some cases in which the stability constant cannot be completely predicted. The model

requires that the ligand can be fully described by its 26 structural descriptors to predict log(K_{MnL}). For ligand types that contain alternative donors or are not represented in their training dataset, such as bispidine-type complexes, the model cannot make suitable attempts to describe them. Additionally, the model is isomer blind as ligands which contain identical structural descriptors produce identical calculated stability constants, despite having different experimental log(K_{MnL}) values. For example, [Mn(3,6-PC2A)(H_2O)] and [Mn(3,9-PC2A)(H_2O)] have experimental log(K_{MnL}) values of 15.53 and 17.09, though are both calculated to have a stability constant of 15.22.

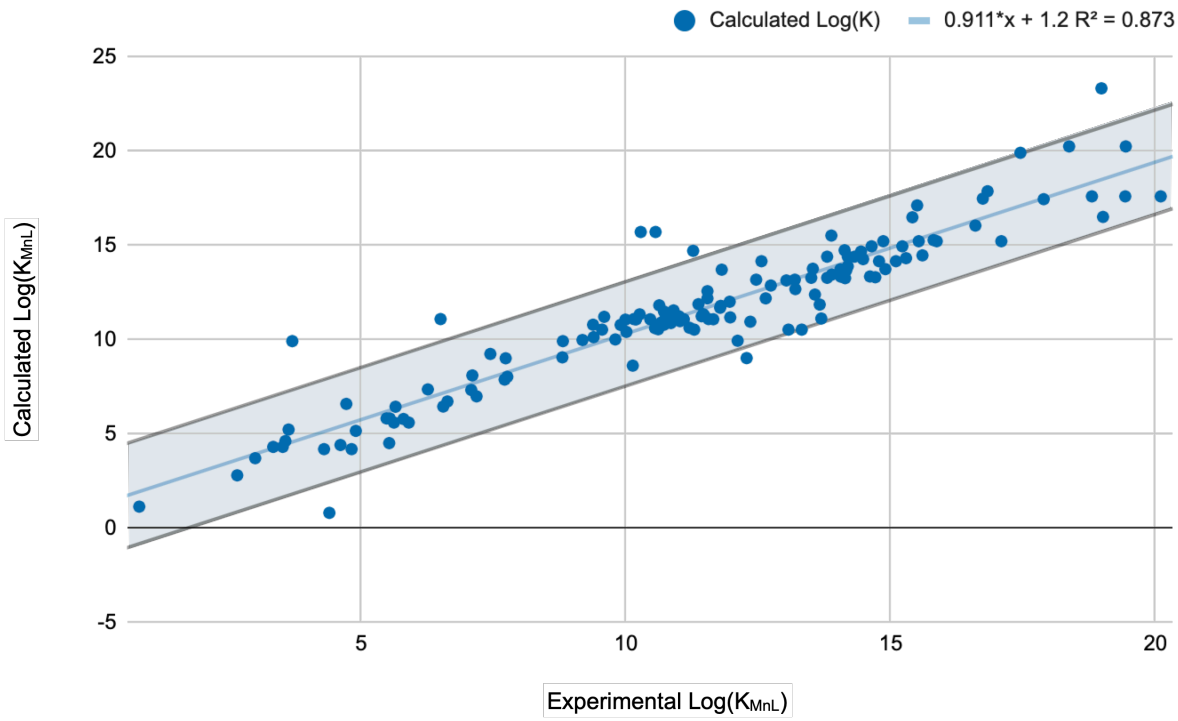
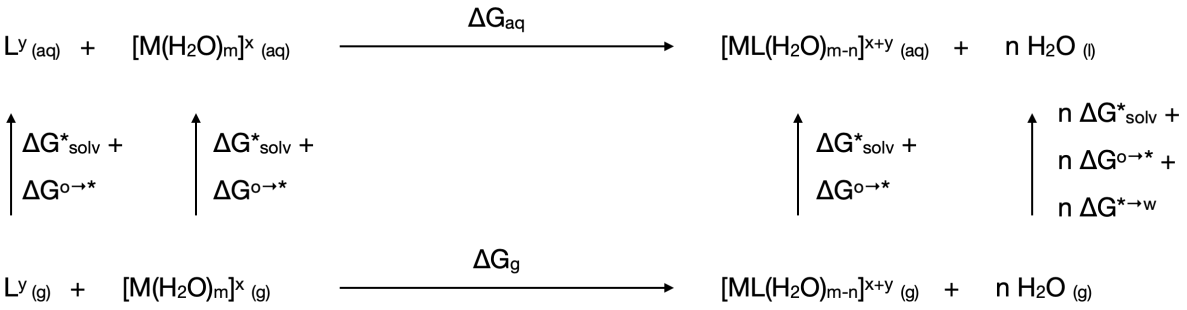


Figure 3.1: Graph of the calculated vs. experimental log(K_{MnL}) values predicted by Uzal-Varela and co-workers in their study of MnBCAs,[7] fitted with a linear trend line and 95% prediction interval.

Computational chemistry, especially a DFT approach, provides an alternative predictive pathway that is derived from quantum mechanical first principles. In theory, DFT can predict the necessary thermochemical values required for thermodynamic stability constant calculations. This ability is limited by the accuracy of the model, which in turn is usually dictated by the assumptions or approximations required to produce a working model. For this application, the problem is compounded by the very small differences in Gibbs energy between log units (e.g. a Δlog(K) of 1 is only 1.36 kcal mol^{-1}), which would require extremely accurate computational modelling to provide the exact thermodynamic stability constants. Instead, the stability constants that are produced from DFT calculations are typically correlated to experimental values to produce equations to transform these calculated log(K_{ML}) values into predicted log(K_{ML}) values that are comparable to experimental log(K_{ML}).[8]

Typically, studies have shown that predicting accurate absolute stability constants is incredibly difficult, with accurate and complete modelling of solvation around ions preventing quantitative log(K_{ML}) prediction.[9,10] Overcoming these inaccuracies can be achieved by explicitly modelling the solute-solvent interaction through the inclusion of explicit water molecules to form a solvation sphere – a concept that is essential in the accurate prediction of Mn-O_{water} bond legnths.[11,12] This however increases the complexity of the calculation (free water is not neatly bound), as well as providing a new problem of adequately sampling the solute-solvent cluster (how many should be included, and where they should be placed).



Scheme 3.1: Thermodynamic cycle as suggested by Bryantsev and co-workers for the determination of ΔG_{aq} via *DFT, where* ^{o},*,^{w} *indicate gas phase, aqueous phase, and solvent water.*

Whilst quantitative values are beyond the reach of what is currently computationally viable, correlation methods have shown promise in predicting log(K_{ML}) with great accuracy.[8,13] In 2015, Bryantsev and co-workers investigated the prediction of stability constants for uranyl complexes using DFT.[8] Their model centred around the thermodynamic cycle in Scheme 3.1. This thermodynamic cycle leverages the ability of DFT to predict the free energy of solvation to construct a cycle that calculates ΔG_{aq} *via* ΔG_{solv} and using Hess' law to navigate from one side of the cycle to the other. In order to use this cycle, certain corrections need to be applied to take into consideration the changes in reference from a 1 atm ideal gas (24.46 L mol^{-1}) to a 1 M solution, seen in Equation 3.2.

$$\Delta G^{o\rightarrow *} = -T\Delta S^{o\rightarrow *} = RT\ln\left({}^{V_0}/_{V^*}\right) = RT\ln(24.46) = 1.89\ kcal\ mol^{-1}$$

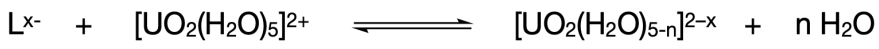
Equation 3.2: Thermodynamic correction to change from 1 atm to 1 M, where T = 298.15 K.

Additionally, a correction needs to be applied to convert the concept of 1 M $H_2O_{(aq)}$ into its correct classification as solvent, $H_2O_{(l)}$, Equation 3.3.

$$\Delta G^{*\rightarrow w} = RT\ln([H_2O]) = 2.38\ kcal\ mol^{-1}$$

Equation 3.3: Thermodynamic correction to change from 1 M water to solvent.

Bryantsev calculated the stability constants for 13 anionic simple ligands according to the equilibrium shown in Scheme 3.2, at the B3LYP/SSC/6-311++G(d,p)/IEFPCM level. This approach was able to produce calculated log(K_{UL}) values with an R^2 of 0.972. Upon further inspection, they identified that segmenting the ligands by their charge provided a marginal increase in R^2 to 0.974 for monoanionic, and an R^2 of 0.958 for dianionic ligands, Figure 3.2. Their approach showed that despite calculated log(K_{UL}) values overestimating by 10-40 log units, linear correlation was able to reliably predict log(K_{UL}) with a RMSE of only 0.83 log units. This approach was also applied in 2016 by the same group to predict the thermodynamic stabilities of VO_2^+ complexes. It provided a regression with R^2 of 0.938, and RMSE of 0.85 log units, indicating the reliable use of a correlation method for log(K) estimation.[14] The overestimation of the stability constant by direct calculation of log(K_{UL}) was attributed to the inherent inadequacy of the IEFPCM solvation model to accurately model charged species, with even greater discrepancy seen when modelled solely in the gas phase.



Scheme 3.2: Reaction scheme modelled by Bryantsev.[8]

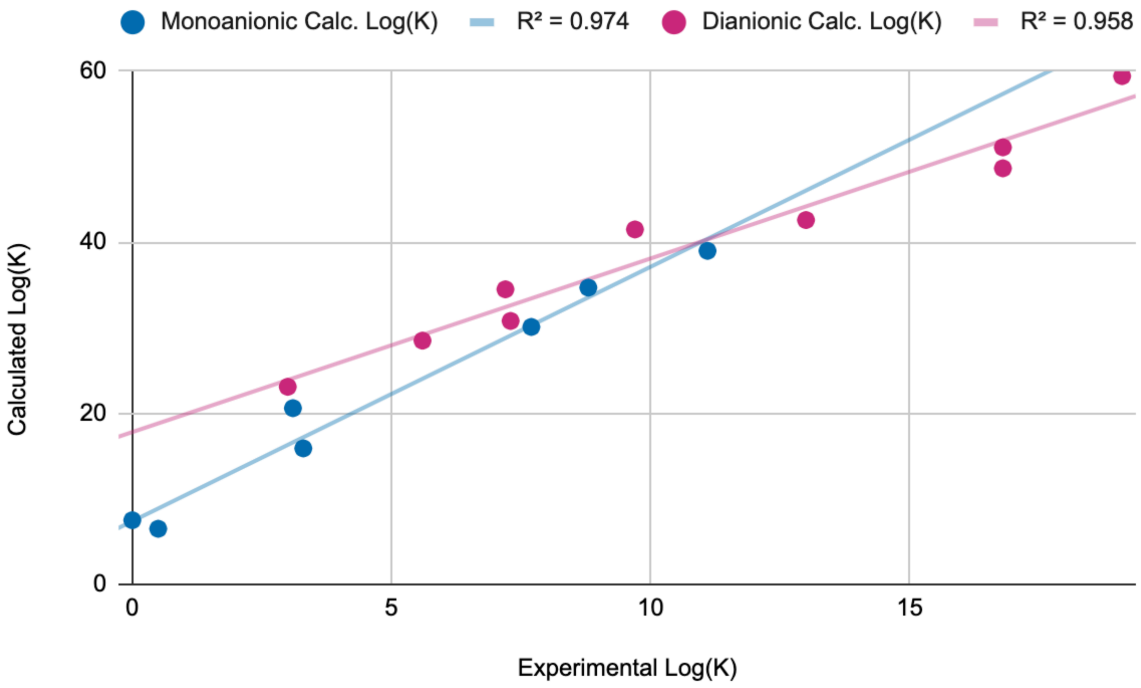
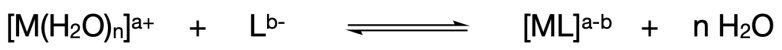


Figure 3.2: Graph of the calculated vs. experimental log(K_{UL}) values predicted by Bryantsev and co-workers in their study of uranium complexes.[8] *More accurate prediction of log(K_{UL}) could be achieved* via *segmenting ligands by charge.*

In 2020, Holland predicted the thermodynamic stability of zirconium radiotracers after creating a calibration curve by modelling the stability constants of 14 metal-desferrioxamine B complexes and Zr(THPN).[13] These calculations were all carried out at the B3LYP/DGDZVP/IEFPCM level, using a

framework with Scheme 3.3, foregoing corrections to change between 1 atm and 1 M, as well as to convert 1 M water into liquid water.



Scheme 3.3: Reaction scheme modelled by Holland.[13]

The calculated stability constants were overestimates, as seen in previous work in the field.[9,13] Least squares linear regression analysis showed a strong linear correlation between calculated and experimental log(K_{ML}) values, with an R^2 of 0.972, which led to Equation 3.4:

$$Log(K)_{estimate} = \frac{Log(K)_{calculated} - 45.48 \pm 2.47}{1.985 \pm 0.095}$$

Equation 3.4: Equation for the prediction of log(K) for DFO complexes via *DFT by Holland.*[13]

Applying the calculated values to Equation 3.4 provides predicted results with a MAE of 1.87 and RMSE of 2.14. Analysis of the residuals also indicated no systematic mistreatment of the data or failure of the method with a mean residual of 0.02.

With the correlation from metal-DFO complexes, Holland then preceded to predict the stability constant of 23 different zirconium complexes, 9 of which were variants of DFO. Whilst this may be overstepping the predictive power of the model to complexes largely independent of the training dataset (and without validation due to difficulty in experimentally measuring Zr stability constants), the ability to provide a prediction, with such lightweight computational requirements of B3LYP/DGDZVP/IEFPCM is highly useful.

Beyond the stability constant, MnBCAs are judged on their contrast enhancement r_1, with both hydration number q, and water exchange rate k_{ex} typically reported in literature. The water exchange rate and hydration number both proportionally contribute to r_1 (Equation 1.5), therefore accurate prediction of the two would aid in providing more accurate methods for predicting r_1 than is currently available – an imprecise semi-empirical method by Peters and Geraldes.[15]

3.1.2 MnBCA DFT in literature

Currently most of the DFT used in the MnBCA field is to provide a structure to analyse – in part due to the difficulty in growing single crystals of aqueous manganese complexes. Most of these analyses are carried out by the research group of Platas-Iglesias. In 2017, they analysed [Mn(1,4-BzDO2AM)(H_2O)]$^{2+}$·2H_2O and [Mn(1,4-DO2AMBz)(H_2O)]$^{2+}$·2H_2O at the M06-2X/TZVP level probing its structure, bond lengths and angles.[16] Additionally, the scalar hyperfine coupling constant

(Equation 1.9) was reported as calculated *via* Gaussian as part of its frequency calculations, which showed good agreement with experimentally derived values. In 2019, the group calculated the DFT optimised geometry of $[Mn(MeNO2A)(H_2O)]\cdot 5H_2O$ vs $[Mn(MeNO2A)(H_2O)_2]\cdot 4H_2O$ at the M06-2X/TZVP level, and these structures were used to provide evidence of a $q = 1$ complex.[17] The Platas-Iglesias group also provided DFT calculations to rationalise experimental differences between 3,6-PC2A and 3,9-PC2A *via* analysing the bond lengths in the coordination sphere at the M11/def2TZVPP level.[11] It was shown that the positioning of the carboxylate, in the 6- or 9-position, affected the coordination of amine atoms around the ring, with 3,9-PC2A providing a less sterically strained structure which allowed for tighter coordination for the pyridyl and secondary amine.

For the water parameters used in MnBCA analysis, very few studies have looked at prediction prior to synthesis and experimental measurement. As referenced in Chapter 1, Peters and Geraldes have devised a formula that can predict the hydration number of a MnBCA by multiplying an experimentally measured r_1 value by a coefficient inversely proportional to its formula weight. Practically speaking the relationship can be judged by eye, with an r_1 (measured at 20 MHz and 298 K) of 0-2.5 mM^{-1} s^{-1} typically for $q = 0$, 2.5-3.8 mM^{-1} s^{-1} for $q = 1$, and greater than 3.8 mM^{-1} s^{-1} for $q = 2$. Experimentally, Gale and co-workers produced a simple method to predict the hydration number by measuring the maximum of transverse relaxation of $H_2{}^{17}O$ in ^{17}O NMR.[18] This is possible as the relaxation of ^{17}O in coordinated water is dominated by the scalar mechanism, and at large enough field strengths $\tau_{SC} = \tau_m = T_{2m}$.[18] Therefore Equation 1.5, 1.9, and 1.12 can be rearranged into Equation 3.5, for the measurement of transverse ^{17}O relaxivity ($r_2{}^o$). From their calculations (with $A_O/\hbar$ fixed at 3.3×10^7 rad s^{-1}) they saw an increase in $r_{2max}{}^o$ of 510 ± 100 mM^{-1} s^{-1} per q, allowing for a calculation of q with a ±0.2 accuracy.

$$q = r^o_{2max}[H_2O]\left(\frac{2}{\sqrt{\frac{S(S+1)}{3}}\cdot\frac{A_O}{\hbar}}\right) \cong \frac{r^o_{2max}}{510}$$

Equation 3.5: Equation for the determination of q via *maximum transverse ^{17}O relaxivity.*

Water exchange rate prediction or analysis has been less studied in MnBCAs with only qualitative assertions made on the relationships certain features have to water exchange rate. In 2011, Drahoš and co-workers developed Mn^{2+} complexes of 1-oxa-4,7-diazacyclononane based ligands, and speculated that an associative mechanism would provide a faster water exchange rate than dissociative, *via* comparison of their experimental results between their ligand with two carboxylate arms vs two phosphonate arms.[19] However, they also noted that $Mn_2[(enota)(H_2O)_2]$ was expected to have an associative mechanism, and yet did not provide high water exchange rates.

Between 2021 and 2022, a flurry of papers speculated on explanations of experimentally derived water exchange rates. In Forgács and co-workers' paper on benzyl functionalised DO2A Mn^{2+} complexes ($[Mn(1,4\text{-}DO2AMBz)]^{2+}$ and $[Mn(1,4\text{-}BzDO2AM)]^{2+}$), they speculated that both a hydrophobic and steric effect from introducing benzyl groups near the coordination sphere acted to improve the water exchange rate compared to $Mn(1,4\text{-}DO2AM)]^{2+}$ under an expected dissociative mechanism.[16] Furthering the idea of steric strain improving water exchange rate, Garda and co-workers suggested that it was steric compression in the vicinity of where water would bind to Mn^{2+} that was responsible for the faster water exchange rate between the structurally similar $[Mn(1,4\text{-}DO2A)(H_2O)]$ and their $[Mn(3,6\text{-}PC2A)(H_2O)]$ complex. They concluded that the introduction of the pyridyl group therefore must lead to a stronger water coordination.[11] Finally in 2021, Kálmán and co-workers reported that the water exchange rate was dependent on complex charge when comparing their oxa-aza macrocycles (*trans*-$O2DO2A^{2-}$ and *trans*-$O2DO2AM^{Pip}$) as the $[Mn(\textit{trans}\text{-}O2DO2AM^{Pip})(H_2O)]^{2+}$ complex had a 35% lower water exchange rate than $[Mn(\textit{trans}\text{-}O2DO2AM)(H_2O)]$ – they suggested that the increase in positive charge would result in a stronger Mn-water bond, and therefore slower exchange.[20]

Outside of the MnBCA space, in 2015, Platas-Iglesias reported an analysis on the water exchange rate, the electron density (ρ_{BCP}) and electron localisation function (ELF) at the bond critical point, and Gd-O_{water} lengths. In this analysis, for dissociative mechanisms (which would be expected for complexes at their saturation) they used electron density at the bond critical point (or where the electron density dips lowest in the bond) to predict the water exchange rate. For Gd-O_{water} there was a logarithmic relationship to the water exchange rate ($0.0144\ \ln(k_{ex}/10^3\ s^{-1}) + 2.34$), with a reasonable R^2 relationship of 0.727, Figure 3.3. They showed that weaker, longer Gd-O_{water} bonds led to faster water exchange rates, as did lower ρ_{BCP} and ELF.

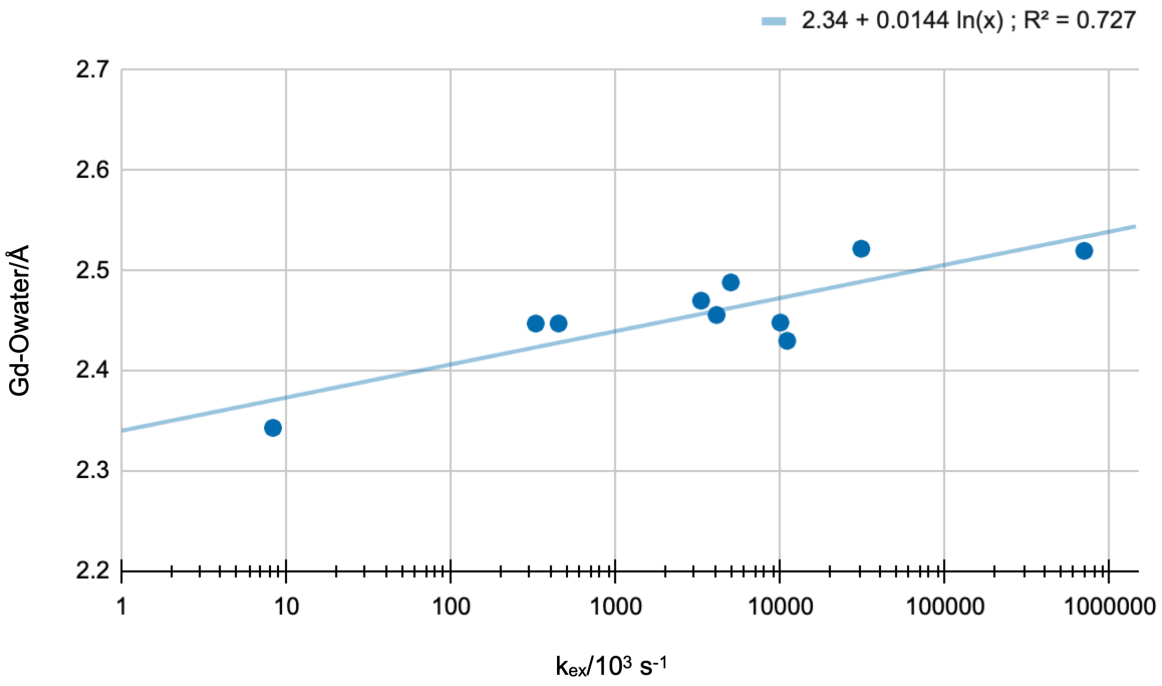


Figure 3.3: Plot of gadolinium-water bond length and water exchange rate for GBCAs studied by Platas-Iglesias.[21]

Overall, DFT calculations in the MnBCA space have not been used to predict the key performance indicators (KPIs) for MnBCAs, instead largely speculating links to experimental values from the calculated structures. This is mirrored in the seemingly ad hoc explanations for water exchange rates and tenuous prediction of r_1, and as such, there is a need for analytics in the space to produce predictive tools that can help MnBCA chemists.

3.1.3 Machine learning

Machine learning (ML) is a collection of applications of artificial intelligence (AI), where a system or collection of algorithms is trained to find a model to explain the output variable of a dataset, and through automatic repetition, improve itself to provide more accurate results. ML is already being used in multiple industries including: healthcare – automated disease diagnosis from medical imaging; civil engineering – optimising designs for efficiency; urban planning – traffic management; and recruitment – learning from recruitment managers' previous appointments and then making automated decisions on future candidates.[22–25] The recruitment example shows the pitfalls of data analytics and machine learning, bad data *in* will lead to bad conclusions *out* – training a model on 30 years of systemic discrimination at a company will teach the model to mirror the biases that it was trained on. Therefore, understanding the dataset that is being analysed (including its inherent biases or weaknesses) is important for useful predictions, and equally understanding the application of the predictions is helpful in being selective with the data a model is trained on.

For the chemical industries, applications of artificial intelligence are not new. Gaussian optimisations, for example, are an implementation of machine learning where algorithms optimise a set of numbers with functionals and basis sets to produce a stable structure. These functionals and basis sets themselves were parameterised (i.e. went through a process of learning) with real world data to improve their accuracy. What is most exciting about the advent of machine learning is the ability to improve the democratisation of data within our industry to provide chemists the tools to analyse the vast quantities of data that they produce to find patterns and insights that they otherwise would not be able to immediately discern themselves.

In this analysis, multi-variate linear regression, neural networks, and support vector machine algorithms were used to analyse a small dataset with many variables. Multi-variate linear regression, Equation 3.6, aims to find a predictive model from a set of independent variables (otherwise known as features). Standard least-squares linear regression is a subset of multi-variate linear regression with least squares used as the optimisation parameter; instead of fitting a single line (1-dimension to link the 2 dimensions of independent and dependent variable) in standard linear regression, multi-variate linear regression attempts to fit a hyperplane through the data points as it considers each feature in its own dimension.

$$y = m_1 x_1 + m_2 x_2 + \cdots + m_n x_n + c$$

Equation 3.6: General equation of a multi-variate linear regression, where each feature (x_1, x_2, x_3, etc.) has a coefficient applied (m_1, m_2, m_3, etc.) to produce an output.

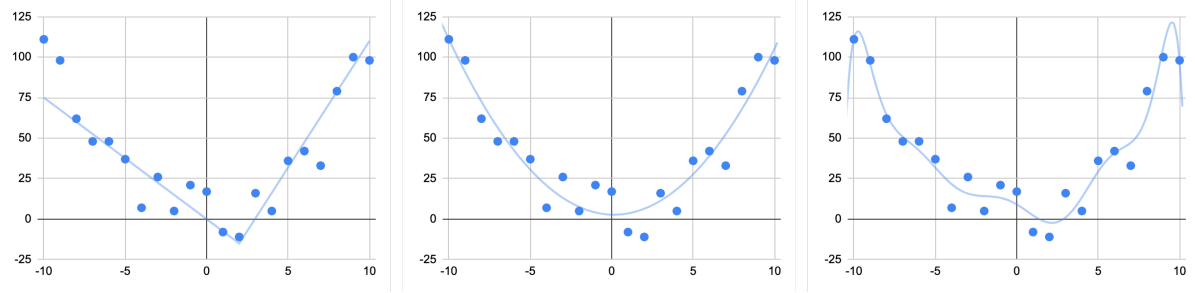


Figure 3.4: Examples of under-fitting, correct fitting, and over-fitting for a quadratic equation with random noise applied.

Alongside this, regularisation can also be applied – this is a technique which penalises the model from having too many 'important' features. Essentially regularisation techniques nudge the model to reduce the weightings of features as much as possible, which can help to reduce complexity. This is important to avoid producing a model which is over-fitted to its training data and is not useful when it encounters new data to make predictions on, Figure 3.4.

This leads into the idea of the importance of a completely separate training and testing dataset. With many regression analyses in Chemistry, the analysis is judged on the ability of the model to predict the data it was trained on – this naturally leads to overinflated R^2 values judging the model. For instance, when the Uzal-Varela multi-variate regression analysis is reproduced with a randomly chosen (but deterministic) 20% sample of the dataset kept aside for testing, the resultant regression model returns an R^2 of 0.594 on the testing dataset (RMSE of 1.736, MAE of 1.349). This is not merely a consequence of reducing the sample size leading to a poorer R^2; an improvement can be made to the coefficients that were published by filtering for complexes with $q \geq 1$, with R^2 of 0.970, RMSE of 0.692, and MAE of 0.532. As such, the best practices applied in the data analytics community should equally be applied for chemical data analytics and predictive models should be tested correctly.

Neural networks are conceptually similar to linear regression in that data is passed from one side (the input), a weighting is applied, and then combined at the end to give the output on the other side. In the case of neural networks, Figure 3.5, the input layer (orange column) has a number of nodes, known as neurons (circles), that are equal to the number of independent variables, also known as 'features'. This input layer passes its information to the middle portion (grey columns) where the computation happens. This middle portion is made up of 'hidden' layers which also comprise of neurons which connect to each of the neurons in the layers before and after them. Each of these connections applies a weighting on the data that it receives before passing it forward to the next neuron. In this way the data travels through the neural network until it ends at the output layer (green column), and the prediction is scored by the program for accuracy. This is when back-propagation occurs and the front end of the neural network is told how well the weightings produced an output when compared to the reality of its training dataset, and it then modifies the weightings over and over until it produces as close a result for the output of the training dataset as it can.

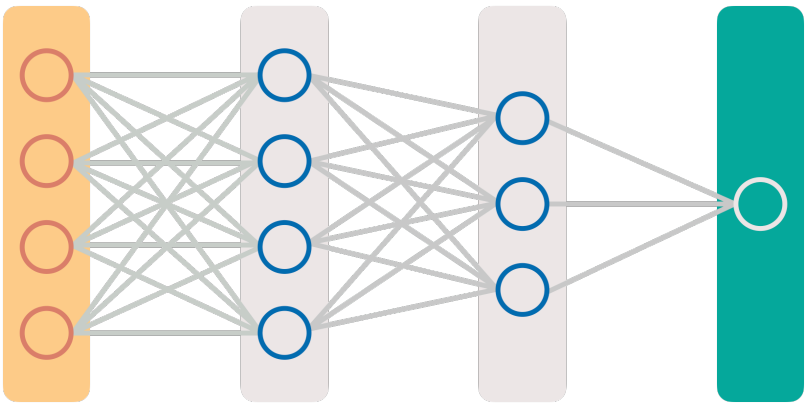


Figure 3.5: Simplified diagram of a neural network with 4 features in its input layer, two hidden layers with 4 and 3 neurons respectively, and a final output layer which produces 1 output. Between each neuron is a connection that applies a weighting to the data it was given. At each neuron these weighted numbers are combined, and this number is then sent to the neurons of the next layer.

Support vector machine regression (SVR) is another machine learning algorithm that can help find the relationship between input variables and an output – the same as any other regression method. However, sometimes finding the relationship between a set of variables and its output is not obvious as it is in the case for a linear regression, and instead would be better conceptualised in a higher dimension. SVM uses various functions to map the data into a higher-dimensional space and produce relationships between variables that were not immediately obvious before. Once in the higher-dimensional space a hyperplane is selected which minimises the error between itself and the actual output value – just like in multi-variate (and simple) linear regression.

3.1.4 SAMPL challenges

SAMPL (Statistical Assessment of the Modelling of Proteins and Ligands) challenges are a set of challenges that score user submissions in developing methods to solve similar challenges as $\log(K_{ML})$ prediction in this study – pK_a, log D, log P (each of these require finding the equilibrium constant between two states). The aim of the SAMPL challenges is to focus efforts in improving the accuracy of computational drug discovery and represent the state-of-the-art approaches to answer these problems. SAMPL0-4 focused on the prediction of the free energy of solvation for a training and testing dataset of small drug-like molecules. With SAMPL5 log D prediction was included, with 76 entries, with an average RMSE of 6.09 and MAE of 4.42. The best entry had a RMSE of 2.10 and MAE of 1.70 log units and used the non-Gaussian SCRF of COSMO-RS, second place used the more standard SMD SCRF.[26,27] 2018's SAMPL6 moved to log P prediction and the best models gave R^2 values just under 0.8, and RMSE values <0.5. As before, the top quantum mechanical approaches used COSMO-RS at BP/TZVP/FINE Single Point level, and SMD with the M06 functional.[28–30]

SAMPL7 in 2020 moved further from what is 'freely available' for improved accuracy. In the challenge overview,[31] it was noted that some of the models required "*extensive knowledge and expertise,*" as such, someone not *au fait* with computational chemistry may struggle in conducting successful free energy calculations. This is the challenge between software development for usability and accuracy, namely to answer the question; is it more important that 100s-1000s of general chemists can use models that provide an acceptable estimate of properties or that 10s of computational chemists can run experiments that perfectly reproduce experimental data?

3.1.5 Aims

The aim of this work is to address the question – can a DFT optimisation predict the key metrics for MnBCAs *before* they have been synthesised? Initially, this will focus on the stability constant, with the aim of providing a general method for MnBCA chemists to run two (or four depending on the method) DFT calculations (one for the ligand, one for the complex), and decide whether their design meets their $\log(K_{MnL})$ requirements. Furthermore, the target user for these calculations should be non-computational chemists. This requires only freely available, general purpose functionals, basis sets and solvation methods for the widely available Gaussian program to be used – using proprietary or niche software will limit the availability of the model for routine use by synthetic chemists. It also means that water cluster modelling should be excluded, as the knowledge required to handle errors or manipulate the input files and the computational power required is prohibitive.

Considering the potential difficulty in accurately finding the ΔG values, we would ideally aim to accurately predict the $\log(K_{MnL})$ values of the test data set. However, failing this, the ability to repeatedly differentiate between isomers *via* calculated $\log(K_{MnL})$ would also be a successful outcome. The first set of information would be able to provide chemists with a simple test before synthesis begins – even when their complexes are novel to the MnBCA space. The second test would help to improve the Uzal-Varela calculation.

The lack of predictive tools for MnBCA chemists in estimating key performance indicators of their complexes is certainly an opportunity for exploratory work, and the second aim is to investigate if more clinically useful data can be predicted from DFT optimised geometries. The aim would be to provide a predictive model using a certain measurement or collection of measurements to calculate the relevant KPIs in MnBCA chemistry. This will initially look at coordination geometry and cavity size to test if a relationship with $\log(K_{MnL})$ exists, before moving towards investigating the factors that have been proposed for fast water exchange in literature, as well as finding a better predictive method for r_1.

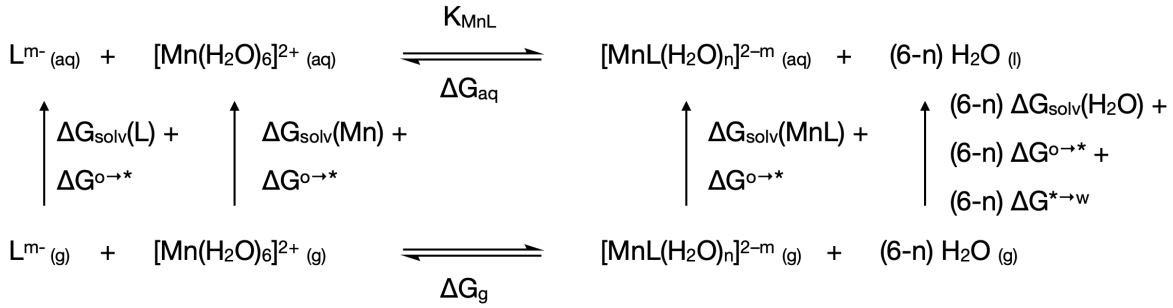
3.2 Results

3.2.1 Computational methods

Density functional theory was used for all calculations with the Gaussian 16 program (version C.01).[32] Structures were geometry optimised without symmetry constraints, in aqueous solution using either IEFPCM or SMD self-consistent reaction fields.[33,34] Each manganese structure with explicit coordinated water, and ligand was confirmed as local minima on the PES *via* frequency calculations. DFT calculations were undertaken with the B3LYP, M06-2X, APFD, ωB97XD, TPSSh, or PW6B95D3 functional, and either 6-31G(d,p), 6-311+G(2d,p), TZVP, or def2TZVPP basis sets.[35–41] Manganese complexes were considered as high spin Mn at the ground state ($3d^5$, multiplicity of 6). Spin contamination was considered for all manganese calculations, by comparing S(S+1) against the calculated $\langle S^2 \rangle$ value, and each calculation produced negligible contamination ($\langle S2 \rangle$-S(S+1)<0.004, i.e. <0.05%). In the cases where first order SCF failed to converge, quadratically convergent SCF was employed (*via* the scf = xqc keyword), in the cases where convergence repeatedly failed for a specific functional/basis set combination, single point energy calculations were employed replacing the electronic energy component of the calculated Gibbs energy. Default values for temperature and pressure were used (298.15 K, 1 atm). The default integration grid (UltraFine, 99 radial shells, 590 angular points per shell) was used throughout. The Orange 3 (version 3.35.0) data visualisation and analysis software was used to perform machine learning tasks *via* python 3.10.10.[42] Orange's neural network widget (a feed-forward network with back-propagation) was used with logistic activation and the Adam optimisation algorithm for optimisation. The neural network contained two hidden layers; the first layer had as many neurons as the dataset had features, the second had two-thirds of this amount +1. Data used in the neural network was pre-processed to have $\mu = 0$ and $\sigma = 1$. Orange's support vector machine (SVM) widget was used with a sigmoid kernel ($\tanh(g \bullet x \bullet y + 1)$) with all other parameters left at default. Orange's linear regression widget was also used with ridge regularisation (L2) (strength alpha set at 0.13).

3.2.2 DFT log(K_{MnL}) calculation

3.2.2.1 Thermodynamic cycle



Scheme 3.4: Thermodynamic cycle modelled in this work.

The reaction modelled is the complexation of a hexaaqua-manganese ion with a fully deprotonated ligand to form a MnBCA, Scheme 3.4. The log of the equilibrium constant K_{MnL} represents the stability of the MnL complex with respect to complete dissociation. The equilibrium constant, K_{MnL}, can be related to the Gibbs energy difference between products and reactants, ΔG, *via* Equation 3.7:

$$\Delta G = -RT \cdot \ln(K_{MnL})$$

$$Log(K_{MnL}) = \frac{-\Delta G}{RT \cdot \ln(10)}$$

Equation 3.7: Manipulation of free energy equation, via *change of base rule for logarithms.*

The value of ΔG is the difference in Gibbs energies of the products and reactants when solvated in water, ΔG_{aq}. The key to calculating this ΔG_{aq} *via* DFT is to calculate the solvation energy, ΔG_{solv}, of each of the reactants and products of the reaction. This creates our thermodynamic cycle for investigating log(K_{MnL}), similar to those seen in literature for metal complex log(K) prediction, Scheme 3.4.[8,14]

To compare the gas calculations (calculated at 1 atm) to the solvated calculations (at 1 M), a correction must be applied, resulting in a 1.89 kcal mol^{-1} stabilisation for the solvated species. As water molecules from the reaction are being explicitly considered in this study, an additional correction of RT·ln([H_2O]) must be applied to each mole of water to account for the conversion between 1 M $H_2O_{(aq)}$ and $H_2O_{(l)}$,[8] as shown in Equation 3.8.

$$\Delta G_{aq} =$$
$$- \left[G\left(L^{m-}{}_{(aq)}\right) + G\left([Mn(H_2O)_6]^{2+}{}_{(aq)}\right) - G\left(L^{m-}{}_{(g)}\right) - G\left([Mn(H_2O)_6]^{2+}{}_{(g)}\right) + 2(1.89\ kcal\ mol^{-1})\right]$$
$$+ \left[G\left([MnL(H_2O)_n]^{2-m}{}_{(g)}\right) + (6-n)\,G\left(H_2O_{(g)}\right) - G\left(L^{m-}{}_{(g)}\right) - G\left([Mn(H_2O)_6]^{2+}{}_{(g)}\right)\right]$$
$$+ \left[G\left([MnL(H_2O)_n]^{2-m}{}_{(aq)}\right) + (6-n)\,G\left(H_2O_{(l)}\right) - G\left([MnL(H_2O)_n]^{2-m}{}_{(g)}\right) - (6-n)\,G\left(H_2O_{(g)}\right) + (7-n)(1.89\ kcal\ mol^{-1}) + (6-n)(RT \cdot \ln(55.5))\right]$$
$$= -G\left(L^{m-}{}_{(aq)}\right) - G\left([Mn(H_2O)_6]^{2+}{}_{(aq)}\right) + G\left([MnL(H_2O)_n]^{2-m}{}_{(aq)}\right) + (6-n)\,G\left(H_2O_{(l)}\right) + (5-n)(1.89\ kcal\ mol^{-1}) + (6-n)(RT \cdot \ln(55.5))$$

Equation 3.8: Gibbs energy calculation of ΔG_{aq} and appropriate corrections, where T = 298.15 K.

Overall, the gas phase Gibbs energies cancel out, but explicitly considering them is required to include the unit and solvent corrections.

3.2.2.2 Testing dataset

The purpose of MnBCA chemistry is to create complexes that have large r_1 contrast enhancement, which is typically achieved with complexes possessing q = 1 or above. Therefore, the dataset that was used in the Uzal-Varela analysis (all ligands shown in Section 8.3) was reduced to remove complexes with q = 0 (less likely to be competitive with GBCAs) and additionally complexes with q = 2 (having multiple coordinated water molecules would require a more rigorous scan of isomers, which would be too computationally expensive for this project). This slimmed the dataset from 150 to 32 complexes, and simplified the thermodynamic cycle, Scheme 3.4, whereby each model only has to deal with one value for n. To categorise complexes without hydration data, complexes were also selected if they had a contrast enhancement of $2.5 < r_1 < 3.8$. This range was chosen after assessing the r_1 values of complexes with a reported hydration number of 1. The testing dataset was then selected to address the limitations in the Uzal-Varela paper, namely dealing with isomers and bispidine-COOH, Table 3.1 and Figure 3.6.

Table 3.1: Testing dataset of ligands used in this study.

Ligand number	Ligand name	Log(K_{MnL})
L05	1,3-H_4CBuDTA	10.78
L14	*trans*-H_4CDTA	14.69
L15	*cis*-H_4CDTA	14.19
L38	3,6-H_2PC2A	15.53
L39	3,9-H_2PC2A	17.09
L76	3,9-H_2PC2A-BP	14.86
L147	1,2-H_4CBuDTA	10.26
L172	Bispidine-COOH	19.47

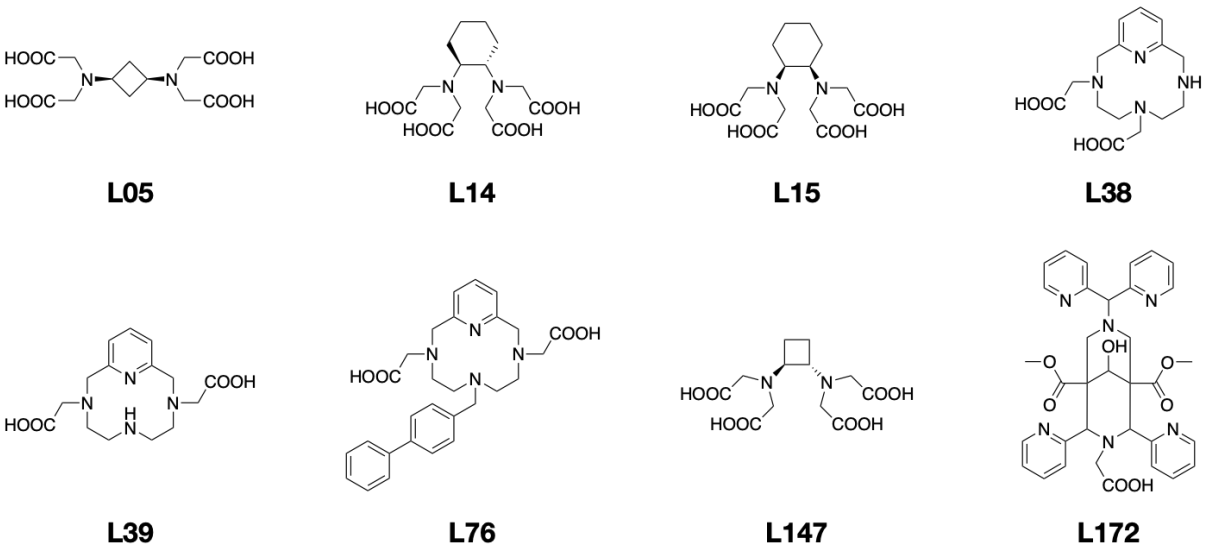


Figure 3.6: Testing dataset of ligands, chosen to test the models on the ability to differentiate between positional and geometric isomers.

This allowed for positional (L05 vs. L147 and L38 vs. L39) and geometric (L14 vs. L15) isomers to be analysed, to test against a complex with a structure that is undefined by the Uzal-Varela model (L172), and to test if DFT could discern a difference between structures with pendant arms (L39 vs. L76).

3.2.2.3 Training dataset

Due to the required cost in running the breadth of calculations required, not all of the remaining 24 complexes could be brought forward for the training dataset. 13 complexes were removed either because their ligand class was already represented, they had extremely low log(K_{MnL}) values, or they were too niche. This left 11 complexes to train the models on, represented in Table 3.2 and Figure 3.7.

Table 3.2: Training dataset of ligands used in this study.

Ligand number	Ligand name	Log(K_{MnL})
L04	H_4EDTA	12.46
L10	H_3PyC3A	14.14
L12	H_3DPASAm	13.53
L13	H_3DPAAA	13.19
L16	H_4PhDTA	11.79
L22	*trans*-H_2O2DO2A	9.38
L29	*trans*-H_2OPC2A	13.03
L45	H_3AAZ3A	11.00
L75	H_2PC2A-DPA	15.87
L121	*trans*-4-HET-CDTA	13.8
L169	Bispidine-py-Me	24.2

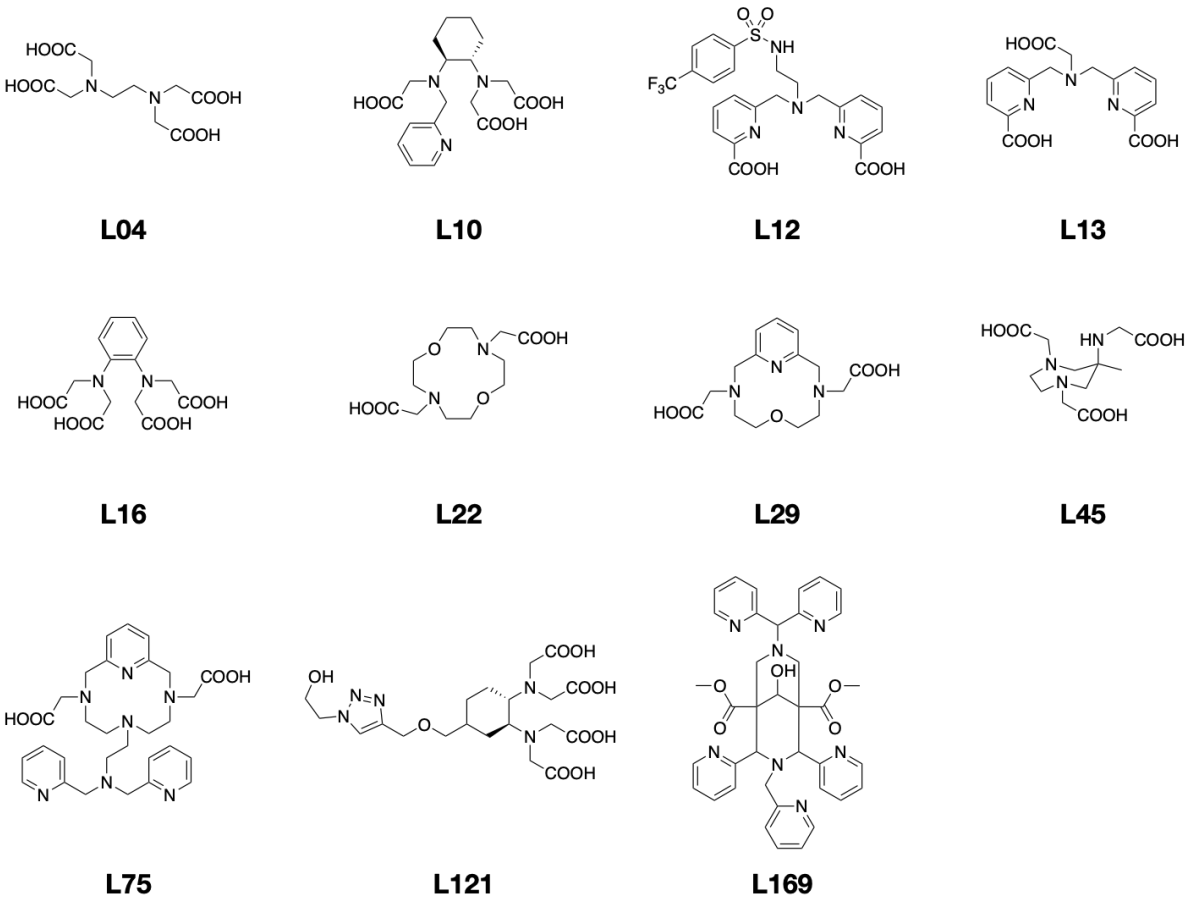


Figure 3.7: Training dataset of ligands.

3.2.2.4 Model chemistries

Three different functional and basis set combinations were tested:

- B3LYP/6-31G(d,p)
- M06-2X/TZVP
- PW6B95D3/TZVP

These were selected to cover a range of cost, with B3LYP/6-31G(d,p) representing a basic functional with a modest basis set, to M06-2X from the Minnesota functionals which showed good performance in the SAMPL challenges with SMD,[30] to a more complex functional in PW6B95D3 that incorporates dispersion, which has performed well for thermochemistry in a benchmarking of over 200 functionals.[43] In a 2019 paper studying the log(K) of terbium complexes, Chen and co-workers found that the effect of the basis set was minimal compared to the functional,[44] so TZVP was chosen for both M06-2X and PW6B95D3.

Most importantly for this work three solvation models available in Gaussian 16 were analysed – IEFPCM, SMD, and IEFPCM(SMD). IEFPCM(SMD) is a model suggested by Foresman and Frisch,[45] in which the total energy component of IEFPCM's Gibbs energy is replaced with the improved total energy from the SMD calculation: $G^{ESMD} = G^{IEFPCM} - E^{IEFPCM} + E^{SMD}$. The SMD model, whilst designed specifically to predict ΔG_{solv}, is not as rigorous as IEFPCM for the thermochemical components of Gibbs energy, and therefore may be improved through replacing the IEFPCM total energy component for the SMD component, whilst keeping all of the more rigorous aspects of IEFPCM.

In addition to solvation and method/basis set, MnBCAs require at least two explicit water molecules to be modelled to have acceptable accuracy in the Mn-O bond length between manganese and the bound water.[12] However, for literature log(K_{MnL}) measurements – Scheme 3.3 and Scheme 3.4, the manganese complex is shown with only the bound water present. Therefore, the complexes were also studied without these extra hydration sphere waters by taking the optimised structure with the water, removing them, and then running a frequency calculation to find the Gibbs values.

A summary is shown in Figure 3.8, which highlights the extent of calculations required for each complex in the training set (2 × 2 × 3 per complex, and 2 × 3 per ligand, resulting in a minimum of 198 calculations for training – in practice there are a lot more due to calculation failure).

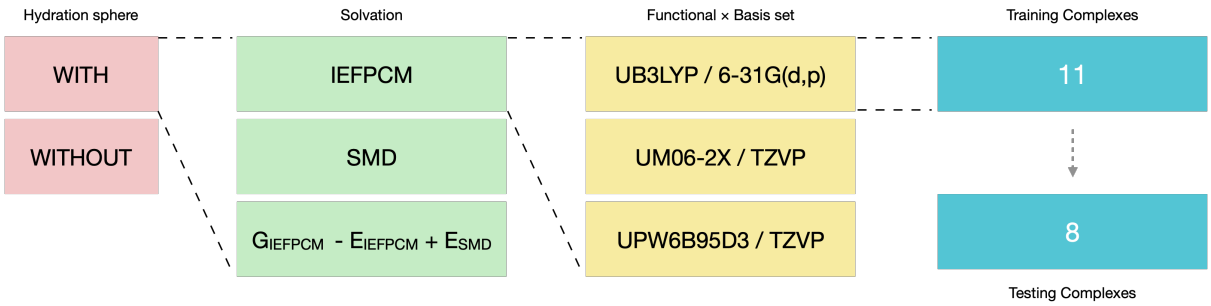


Figure 3.8: Framework for the experiment design – each model is comprised of a hydration sphere, solvation method, and functional/basis set combination, and trained on 11 MnBCAs before validation on 8 different MnBCAs.

3.2.2.5 Optimisations

$$\Delta G_{aq} = -G\left(L^{m-}{}_{(aq)}\right) - G\left([Mn(H_2O)_6]^{2+}{}_{(aq)}\right) + G\left([MnL(H_2O)_n]^{2-m}{}_{(aq)}\right) + (6-n)\,G\left(H_2O_{(aq)}\right) + (5-n)(1.89\ kcal\ mol^{-1}) + (6-n)(RT \cdot \ln(55.5))$$

As shown in the shortened version of Equation 3.8 (reproduced above), four optimisations were required for each model studied: ligand, hexaaqua-manganese, the complex, and water. Of these, hexaaqua-manganese and water were used in each model.

The speciation of Mn^{2+} in solution has been investigated, with EXAFS, crystallography, and relaxometry studies confirming in solution that manganese exists as a hexaaqua octahedral complex, with minimal hydrolysis ($pK_a = 10.56$).[46–49] The optimised structure for $[Mn(H_2O)_6]^{2+}$ was sought with an input file derived from a hexaaqua-manganese(II) ethylenediaminetetraacetate crystal structure (CShM was computed for this at 0.138 for octahedral, essentially perfectly octahedral).[50] The optimisations all fell into difficulty converging, as intermolecular H-bonding consistently distorted the desired octahedral shape, Figure 3.9. UB3LYP/3-21G without solvation was able to produce a CShM of 2.049 for octahedral, however clearly showed distortion visually. Solvation *via* IEFPCM, and then increasing functional and basis set complexity was attempted without success (O_h CShM: 12.590 and 9.130 respectively; trigonal prismatic CShM: 2.417 and 2.071 respectively). Difficulty in finding the optimised geometry of $[Mn(H_2O)_6]^{2+}$ has been found in literature as well.[51–53] Knapp and Galstyan found that transition metal (II) hexaaqua complexes needed to have their bond angles constrained in order to provide the correct geometries.[51] Without such geometry constraints, intermolecular H-bonds between water ligands caused a significant distortion of the shape from the expected 90° bond angles. They posited that the difficulty in optimising correctly was due to the directionality of the d-orbitals meaning that the 90° bond angles were easier to obtain in transition metal (III) complexes over (II) complexes.

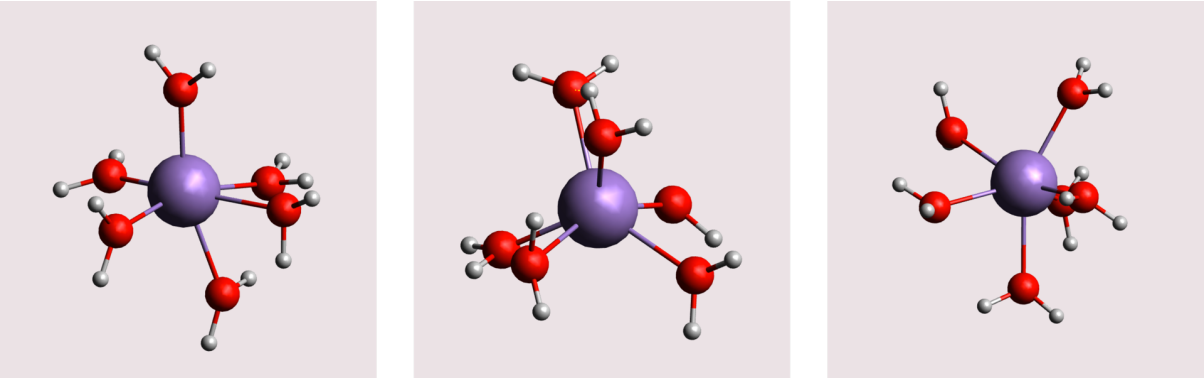


Figure 3.9: Attempted optimisations of $[Mn(H_2O)_6]^{2+}$. Left: UB3LYP/3-21G; centre: UB3LYP/3-21G/IEFPCM; right: UM06-2X/TZVP/IEFPCM.

To overcome this, a geometry optimised manganese structure was sought in the literature. Rudolph and Irmer optimised a $[Mn(H_2O)_6(H_2O)_{12}]^{2+}$ complex, with 12 explicit water molecules in the outer hydration sphere with UB3LYP/6-311+G(d,p)/PCM model chemistry.[48] This structure was used (with the 12 explicit water molecules removed) as the structure that frequency calculations were carried out on for the Gibbs energies with the three functional/basis set combinations and two solvation models. Water, ligands, and complexes were optimised with no imaginary frequencies to confirm minima. For structures that were especially difficult to optimise, single-point energy (SPE) calculations were carried out on geometries that were optimised with a different functional and the zero-point energy was replaced in the original Gibbs energy. All complexes were tested for spin contamination (S^2 - S(S+1)), and all showed <0.05% spin contamination.

3.2.2.6 Model training

The inclusion of L169 into the training dataset quickly presented problems for the experiment. The calculated $\log(K_{MnL})$ values were quite a distance below what was expected, Figure 3.10, indicating a systematic over-stabilisation of the ligand (or under-stabilisation of the complex). From the trend the rest of the data exhibits, it would be expected that L169 (actual $\log(K_{MnL})$ of 24.2) would have a calculated $\log(K_{MnL})$ of around 250.00 instead of a value around 200.00 for M06-2X/TZVP/IEFPCM. Considering this anomaly with respect to the rest of the data giving a positive gradient, L169 was removed from the training dataset.

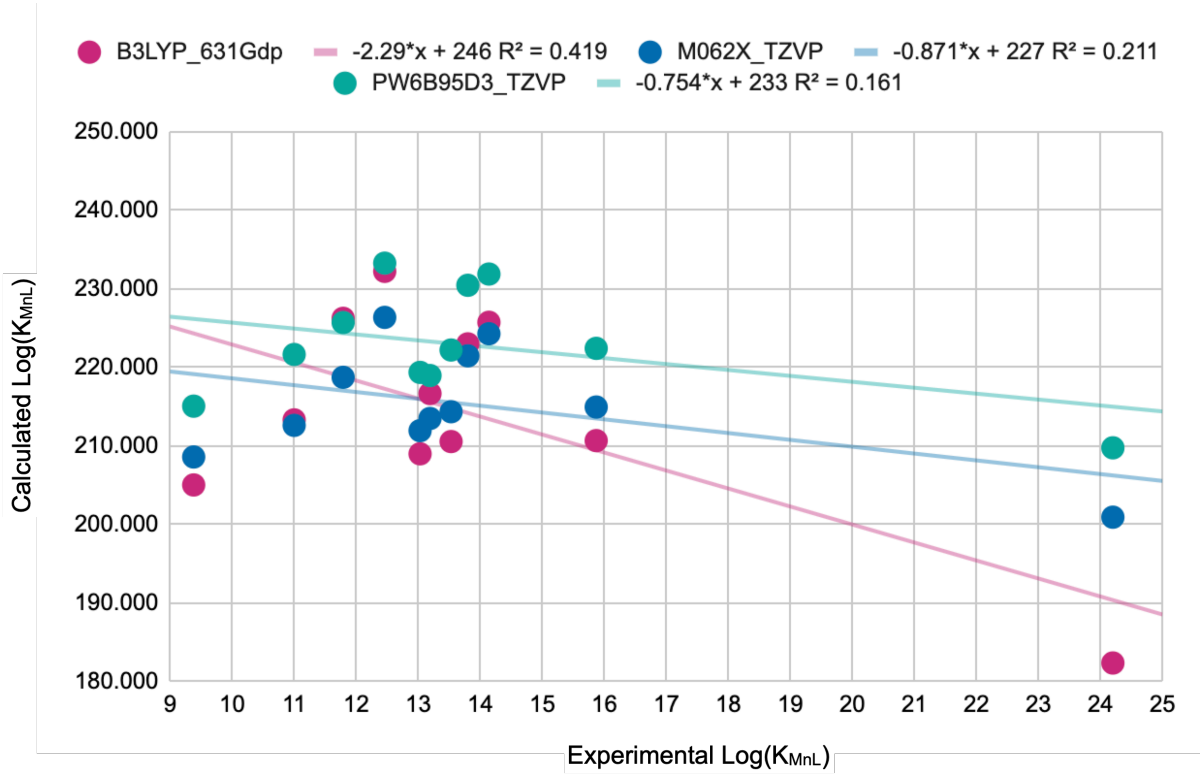
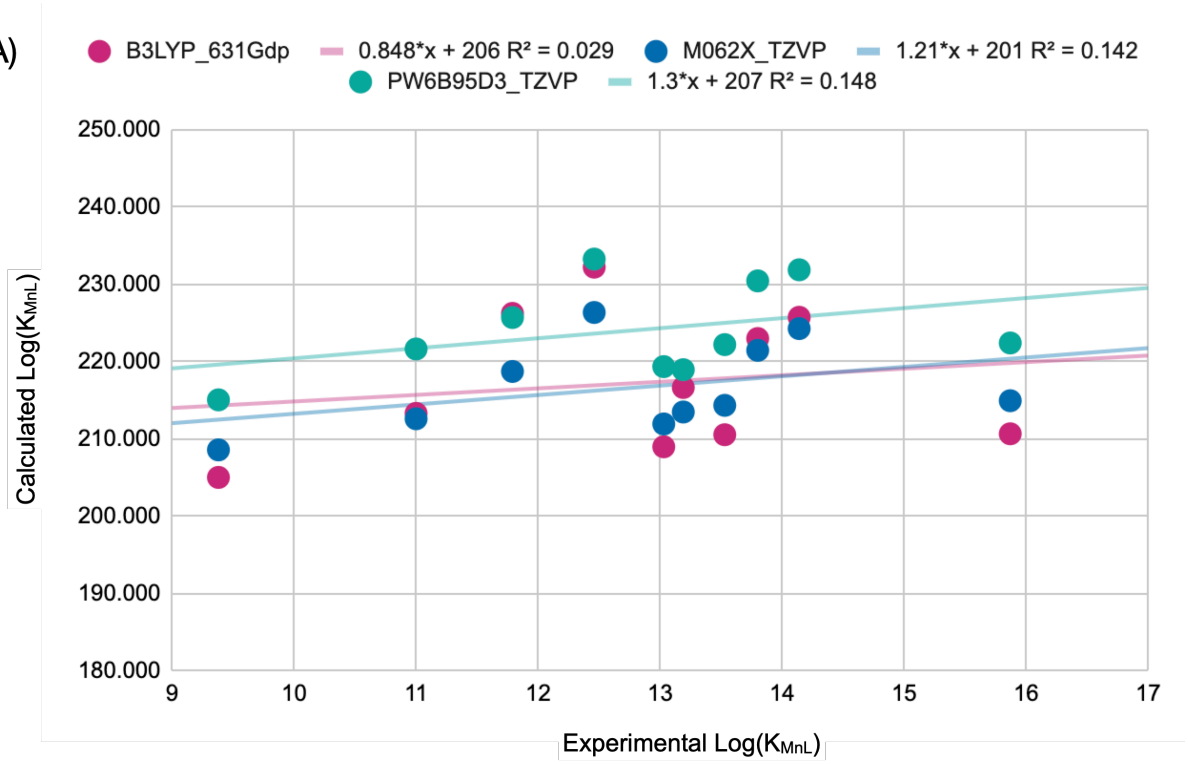


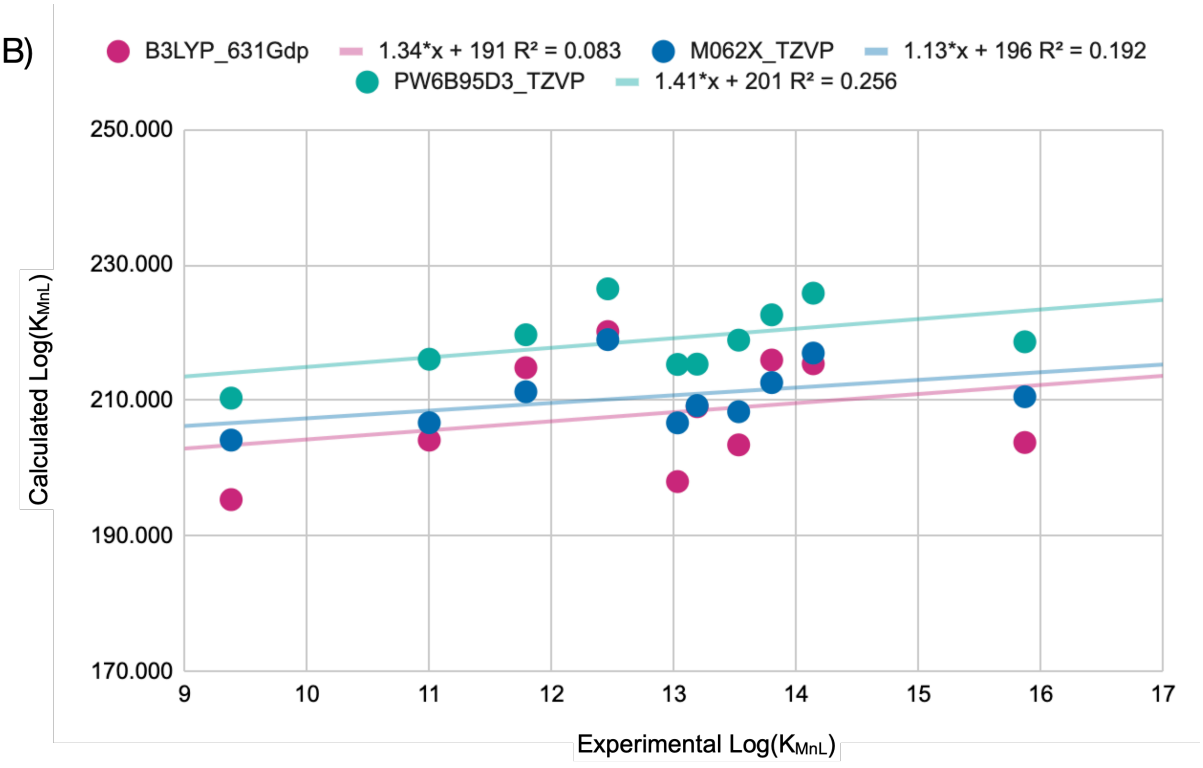
Figure 3.10: Calculated log(K_{MnL}) values from 'With water × IEFPCM' with L169 included in the training dataset.

The calculated log(K_{MnL}) was plotted against literature log(K_{MnL}) and a least-squares method linear regression trend line was calculated for each model, these are shown in Figure 3.11, with their R^2 values summarised in Table 3.3. The top three models are: 'without water × IEFPCM × PW6B95D3/TZVP' (Entry 12, model 1, R^2 0.256), 'with water × IEFPCM(SMD) × PW6B95D3/TZVP' (Entry 9, model 2, R^2 0.232), 'without water × IEFPCM(SMD) × PW6B95D3/TZVP' (Entry 18, model 3, R^2 0.218).

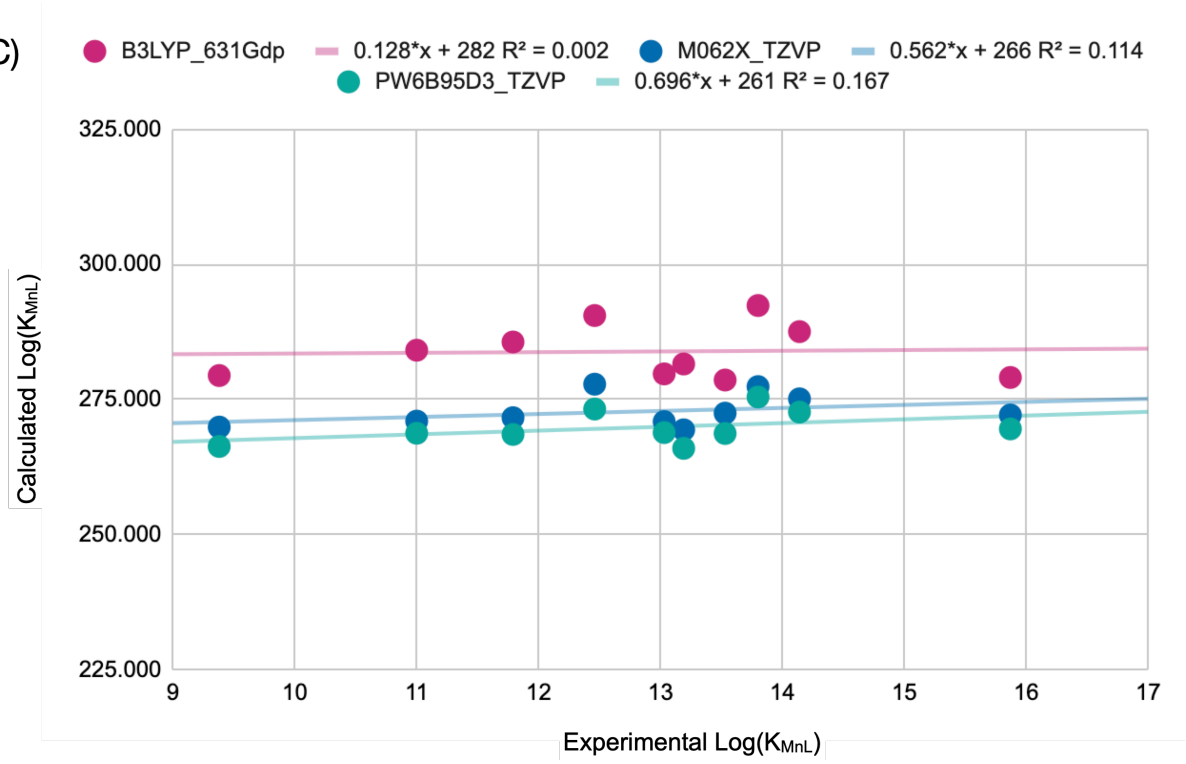
A)


B3LYP_631Gdp
0.848*x + 206 R² = 0.029
M062X_TZVP
1.21*x + 201 R² = 0.142
PW6B95D3_TZVP
1.3*x + 207 R² = 0.148
250.000
240.000
230.000
220.000
210.000
200.000
190.000
180.000
9
10
11
12
13
14
15
16
17
Calculated Log(K_MnL)
Experimental Log(K_MnL)

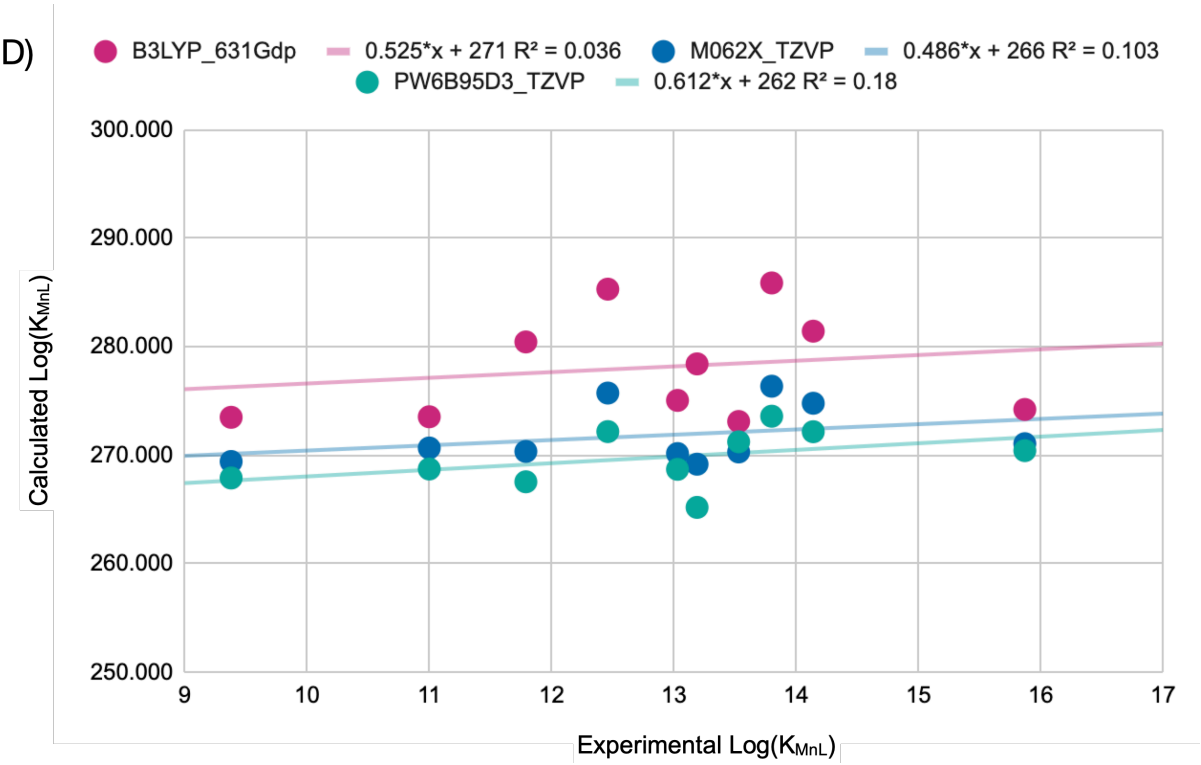
B)

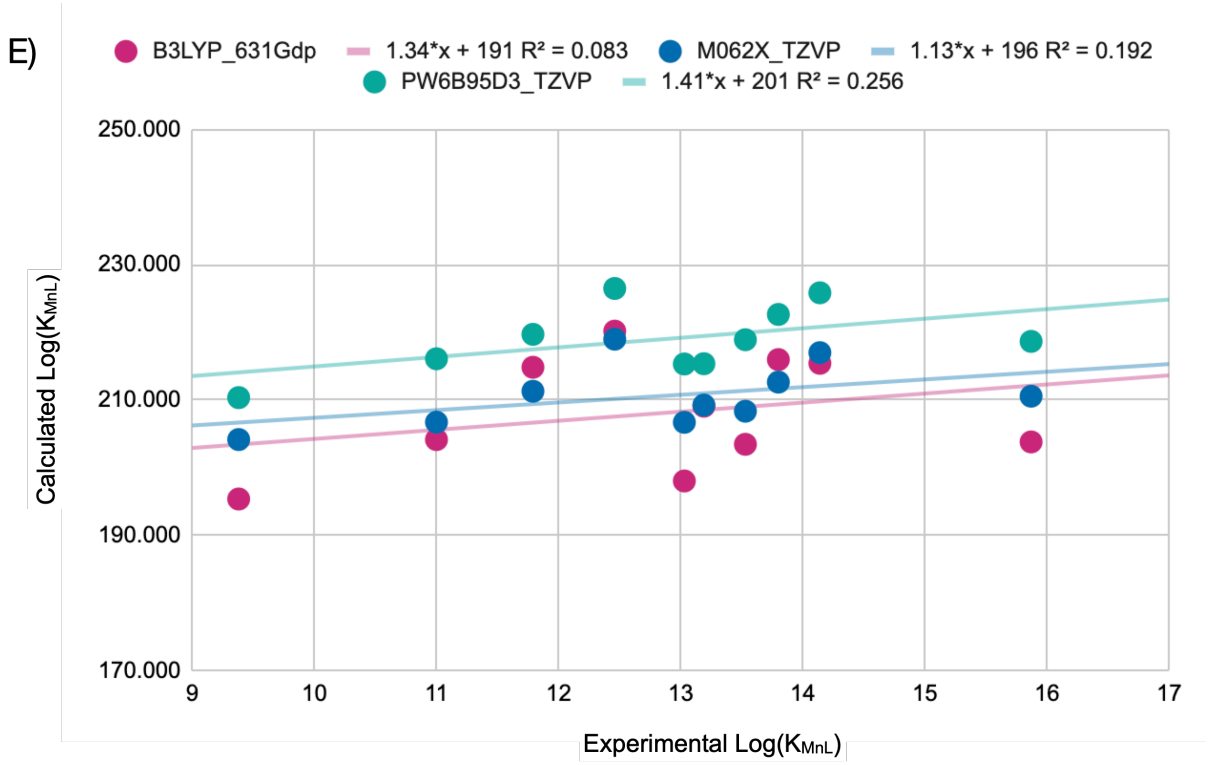

B3LYP_631Gdp
1.34*x + 191 R² = 0.083
M062X_TZVP
1.13*x + 196 R² = 0.192
PW6B95D3_TZVP
1.41*x + 201 R² = 0.256
250.000
230.000
210.000
190.000
170.000
9
10
11
12
13
14
15
16
17
Calculated Log(K_MnL)
Experimental Log(K_MnL)

C)


B3LYP_631Gdp
0.128*x + 282 R² = 0.002
M062X_TZVP
0.562*x + 266 R² = 0.114
PW6B95D3_TZVP
0.696*x + 261 R² = 0.167
325.000
300.000
275.000
250.000
225.000
9
10
11
12
13
14
15
16
17
Calculated Log(K_MnL)
Experimental Log(K_MnL)

D)


B3LYP_631Gdp
0.525*x + 271 R² = 0.036
M062X_TZVP
0.486*x + 266 R² = 0.103
PW6B95D3_TZVP
0.612*x + 262 R² = 0.18
300.000
290.000
280.000
270.000
260.000
250.000
9
10
11
12
13
14
15
16
17
Calculated Log(K_MnL)
Experimental Log(K_MnL)



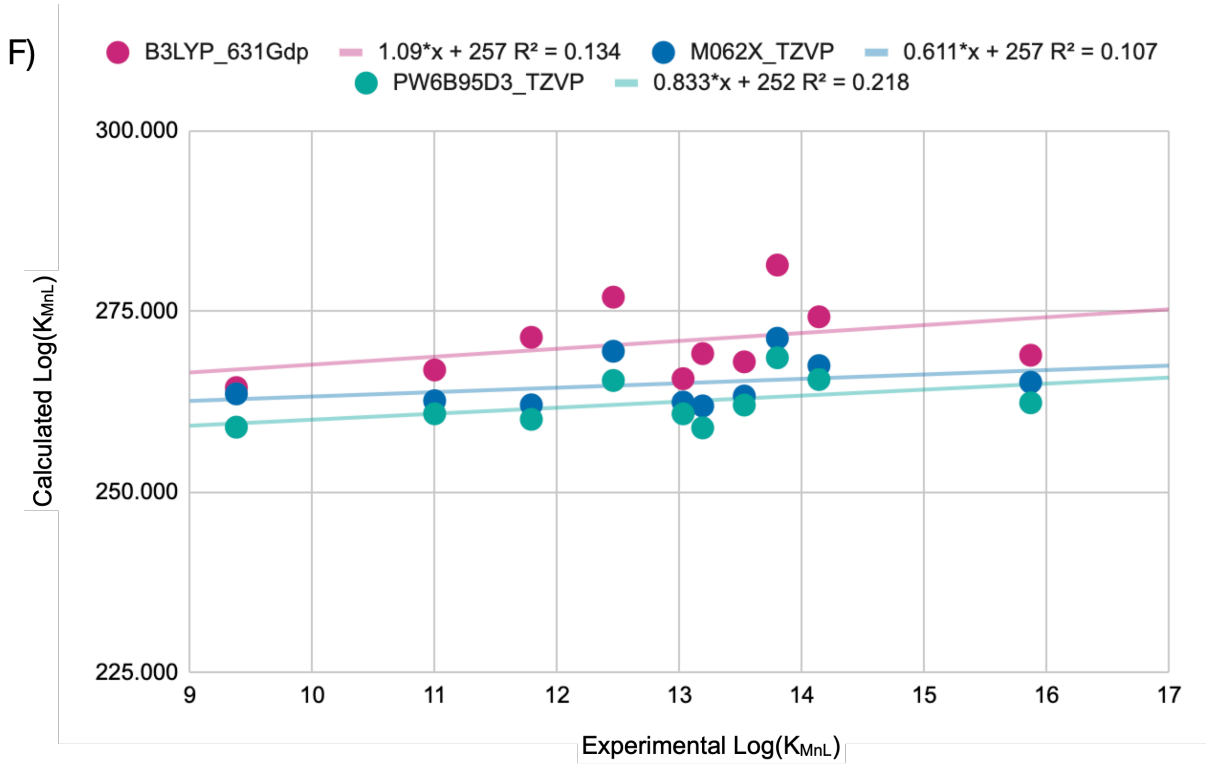


Figure 3.11: Plots of calculated vs. experimental log(K_{MnL}) values. Pink circles: UB3LYP/6-31G(d,p); blue circles: UM06-2X/TZVP; green circles: UPW6B95D3/TZVP. Model A: 'with water × IEFPCM'; model B: 'without water × IEFPCM'; model C: 'with water × SMD'; model D: 'without water × SMD'; model E: 'with water × IEFPCM(SMD)'; model F: 'without water × IEFPCM(SMD)'.

Table 3.3: R^2 values from the least squares linear regression between calculated and experimental log(K_{MnL}) by model.

Entry	Hydration sphere	Solvation	Functional × Basis set	R^2
1	With	IEFPCM	B3LYP/6-31G(d,p)	0.029
2	With	IEFPCM	M06-2X/TZVP	0.142
3	With	IEFPCM	PW6B95D3/TZVP	0.148
4	With	SMD	B3LYP/6-31G(d,p)	0.020
5	With	SMD	M06-2X/TZVP	0.114
6	With	SMD	PW6B95D3/TZVP	0.167
7	With	IEFPCM(SMD)	B3LYP/6-31G(d,p)	0.020
8	With	IEFPCM(SMD)	M06-2X/TZVP	0.118
9	With	IEFPCM(SMD)	PW6B95D3/TZVP	0.232
10	Without	IEFPCM	B3LYP/6-31G(d,p)	0.083
11	Without	IEFPCM	M06-2X/TZVP	0.192
12	Without	IEFPCM	PW6B95D3/TZVP	0.256
13	Without	SMD	B3LYP/6-31G(d,p)	0.036
14	Without	SMD	M06-2X/TZVP	0.103
15	Without	SMD	PW6B95D3/TZVP	0.180
16	Without	IEFPCM(SMD)	B3LYP/6-31G(d,p)	0.134
17	Without	IEFPCM(SMD)	M06-2X/TZVP	0.107
18	Without	IEFPCM(SMD)	PW6B95D3/TZVP	0.218

The performance of each component of the models is analysed in the average ranks in Table 3.4. Unsurprisingly, the most adept functional PW6B95D3 outperforms M06-2X and B3LYP, with an average rank of 4.0 compared to 9.7 and 14.7. The additional explicit water molecules in the hydration sphere had a lower average rank than without (10.4 vs. 8.4). The SMD solvation method, on average, performed the worst of the three solvation methods (11.2), with IEFPCM performing the best (8.3), and the replacement energy method performed roughly as well (8.8).

Table 3.4: Average ranking of each component of the models.

Hydration sphere	With	Without
Average rank	10.4	8.4

Functional × Basis Set	B3LYP	M06-2X	PW6B95D3
Average rank	14.7	9.7	4.0

Solvation	IEFPCM	SMD	IEFPCM(SMD)
Average rank	8.3	11.2	8.8

The low R^2 for all of these models indicate that the link between the calculated and the actual $\log(K_{MnL})$ is not accurately described by a linear regression. This would suggest that the models would all fail in accurately predicting the $\log(K_{MnL})$ values of the testing dataset, however could still be capable of distinguishing between isomers. These top three performing models were taken forward to analyse the testing dataset.

3.2.2.6.1 Clustering

Within the graphs in Figure 3.11, it visually appears as if the calculated log(K) data shows clustering. Clustering was found in Bryantsev's DFT prediction of uranyl complexes' stability constants, in which they found that splitting their data by complex charge provided higher R^2 values than when combined.[8] The top three models were segmented by charge, Figure 3.12, but did not show appreciable difference between the lines of best fit per each charge, nor were all of the points of a specific charge contained in one cluster.

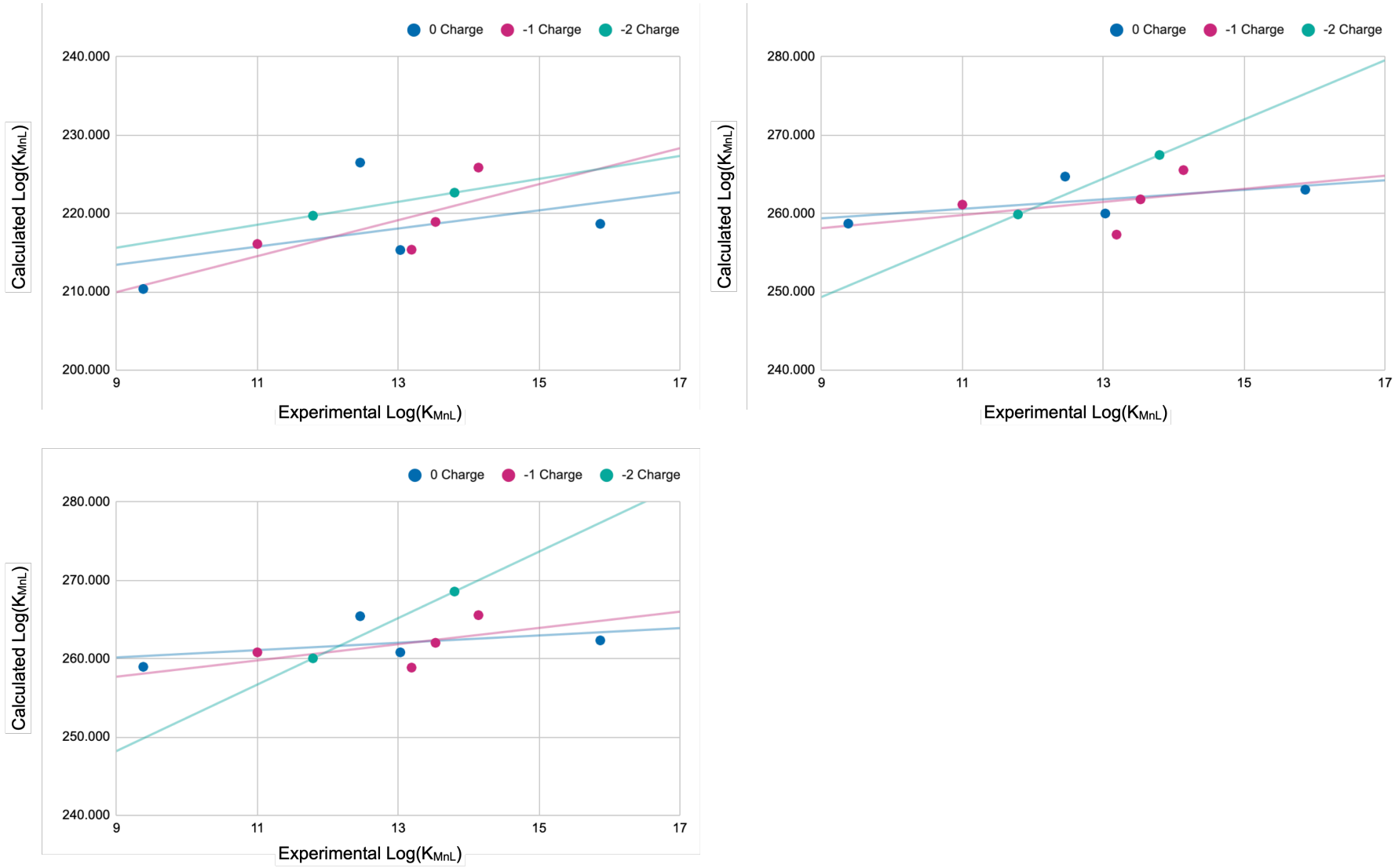


Figure 3.12: Calculated vs. Experimental log(K_{MnL}) plots of the top three performing models segmented by complex charge, top left – model 1, top right – model 2, bottom left – model 3.

3.2.2.7 Model testing

Table 3.5: Experimental and predicted log(K_{MnL}) values for each ligand in the testing dataset.

Ligand	Experimental	Model 1	Model 2	Model 3
1,3-H_4CBuDTA	10.78	14.39	13.44	11.95
1,2-H_4CBuDTA	10.26	15.46	16.19	15.00
trans-H_4CDTA	14.69	20.23	18.18	20.69
cis-H_4CDTA	14.19	19.39	16.78	17.32
3,6-H_2PC2A	15.53	10.91	12.48	11.97
3,9-H_2PC2A	17.09	12.06	13.53	13.94
3,9-H_2PC2A-BP	14.86	10.23	10.85	11.07
Bispidine-COOH	19.47	10.24	9.06	10.64

As suggested by their low R^2 values, each of the models had significant errors, Table 3.5 and Figure 3.13. Model 2 and 3 performed better than 1 in predicting the log(K_{MnL}) values of the testing dataset complexes. This was unexpected as their training sets gave lower R^2 values than model 1. As before, the bispidine complex heavily contributes towards the prediction error. Removing L172, the prediction RMSE and MAE both improve, with MAE doing more so (as it is less sensitive to outliers than RMSE, which squares its errors), Table 3.6. Model 3 gave a MAE of 1.79 and RMSE of 3.92 for its log(K_{MnL}) prediction, this is on a similar magnitude as Uzal-Varela's MAE of 0.99 and RMSE of 1.48 (based on 150 complexes and with bispidines also removed).

Table 3.6: Error analysis results for each model's predicted log(K_{MnL}) values.

Model	RMSE	RMSE (without L172)	MAE	MAE (without L172)
Model 1	5.73	4.91	5.85	2.23
Model 2	5.23	3.72	4.36	1.84
Model 3	4.93	3.92	4.25	1.79

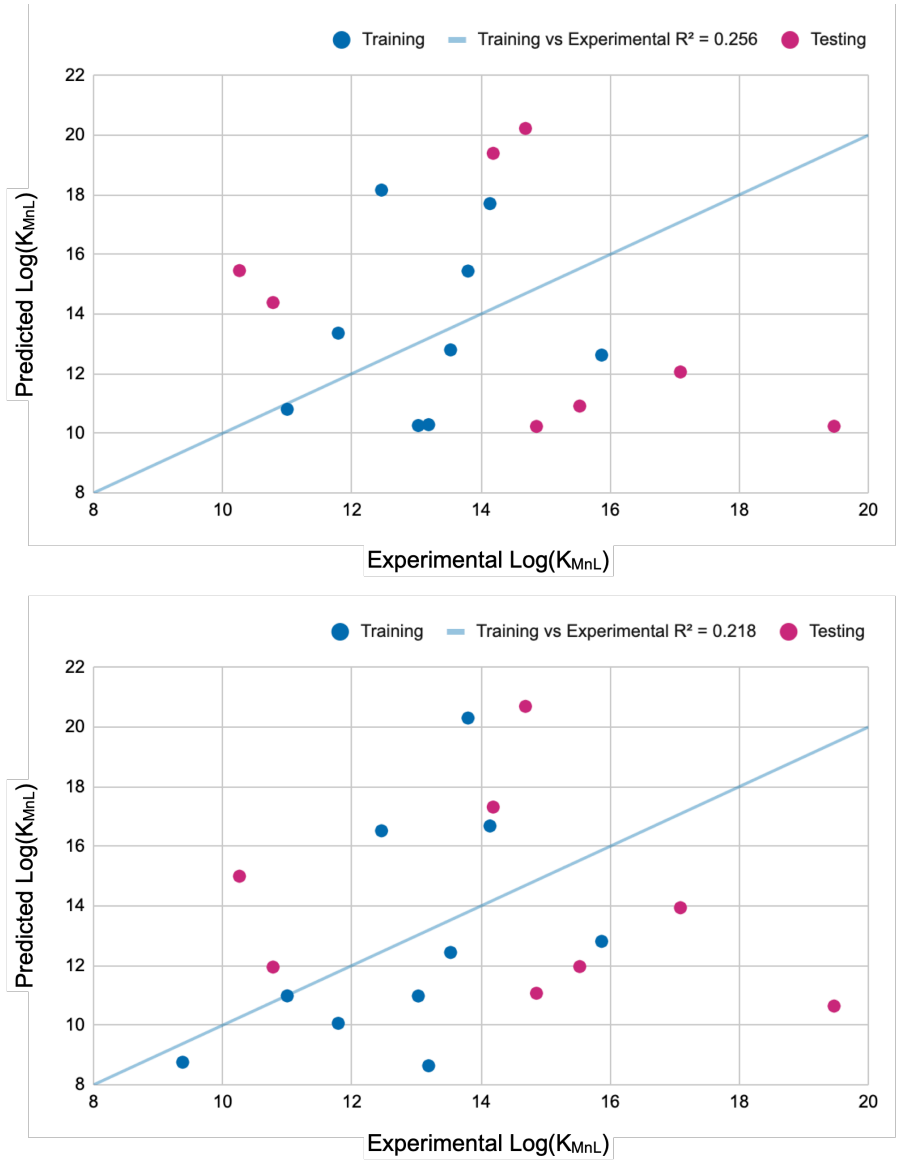




Figure 3.13: Predicted log(K_{MnL}) values for the testing and training datasets per model, top left – model 1, top right – model 2, bottom left – model 3.

3.2.2.8 Model prediction

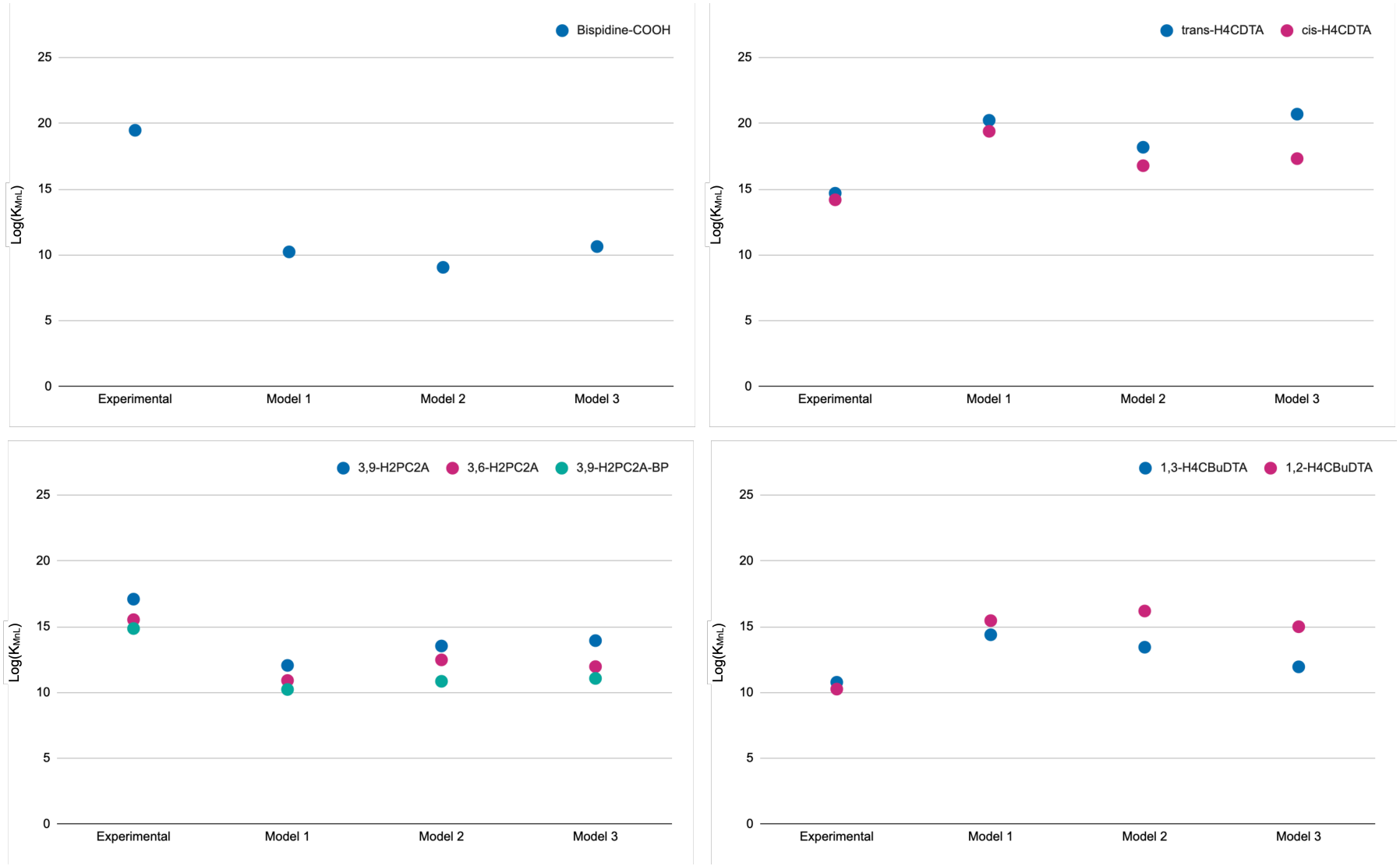


Figure 3.14: Predicted log(K_{MnL}) values by ligand family vs. experimental log(K_{MnL}) values

3.2.2.8.1 Bispidine-COOH

As seen with the removal of **L169** from the training set, the DFT calculations result in bispidine-COOH being attributed a much lower log(K_{MnL}) than its literature value. Bispidine complexes are structurally different to the majority of the complexes that have been published as MnBCAs. They are the first published example of a MnBCA with a heptadentate ligand *and* possessing a hydration number of 1 as a result of its specialised and rigid cavity. In terms of the model, these factors may have proved to be too different from the rest of the training set to approach a correct prediction even slightly, Figure 3.14.

3.2.2.8.2 Isomer prediction

The top performing models were each able to correctly predict both that a *trans*-CDTA complex would be more stable than a *cis*-CDTA one, and that a 3,9-PC2A complex would be more stable than one with 3,6-PC2A. Additionally, each of the models were able to predict that 3,9-PC2A-BP would be less stable than both 3,6- and 3,9-PC2A – all three are treated identically in the Uzal-Varela method. As expected from the low R^2s, predicted values were on average 4.3 and 3.8 log units away from the experimental, respectively. The predicted values for the CBuDTA isomers were unable to match with experimental. Interestingly, model 2 and 3 provided values that had a much larger energy difference between the CBuDTA isomers than for the others. As model 1 is purely IEFPCM, and 2 and 3 replace the Gibbs energy of IEFPCM with SMD, this indicates that the SMD optimisation has produced a much lower energy geometry for $[Mn(1,3\text{-CBuDTA})(H_2O)]^{2-}$ than in IEFPCM (or much higher energy ligand).

3.2.3 DFT geometry and MnBCA metrics

3.2.3.1 Hypotheses and model conditions

In this section, the aim was to investigate if the optimised geometry from DFT calculations could be correlated to the KPIs for MnBCAs. The first avenue of investigation was to test the idea of whether geometry could have an effect on the stability of the complex. Deviations from established 'stable' coordination geometries will lead to non-optimal coordination, and if this influences stability it might in turn be correlated to the stability constant.

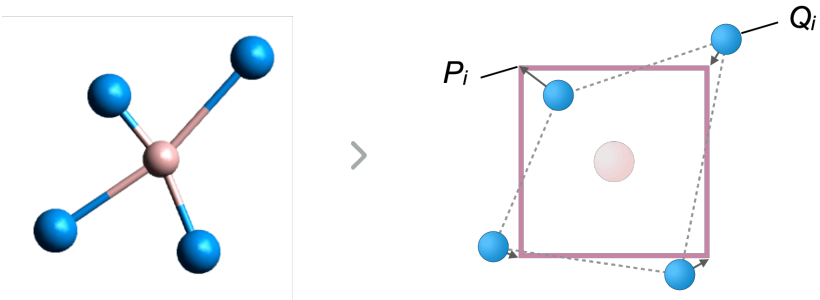


Figure 3.15: Schematic for the calculation of CShM for a fictional NiL_4 complex; the coordinates from its DFT optimised geometry can be compared to a reference polyhedra (in this case a square for square planar), via *Equation 3.9.*

$$CShM(P) = \min\left(\frac{\sum_{i=1}^{n}|\vec{Q}_i - \vec{P}_i|^2}{\sum_{i=1}^{n}|\vec{Q}_i - \vec{Q}_0|^2}\right) \times 100$$

Equation 3.9: Equation for CShM for reference geometry P, where Q_0 is the coordinate vector of the centre of the complex under study, for n donors in the complex.

The ability to measure how well a structure adheres to an optimal coordination polyhedron is provided by the continuous shape measure (CShM), which measures the deviation of a set of coordinates from reference polyhedra – essentially a measure of their shape, Equation 3.9 and Figure 3.15.[54,55] The validation of this theory would be a correlation where low CShM scores (i.e., very little deviation) led to higher stability constants, and increasing deviation destabilised the complex.

Alongside this, the potential effect of the donor bond lengths on log(K_{MnL}), k_{ex}, and r_1 was investigated. This part explored whether specific bonds could be used to predict properties and tested the assumption that a ligand's 'cavity size' impacts a complex's stability constant. The claims in literature surrounding the water exchange rate – metal-water bond length, charge, steric 'pressure' – were also investigated *via* correlating the size of these effects with k_{ex}. For steric 'pressure', the overlap between Van der Waals spheres was calculated between the water molecule and the complex, Figure 3.16, *via* Equation 3.10, and Van der Waal radii as reported by Batsanov.[56]

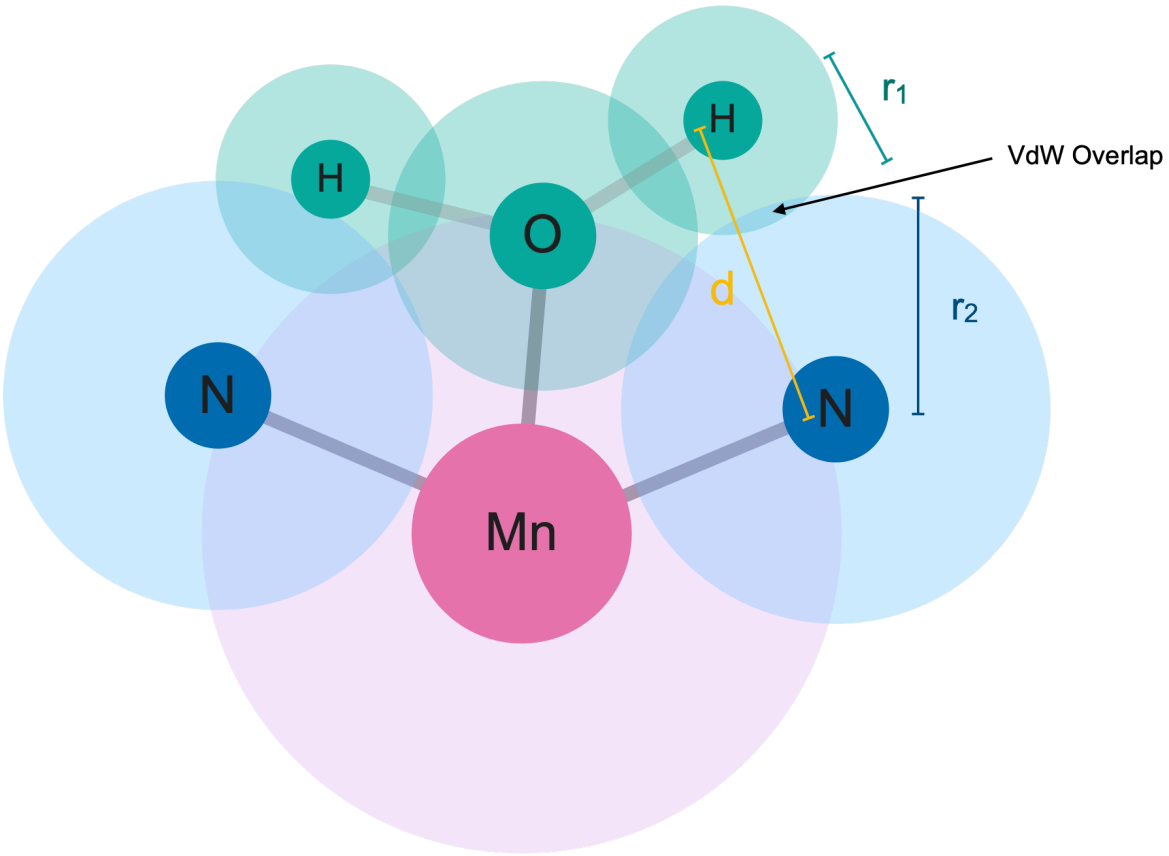


Figure 3.16: Schematic of the variables used in the calculation of two sphere overlap volume.

$$v_{overlap} = \frac{\pi}{12d}(r_1 + r_2 - d)^2(d^2 + 2d(r_1 + r_2) - 3(r_1 - r_2)^2)$$

Equation 3.10: Equation for the volume between two overlapping spheres, where $v_{overlap} = 0$ if $d \geq r_1 + r_2$.

The calculated overlaps were then used to calculate a fractional total overlap with the Van der Waals volume of the water molecule. The water molecule volume was treated as two non-interacting O-H bonds with the three-sphere volume overlap not considered due to the extremely large increase in complexity in considering a three-sphere overlap compared to a two-sphere overlap. This assumption means that the water volume is slightly overestimated as the area where all three atoms overlap is counted twice, though this 'extra' overlap is assumed to be small relative to the overall volume. Next, the overlap with atoms from the complex were calculated between each of the three atoms from the water molecule. Again, to avoid the three-sphere problem, overlaps were calculated as two-sphere overlaps. In a similar fashion, the overlap between ligand atoms and hydrogen is infrequent and minor compared to the overlap between ligand atoms and oxygen (due to the oxygen naturally being closer to the complex *via* the Mn-O bonding).

3.2.3.1.1 Evaluating model chemistries

Choosing the correct model chemistry in DFT is very important in producing accurate optimised geometries, which in turn can have better results when calculating thermochemical properties – or in the case of this analysis to correlate the geometries to MnBCA parameters. This required evaluating the model chemistries against a benchmark structure. Ideally, the benchmark would be as close to the real situation that we are trying to model with DFT. In this case, the averaged structure of a MnBCA in solution would be perfect, however, this data is not available and so the next best candidate is to use crystal structures. Therefore, three crystal structures were chosen to benchmark against: K[Mn(DPAAA)(H_2O)], [Mn(H_2O)$_5$][Mn(PhDTA)(H_2O)], [Mn(trans-O2DO2A)(H_2O)].[20,57,58] The ligands of these complexes are shown in Figure 3.17. DFT geometry optimisations with nine different model chemistries were undertaken with IEFPCM as the SCRF method. These models were chosen to assess model chemistries used in MnBCA literature (M06-2X/TZVP and TPSSh/def2TZVPP), selected functionals using TZVP (APFD/TZVP, PW6B95D3/TZVP, ωB97XD/TZVP) chosen for their good performance in previous benchmarks,[43,45] two model chemistries to evaluate the benefit of def2TZVPP over TZVP (M06-2X/def2TZVPP and PW6B95D3/def2TZVPP), one model chemistry suggested by Foresman and Frisch (UAPFD/6-311+G(2d,p)),[45] and finally B3LYP/6-31G(d,p). The optimised geometries were evaluated *via* mean absolute error (MAE), comparing the crystal structure geometry with the DFT optimised geometry. As well as the entire structure (heavy atoms only), the ability to replicate the coordination sphere is very important in this analysis – therefore, the MAE of all donors to Mn, the water to Mn, N donors to Mn, and O donors to Mn were also calculated,

Table 3.7.

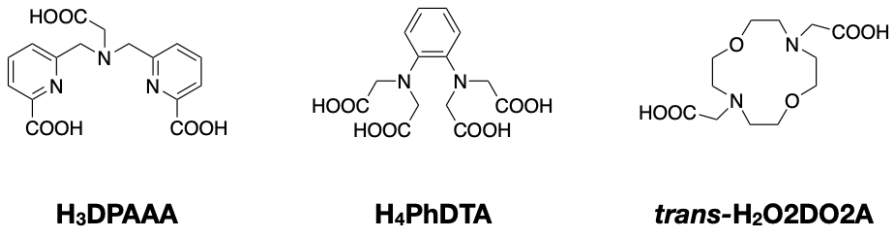


Figure 3.17: Ligands for the MnBCAs used to benchmark the DFT model chemistries.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |r_{XRD,i} - r_{DFT,i}|$$

Equation 3.11: Equation for MAE between crystal structure (XRD) and DFT geometry optimisation (DFT), where n is the total number of bonds, and r is the bond length.

Table 3.7: Results of the MAE analysis with MAE averaged for each benchmarking MnBCA, then averaged between the three benchmarking MnBCAs.

Method and Basis Set[a]	MAE (all bonds, average)[b]/Å	MAE (all donors, average)/Å	MAE (Mn-O (water), average)/Å	MAE (all N donors, average)/Å	MAE (all O donors, average)/Å
UAPFD/TZVP	0.0368	0.0481	0.0263	0.0564	0.0452
UM062X/TZVP	0.0234	0.0302	0.0081	0.0427	0.0219
UPW6B95D3/TZVP	0.0339	0.0338	0.0209	0.0342	0.0335
UωB97XD/TZVP	0.0345	0.0415	0.0191	0.0527	0.0350
UM062X/def2TZVPP	0.0247	0.0340	0.0054	0.0506	0.0235
UPW6B95D3/def2TZVPP	0.0262	0.0363	0.0238	0.0257	0.0398
UTPSSh/def2TZVPP	0.0281	0.0513	0.0310	0.0408	0.0551
UB3LYP/6-31G(d,p)	0.0329	0.0637	0.0157	0.0812	0.0535
UAPFD/6-311+G(2d,p)	0.0287	0.0475	0.0319	0.0577	0.0440

[a] The U before the model chemistry indicates that the system is unrestricted – i.e. there are unpaired electrons in the system, known as an open-shell system. [b] Mean absolute error of the bonds between all atoms apart from C-H bonds.

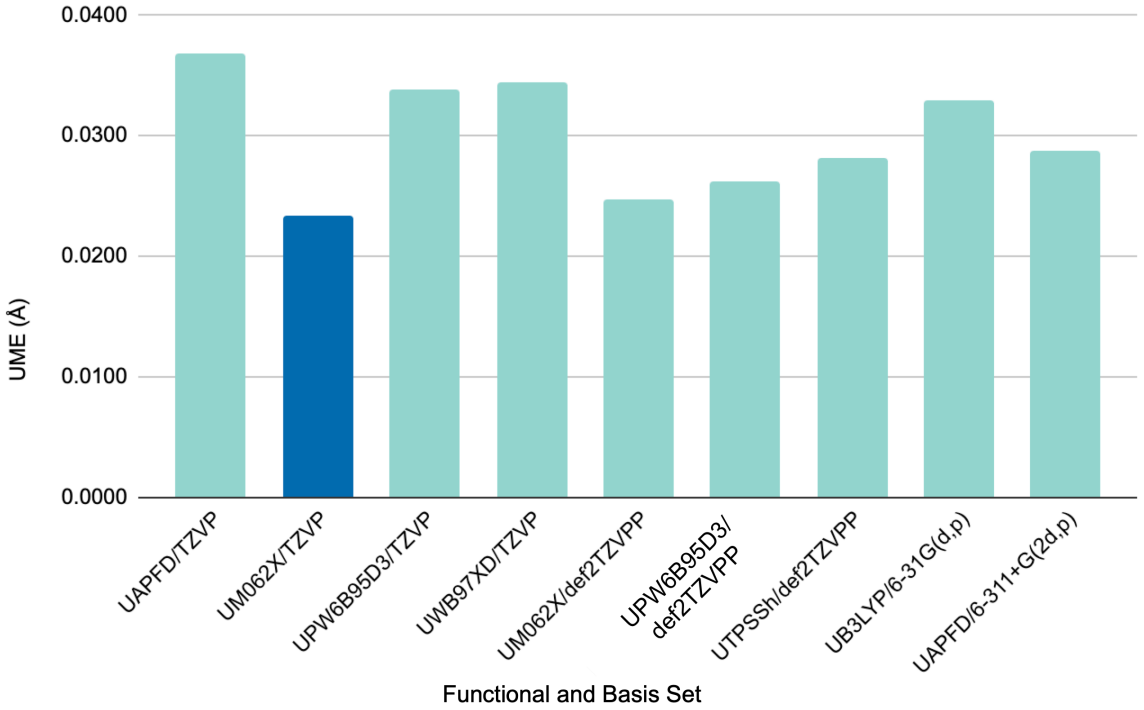


Figure 3.18: Chart of MAEs (all bonds except C-H, averaged for all three benchmarking complexes) per model chemistry, lowest scoring functional/basis set highlighted.

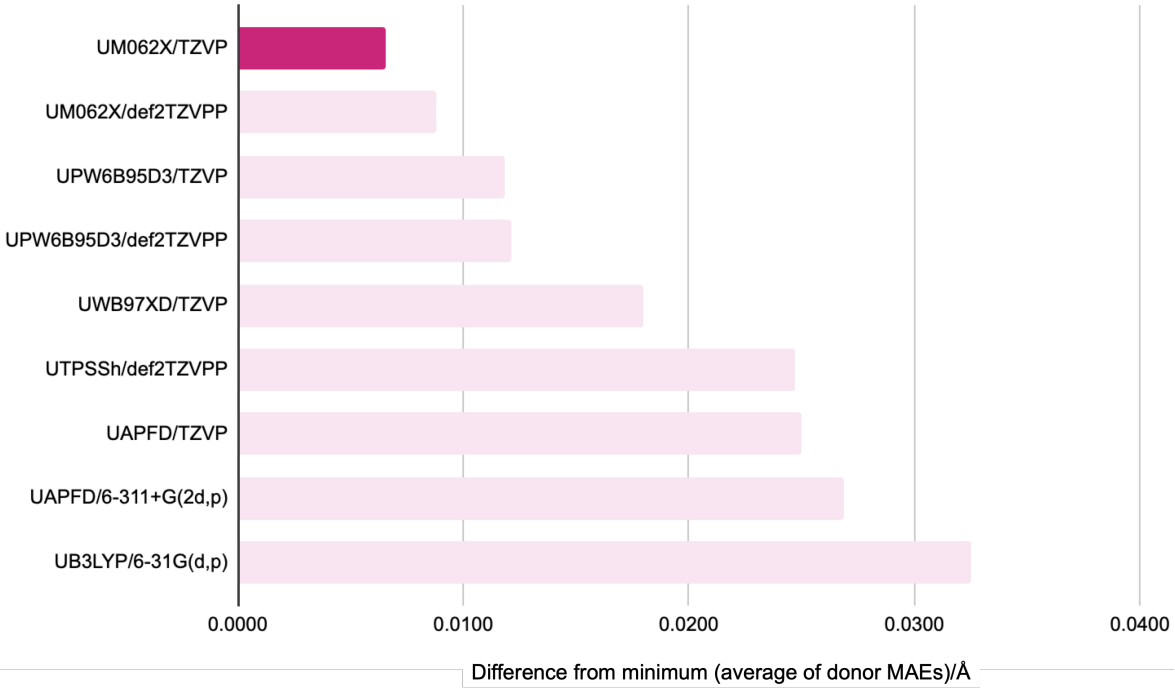


Figure 3.19: Chart of difference from minimum for donor MAEs per model chemistry, lowest scoring functional/basis set highlighted.

The MAE for all bonds (C-H removed) was plotted in Figure 3.18. The M06-2X/TZVP functional/basis set provided the lowest deviation for the nine chemistries. Looking through the other scores for the coordination MAEs, there was no one clear consistent lowest error scoring model chemistry. Figure 3.19 plots the difference a model chemistry has between itself and the lowest error model chemistry (for example in N donors, APFD/TZVP (0.0564 Å) is 0.0307 Å above the minimum in that column (0.0257 Å), averaged between the three donor MAEs (Mn-Water, N donors, O donors). Despite M06-2X not being the lowest error functional/basis set for all of the donor MAEs it performs best when averaging its difference to the minimum. Interestingly, in Figure 3.19, the older TZVP basis set outperformed the newer, larger def2TZVPP basis set for both M06-2X and PW6B95D3; this was a welcome result as the smaller basis set calculations complete significantly faster than def2TZVPP calculations – extremely important in a project like this which is operating at quite a large scale (46 MnBCAs). Therefore M06-2X/TZVP/IEFPCM was chosen as the model chemistry that was used in the project.

3.2.3.1.2 Data collection

As the collection of ligands in this analysis are the same set (essentially the exhaustive list of all published MnBCAs before 2022) that are in the Uzal-Varela analysis, their naming system was carried over to aid in direct comparison and managing the project (Section 8.3). Each of these ligands were investigated to find published data on water exchange rate, r_1, q, and coordination number to add to the collected log(K_{MnL}) data from the Uzal-Varela analysis. This data helped to decide on the required hydration number to accurately optimise a MnBCA *via* DFT. With the model chemistry in hand, the 50 MnBCAs that were used in the Uzal-Varela article (100 more, less relevant MnBCAs were in the supporting information which also contributed to their analysis) were optimised at M06-2X/TZVP/IEFPCM level. Four of these failed to optimise – **L7** (H_3NTA) was removed due to being an extremely poor MnBCA, **L34** (H_4DOTP), **L42** (H_2NO2A), and **L47** (1,4,7,10,13-pentaazacyclopentadecane) were removed due to difficulty in finding a minimum on the potential energy surface (PES) during optimisation.

3.2.3.1.3 Success parameters

In the analysis, the aim was to find a correlation between a MnBCA KPI such as the stability constant and a measured length, or metric calculated from the DFT geometry optimisation. The relationship between the two was defined either *via* a linear or quadratic fit to the data, and then assessed *via* the coefficient of determination, R^2. This statistic allows evaluation of predictive models as it measures how well the actual values match to the regression line from the model. Essentially, it measures the variance or distance the actual values have, Figure 3.20. Equation 3.12 shows that it measures the ratio between the sum of pink squares and the sum of the blue squares from Figure 3.20, with smaller pink squares (i.e., a very well-fitting trend line) providing a smaller number to remove from 1, and so a larger R^2 value.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - y_{average})^2}$$

Equation 3.12: Equation for the coefficient of determination, R^2, pink highlighting the sum of squares of the difference to the fitted trend line, blue highlighting showing the sum of the squares of the difference to $y_{average}$, seen in Figure 3.20 .

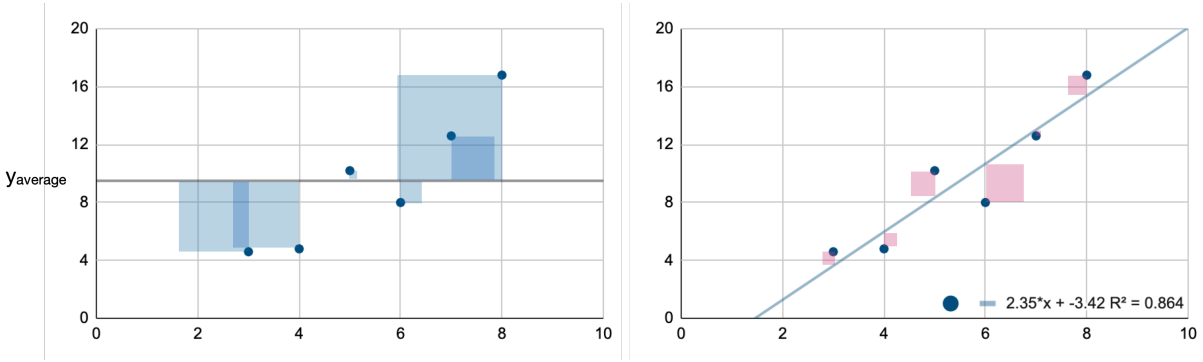
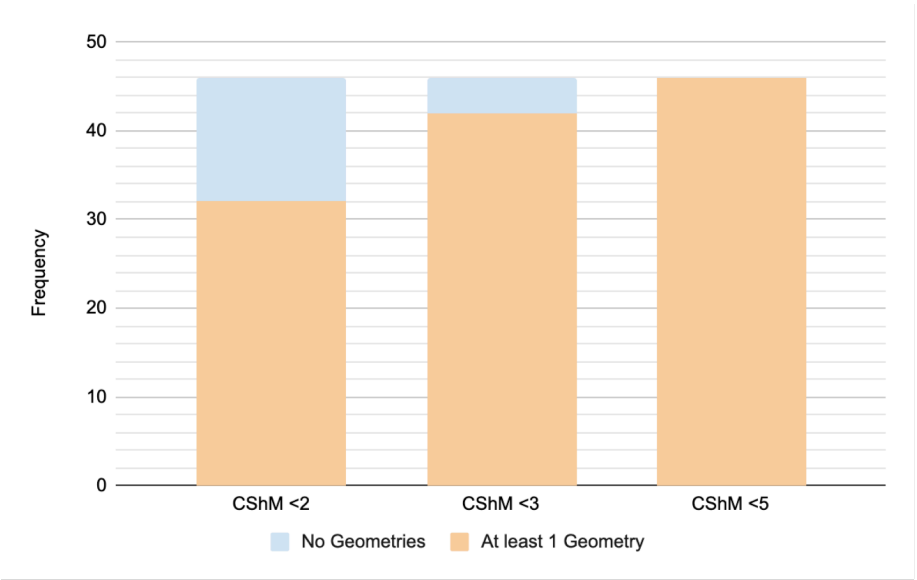
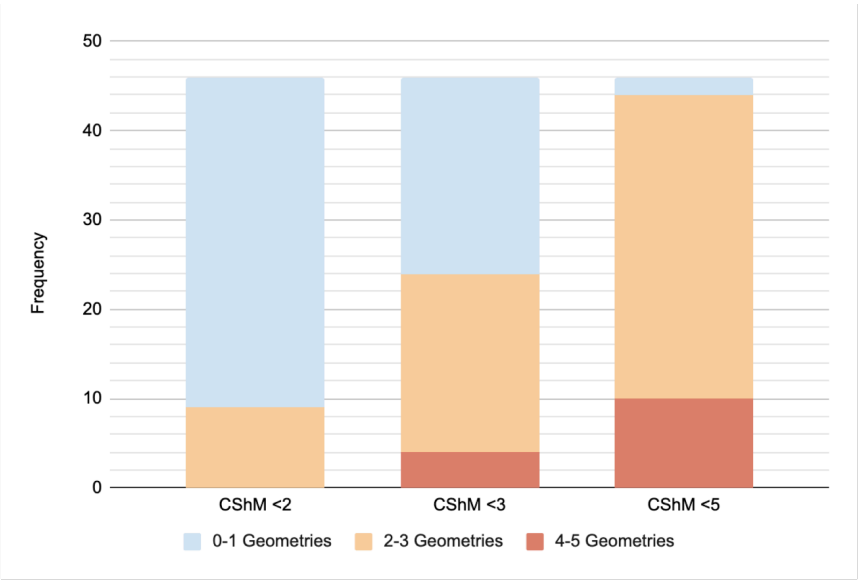
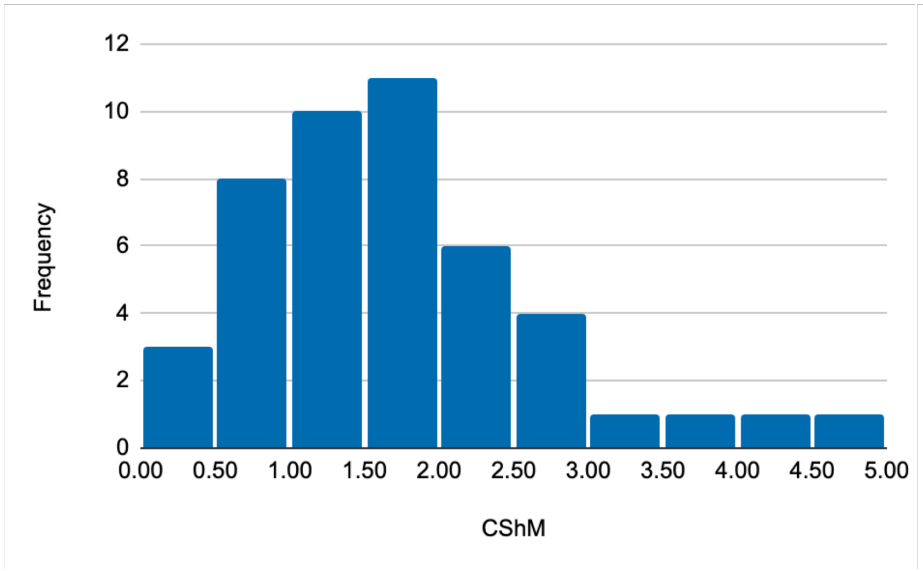


Figure 3.20: Plots showing how the coefficient of determination, R^2, is calculated in tandem with Equation 3.12.

3.2.3.2 Continuous Shape Measure and log(K_{MnL})







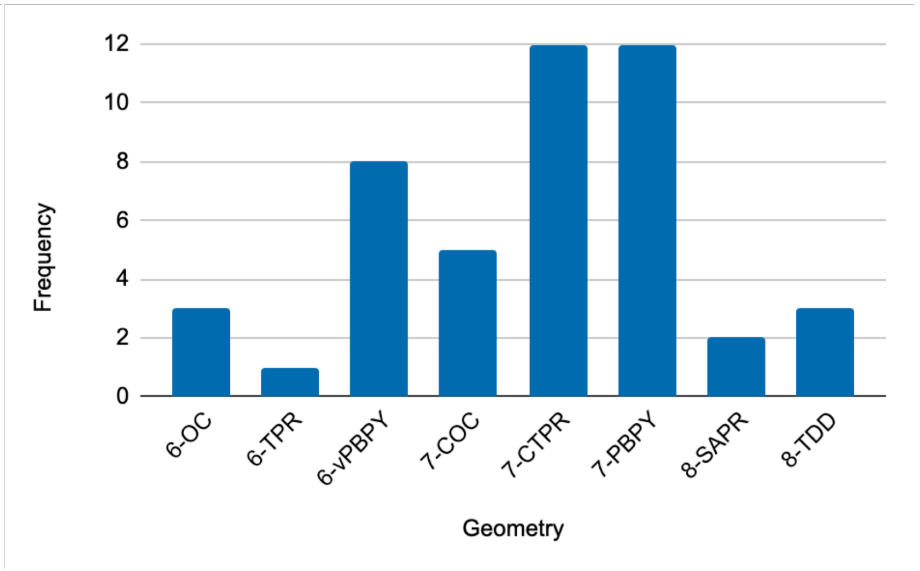


Figure 3.21: Figures of data exploration around the MnBCA dataset with respect to CShM and geometries. Top left: stacked column chart of the proportion of MnBCAs that can be described by standard geometries below a certain CShM threshold; top right: stacked column chart of the proportion of MnBCAs that can be described by standard geometries below a certain CShM threshold, split by the number of geometries that fall within the threshold; bottom left: histogram of the MnBCAs' lowest CShM; bottom right: frequency of geometries (with lowest CShM per MnBCA) optimised for octahedral (6-OC), trigonal prism (6-TPR), vacant pentagonal bipyramidal (6-vPBPY), capped octahedral (7-COC), capped trigonal prism (7-CTPR), pentagonal bipyramidal (7-PBPY), square antiprism (8-SAPR), and triangular dodecahedron (8-TDD).

The first area analysed for trends to MnBCA KPIs was in geometry and shape. Of the 46 MnBCAs analysed, 70% can be described by a CShM score below 2, Figure 3.21 (top left), rising to 91% for below 3, and 100% with a score below 5 for at least one geometry. Expanding this out, Figure 3.21 (top right) we can see that most of the complexes are described by multiple geometries depending on the threshold – 52% are described by more than one geometry with scores below 3, and when the cut off is loosened to a CShM below 5, 96% of MnBCAs have multiple geometries that describe them. For the wider analysis, only the lowest scoring geometry and its CShM was used to correlate to its MnBCA's stability constant.

The distribution of the lowest CShM for each MnBCA was analysed (Figure 3.21 (bottom left)) and revealed that the dataset is negatively skewed. It is expected that these published complexes would, on the whole, be well described by standard geometries, and for their CShMs to more likely be below 2 than above it. In terms of the geometries themselves, Figure 3.21 (bottom right), 7-coordinate species (water coordination included) are most popular with an equal number of MnBCAs optimising as capped trigonal prism or pentagonal bipyramidal. Into the 6-coordinate species, vacant pentagonal bipyramidal (vacant on the equatorial ring) was seen most often. The average value for the 6-coordinate species was 2.78, suggesting that the ligands used to complex the manganese are too restrictive, or bulky, to allow the donors to position themselves in a position that minimises steric clash and maximises orbital overlap. Removing the option of a vacant pentagonal bipyramidal geometry changes the average to 3.86, further indicating the mismatch between ligand design and optimal coordination around the manganese.

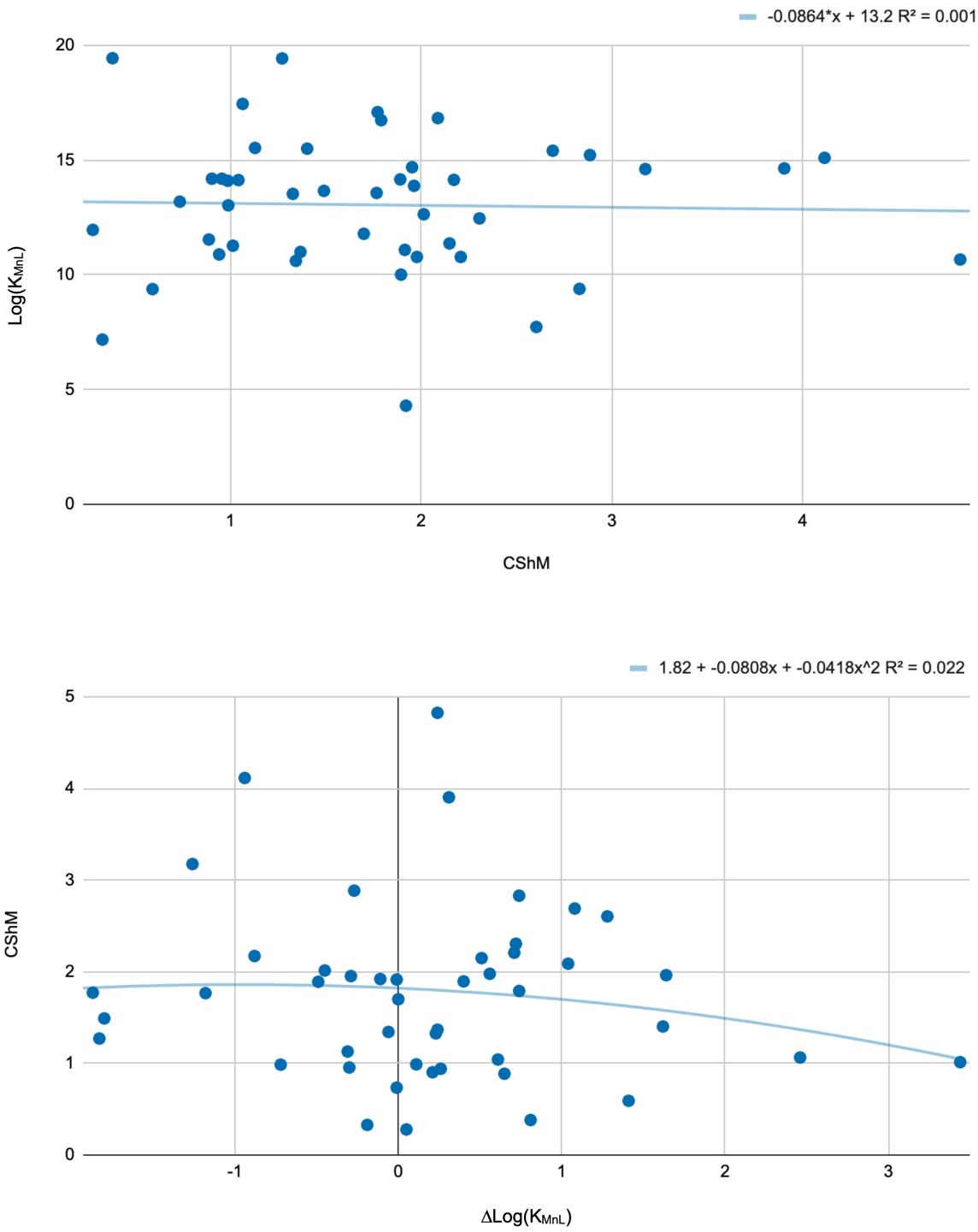
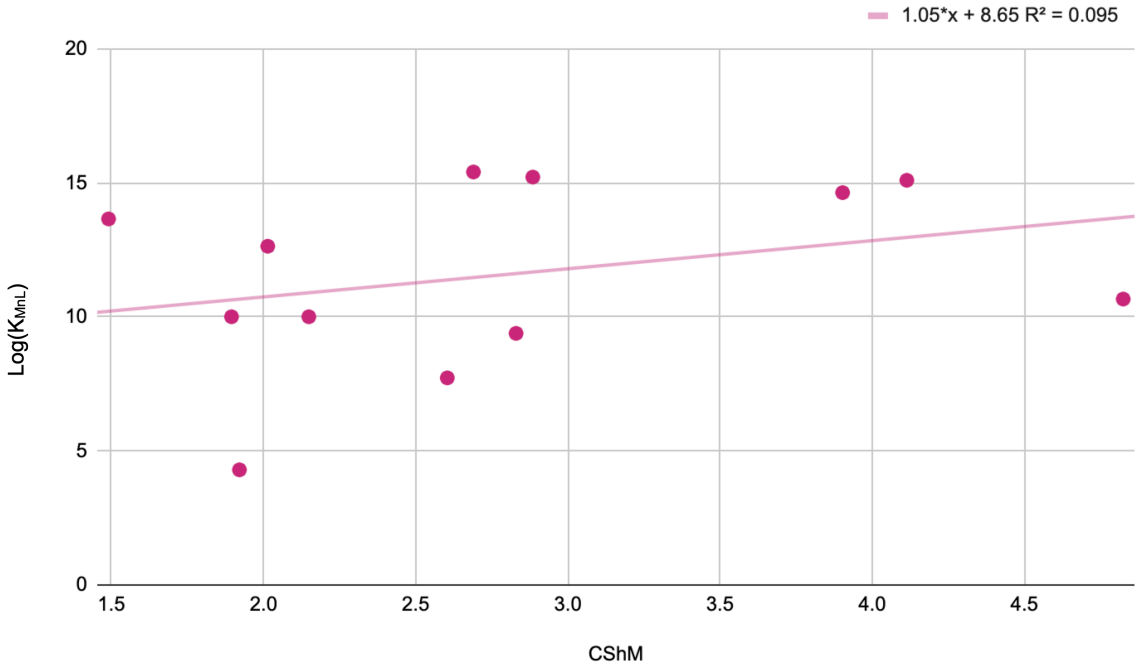
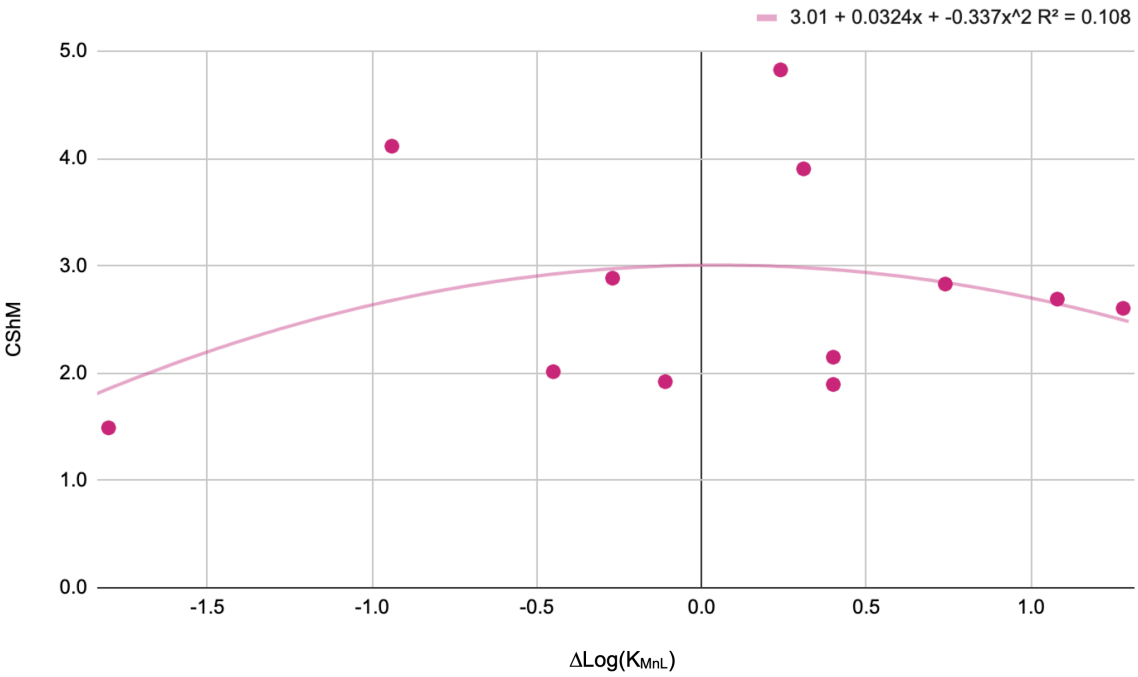


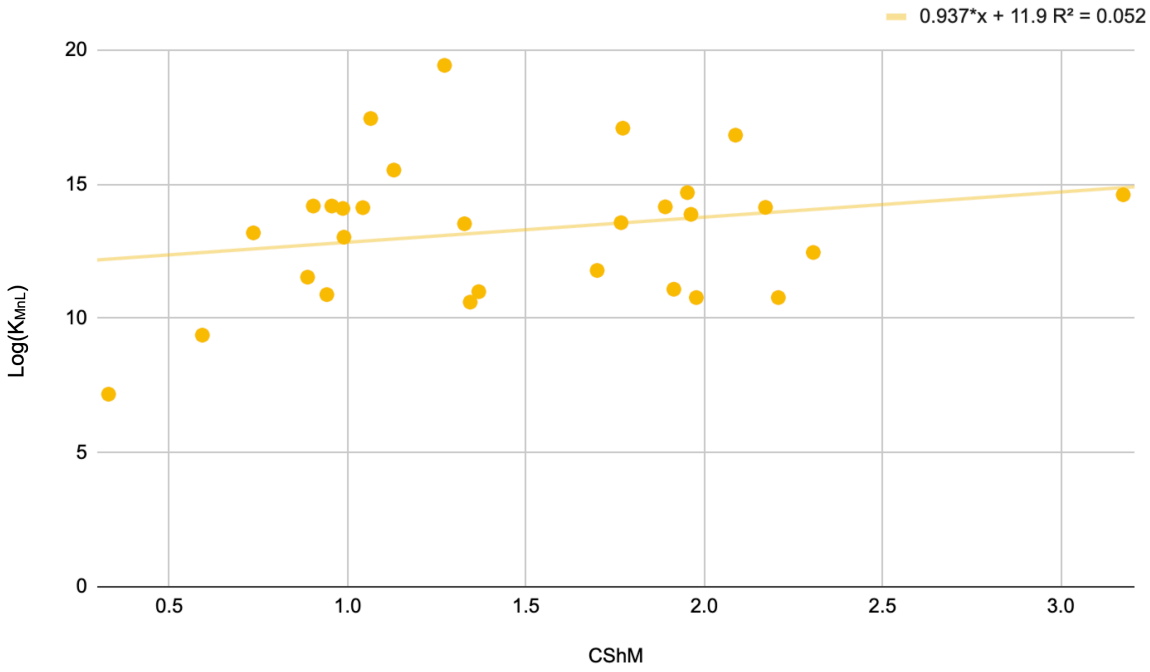
Figure 3.22: Top: graph of log(K_{MnL}) vs lowest CShM, R^2 for the trend line is 0.001; bottom: lowest CShM vs the residual between Uzal-Varela's log(K_{MnL}) prediction[7] and experimental, R^2 for the quadratic trend line is 0.022.

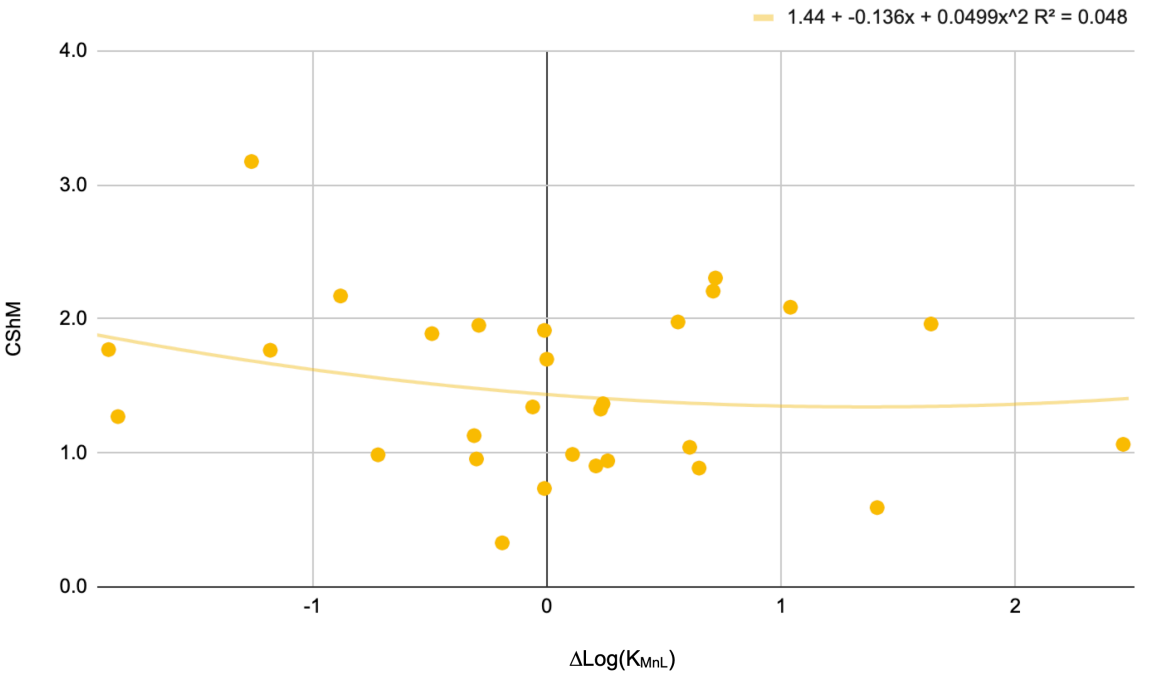
The experimental log(KMnL) values were plotted against lowest CShM in Figure 3.22 (top). The data shows almost complete lack of correlation between the two variables, with the gradient of the trend line ($\log(K_{MnL})$ = -0.0864(CShM) + 13.2) essentially being flat. The residuals from the Uzal-Varela prediction model were also plotted against CShM, Figure 3.22 (bottom). This would reveal if an extra term dependent on the shape of the complex could bridge the gap between the Uzal-Varela prediction

and experimental. This also showed no correlation between the two variables – if the relationship were true the expected graph would be positive quadratic with a minimum at $\Delta\log(K_{MnL})$ 0, with increasing lowest CShM as $|\Delta\log(K_{MnL})|$ increases. Instead, there does not appear to be any relationship, as the R^2 for the second order polynomial line is 0.022, and the x^2 coefficient is negative.






0.937*x + 11.9 R² = 0.052
20
15
10
5
0
0.5
1.0
1.5
2.0
2.5
3.0
Log(K_MnL)
CShM


1.44 + -0.136x + 0.0499x^2 R² = 0.048
4.0
3.0
2.0
1.0
0.0
-1
0
1
2
CShM
ΔLog(K_MnL)

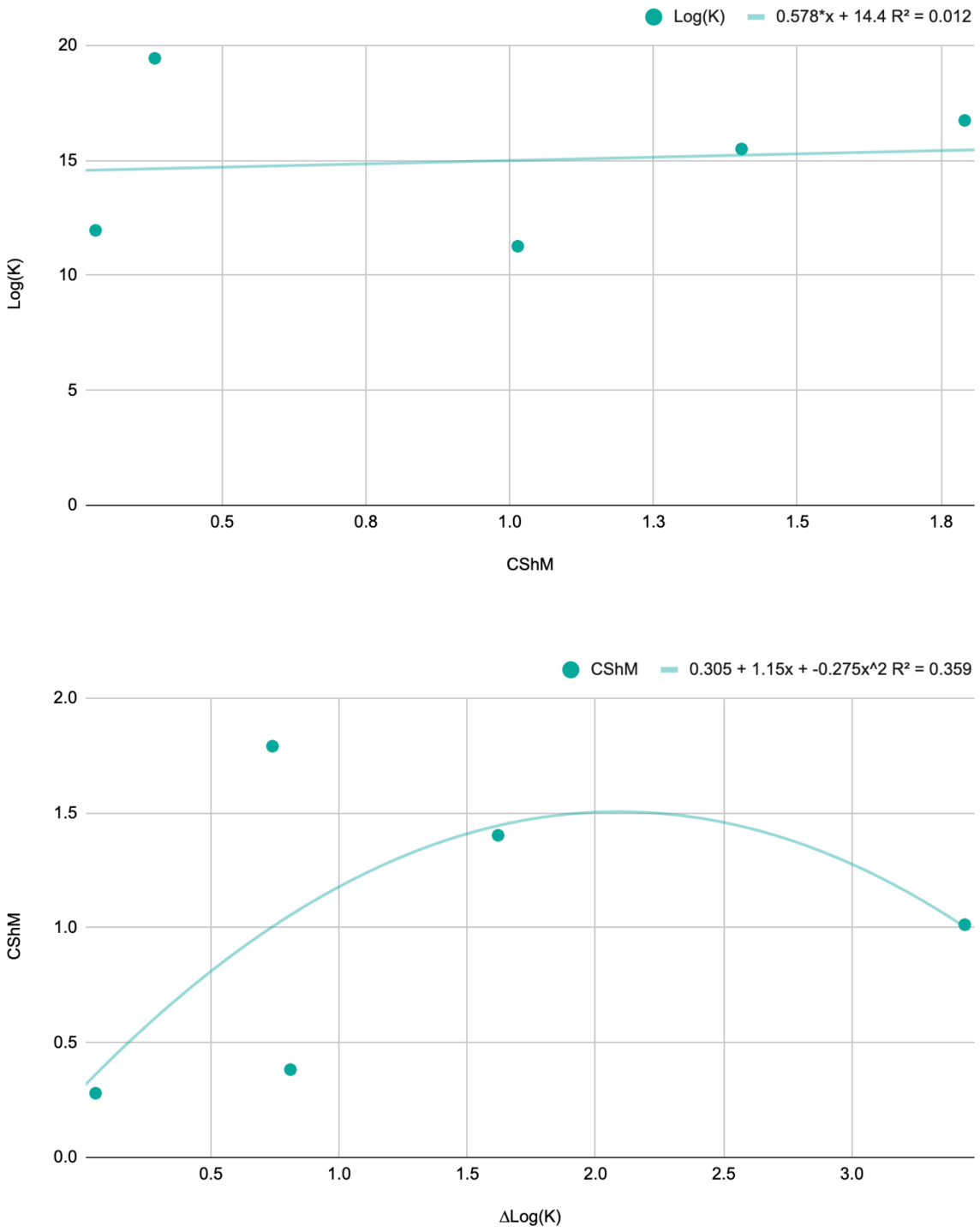


Figure 3.23: Plots of log(K_{MnL}) at varying CShM, and CShM at Δlog(K_{MnL}) values for 6-coordinate (pink), 7-coordinate (yellow), and 8-coordinate (green).

The dataset was then segmented by coordination number in case the graph in Figure 3.22 contains clustered data that is not immediately obvious. In a similar fashion to the unsegmented data, CShM is not correlated to log(K_{MnL}) (R^2 below 0.1 for all three) when cut by coordination number, Figure 3.23. The low R^2 for 7-coordinate (which has 29 complexes) suggests that even with more data the R^2 values

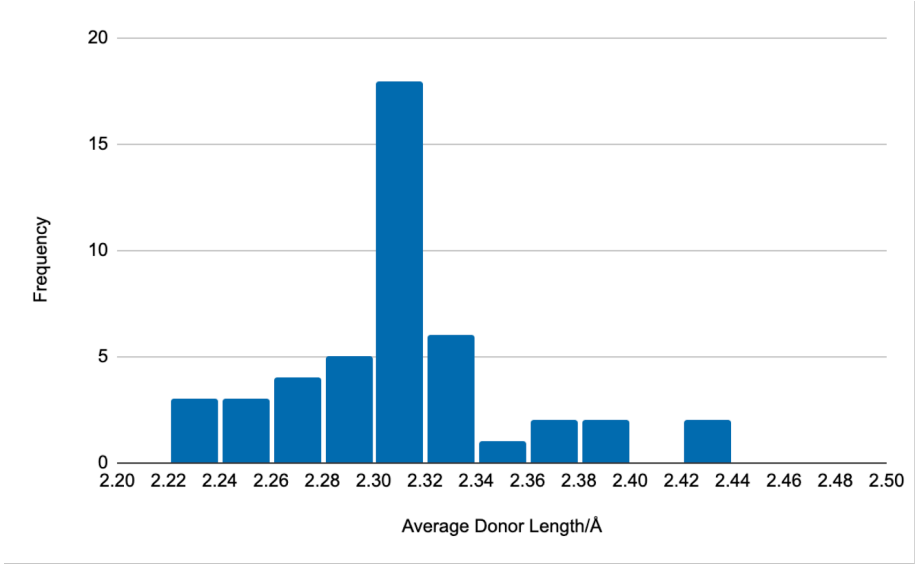
for the 6- and 8-coordinate plots would not improve. Equally, there does not appear to be a trend for $\Delta\log(K_{MnL})$ either.

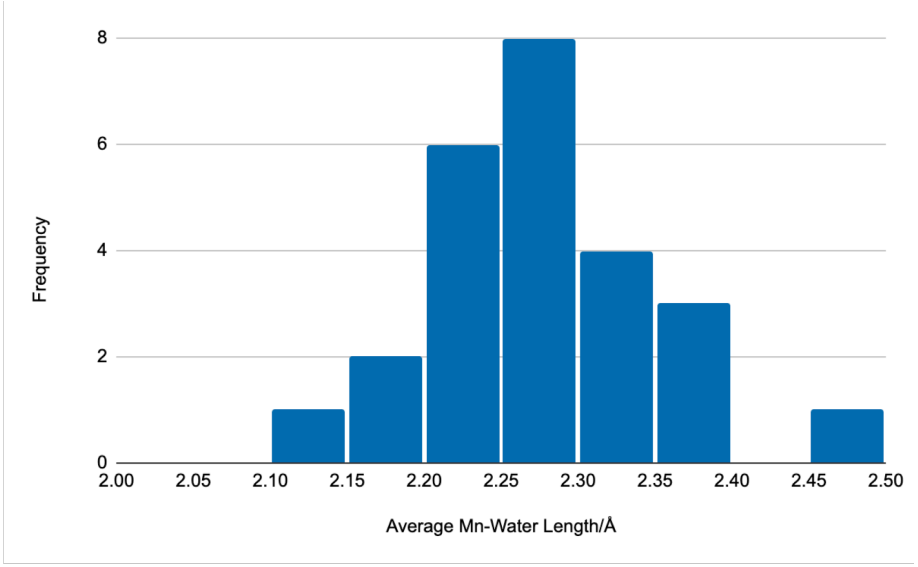
This indicates that the stability constant of a manganese complex is unlikely to be affected by its geometry (or alternatively how well its geometry can be described by the regular polyhedra that we assign to complexes to describe them). The outcome of the Uzal-Varela analysis, in being able to correlate a donor to an increase in stability constant, suggests that the strength of the bond (which is a function of MO overlap) is what is most important. It may be that only at very high CShM values would shape be an indicator of stability, values that are realistically unlikely to occur as ligands are fluidly arranged around a metal, always tending towards a more stable configuration. The thought experiment for this would be to imagine a highly strained ligand with four carboxylates and two tertiary amines where each donor coordinates on the same side of the manganese ion – such a complex would be extremely unstable, yet the Uzal-Varela model would predict an identical $\log(K_{MnL})$ as CDTA for it.

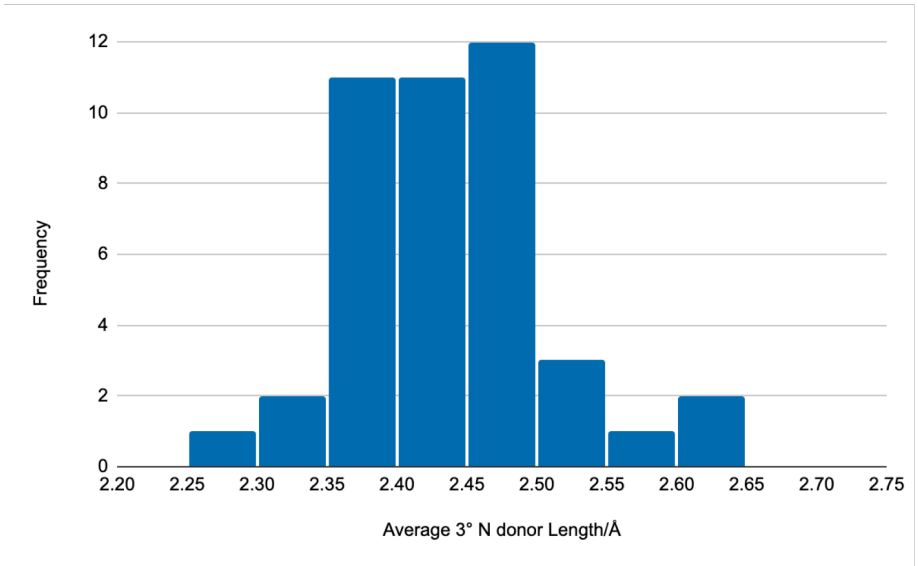
In Figure 3.21 (top right), there were two complexes which had a lowest CShM between 3 and 5 (4.83 and 4.11). These complexes, **L46** and **L11**, have a $\log(K_{MnL})$ of 10.67 and 15.10 – a much lower $\log(K_{MnL})$ would be needed to begin to validate the hypothesis. Furthermore, **L50**, described by a pentagonal bipyramidal geometry (0.33) has one of the lowest $\log(K_{MnL})$ values in the dataset (7.18). This hints at another limitation of the hypothesis – suppose three ethylenediamine ligands were to perfectly provide an octahedral complex with Mn^{2+}, the hypothesis would expect that it would have a high $\log(K_{MnL})$, even though it does not. The idea that an extremely simple ligand could invalidate the theory led to questioning whether instead CShM could improve on the already very good predictions by the Uzal-Varela model, which ultimately also showed it was unable to do.

Ultimately, it may be that the determination of the stability constant is too dynamic to be modelled by something as simple and static as a DFT geometry optimisations shape. CShM, whilst an easy metric to quote to communicate the shape of an optimised geometry, has been shown to be limited in its application in predicting something meaningful to the field such as the MnBCA stability constant.

3.2.3.3 Bond length distribution and log(K_{MnL})







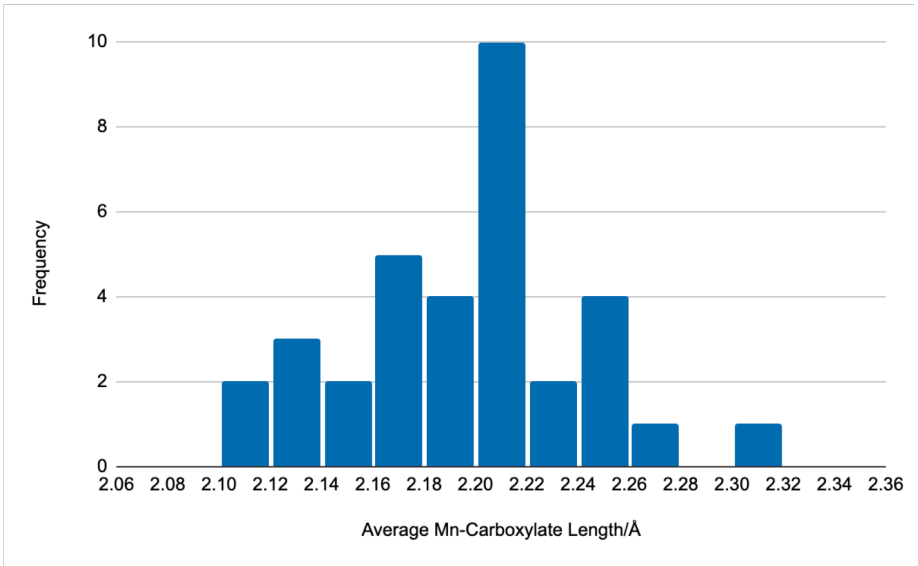


Figure 3.24: Figures of data exploration around the MnBCA dataset with respect to bond lengths. Top left: histogram of the average donor bond length for each MnBCA; top right: histogram of the average Mn-water bond length for each $q \geq 1$ *MnBCA; bottom left: histogram of the average tertiary amine donor length for each MnBCA; bottom right: histogram of the average Mn-carboxylate bond length for each MnBCA.*

The dataset from the CShM analysis generates plenty of data that may have a relationship with MnBCA KPIs. From these DFT optimised geometries, the bond length distribution (BLD) can be analysed. As with the CShM analysis, only the coordination sphere is being evaluated. Figure 3.24 (top left) shows the distribution of average Mn-donor lengths for the MnBCAs. It would appear to suggest that most bonds will be between 2.30-2.32 Å in length, and equally, that there might not appear to be enough variation in bond lengths to correlate to a KPI. However, as this is an 'average' donor length, the true nature of donor bond lengths is revealed when the individual donors are analysed, Figure 3.24 (top right, bottom left and right). From this, we can see that Mn-water bonds are normally distributed about a mean of 2.27 Å, σ of 0.07 Å; the average of 2.30-2.32 Å seen in Figure 3.24 (top left) is made up from averaging tertiary amines with a mean of 2.44 Å (σ of 0.07 Å), and carboxylates with a mean of 2.20 Å (σ of 0.05 Å) – the majority of MnBCAs are based on designs with tertiary amines and carboxylates. The idea of a cavity size implies a uniform space within which the metal can sit inside of. The difference between tertiary amine and carboxylate distributions suggests that a ligand does not maintain all of its donors at the same distance – some of the donors will sink further into the metal than others.

For this analysis, the expected relationship would be a negative quadratic. If the donor length is too small or too large, the stability would reduce. Consequently, there would need to be a zone in the middle that is optimal for the most stable bond. In Figure 3.25, the average donor length was plotted against $\log(K_{MnL})$, and a second order polynomial was fitted to the data. This fit provided an R^2 of 0.154 for a negative quadratic. Looking over the plot, it appears to have a very weak correlation with increasing donor length leading to more stable complexes. The predictive power is hampered by the high variance in the dataset; R^2 can be thought of as a function of spread – in Figure 3.25 the low R^2 is a symptom of the large variance in the MnBCA dataset. Using the fitted equation would suggest a theoretical maximum stability constant at 2.42 Å of 14.88 (clearly incorrect with values approaching 20 within the dataset). It appears that in this instance only a qualitative assessment can be made – the larger the average the donor length, the *more likely* that a higher stability constant will be achieved. This is not something that is altogether unexpected, as the bispidine complex, **L172**, has a large rigid cavity that provides a very high stability constant of 19.47 and shows kinetic inertness to transmetalation due to having a cavity that is too large for Zn.[59] As explained, however, caution should be applied – Figure 3.25 shows complexes with average donor length greater than 2.35 Å yet $\log(K_{MnL})$s below 13.

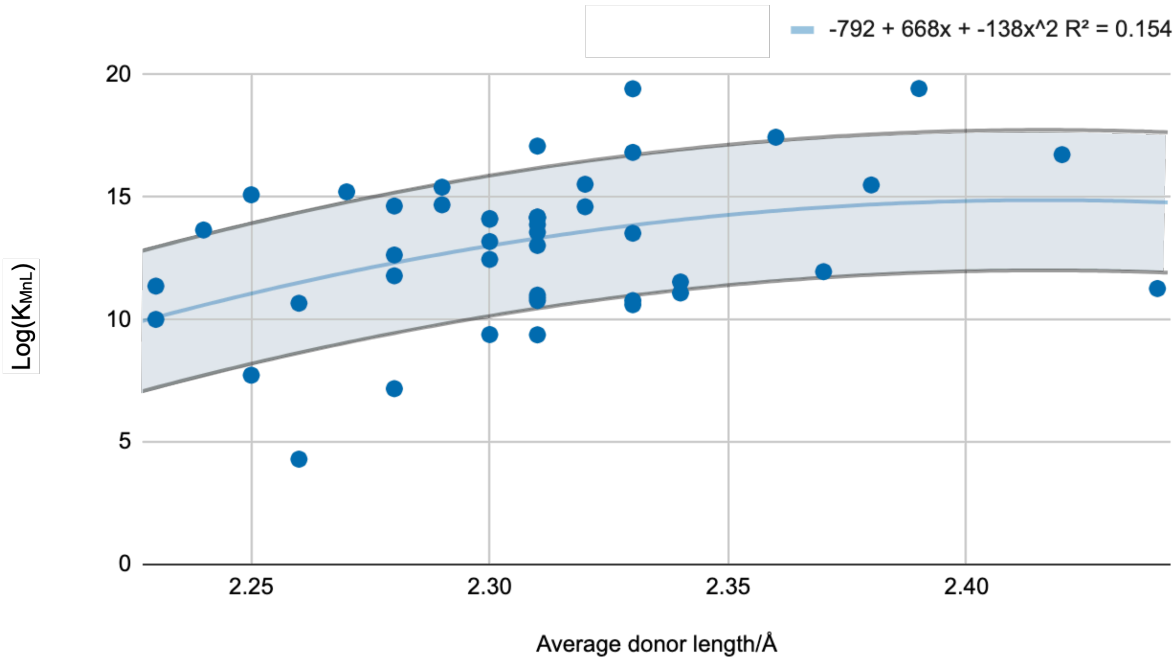


Figure 3.25: Plot of log(K_{MnL}) vs average donor length, fitted with a second order polynomial trend line and ±1σ prediction interval.

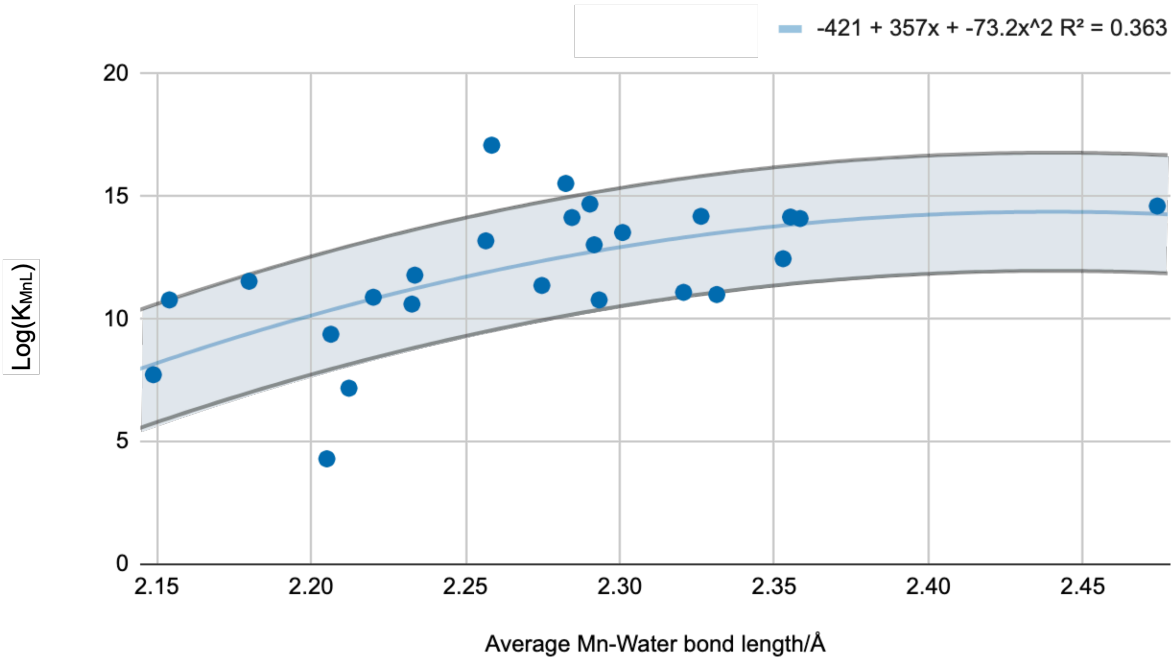


Figure 3.26: Plot of log(K_{MnL}) vs average Mn-water bond length, fitted with a second order polynomial trend line and ±1σ prediction interval.

This treatment was also applied to the average Mn-water bond length, Figure 3.26, which showed a tighter fit to a negative quadratic equation ($\log(K_{MnL}) = -73.2x^2 + 357x - 421$), R^2 of 0.363, and a σ of ± 2.41 log units. Despite this improvement, the polynomial trend line is still overfitting the data – **L172**, for instance, is a q = 1 complex with $\log(K_{MnL})$ of 19.47, which would be outside of the prediction

possible with this model. Instead, it appears that it works reasonably well in being able to predict its training dataset within 1 standard deviation but again only to a qualitative degree (± 2.41 log units is still a wide window of prediction).

Of the individual donors, only tertiary amines and carboxylates had enough data points to allow for an analysis, Figure 3.27. Both were fitted with a quadratic trend line (expecting to find an optimal bond length for stability), and both gave low R^2 (below 0.1) indicating that their models would not be useful in predicting the stability constant.

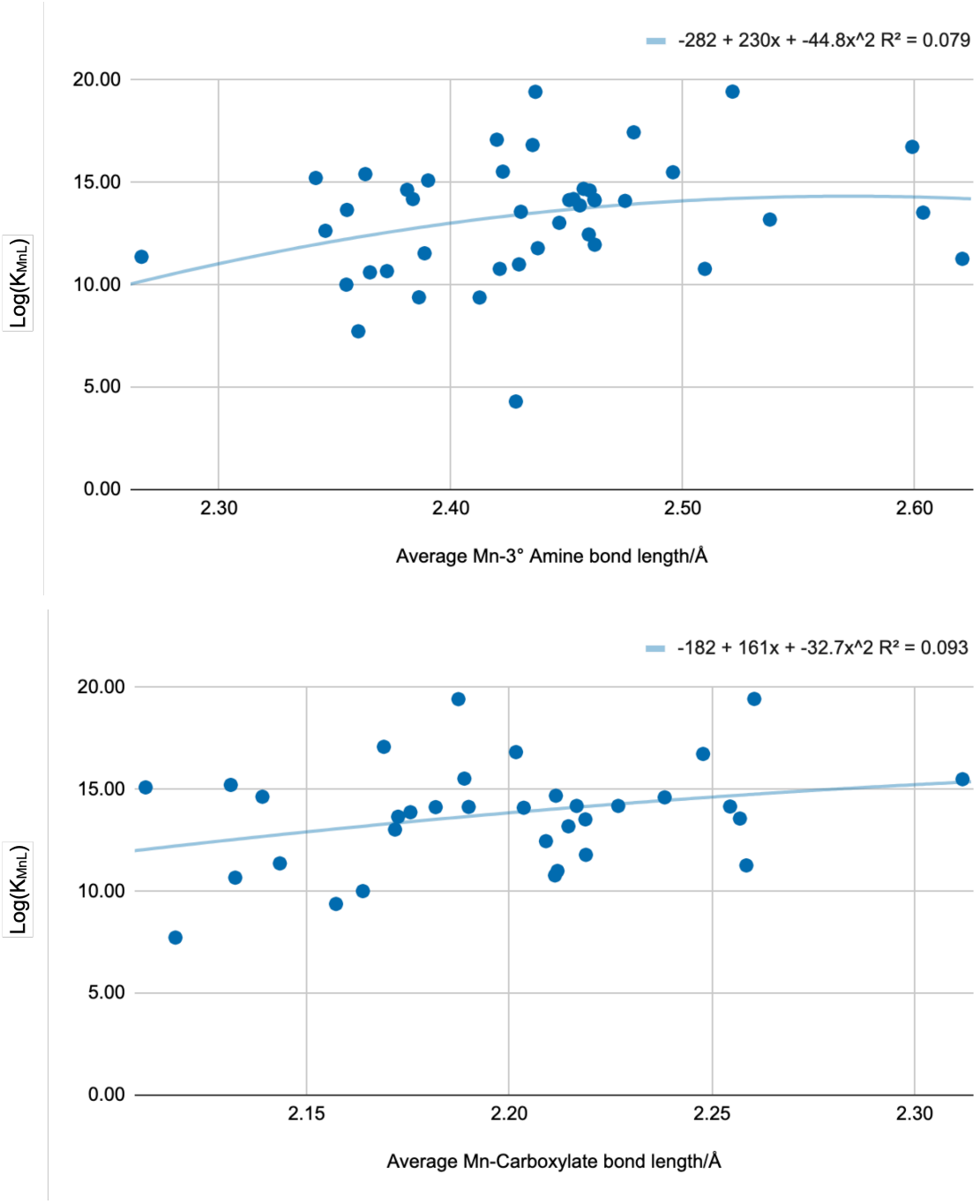


Figure 3.27: Top: plot of log(K_{MnL}) vs average Mn-tertiary amine bond length, fitted with a second order polynomial trend line; bottom: plot of log(K_{MnL}) vs average Mn-tertiary amine bond length, fitted with a second order polynomial trend line.

3.2.3.4 Water parameters

3.2.3.4.1 Bond length distribution

As in Section 3.2.3.3 which looked at the potential effect of optimised geometries on stability constant, the same analysis was carried out on metrics used to assess the Mn-water bond for MnBCAs. None of the bonds correlated to the water exchange rate (R^2 were all below 0.1 for a logarithmic trend line), Figure 3.28. The data did not appear to show any pattern, indicating that unlike for Gd-O_{water} bonds (and the accompanying analyses on the ρ_{BCP} and ELF and the bond critical point),[21] water exchange cannot be predicted *via* optimised bond lengths. This treatment was also applied to r_1 data for complexes with complete data, Figure 3.29, and a similarly uncorrelated result was obtained with all of the R^2s less than 0.1.

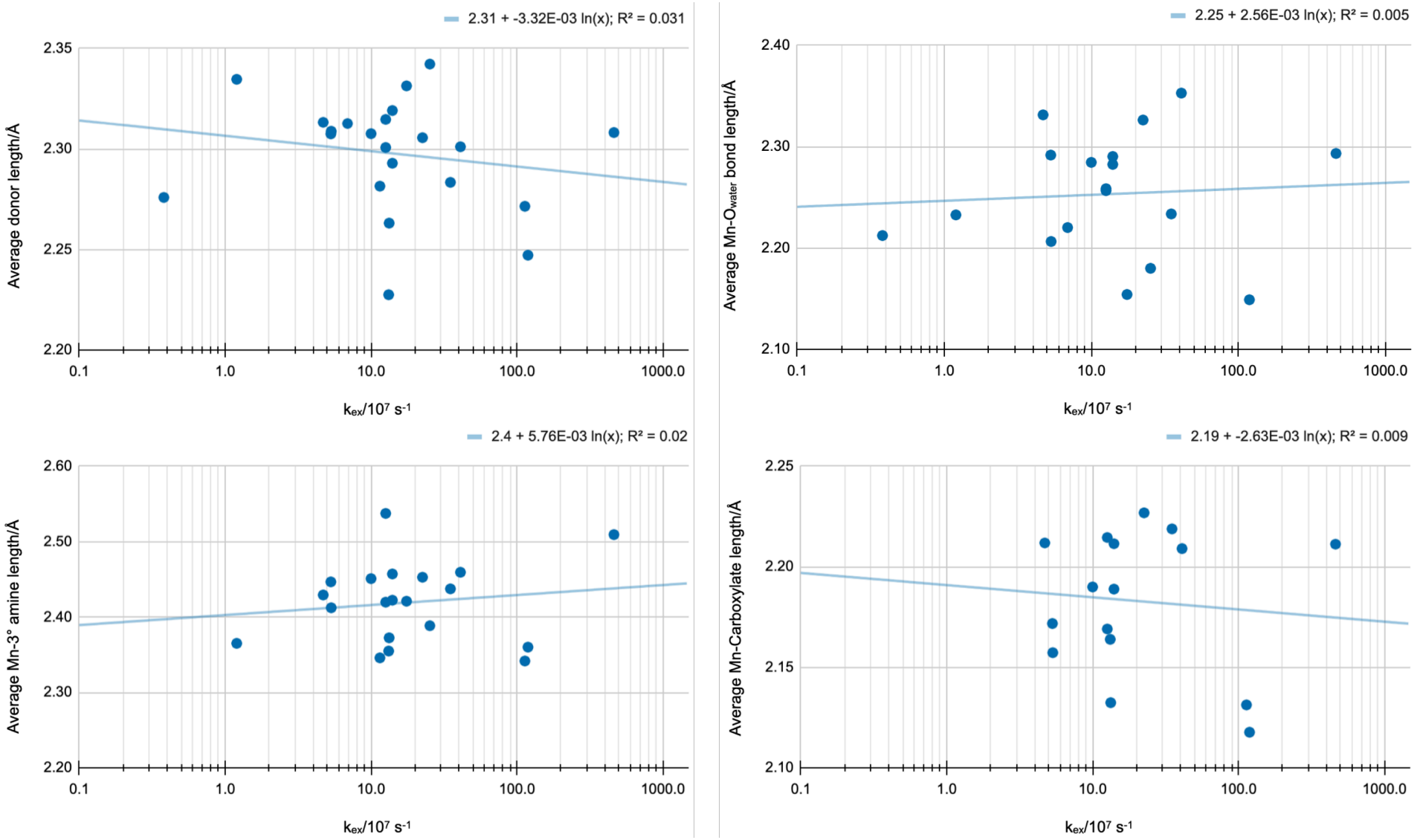


Figure 3.28: Plots of water exchange rate vs bond lengths. Top left: against average donor length; top right: against average Mn-water bond length; bottom left: against average Mn-tertiary amine bond length; bottom right: against average Mn-carboxylate bond length. All graphs fitted with a logarithmic trend line; k_{ex} is on a logarithmic scale.

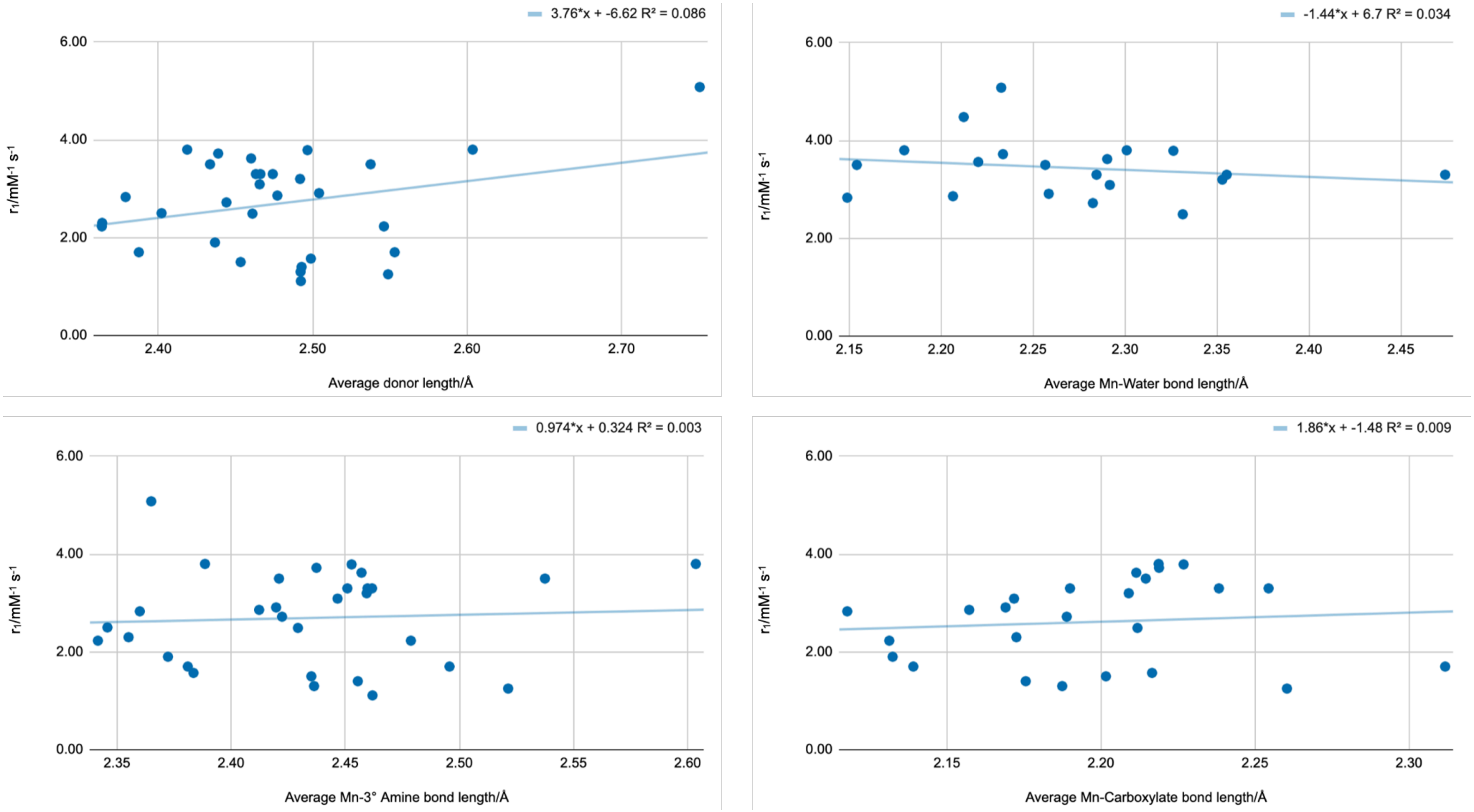


Figure 3.29: Plots of r_1 *vs bond lengths. Top left: against average donor length; top right: against average Mn-water bond length; bottom left: against average Mn-tertiary amine bond length; bottom right: against average Mn-carboxylate bond length. All graphs fitted with a linear trend line.*

3.2.3.4.2 Charge

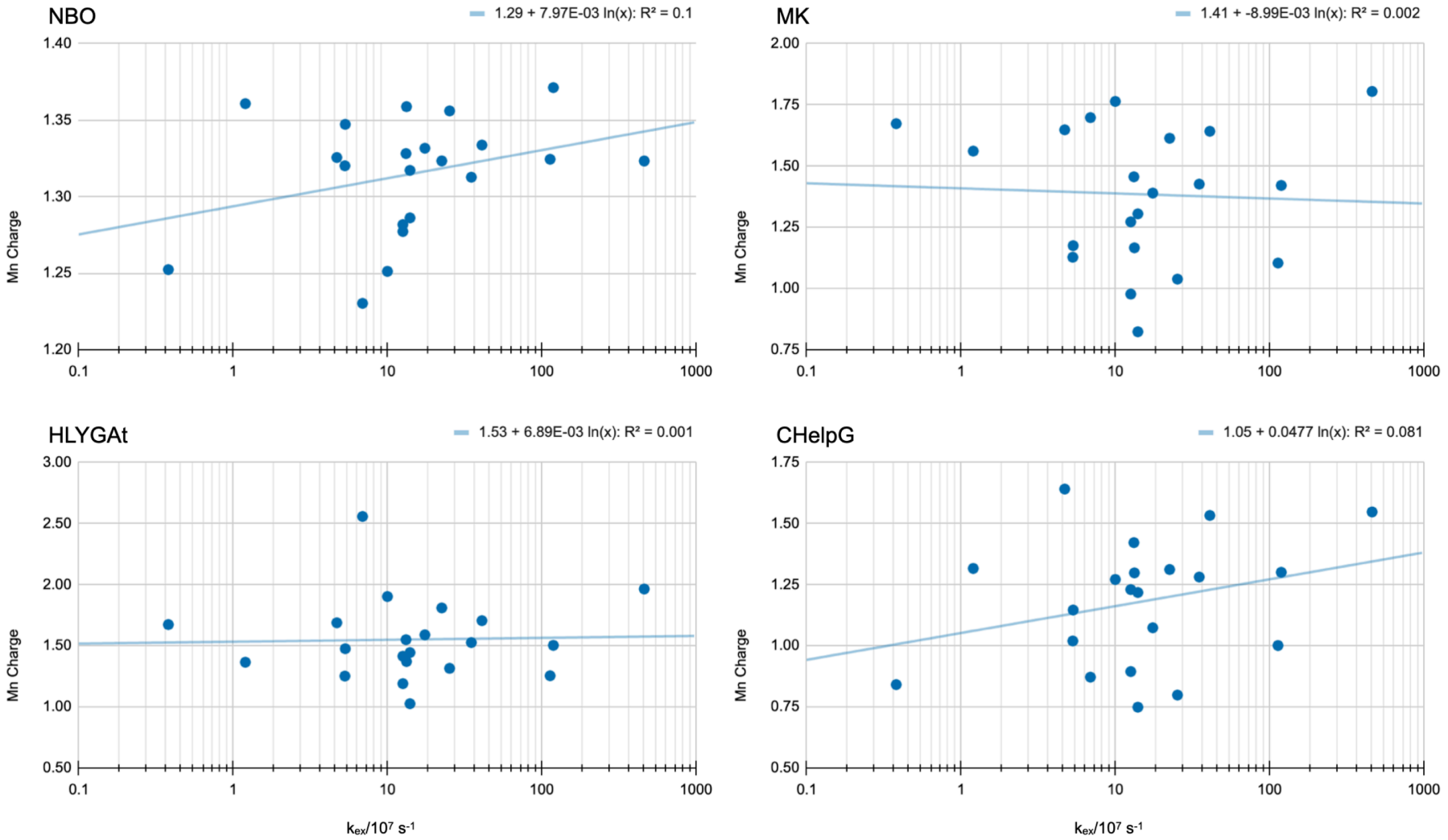
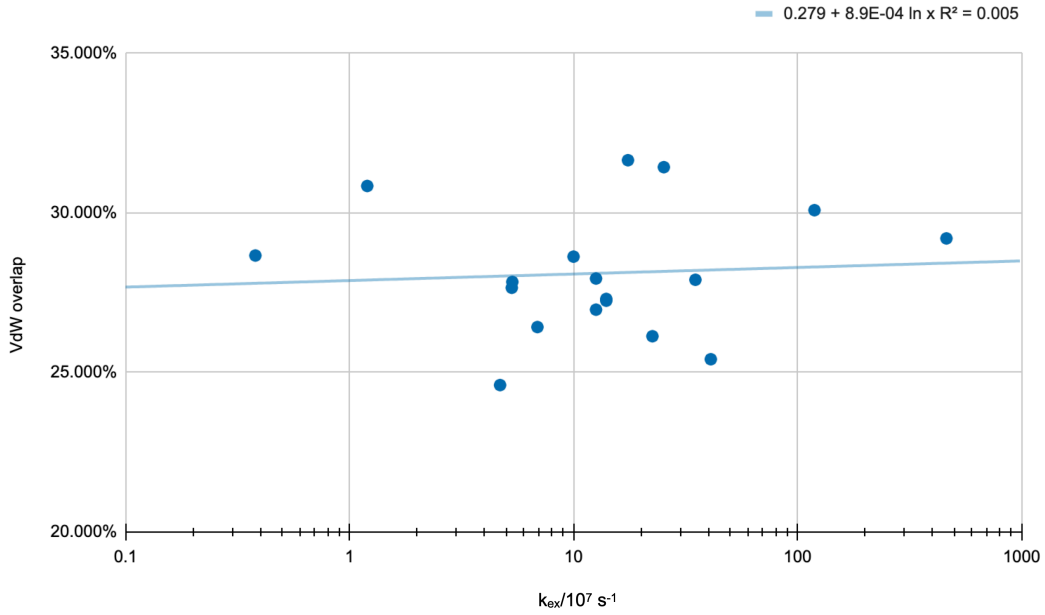


Figure 3.30: Plots of water exchange rate vs Gaussian calculated Mn charges. Top left: NBO; top right: MK; bottom left: HLYGAt; bottom right: CHelpG. All graphs fitted with a logarithmic trend line; water exchange rate is on a logarithmic scale.

Gaussian is also able to calculate non-integer 'charges' derived from the electrostatic potential of the optimised molecule, these are often used to describe or rationalise bonding behaviour (qualitatively and quantitatively).[60–62] These were calculated for the MnBCA dataset with four different charge fitting methods, three based on electrostatic potential (MK, HLYGAt, and CHelpG)[63–65] and one on natural bond orbital analysis (NBO).[66] The calculated charges for each MnBCA were plotted against water exchange rate for each charge fitting method in Figure 3.30. All four were uncorrelated to water exchange rate (logarithmic trend line, R^2 below 0.1).

3.2.3.4.3 Steric overlap



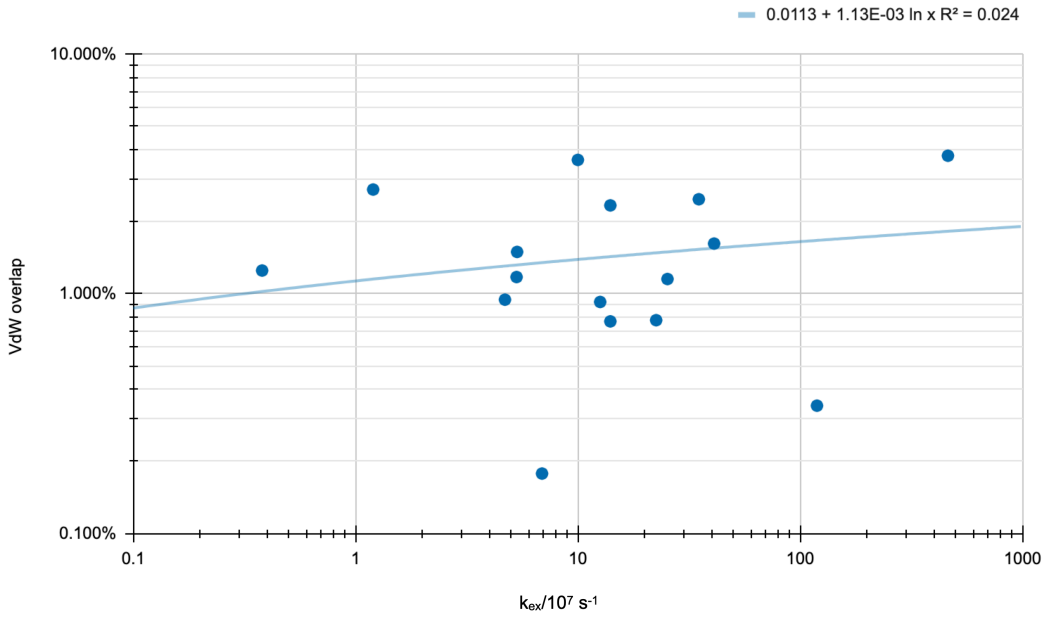


Figure 3.31: Plots of water exchange and VdW overlap, top: VdW overlap measured including the Mn-water interaction; bottom: VdW overlap with the Mn-water interaction removed. In both plots water exchange rate is plotted on a logarithmic scale, and in the bottom plot, the VdW overlap is plotted on a logarithmic scale.

The ‘steric pressure' that the bound water experiences was captured by using a proxy measurement of 'Van der Waals overlap’. This was calculated for all $q \geq 1$ complexes and plotted against their water exchange rate, Figure 3.31. Within the data, it was evident that the majority (95% on average) of the VdW overlap was due to the Mn-water overlap, with a minority due to the presence of the ligand. In Figure 3.31 (top) the full overlap of the VdW spheres from the complex on the water is shown, and there does not appear to be a relationship between VdW overlap and water exchange rate – the plotted logarithmic trend line has an R^2 of 0.005). Alternatively, to purely capture the effect that the ligand is having on the water, the steric overlap purely from the Mn^{2+} ion was removed – Figure 3.31 (bottom). The R^2 for a logarithmic trend line for this relationship was also extremely low, 0.024, indicating no correlation.

These two results and those for the charges go against the expected relationship discussed in literature, with the water exchange rate being unaffected by high or low charges, and whether or not the water is being 'pressured' or not. In the examples given in Section 3.1.2, these conclusions are often being drawn after comparing 2-3 data points and predicting a full trend, essentially overfitting a trend on a very small sample size.

3.2.4 Leveraging the large dataset from Section 3.2.3

In calculating the optimised geometries in Section 3.2.3, 46 complexes were optimised at M06-2X/TZVP/IEFPCM level. This provided a large amount of already computed data which could be mined for log(K_{MnL}) prediction *via* ΔG_{aq}. The M06-2X/TZVP/IEFPCM optimisations of the ligands from Section 3.2.3 were carried out, with difficult to optimise ligand/complex pairs removed, and replaced with L75, L121, L147, and L167 (used in Section 3.2.2.5). Unlike in Section 3.2.2.5, the data needed to be treated to accommodate complexes with hydration numbers of 0-2 – this meant that ΔG_{aq} values were calculated using Equation 3.13:

$$\Delta G_{aq} = -G\left(L^{m-}{}_{(aq)}\right) - G\left([Mn(H_2O)_6]^{2+}{}_{(aq)}\right) + G\left([MnL(H_2O)_n]^{2-m}{}_{(aq)}\right) + (6-n)\,G\left(H_2O_{(aq)}\right) + (5-n)(1.89\ kcal\ mol^{-1}) + (6-n)(RT \cdot \ln(55.5))$$

Equation 3.13: Equation for the calculation of ΔG_{aq}, where n = 0, 3, or 6, for q = 0, 1, or 2).

The calculated log(K_{MnL}) values were plotted against experimental log(K_{MnL}) values, as can be seen in Figure 3.32. The data was fitted with a linear regression line with 95% prediction intervals, which produced a slope of 1.70 ± 0.51, and intercept of 191 ± 7, Equation 3.14.The sign and magnitude of the

gradient, (and visually the shape of the data) indicates the desired trend where the calculated log(K_{MnL}) scales proportionally with experimental log(K_{MnL}). This follows the literature precedent with DFO complexes and uranyl complexes.[8,13] The difference between the M06-2X/TZVP/IEFPCM MnBCA dataset and those two are the much lower R^2 values – Holland's DFO study R^2: 0.97, Bryantsev's UO_2 study R^2: 0.95 and 0.967.

$$Log(K_{MnL})_{predicted} = \frac{Log(K_{MnL})_{calculated} - 191 \pm 7}{1.70 \pm 0.51}$$

Equation 3.14: Equation for the prediction of log(K_{MnL}) as provided by the linear regression between calculated and experimental log(K_{MnL}).

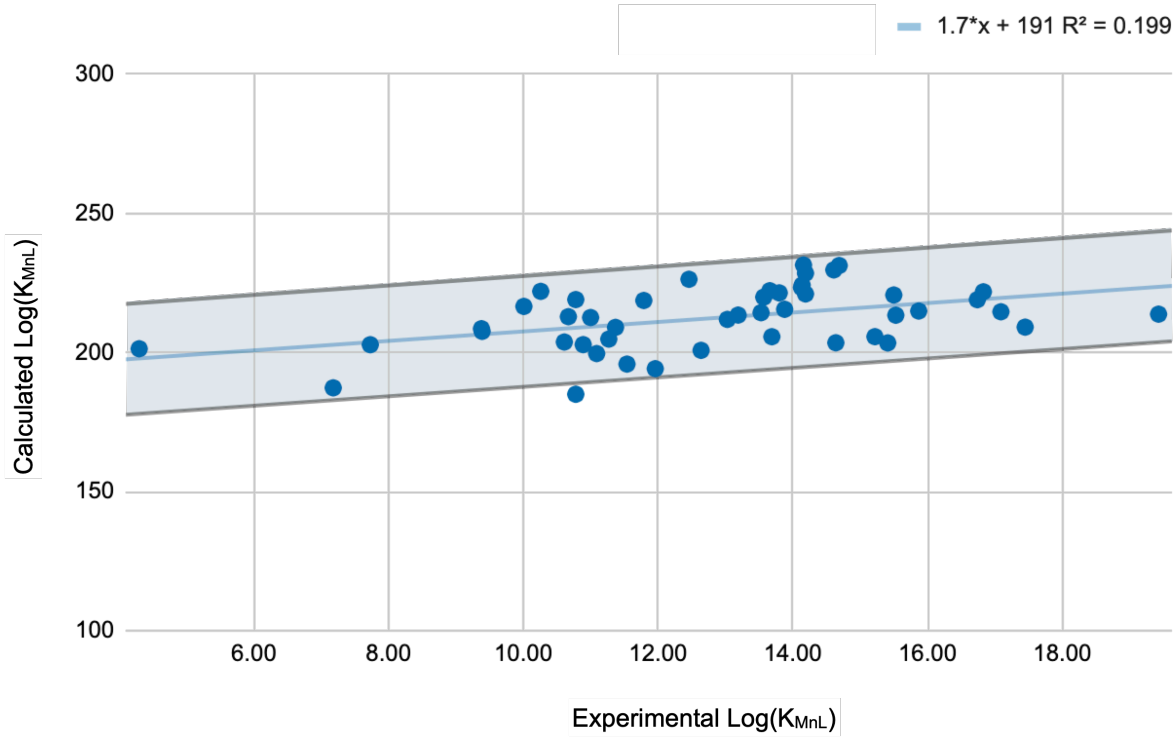
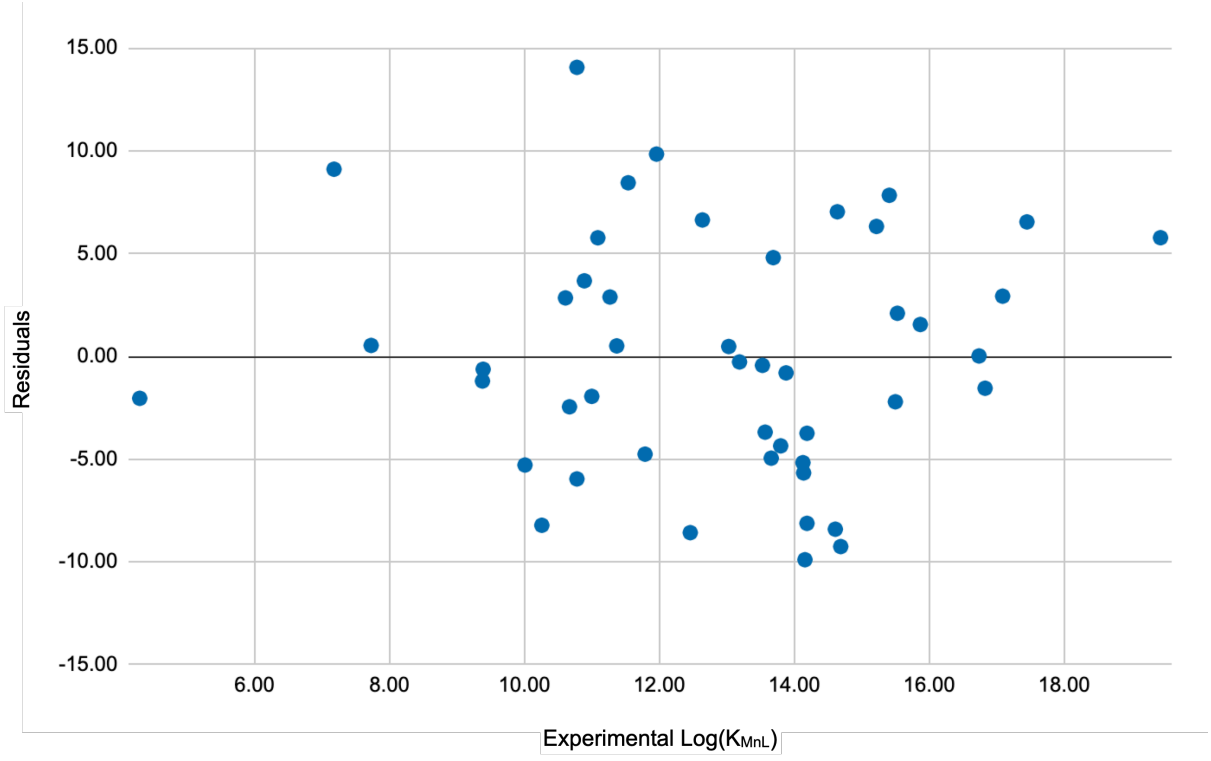


Figure 3.32: Calculated vs. experimental log(K_{MnL}) for MnBCAs calculated at the M06-2X/TZVP/IEFPCM level including 95% prediction interval.

When restricted to the training dataset seen earlier on in this chapter, the 'with water × M06-2X × TZVP × IEFPCM' model was only able to produce an R^2 of 0.142 (Table 3.3, Entry 2). Interestingly, with an increased amount of data available, the model is able to improve its R^2 value to 0.199. This improvement in R^2 with greater sample size speaks to the sample size constraints that were necessary to feasibly carry out the study in Section 3.2.2.6. Nonetheless, in an ideal scenario the dataset would be providing R^2 values approaching and exceeding 0.7 for good predictive power; the small R^2 value indicates that the predictive power of the model would be low if it were to be tested with a complex not in its training dataset.

Applying Equation 3.14 on the training dataset, provided predicted log(K_{MnL}) values with residuals ($\hat{y}$ - y_{expt}) with σ of 5.76 log units and $\bar{y}$ of 0.00, Figure 3.33. The sample mean being distributed about 0.00 indicates that there is no inherent error in the approach, but with a large σ of 5.76 log units, the MnBCA dataset itself has a lot of variance, which makes prediction *via* linear regression difficult. This is seen when plotting predicted log(K_{MnL}) against experimental log(K_{MnL}), Figure 3.33. The 95% prediction interval of the estimate is ±11.73 log units, essentially giving very low confidence in the prediction, as would be expected from an R^2 of 0.199.

This perhaps helps to explain the incorrect isomer prediction for CBuDTA (Section 3.2.2.8.2) – a small R^2 is a result of inherently having large variance, and even with much larger samples the R^2 does not drastically improve. As for CBuDTA, DFT is able to produce a model with a positive gradient as with the Holland and Bryantsev papers, but in this case, it appears that DFT is systematically mistreating the $[Mn(1,3\text{-CBuDTA})(H_2O)]^{2-}$ (or alternatively over-stabilising $[Mn(1,2\text{-CBuDTA})(H_2O)]^{2-}$) compared to the rest of the ligands.



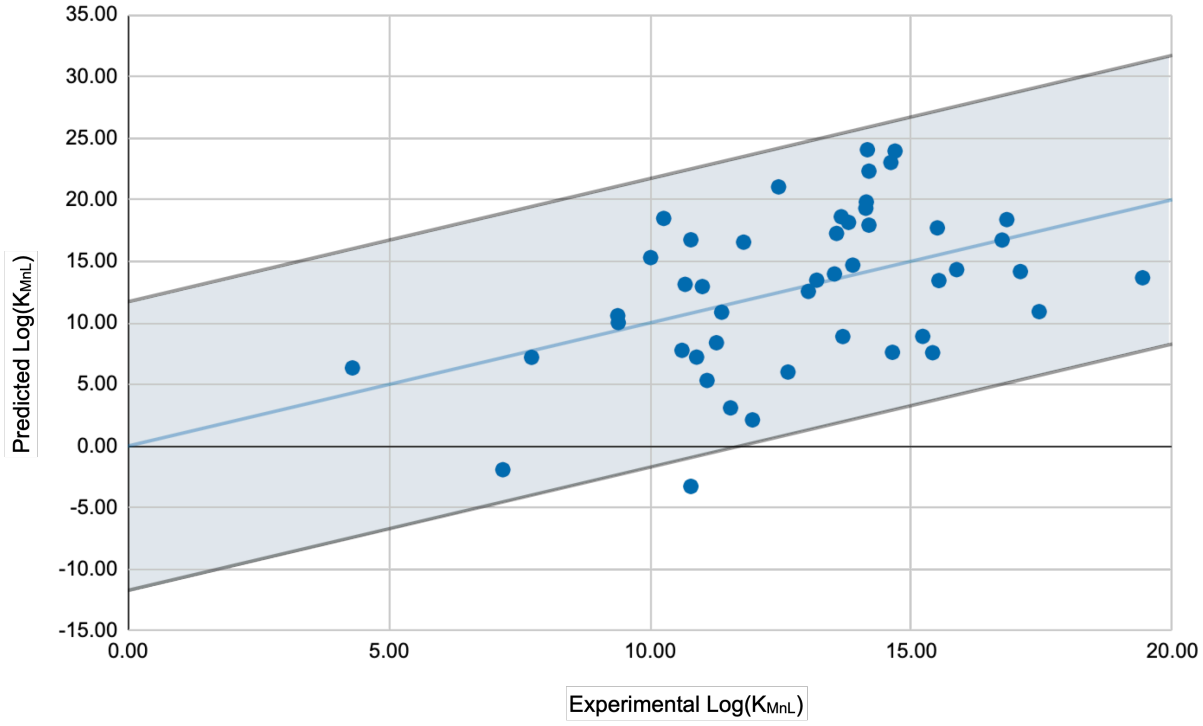


Figure 3.33: Top: plot of the residuals vs experimental log(K_{MnL}), σ = 5.76 log units and $\bar{y}$ = 0.00 log units; bottom: predicted vs experimental log(K_{MnL}) values from the 'with water × M06-2X/TZVP × IEFPCM' model.

3.2.4.1 Clustering

With a larger dataset, clustering should become more evident than in the previous experiment which had a much smaller training dataset. The data from Figure 3.32 was segmented by complex charge in Figure 3.34, showing a clear clustering between -2, 0, and +2 charged complexes, with -1 complexes laying somewhere between -2 and 0 as would be expected if negative charge had a proportional effect on log(K). These plots were then exploded in Figure 3.35, with a least squares linear regression applied (-3 was removed as there is only one datum, and these are not attractive MnBCAs). The resultant graphs show an improvement in R^2 for -2, 0, and +2 charged complexes – the combined graph in Figure 3.34 gave an R^2 of 0.199, vs -2: 0.228, 0: 0.319, +2: 0.313. These values are higher than the best three models in Section 3.2.2.6 which used more powerful functionals and more accurate methods in dealing with solvation. This mutually exclusive trend is mirrored in the average calculated log(K_{MnL}) values: -2: 222.2, -1: 215.3, 0: 207.6, 2: 195.0 log units. Segmenting by charge negatively affected the -1 complexes to give a very flat (but still positive) relationship – giving an R^2 of 0.006. The expected relationship that was seen in the other charges could possibly be teased out with more data. This would reduce the reliance on the two highest experimental log(K_{MnL}) values (DFT appears to be handling them anomalously) – removing them gives a gradient of 1.67 and R^2 of 0.547 (and R^2 of 0.247 for the whole dataset). From these results, it would be expected that +1 charged species would lie between the regression lines of 0 and +2 charged complexes.

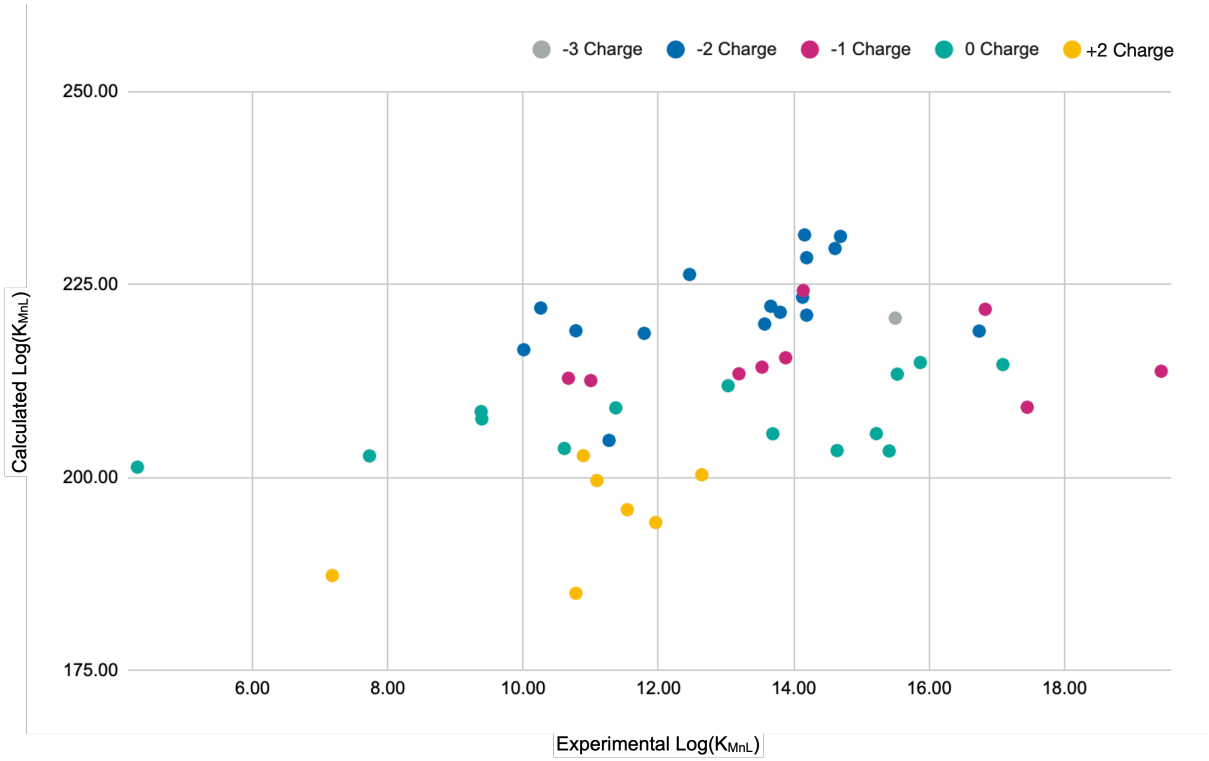


Figure 3.34: Calculated vs. experimental log(K_{MnL}) for the 'with water × M06-2X/TZVP × IEFPCM' model segmented by complex charge.

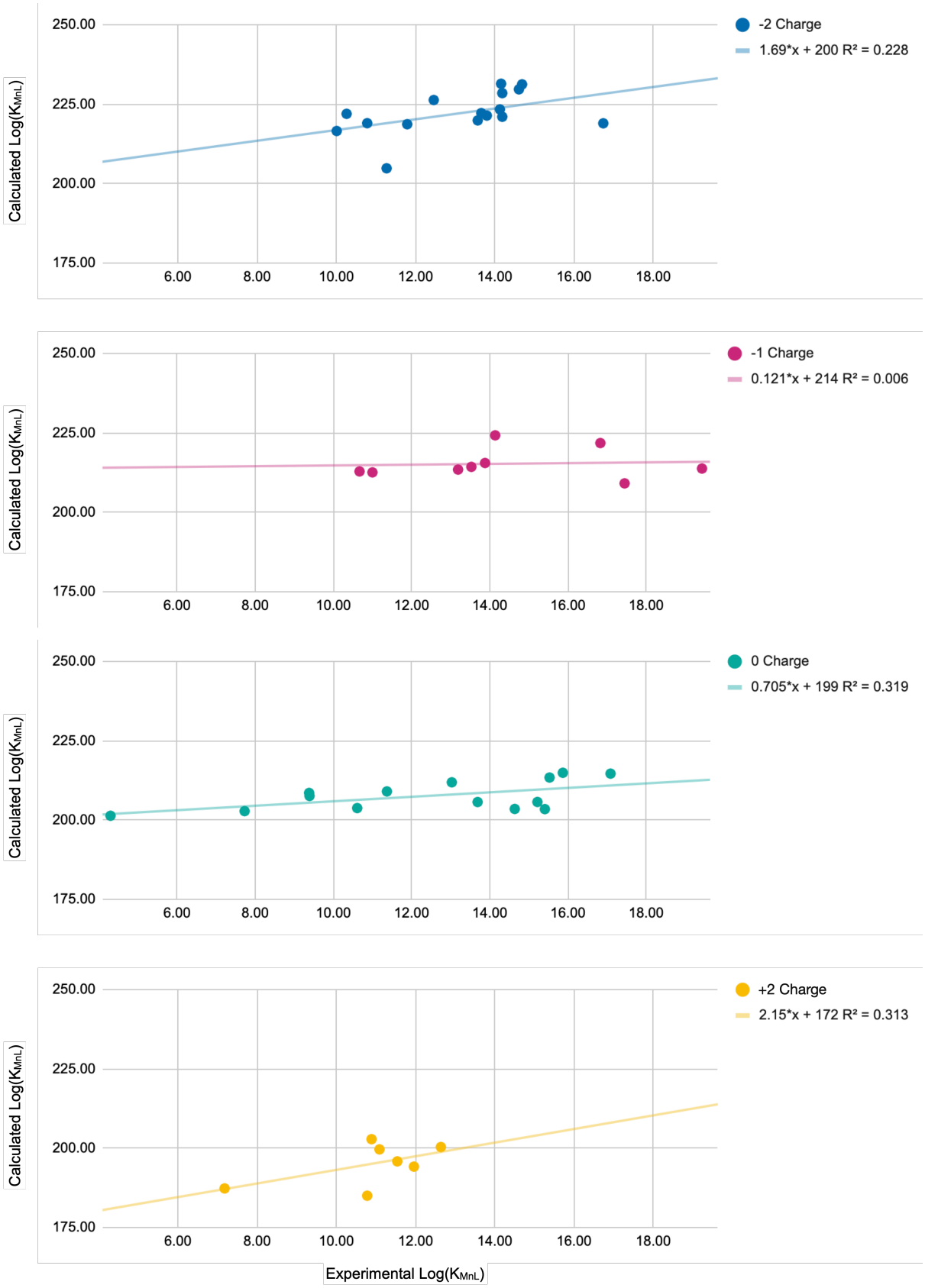


Figure 3.35: Exploded plots of Figure 3.34, separated by complex charge.

The data was also cut by hydration number, mirroring the experiment in Section 3.2.2.6.1 and testing the effect of increasing the dataset for the $q = 1$ complexes, Figure 3.36. Complexes with $q = 0$ showed the smallest gradient, which in turn meant that the R^2 for this subset of such a high variance dataset was very low (0.06). For $q = 1$ complexes, as the dataset has expanded from 10 to 23 complexes, the R^2 value has increased from 0.142 in Section 3.2.2.6 to 0.237, which would have positioned it as one of the best models. The lower number of datapoints for $q = 2$ complexes meant that the relationship is likely to have a strong degree of error compared to the other two. Clustering by hydration number does not segment the data as cleanly as complex charge – possibly due to the high variation, so caution should be used in interpreting the significance of these regression lines for prediction of $\log(K_{MnL})$ by hydration number.

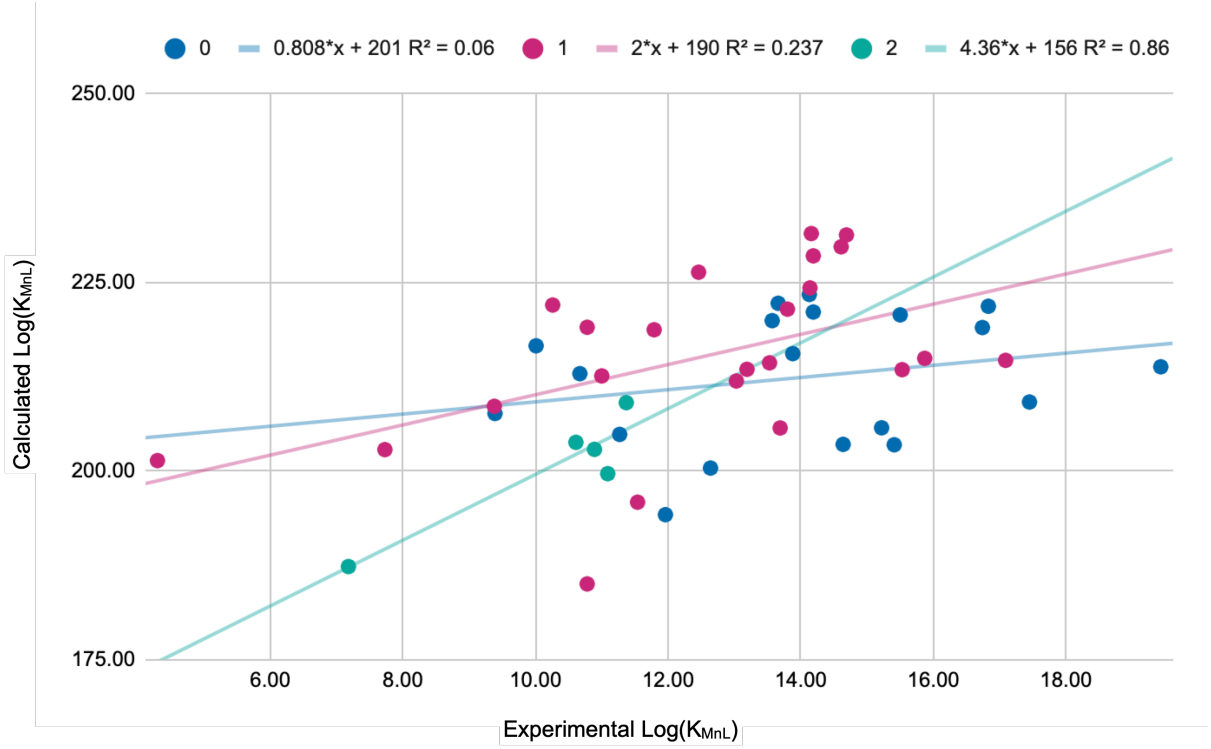


Figure 3.36: Calculated vs. experimental log(K_{MnL}) for the 'with water × M06-2X/TZVP × IEFPCM' model segmented by hydration number.

3.2.5 Machine learning prediction of MnBCA KPIs

In this section, the datasets from Section 3.2.2 and 3.2.3 were combined for 43 MnBCAs with largely complete data for properties such as molecular weight and q as well as the computationally derived properties such as thermal and Gibbs energies of the Mn^{2+} complex and ligand, CShM, bond lengths, charge, and VdW overlaps (**L6**, **L7**, **L11**, **L34**, **L35**, **L42**, **L47** were removed due to incomplete data). This data was then randomly (but deterministically) split with 80% in the training dataset, and 20% in the testing dataset, before the ML models were trained to predict the stability constant for the MnBCAs. The models were trained on three datasets, one with all variables that have been produced in this work;

one with only the DFT energy variables (thermal and Gibbs energies of ligand and complex, calculated $\log(K_{MnL})$, and the ΔG used in the $\log(K_{MnL})$ calculation); and one with the variables used in the bond length distribution and geometry study (coordination number; geometry; bond lengths (minimum, maximum, average) for donors, Mn-carboxylate, Mn-3° amine, Mn-water; VdW overlaps, CShM).

Table 3.8: R^2 results from the regression tasks split by ML model, and variables each model was given to perform the regression with.

Dataset	Linear Regression	Neural Network[a]	SVR
All variables	-0.284	-0.601	-0.136
DFT energies	0.215	0.253	0.371
Geometry measurements	-0.423	-0.515	-0.242

[a] The number of neurons in the hidden layers of the neural network was varied depending on the number of features in the input layer: the 'all variables' set had 43 and 30 neurons in its hidden layers; the 'DFT energies' set had 5 and 4; the 'geometry measurements' set had 20 and 14.

The result of the testing is shown in Table 3.8. With all of the variables provided, none of the ML methods were able to produce an R^2 value above 0 – indicating that using the average $\log(K_{MnL})$ value would have provided a better prediction (Equation 3.12). This suggests that either there is no relationship between the collected data from the DFT calculations, or that the collection of dependent variables in the data is too large to work with, with too many 'red herring' variables being looked at by the models. This rings true when slimming the number of variables given to the models – the DFT energies dataset was able to produce positive R^2 values across all three models, and in fact, all three methods were able to produce a better prediction of $\log(K_{MnL})$ than when the calculated $\log(K_{MnL})$ was produced in Section 3.2.4, which gave an R^2 of 0.199. The 0.371 produced by the SVR model is better than any of the R^2s produced in Section 3.2.2, highlighting the usefulness of SVR in finding relationships in potentially very complex situations (its predictions had an RMSE of 1.95 log units, and MAE of 1.44 log units). Finally, the data from the geometry measurements was input to the models and again, the models were unable to find a satisfactory relationship between input and output, further underlining the lack of relationship between DFT geometry optimisations and stability constant.

As highlighted in Section 3.1.3, understanding what is needed from a model, and understanding the sort of data that would be used as input in a real-world use of your model is important for deciding what it should be trained on. In this case, the model should be designed to be helpful for clinically viable MnBCAs. Looking at the dataset, it was evident that certain MnBCAs were not actually providing much use in training the models because they are over-representing 'junk data' that would not be able to produce a clinically useful MnBCA. The first filter that was applied was to remove MnBCAs that did not have $q \geq 1$, Table 3.9.

Table 3.9: R^2 results from the regression tasks split by ML model (on MnBCAs that have $q \geq 1$), and variables each model was given to perform the regression with.

Dataset	Linear Regression	Neural Network[a]	SVR
All variables	0.752	0.695	0.236
DFT energies	0.336	-0.055	0.335
Geometry measurements	0.018	-0.424	0.056
Handpicked	0.740	0.671	0.216

[a] The number of neurons in the hidden layers of the neural network was varied depending on the number of features in the input layer: the 'all variables' set had 43 and 30 neurons in its hidden layers; the 'DFT energies' set had 5 and 4; the 'geometry measurements' set had 20 and 14; the handpicked set had 19 and 14.

Removing the MnBCAs that did not have coordinated water molecules massively improved the ability of the linear regression model (25 remaining MnBCAs). For the 'all variables' dataset it produced an R^2 of 0.752 (RMSE of 1.98 log units, MAE of 1.32 log units – the larger RMSE indicates a small number of MnBCAs had a prediction with large deviation), and the neural network also was able to produce reasonably good stability constant predictions with an R^2 of the testing dataset of 0.695. In this, it appears that with a smaller, more focused dataset both models are able to perform better – in the case of linear regression, this total dataset performs better than its subsets (seen in the R^2 values of 'DFT energies' and 'Geometry measurements'). This indicates that a mixture of variables is needed, and that the regularisation is working to suppress some of the less important features. This improved result led to defining a new dataset with handpicked features depending on what is known to affect the stability constant, as well as the individual features that contributed most to the good R^2 with the 'all variables' dataset with linear regression – CShM, coordination number, geometry type, molecular weight, the six DFT features, *q*, the four charge calculations, and average Mn-carboxylate, Mn-donor, Mn-water, Mn-3° amine bond lengths. This slimmed combined dataset was able to produce an equally good R^2 in the linear regression (0.740) and neural network (0.671) models.

Table 3.10: R^2 results from the regression tasks split by ML model (on MnBCAs that have $q \geq 1$, and stability constant ≥10), and variables each model was given to perform the regression with.

Dataset	Linear Regression	Neural Network[a]	SVR
All variables	0.851	0.073	-0.27
DFT energies	0.873	0.472	0.397
Geometry measurements	-0.107	0.756	-0.011
Handpicked	0.879	0.533	-0.098

[a] The number of neurons in the hidden layers of the neural network was varied depending on the number of features in the input layer: the 'all variables' set had 43 and 30 neurons in its hidden layers; the 'DFT energies' set had 5 and 4; the 'geometry measurements' set had 20 and 14; the handpicked set had 19 and 14.

At the risk of overfitting the model to the training dataset, a final additional filter was included to remove MnBCAs with stability constants below 10 (these would be instantly rejected for clinical applications), which slimmed the dataset to 21 remaining complexes. Removing this 'bad faith' data produced excellent R^2 values for the linear regression model on the 'all variables', 'DFT energies', and 'handpicked' datasets (0.851, 0.873, and 0.879 respectively). The models produced were able to predict the stability constants of the testing dataset very accurately (RMSE of 'handpicked'/linear regression was 0.52 log units, MAE was 0.50 log units). Somewhat surprisingly, the neural network was able to produce a model which could predict the stability constant from the geometry measurements with R^2 of 0.756 – the largest value of the neural network models (RMSE of 0.74, MAE of 0.70 log units).

3.3 Discussion

The overall QM model we use has a range of constraints in order for its application to be feasible:

- SCRFs are not hyper accurate, and due to computational cost, we chose not to model a cluster of water surrounding the ligand, hexaaqua manganese, and complex.
- Conformers were not considered for any of the ligands, and only *cis/trans* isomerism was considered for $q = 2$ complexes in their water coordination.
- In reality, the experimental $\log(K_{MnL})$ will likely be surveying an abundance of protonation states on the ligand, water, hexaaqua manganese, and complex.
- The manganese optimisation assumes a bound water molecule, when in fact MnBCAs will be existing in a spectrum of how bound the water is as it rapidly exchanges water with its hydration sphere.
- The hydration number was simplified to an integer, when in reality complexes will exist in a distribution of $q = 0$ (where we could have alternative donors present) to $q = 6$ (uncomplexed).
- Ionic strength was set to 0, whereas the experimental $\log(K_{MnL})$ values used in this study were measured in ionic strengths between 0.1 and 0.15 M solutions of either NaCl, KCl, KNO_3, NMe_4Cl, $NMe_4(NO_3)$, $NaClO_4$, $NaNO_3$ depending on the publication.
- Experimental $\log(K_{MnL})$ values, strictly speaking, are calculated using activity rather than concentration.

Taken together, these likely contribute to the difference in calculated $\log(K_{MnL})$ and the experimental $\log(K_{MnL})$.

Additionally, the experiment design was constrained to reduce the scope of what could be tested and how, because the desire was for calculations to be run by non-computationally literate chemists (therefore restricted in terms of software). For the model identification task, each complex was categorised as $q = 0$, 1, or 2 based on available published data (rounded), else r_1: $q = 0 < 2.5 < q = 1 <$

3.8 < q = 2. These numbers were chosen by surveying the literature r_1 values of q = 0, 1, 2 complexes, and simplifies the problem of hydration numbers so that the work could feasibly be carried out. Potentially both q = 0 and q = 1 calculations could be run and then weighted in accordance with their hydration data for improved accuracy (in the case of q > 1, you would do the same but with q = 1, q = 2). This would still be a simplification, as a more realistic situation is a spectrum of positions that the water is bound at during log(K_{MnL}) measurement.

As was seen with the success in calculating separate models for each charge in Section 3.2.4.1, it is possible that every class of ligands would need its own model to accurately predict log(K_{MnL}) – this is most evident for the bispidine complexes which were completely unpredictable when the models were trained on ligands that were far less restrictive, and far less effective at complexing manganese. It is important to note that in both the case of L169 and L172, DFT models produced a lower calculated value than expected implying that if the models were trained with a bispidine-only training set, better predictions would likely be achieved. Taking this concept further, it is also possible that other ways of clustering the data would produce higher R^2 – segmenting by acyclic/macrocyclic, or each ligand type being modelled separately (PC*X*A, DO*X*A, ED*X*A, CD*X*A families).

Given the number of simplifications required for practicality, it is remarkable that predicted log(K_{MnL}) values can achieve MAEs of < 2 log units (± 2.73 kcal mol^{-1}). This is on an equivalent magnitude to that provided by Uzal-Varela, who – on their dataset of 150 complexes – were able to produce a phenomenological model between donors and log(K_{MnL}) that provided an RMSE and MAE of 1.48 and 0.99 log units.[7]

One of the other problems this analysis suffered is the variance in the MnBCA dataset, which is in part due to the small window in which stability constants exist for MnBCAs – the vast majority of MnBCAs in the dataset exist in a 5 log unit window. Having a smaller window in which to work within means that any deviations from reality – whether the geometries are incorrectly set up, or ignoring conformers/isomers, or even the inherent error of Gaussian – become more significant when trying to find a correlation relationship (the flatter a gradient, the closer to $\bar{y}$, and the more important low variance is for a good R^2 (Equation 3.12)).

The lack of a strong correlation between donor bond lengths – tertiary amines, carboxylates, average donor length – and stability constant speaks against the assumption that ligands necessarily have a cavity size. Unless extremely constrained, as in the case for **L172** or cryptands, for instance, most ligands are far too flexible to have a specific cavity size – more realistic is a cavity range. Yet, the lack of correlation in the CShM analysis shows that throughout a ligand's cavity range whichever shape is

acquired will not be what determines its stability constant. Instead, as seen in the Uzal-Varela analysis, specific Mn-donor interactions are more important.

3.4 Conclusion

The first set of experiments investigated if the DFT methods that a non-expert level chemist could reasonably be expected to use can predict $\log(K_{MnL})$ values. Gaussian 16 was used to run optimisation and frequency calculations to find a minimum energy geometry and relevant thermodynamic values. These values were then used to create calculated $\log(K_{MnL})$ values which were correlated to experimental values *via* linear least squares regression. The models with the highest R^2 values were taken forward and validated against a testing dataset, testing the models' accuracy and ability to differentiate between structures that appear identical to other $\log(K_{MnL})$ prediction methods. The three models had predictions that gave MAEs of 2.23, 1.84, 1.79 log units, and RMSEs of 4.91, 3.72, 3.92 log units (MAEs will give a broader idea of the error level due to RMSE being sensitive to outliers). In addition to this, the models were able to predict the relative positions of isomers in the CDTA family, and three complexes in the PC2A family that are identical to the Uzal-Varela method. The models struggled to correctly predict the relative positions for the CBuDTA family, implying that there may have been systematic mistreatment of either of the ligand/complex pairs. The three models also made poor attempts at predicting bispidine-COOH when trained on a dataset which likely did not resemble the bispidine family enough, which was also the reason why Uzal-Varela removed three bispidine complexes from their final model.

The main aim of the geometry section was to find if any correlations existed between the DFT calculated geometry of a MnBCA, and the KPIs typically used to assess their suitability in literature. Whilst some analyses produced somewhat promising relationships – i.e. $\log(K_{MnL})$ vs average Mn-water bond length with a negative quadratic trend line – overall the consensus between the shape measures and bond length distributions showed no significant correlation to the stability constant, water exchange rate, or hydration number. This gave some evidence towards disproving some of the conjecture around water exchange rate and design considerations for MnBCAs that can be found in literature. The inability to find a relationship between what is essentially a set of optimised coordinates and the MnBCA properties, may be due to the reality being considerably more dynamic – there is not just one static conformation or species that is being measured when r_1 or $\log(K_{MnL})$ is measured experimentally. However, in this instance, it would be unfeasible to use multiple conformers/isomers/hydration numbers to find a more accurate species distribution in which measurements are then taken for the correlation.

The dataset from Section 3.2.3 was used to investigate the effect of a larger dataset on the predictive ability of one of the models from Section 3.2.2.6. This was able to provide a larger R^2 than when only tested on the training dataset (0.199 compared to 0.142), and a positive slope as expected from the experiment in Section 3.2.2.6 and similar experiments in literature.[8,13] Analysis of the residuals showed no inherent error in the method, but highlighted the high variance that the MnBCA dataset produces when studied by DFT. The data showed that clustering by complex charge improved the R^2 value of the linear fit from 0.199 to -2: 0.228, 0: 0.319, +2: 0.313, and highlighted that the -1 charge subset of complexes may require more data to produce the desired fit. When segmented by hydration number, there appears to be a weak clustering between $q = 1$ and $q = 0$ complexes, with $q = 0$ complexes providing higher $\log(K_{MnL})$ values, as would be expected with saturating the coordination sphere.

The question at the centre of any predictive model is 'what is an acceptable level of accuracy?' As a synthetic MnBCA chemist, a ±2.5 log unit estimate would certainly be enough to decide whether or not to start or abandon synthetic efforts. Model 3 is able to provide a MAE of 1.79 and RMSE of 3.92 (indicating the model can sometimes predict values with a large error). This does not improve on Uzal-Varela's MAE of 0.99 and RMSE of 1.48. As such, as long as the complex can be summarised by their descriptors,[7] it would be better to use Equation 3.1 to predict $\log(K_{MnL})$, and in the cases where the complex cannot, as long as it does not deviate too far from our training set of complexes – i.e. an 8-coordinate, rigid bispidine – then the extra effort of optimisation to provide a ±1.79 log unit estimate (with the knowledge that the predictive model has an R^2 of 0.218) is better than having no estimate at all. As for isomer prediction, whilst the low R^2 values of the models are unable to accurately predict $\log(K_{MnL})$, they offer a complementary analysis to the Uzal-Varela model in terms of qualitatively differentiating between complexes which appear identical to that model.

When individually assessed for correlation, none of the features of the MnBCA dataset were able to provide a high R^2 estimate for the stability constant (or indeed any other properties for that matter). However, these features were able to produce ML models that could combine and weight each of the features to provide predictions that were far more accurate. Even without pruning the dataset of complexes that would not qualify as clinically viable MnBCAs, ML models produced higher R^2 values than simple linear regression – SVR with the 'DFT energies' dataset produced an R^2 of 0.371. Once the dataset was cleaned, these models were able to further improve their predictive power, with the end result of the linear regression model being able to provide an excellent R^2 of 0.879, RMSE of 0.52 log units, and MAE of 0.50 log units for the 'handpicked' dataset. This level of accuracy would surpass the level required for synthetic chemists in deciding whether or not a hypothetical complex is worth pursuing synthetically; and most importantly relied on data that can be generated from two DFT calculations once a sensible integer estimation of q is provided.

3.4.1 Future work

As seen with the 'with water × M06-2X × TZVP × IEFPCM' model, the multi-model study could be expanded to q = 0-2 complexes, and a modest increase in R^2 was obtained when comparing the result from 10 complexes and 23 complexes for q = 1. In terms of improving the overall model itself, the self-imposed constraints of the project could be addressed. Attempting the entire project with the COSMO-RS solvation model *via* COSMOtherm software could possibly increase the accuracy of the solvation energies and produce calculated values with lower variance. Additionally, explicitly adding water molecules into the optimisations to help transition from the bulk solvent field to the solute is likely to produce better solvation energies. In a similar fashion, more accurately sampling the true state of the manganese complex during experimental $\log(K_{MnL})$ measurement may lead to an increase in accuracy – this could be achieved by identifying and optimising conformers of the ligands and complexes, different protonation states by considering pK_a, or different hydration numbers (as the ligand and complex will exist in a spectrum between complexed and free). Then a Boltzmann distribution with the acquired Gibbs energies could be calculated to place a weighting in front of every state to model the actual environment more accurately. Practically speaking, employing all of these suggestions alone would take a year *at least* to finalise for *one* model alone, for potentially minimal increases in accuracy. As such, testing would need to be carried out to determine which changes contribute most to a reduction in variance.

The precedent of the DFO and UO_2 papers suggests that manganese itself is the barrier to a usable prediction as opposed to the approach. In terms of aiming for a useful predictive model, moving to a metal that is easier to model, such as one with a closed shell like Zr(IV) or Ga(III), with plenty of available literature complexes would be more likely to provide success.

Most exciting was the ability of ML models to take the data produced in this work and combine it to find patterns in the data that are more complex than can be analysed with simple data analysis tools. Using the above recommendations to provide more accurate DFT derived properties should be able to improve on the result from Table 3.8. However, providing all of the models more data to train on (and test with) could improve the R^2 values, potentially reducing the effect of outliers within the dataset that was studied. As noted, there are 100 other complexes in the Uzal-Varela dataset which could be geometry optimised and added to the dataset; of these, 21 have r_1 values. The logical next step would be to take the DFT calculated data for all 150 complexes, filter for complexes with complete r_1 data and train a ML model that can predict the ultimate KPI for MnBCA – its ability to provide useful contrast.

3.5 References

1 A. Gupta, P. Caravan, W. S. Price, C. Platas-Iglesias and E. M. Gale, *Inorg. Chem.*, 2020, **59**, 6648–6678, DOI: 10.1021/acs.inorgchem.0c00510.

2 J. Wahsner, E. M. Gale, A. Rodríguez-Rodríguez and P. Caravan, *Chem. Rev.*, 2019, **119**, 957–1057, DOI: 10.1021/acs.chemrev.8b00363.

3 S. Laus, R. Ruloff, É. Tóth and A. E. Merbach, *Chem. - A Eur. J.*, 2003, **9**, 3555–3566, DOI: 10.1002/chem.200204612.

4 C. Kies, in *Nutritional Bioavailability of Manganese*, 1987, pp. 1–8, DOI: 10.1021/bk-1987-0354.ch001.

5 NIH, Dietary Supplement Fact Sheet on Manganese for Health Professionals.

6 X. G. Kondakis, N. Makris, M. Leotsinidis, M. Prinou and T. Papapetropoulos, *Arch. Environ. Heal. An Int. J.*, 1989, **44**, 175–178, DOI: 10.1080/00039896.1989.9935883.

7 R. Uzal-Varela, F. Pérez-Fernández, L. Valencia, A. Rodríguez-Rodríguez, C. Platas-Iglesias, P. Caravan and D. Esteban-Gómez, *Inorg. Chem.*, 2022, **61**, 14173–14186, DOI: 10.1021/acs.inorgchem.2c02364.

8 S. Vukovic, B. P. Hay and V. S. Bryantsev, *Inorg. Chem.*, 2015, **54**, 3995–4001, DOI: 10.1021/acs.inorgchem.5b00264.

9 O. Gutten and L. Rulíšek, *Inorg. Chem.*, 2013, **52**, 10347–10355, DOI: 10.1021/ic401037x.

10 O. Gutten, I. Beššeová and L. Rulíšek, *J. Phys. Chem. A*, 2011, **115**, 11394–11402, DOI: 10.1021/jp205442p.

11 Z. Garda, E. Molnár, N. Hamon, J. L. Barriada, D. Esteban-Gómez, B. Váradi, V. Nagy, K. Pota, F. K. Kálmán, I. Tóth, N. Lihi, C. Platas-Iglesias, É. Tóth, R. Tripier and G. Tircsó, *Inorg. Chem.*, 2021, **60**, 1133–1148, DOI: 10.1021/acs.inorgchem.0c03276.

12 V. Patinec, G. A. Rolla, M. Botta, R. Tripier, D. Esteban-Gómez and C. Platas-Iglesias, *Inorg. Chem.*, 2013, **52**, 11173–11184, DOI: 10.1021/ic4014366.

13 J. P. Holland, *Inorg. Chem.*, 2020, **59**, 2070–2082, DOI: 10.1021/acs.inorgchem.9b03515.

14 A. S. Ivanov and V. S. Bryantsev, *Dalton Trans.*, 2016, **45**, 10744–10751, DOI: 10.1039/c6dt01752e.

15 J. A. Peters and C. F. G. C. Geraldes, *Inorganics*, 2018, **6**, 116, DOI: 10.3390/inorganics6040116.

16 A. Forgács, L. Tei, Z. Baranyai, D. Esteban-Gómez, C. Platas-Iglesias and M. Botta, *Dalton Trans.*, 2017, **46**, 8494–8504, DOI: 10.1039/C7DT01508A.

17 R. Pujales-Paradela, F. Carniato, D. Esteban-Gómez, M. Botta and C. Platas-Iglesias, *Dalton Trans.*, 2019, **48**, 3962–3972, DOI: 10.1039/C9DT00211A.

18 E. M. Gale, J. Zhu and P. Caravan, *J. Am. Chem. Soc.*, 2013, **135**, 18600–18608, DOI: 10.1021/ja4094132.

19 B. Drahoš, M. Pniok, J. Havlíčková, J. Kotek, I. Císařová, P. Hermann, I. Lukeš and É. Tóth, *Dalton Trans.*, 2011, **40**, 10131, DOI: 10.1039/c1dt10543d.

20 F. K. Kálmán, V. Nagy, R. Uzal-Varela, P. Pérez-Lourido, D. Esteban-Gómez, Z. Garda, K. Pota, R. Mezei, A. Pallier, É. Tóth, C. Platas-Iglesias and G. Tircsó, *Molecules*, 2021, **26**, 1524, DOI: 10.3390/molecules26061524.

21 M. Regueiro-Figueroa and C. Platas-Iglesias, *J. Phys. Chem. A*, 2015, **119**, 6436–6445, DOI: 10.1021/acs.jpca.5b01728.

22 J. M. Puaschunder, *SSRN Electron. J.*, 2020, **6**, 94–98, DOI: 10.2139/ssrn.3525037.

23 A. K. Jha, A. Ghimire, S. Thapa, A. M. Jha and R. Raj, in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, IEEE, 2021, pp. 937–944, DOI: 10.1109/ICICT50816.2021.9358548.

24 P. Lu, S. Chen and Y. Zheng, *Math. Probl. Eng.*, 2012, **2012**, 1–22, DOI: 10.1155/2012/145974.

25 A. L. Hunkenschroer and C. Luetge, *J. Bus. Ethics*, 2022, **178**, 977–1007, DOI: 10.1007/s10551-022-05049-6.

26 C. C. Bannan, K. H. Burley, M. Chiu, M. R. Shirts, M. K. Gilson and D. L. Mobley, *J. Comput. Aided. Mol. Des.*, 2016, **30**, 927–944, DOI: 10.1007/s10822-016-9954-8.

27 A. Klamt, F. Eckert, J. Reinisch and K. Wichmann, *J. Comput. Aided. Mol. Des.*, 2016, **30**, 959–967, DOI: 10.1007/s10822-016-9927-y.

28 M. Işık, T. D. Bergazin, T. Fox, A. Rizzi, J. D. Chodera and D. L. Mobley, *J. Comput. Aided. Mol. Des.*, 2020, **34**, 335–370, DOI: 10.1007/s10822-020-00295-0.

29 C. Loschen, J. Reinisch and A. Klamt, *J. Comput. Aided. Mol. Des.*, 2020, **34**, 385–392, DOI: 10.1007/s10822-019-00259-z.

30 D. Guan, R. Lui and S. Matthews, *J. Comput. Aided. Mol. Des.*, 2020, **34**, 511–522, DOI: 10.1007/s10822-020-00278-1.

31 M. Amezcua, L. El Khoury and D. L. Mobley, *SAMPL7 Host–Guest Challenge Overview: assessing the reliability of polarizable and non-polarizable methods for binding free energy calculations*, Springer International Publishing, 2021, vol. 35, DOI: 10.1007/s10822-020-00363-5.

32 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G.

Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian, Inc., Wallingford, CT, 2016.

33 S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129, DOI: 10.1016/0301-0104(81)85090-2.

34 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396, DOI: 10.1021/jp810292n.

35 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652, DOI: 10.1063/1.464913.

36 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241, DOI: 10.1007/s00214-007-0310-x.

37 Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656–5667, DOI: 10.1021/jp050536c.

38 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728, DOI: 10.1063/1.1674902.

39 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571–2577, DOI: 10.1063/1.463096.

40 A. Austin, G. A. Petersson, M. J. Frisch, F. J. Dobek, G. Scalmani and K. Throssell, *J. Chem. Theory Comput.*, 2012, **8**, 4989–5007, DOI: 10.1021/ct300778e.

41 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615, DOI: 10.1039/b810189b.

42 J. Demšar, T. Curk, A. Erjavec, C. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik and B. Zupan, *J. Mach. Learn. Res.*, 2013, **14**, 2349–2353.

43 N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, 115, 2315–2372, DOI: 10.1080/00268976.2017.1333644.

44 H. Chen, R. Shi and H. Ow, *ACS Omega*, 2019, **4**, 20665–20671, DOI:

10.1021/acsomega.9b02851.

45 J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc., Wallingford, CT, 3rd edn., 2015.

46 H. Konieczna, D. Lundberg and I. Persson, *Polyhedron*, 2021, **195**, 114961, DOI: 10.1016/j.poly.2020.114961.

47 D. Esteban-Gómez, C. Cassino, M. Botta and C. Platas-Iglesias, *RSC Adv.*, 2014, **4**, 7094, DOI: 10.1039/c3ra45721d.

48 W. W. Rudolph and G. Irmer, *Dalton Trans.*, 2013, **42**, 14460, DOI: 10.1039/c3dt51493e.

49 C. F. Baes and R. S. Mesmer, *The Hydrolysis of Cations*, Wiley, New York, 1976, DOI: 10.1002/bbpc.19770810252.

50 H. Ma and D. Niu, 2006, DOI: 10.5517/cc9ldnx.

51 G. Galstyan and E.-W. Knapp, *J. Comput. Chem.*, 2015, **36**, 69–78, DOI: 10.1002/jcc.23764.

52 S. V Jerome, T. F. Hughes and R. A. Friesner, *J. Phys. Chem. B*, 2014, **118**, 8008–8016, DOI: 10.1021/jp501086h.

53 R. Gilson and M. C. Durrant, *Dalton Trans.*, 2009, 10223, DOI: 10.1039/b911593e.

54 J. Cirera, E. Ruiz and S. Alvarez, *Organometallics*, 2005, **24**, 1556–1562, DOI: 10.1021/om049150z.

55 S. Alvarez and J. Echeverría, *J. Phys. Org. Chem.*, 2010, **23**, 1080–1087, DOI: 10.1002/poc.1735.

56 S. S. Batsanov, *Inorg. Mater.*, 2001, **37**, 871–885, DOI: 10.1023/A:1011625728803.

57 A. Forgács, R. Pujales-Paradela, M. Regueiro-Figueroa, L. Valencia, D. Esteban-Gómez, M. Botta and C. Platas-Iglesias, *Dalton Trans.*, 2017, **46**, 1546–1558, DOI: 10.1039/C6DT04442E.

58 K. Pota, Z. Garda, F. K. Kálmán, J. L. Barriada, D. Esteban-Gómez, C. Platas-Iglesias, I. Tóth, E. Brücher and G. Tircsó, *New J. Chem.*, 2018, **42**, 8001–8011, DOI: 10.1039/C8NJ00121A.

59 D. Ndiaye, P. Cieslik, H. Wadepohl, A. Pallier, S. Même, P. Comba and É. Tóth, *J. Am. Chem. Soc.*, 2022, **144**, 22212–22220, DOI: 10.1021/jacs.2c10108.

60 B. E. Osborne, T. T. C. Yue, E. C. T. Waters, F. Baark, R. Southworth and N. J. Long, *Dalton Trans.*, 2021, **50**, 14695–14705, DOI: 10.1039/d1dt02224e.

61 L. Carreras, M. Serrano-Torné, P. W. N. M. van Leeuwen and A. Vidal-Ferran, *Chem. Sci.*, 2018, **9**, 3644–3648, DOI: 10.1039/C8SC00233A.

62 T. Csupász, D. Szücs, F. K. Kálmán, O. Hollóczki, A. Fekete, D. Szikra, É. Tóth, I. Tóth and G. Tircsó, *Molecules*, 2022, **27**, 371, DOI: 10.3390/molecules27020371.

63 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145, DOI: 10.1002/jcc.540050204.

64 H. Hu, Z. Lu and W. Yang, *J. Chem. Theory Comput.*, 2007, **3**, 1004–1013, DOI: 10.1021/ct600295n.

65 C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, 1990, **11**, 361–373, DOI: 10.1002/jcc.540110311.

66 J. P. Foster and F. Weinhold, *J. Am. Chem. Soc.*, 1980, **102**, 7211–7218, DOI: 10.1021/ja00544a007.

Chapter 4

What Zr(IV) species is being formed within DOXIL?

4.1 Introduction

Doxorubicin, Figure 4.1, is a chemotherapy medicine used in the treatment of multiple cancer types including breast cancer, pancreatic cancer, ovarian cancer, multiple myeloma, and Kaposi's sarcoma.[1–6] As an anticancer drug, doxorubicin intercalates with DNA inhibiting topoisomerase II – an enzyme used in DNA transcription, which stops DNA replication. Subsequent attempts at DNA repair fail, which triggers cell death.[6] Doxorubicin, however, is a nonspecific chemotherapy agent. As such, it suffers from dose dependent risk of toxicity to multiple organs throughout the body – NHS guidelines limit the total cumulative dose of doxorubicin to 450 mg/m^2.[7] The greatest of these is the risk of aggressive cardiac damage *via* enlargening cardiomyocytes, and adversely affecting mitochondria in the heart.[6,8] Alongside this, doxorubicin directly and indirectly causes toxicity in the brain, liver, and kidneys, as well other common chemotherapy side effects such as hair loss, immune system suppression, and gastrointestinal problems.[9]

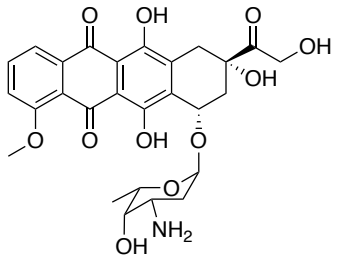


Figure 4.1: Skeletal diagram of the non-specific anti-cancer drug doxorubicin.

4.1.1 Liposomal drug delivery

An approach to limit the possible toxicity of free doxorubicin is to use a drug delivery method, such as encapsulating it within a liposome, Figure 4.2. Liposomal drug delivery is typically used to provide a longer circulation time than the free drug, passive targeting of the disease site, and excellent biocompatibility.[10] The EMA and FDA have approved 14 liposomal products which cover chemotherapeutics (such as DOXIL, Onivyde, Vyxeos), vaccine delivery (Shingrix), and sustained release analgesia (Exparel and DepoDur).[10] Currently there are two liposomal formulations of doxorubicin (Myocet and DOXIL (in the US, Caelyx in the EU)) that are authorised, which differ in their composition and method of doxorubicin loading. For doxorubicin, liposomal delivery effectively hides the drug from the body until the liposome has passively accumulated in a tumour - typically due to extravasation of the liposome through the tumour's 'leaky' blood vessels.[11] At which point the doxorubicin can diffuse out, or the liposome membrane is compromised. This increase in the delivery of doxorubicin to the tumour site reduces the risk of cardiotoxicity and side effects.[6]

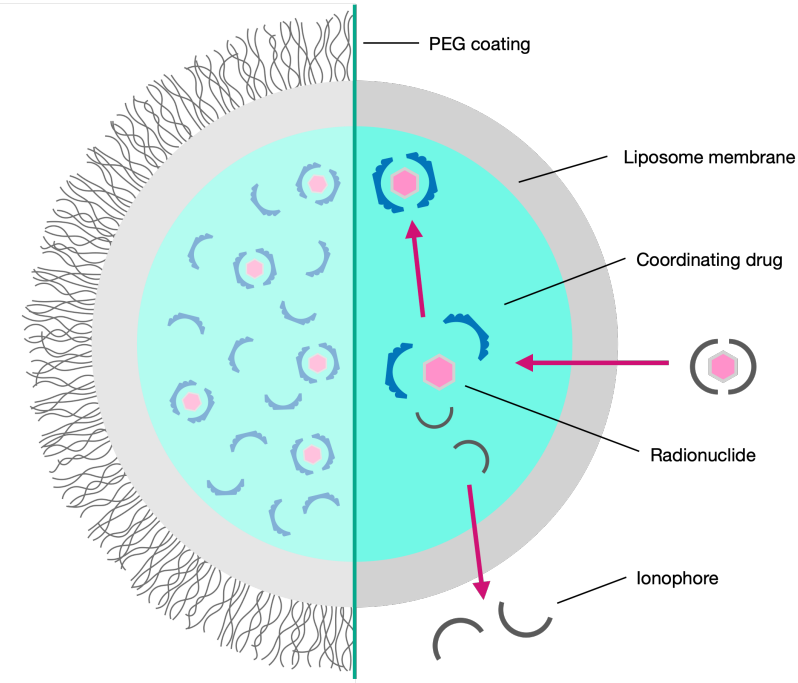


Figure 4.2: Diagram of a PEGylated liposome being radiolabelled via *ionophore-drug binding. Ionophores transport the radionuclide across the liposome membrane, where it is bound by the encapsulated drug, trapping it within the liposome.*

Liposomal doxorubicin has been shown by Gabizon and co-workers to significantly improve the accumulation of doxorubicin in malignant effusions tapped from cancer patients compared to free doxorubicin.[12] Additionally, the amount of doxorubicin detected in these effusions peaked between 3 to 7 days from administration, suggesting that liposomal doxorubicin both increases the size and duration of exposure compared to free doxorubicin.[13] However, most importantly, liposomal formulations of doxorubicin offer significantly reduced toxicity and side effects, as demonstrated by its phase III clinical trial results. As a first-line treatment of metastatic breast cancer, doxorubicin patients were significantly more likely to develop cardiotoxicity than those treated with DOXIL ($p<0.001$, hazard ratio of 3.16).[14]

Understanding the pharmacokinetics is crucial to efficiently develop improvements to targeting and understanding the mechanism of drug release – this can be achieved *via* radiolabelling these liposomes allowing for non-invasive imaging. PET radiolabelling overcomes many of the disadvantages of other methods – MRI suffers from low sensitivity, ultrasound is impractical for whole-body imaging, whole-body CT requires larger radiation doses than PET, and fluorescence tomography suffers from poor depth penetration. PET is able to be administered at nanomolar concentrations, can be used for whole-body imaging, and does not suffer from tissue penetration issues. Multiple strategies can be employed to image liposomes *in vivo via* radiolabelling – either radiolabelling the surface, or through intraliposomal labelling. The most common form uses weakly coordinating ligands (known as ionophores) that are able to transport the radionuclide across the liposome membrane (due to their lipophilic character).

Once inside, encapsulated chelators within the liposome can bind to the radionuclide, reducing the ability of the radionuclide to escape, Figure 4.2. Studies have shown success in using oxine as an ionophore to deliver a radiometal inside a liposome for *in vivo* PET imaging, Figure 4.2. In three separate publications, T. M. de Rosales and co-workers have used [^{89}Zr]Zr(oxinate)$_4$, [^{52}Mn]Mn(oxinate)$_2$, and [^{64}Cu]Cu(2-hydroxyquinoline)$_2$ in radiolabelling liposomal drug delivery formulations (Figure 4.3).[15–17] The first of these explored methodologies in radiolabelling liposomes containing metal-chelating drugs – DOXIL containing doxorubicin, or PLA (PEGylated liposomal alendronate) containing ammonium alendronate (Figure 4.4). Radiolabelled PLA with all three radionuclides showed excellent *in vitro* radiolabelling stability in human serum stabilities (≥ 95%), whilst radiolabelled DOXIL exhibited better stability with ^{52}Mn compared to ^{89}Zr. In preclinical full-body PET-CT scanning of a mouse model of breast and ovarian cancer both ^{89}Zr-PLA and ^{64}Cu-PLA were able to image the expected biodistribution of liposomes at the primary tumour sites.[16] Furthermore in 2020, the de Rosales group used [^{89}Zr]Zr(oxinate)$_4$ to radiolabel a liposomal glucocorticoid, demonstrating the ability to track the biodistribution of the liposome and uptake at inflamed sites.

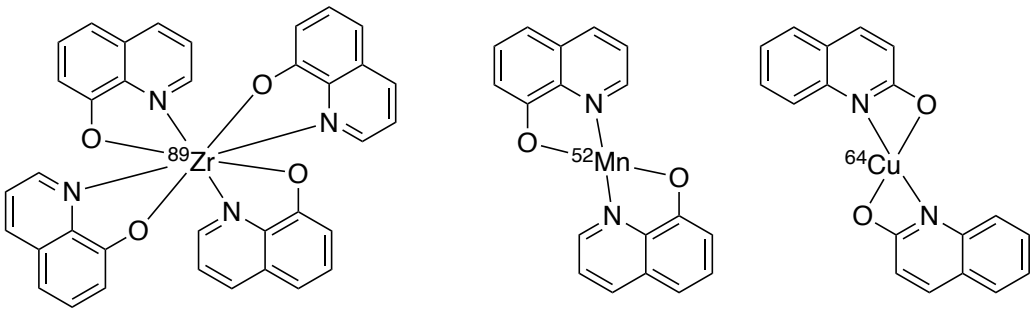


Figure 4.3: [^{89}Zr]Zr(oxinate)$_4$, [^{52}Mn]Mn(oxinate)$_2$, and [^{64}Cu]Cu(2-hydroxyquinoline)$_2$ used by T. M. de Rosales and co-workers for liposomal radiolabelling.

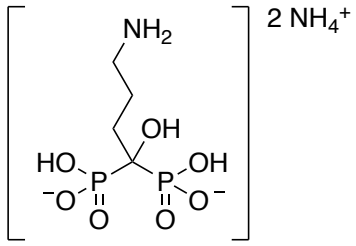


Figure 4.4: Ammonium alendronate, the active pharmaceutical ingredient of PLA.

4.1.2 Zr DFT studies

The exact species that forms within the liposome once oxine has transferred ^{89}Zr into the liposome has yet to be characterised. The *in vivo* stability of radiolabelled liposomes suggests that [^{89}Zr]Zr(oxinate)$_4$ does not passively diffuse out of the liposome as would be expected for a neutral lipophilic complex – implying a conversion of [^{89}Zr]Zr(oxinate)$_4$ into something else.

DFT modelling can help to shed light on the possible reactivity of Zr(oxinate)$_4$ in an environment containing doxorubicin under different simulated pH ranges. In a similar fashion, Holland and co-workers have modelled the conversion of $[Zr(C_2O_4)_4]^{4-}$ to Zr(MeAHA)$_4$ through ligand substitution *via* a dissociative pathway under simulated basic and acidic conditions.[18] N-methyl acetohydroxamate (MeAHA) approximates the coordinating groups of desferrioxamine B (DFO), with a further MeAHA unit to fully saturate the zirconium coordination sphere, Figure 4.5. Before the authors calculated the energies of each formed complex, a range of functionals and basis sets were evaluated *via* comparison of the bond lengths and angles of optimised geometries to the X-ray structure of Zr(MeAHA)$_4$. The authors settled on the B3LYP functional with the LANL2DZ/IEFPCM basis set as this combination performed best when the overall root mean square deviation between optimised geometry and crystal structure coordinates, and computational cost were considered.

The Gibbs energies at each step of the dissociative ligand substitution was calculated according to Equation 4.1. Relative Gibbs energy changes from the starting arrangement of $[Zr(C_2O_4)]^{4-}$ and 4 MeAHA were then calculated. Holland and co-workers simulated acidic conditions *via* optimising the protonated complex, removing $C_2O_4H^-$, and then complexation with MeAHA. Basic conditions were simulated with all free ligands being fully deprotonated.

The complete ligand substitution reaction was found to be thermodynamically feasible under the simulated basic conditions (ΔG_{solv} = -81 kJ mol^{-1}), however this became infeasible under simulated acidic conditions (ΔG_{solv} = +124 kJ mol^{-1}), Figure 4.6. Whilst the final structure in the basic environment is 81 kJ mol^{-1} below the starting Zr(oxalate)$_4$, an average energy gap between steps in the forward direction of 174.5 kJ mol^{-1} indicates that the 6-coordinate intermediate structure would be very difficult to achieve *via* dissociation (for a barrier-less reaction, at 100 °C, a 174.5 kJ mol^{-1} energy would have a half-life of 9.2 ×10^{13} hours). As the overall DFT result mirrored experimental results, it suggests there is a problem in the mechanism or in the DFT methods – polarisable continuum model (PCM) based methods become less accurate as the overall charge deviates from neutrality (the complexes range from 4– to 1+) as solvation becomes less well approximated by a solute sitting in a neat cavity inside a field of charge; and secondly, the choice of B3LYP and LANL2DZ may have contributed to the error in accurately predicting ΔG_{solv}.[19]

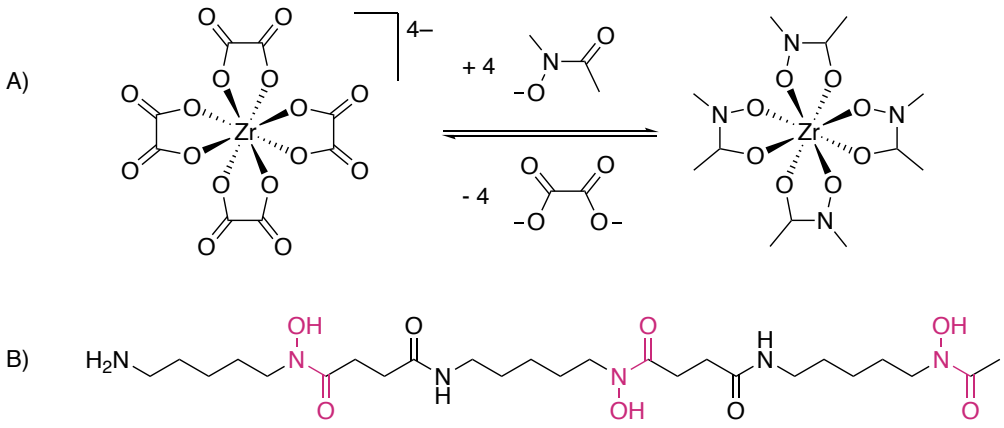
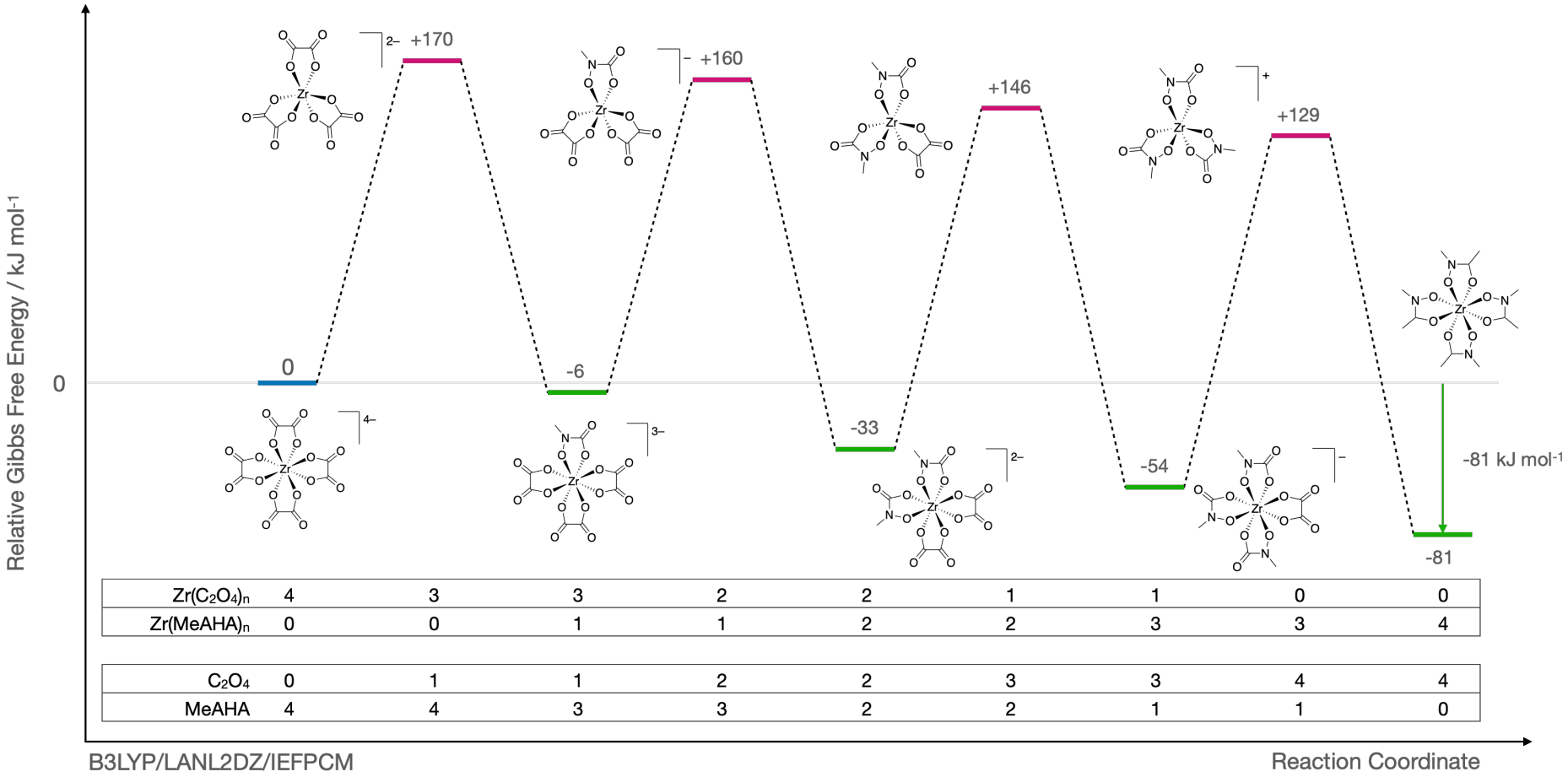


Figure 4.5: A) Conversion of $[Zr(C_2O_4)]^{4-}$ to $Zr(MeAHA)_4$ modelled by Holland and co-workers;[18] B) Desferrioxamine B, with coordinating groups highlighted in pink.

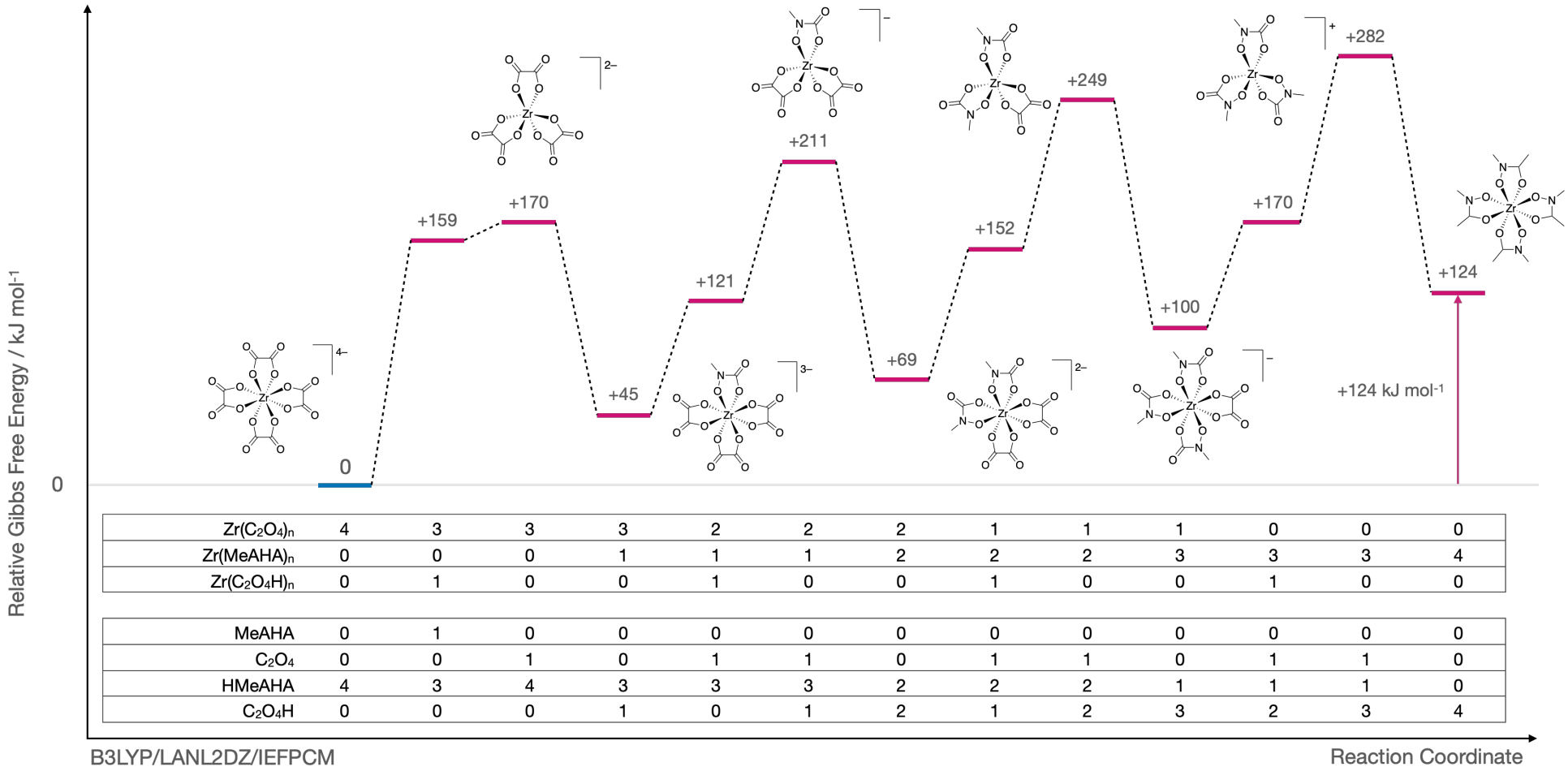
$$G(step) = G(complex) + \sum_{0}^{a} G(MeAHA) + \sum_{0}^{b} G(C_2O_4) + \sum_{0}^{c} G(HMeAHA) + \sum_{0}^{d} G(C_2O_4H)$$

Equation 4.1: Equation for the calculation of Gibbs energy at each step in the dissociative mechanism as a summation of complex and free ligands, where a, b, c, and d are the stoichiometric equivalents of unchelated MeAHA, C_2O_4, HMeAHA, and C_2O_4H existing in solution respectively.



$Zr(C_2O_4)_n$	4	3	3	2	2	1	1	0	0
$Zr(MeAHA)_n$	0	0	1	1	2	2	3	3	4

C_2O_4	0	1	1	2	2	3	3	4	4
MeAHA	4	4	3	3	2	2	1	1	0



$Zr(C_2O_4)_n$	4	3	3	3	2	2	2	1	1	1	0	0	0
$Zr(MeAHA)_n$	0	0	0	1	1	1	2	2	2	3	3	3	4
$Zr(C_2O_4H)_n$	0	1	0	0	1	0	0	1	0	0	1	0	0

MeAHA	0	1	0	0	0	0	0	0	0	0	0	0	0
C_2O_4	0	0	1	0	1	1	0	1	1	0	1	1	0
HMeAHA	4	3	4	3	3	3	2	2	2	1	1	1	0
C_2O_4H	0	0	0	1	0	1	2	1	2	3	2	3	4

Figure 4.6: DFT calculated reaction coordinate of the relative solvated free energy in the conversion of $[Zr(C_2O_4)_4]^{4-}$ into $Zr(MeAHA)_4$ under simulated basic (top) and acidic (bottom) conditions by Holland and co-workers.[18] The upper table insets show the number of coordinated ligands in the Zr(IV) complex for that arrangement, the lower inset shows the uncoordinated ligands accounting for stoichiometric equivalence between arrangements. Above each energy level is the relative change in Gibbs energy/kJ mol^{-1} to the starting arrangement of $[Zr(C_2O_4)_4]^{4-}$ in aqueous solution. Calculations carried out at B3LYP/LANL2DZ/IEFPCM.

4.1.3 Aim of this work

The aim of this work is to model zirconium oxinate within an environment designed to imitate the inside of a liposome used to transport doxorubicin sulfonate. This requires calculating the Gibbs energies of arrangements of successive ligand substitutions of oxinate with a fragment of doxorubicinate sulfonate (HDox) to understand the overall ΔG of the reaction coordinate as the zirconium moves from being complexed by oxinate to Dox. These calculations will be investigated under simulated basic, neutral, and acidic pHs, also investigating the potential of 6-coordinate vs. 8-coordinate zirconium with two coordinated water molecules. Finally, understanding the consequences of the ΔG values will be realised *via* Boltzmann population distribution calculations, and making tentative statements about the kinetic feasibility *via* the Eyring-Polanyi equation.

4.2 Results and Discussion

4.2.1 Method

Density functional theory was used for all calculations with the Gaussian 16 program (version C.01).[20] Structures were geometry optimised without symmetry constraints, in aqueous solution using the SMD self-consistent reaction fields.[21] Each optimised structure was confirmed as a local minima on the PES *via* the presence of no imaginary frequencies. DFT calculations were undertaken with the B3LYP, M06-2X, ωB97xd, or PW6B95D3 functional, and either 6-31G(d,p), LANL2DZ, LANL2TZ, DGDZVP, Def2-TZVP, or Sapporo-TZP-2012 basis sets (sourced from the online Basis Set Exchange) for all atoms.[22–30] Mixed basis sets were used *via* the 'gen' keyword, with the Sapporo-TZP-2012 basis set for Zr and 6-31G(d,p) for ligand atoms. In the cases where first order SCF failed to converge, quadratically convergent SCF was employed (*via* the 'scf = xqc' keyword). Default values for temperature and pressure were used unless otherwise noted (298.15 K, 1 atm). The default integration grid (UltraFine, 99 radial shells, 590 angular points per shell) was used throughout.

4.2.2 Benchmarking each model chemistry

In order to better understand the identity of the zirconium species within the liposome, density functional theory calculations were undertaken to find the Gibbs energy of these possible Zr(IV) complexes. To accurately find these energies, accurate geometry optimisations were needed.[19] This required benchmarking of both the basis set and the functional's ability to reproduce atomic bond lengths and angles from a crystal structure. The crystal structure chosen was of $Zr(oxinate)_4$, which would be the starting point of all of the substitution reactions towards any Zr-doxorubicinate complexes.[31]

4.2.2.1 Chosen basis sets and functionals to be evaluated

First to be decided was the basis set required. In the DFT calculations of $Zr_4(methacrylate)_{12}$ by Kreutzer and co-workers, they recommended that at least a triple zeta basis set should be used to accurately describe the geometry of Zr_4 clusters.[32] Holland and co-workers tested nine different basis sets: LANL2DZ, DGDZVP, CEP-4G, CEP-31G, CEP-121G, SDDALL, LANL2DZ/6-31+G(d), Sapporo-DZP/6-31+G(d), and Sapporo-TZP/6-31+G(d). Their best RMSE result came with the Sapporo-TZP/6-31+G(d), and they used the double zeta basis sets LANL2DZ and DGDZVP in their study.

Taking these results into consideration, the chosen basis sets for benchmarking were LANL2DZ, LANL2TZ (a triple zeta basis set in the same family as the widely used LANL2DZ), DGDZVP, Def2-TZVP, and the updated Sapporo-TZP-2012 basis set for Zr and 6-31G(d,p) for the ligand atoms. These follow the suggested basis sets from Holland and co-workers, whilst updating them and adding the very popular Def2-TZVP basis set by Truhlar and co-workers.[30]

For the functional, the three functionals in Chapter 3 were used as well as ωB97xd, which scored well in Mardirossian and co-workers' study of 200 DFT functionals (13th out of 200 for 'easy thermochemistry'; B3LYP (79/200), M06-2X (10/200), PW6B95D3 (3/200)) and used throughout our research group.[33]

4.2.2.2 Evaluation parameters

Root mean square error (RMSE), mean absolute error (MAE), and continuous shape measurement (CShM) were chosen to analyse the optimised geometries, Equation 4.2, Equation 4.3, and Equation 4.4.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{(x_{XRD,i} - x_{DFT,i})^2 + (y_{XRD,i} - y_{DFT,i})^2 + (z_{XRD,i} - z_{DFT,i})^2}\right)^2}$$

Equation 4.2: Equation for RMSE between crystal structure (XRD) and DFT geometry optimisation (DFT), where n is the total number of atoms.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|r_{XRD,i} - r_{DFT,i}|$$

Equation 4.3: Equation for MAE between crystal structure (XRD) and DFT geometry optimisation (DFT), where n is the total number of bonds, and r is the bond length.

$$CSHM(P) = \min\left(\frac{\sum_{i=1}^{n}|\vec{Q}_i - \vec{P}_i|^2}{\sum_{i=1}^{n}|\vec{Q}_i - \vec{Q}_0|^2}\right) \times 100$$

Equation 4.4: Equation for CSHM for reference geometry P, where $\vec{Q}_0$ is the coordinate vector of the centre of the complex under study, for n donors in the complex.

MAE examines the difference between bond lengths, therefore it is sensitive to the size of the complex as it is a 1-dimensional measure of error. For the MAE analysis, the deviation for heavy atom bonds (without carbon-hydrogen bonds) and Zr-donor bonds were calculated. RMSE is more sensitive to angle deformation as it is a 3-dimensional measure of error, concerned with the positioning of the atoms in space. CShM is a measurement of a 3-dimensional property, as it calculates a score for how well the geometry compares to a reference polyhedron – this means that it is completely blind to size, it only measures geometry. In addition to these conceptual properties of the geometry optimisation, the computational cost was also used to determine whether or not a basis set or functional would be practical for the project. CShM was calculated using the triangular dodecahedron geometry, D_{2d}, a stable geometry seen in Holland's DFT study of $Zr(MeAHA)_4$ and compared to the CShM of the $Zr(oxinate)_4$ XRD structure.[18,31] For CShM analysis, typically anything below 3 can be used as an acceptable

descriptor of a geometry, with anything below 1 indicating very good agreement with the polyhedron.[34,35] The larger the system – and especially one where rigidities are introduced from the ligands (i.e. Zr with 4 bi-dentate aromatic ligands) – the harder it will be to match the reference polyhedron. Thus, a triangular dodecahedron CShM of 1.337, Table 4.1, indicates that it is a very suitable description of the geometry of $Zr(oxinate)_4$).

Table 4.1: Comparison of error parameters for the DFT optimised geometries of $Zr(oxinate)_4$ in aqueous solution for different basis sets and the B3LYP functional.

Basis Set[a]	Rel. opt. time	MAE (all)[b]/Å	MAE (Zr-Donor)/Å	MAE (Zr-N)/Å	MAE (Zr-O)/Å	RMSE/Å	CShM
LANL2DZ	1.0	0.135	0.144	0.113	0.169	0.534	1.755
LANL2TZ	1.4	0.115	0.194	0.226	0.155	0.474	1.499
DGDZVP	2.2	0.121	0.202	0.218	0.185	0.507	1.701
Def2-TZVP	15.4	0.106	0.182	0.241	0.091	0.470	1.626
Sapporo-TZP-2012/6-31G(d,p)	2.1	0.112	0.182	0.242	0.086	0.465	1.607
Experimental XRD[c]							1.337

[a] Calculated with the B3LYP functional. [b] Mean absolute error of the bonds between all atoms apart from C-H bonds. [c] Experimental single-crystal X-ray data from Cambridge Crystallography Data Centre deposition number 798048.

Table 4.2: Comparison of error parameters for the DFT optimised geometries of Zr(oxinate)4 in aqueous solution for different functionals with Sapporo-TZP-2012/6-31G(d,p).

Functional[a]	Rel. opt. time	MAE (all bonds)[b]/Å	MAE (Zr-Donor)/Å	MAE (Zr-N)/Å	MAE (Zr-O)/Å	RMSE/Å	CShM
B3LYP	1.0	0.112	0.182	0.242	0.086	0.465	1.607
M06-2X	1.3	0.093	0.101	0.121	0.076	0.671	1.356
ωB97xd	1.4	0.093	0.114	0.141	0.077	0.610	1.357
PW6B95D3	1.1	0.091	0.107	0.134	0.070	0.839	1.778
Experimental XRD[c]							1.337

[a] Calculated with the Sapporo-TZP-2012/6-31G(d,p) basis set. [b] Mean absolute error of the bonds between all atoms apart from C-H bonds. [c] Experimental single-crystal X-ray data from Cambridge Crystallography Data Centre deposition number 798048.

The basis sets were benchmarked with calculations using the B3LYP functional, Table 4.1. Def2-TZVP and Sapporo-TZP-2012/6-31G(d,p) provided geometries with the lowest error for RMSE and MAE all bonds, unsurprising considering the age of the LANL basis sets, and only double-zeta nature of LANL2DZ and DGDZVP in describing Zr. The closeness between Def2-TZVP and Sapporo-TZP-2012/6-31G(d,p) was also seen in their CShM for triangular dodecahedron (1.626 and 1.607) – LANL2TZ with B3LYP provided the closest to experiment at 1.499. Beyond the error metrics, the relative optimisation time was used to determine the basis set, with Sapporo-TZP-2012/6-31G(d,p) optimising 7 times faster than Def2-TZVP – a crucial metric to consider with the large size of the system under study.

Functional testing was then carried out, Table 4.2. Somewhat surprisingly, B3LYP managed to provide the best RMSE (0.465 Å) and a MAE that is very close to the higher power functionals (all bonds: 0.112 Å). It is let down by its treatment of the zirconium atom – seen in the poorer Zr-Donor and Zr-N scores. The expected best functional, PW6B95D3, did not perform as well in the 3-dimensional error metrics, with an RMSE of 0.839 Å and CShM of 1.778, indicating the bond lengths were accurate but the optimised geometry itself was not. ωB97xd and M06-2X provided near identical scores for their error metrics, relative optimisation time, and CShM; at this level of closeness the inherent error from Gaussian software will be more significant. The deciding factor was in fact the SCRF chosen for the experiment. The recommended SCRF method for calculating ΔG_{solv} in Gaussian is the SMD variation of IEFPCM by Truhlar and co-workers, which was parameterised using their Minnesota functionals (M05-2X).[21,36] Because of this, M06-2X was chosen under the assumption that as a better match with SMD, more accurate solution ΔG values would be calculated.

4.2.3 Model conditions

The real-world experiment that was modelled involves a liposome with an aqueous core, a 10^4 excess of doxorubicin to [^{89}Zr]Zr(oxinate)$_4$, and acidic pH. The identity of the Zr species once [^{89}Zr]Zr(oxinate)$_4$ passes into the liposome is unknown and understanding its identity was the aim of the experiment. Overall, the experimental conditions are too complex to simulate accurately. Numerous factors would have to be considered including; the aqueous solvent (and the ammonium sulfate gradient), the large size of doxorubicin, the multiple potential coordination sites on doxorubicin, rotations/conformations/isomerism, the huge excess of doxorubicin, or even considering the unknown and unexpected mechanisms and side reactions that could also be occurring simultaneously. This naturally leads to choosing approximations and limitations to feasibly carry out the study. One of which has already been disclosed – in this study the SMD solvent reaction field was used to approximate the effect of having an aqueous solvent. This significantly expedites the calculation as creating a cluster of explicit water molecules to model the first and second solvation spheres (and their interaction with bulk water) would be extremely prohibitively expensive – almost impossible at the chosen M06-2X/Sapporo-TZP-2012/6-31G(d,p) level with the resources available.

4.2.3.1 Approximations

4.2.3.1.1 Fragments

To minimise the computational cost of the study, the doxorubicin molecule was fragmented. Coordination of doxorubicin to zirconium was expected to be through the aldol unit on the side of the fragment with fewest substituents.[37,38] Therefore, a chemically relevant fragment of doxorubicin was chosen to keep the parts of the molecule that are most likely responsible for coordination, and also to keep the electronic system largely intact, Figure 4.7. This reduces the electron count from 286 to 140, simplifying the calculations that Gaussian undertakes in optimising a zirconium species with four chelated fragments. This fragment will from now on be referred to as HDox when protonated, and Dox when deprotonated; similarly, the oxine ligand will be referred to as HOx and Ox for brevity.

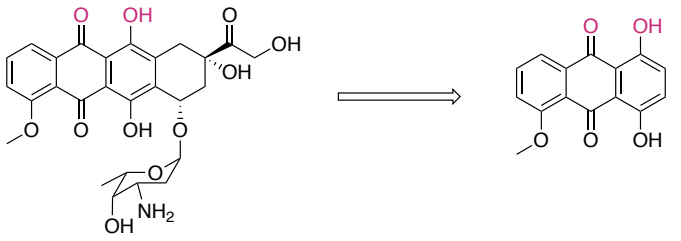


Figure 4.7: Fragment of doxorubicin, HDox, with binding aldol unit highlighted in pink.

4.2.3.1.2 Isomers/conformers

As with all applications of DFT geometry optimisations to a real-world problem, the global (or local) minimum may be within reach of nearby local minima (isomers or conformers) that are within an achievable energy gap. For a large system like this, there will be 10s-100s of conformations for each complex isomer. Sampling all isomers and conformers would be unaffordable for minimal benefit, as only a very small number will dominate due to steep Boltzmann weighting from having the lowest Gibbs energy. As such, the converged geometry provided by Gaussian was taken, whilst appreciating potentially non-negligible energy differences may occur.[39]

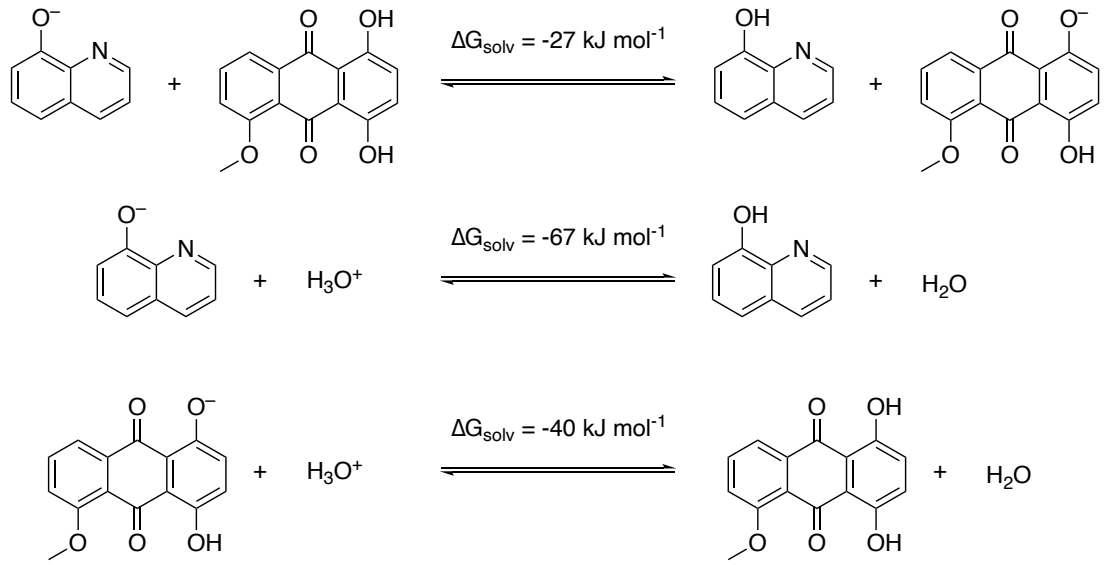
Isomerism is expected to provide non-negligible energy differences, however whilst comprehensively dealing with isomerisation would provide the most accurate picture, it would come at exorbitant computational cost. Thus, isomer sampling was conducted which varied the number of isomers considered with the possible number of isomers of the complex: $Zr(Ox)_4$ crystal structure (1 isomer considered), $[Zr(Ox)_3]^+$ (2 isomers considered), $Zr(Ox)_3(Dox)_1$ (2 isomers considered), $[Zr(Ox)_2(Dox)_1]^+$ (4 isomers considered), $Zr(Ox)_2(Dox)_2$ (7 isomers considered), $[Zr(Ox)_1(Dox)_2]^+$ (4 isomers considered), $Zr(Ox)_1(Dox)_3$ (2 isomers considered), $[Zr(Dox)_3]^+$ (2 isomers considered), and finally $Zr(Dox)_4$ (TDD-8, 1 isomer considered). This strategy allowed us to sensibly sample the isomers at each level, *and* with 7 $Zr(Ox)_2(Dox)_2$ isomers, investigate whether the difference in energy between

isomers was comparable to the difference between steps in the mechanism. Regardless of the attempts in sampling for the most stable isomer, in similar Zr systems ligand exchange and isomerisation is expected to be facile under experimental conditions.[40–42] Therefore, at the reaction energies available to the ligands and complexes, missing the desired lowest energy isomer is not expected to be detrimental to the outcome of the experiment, as the aim of the experiment is not to find which isomer is most stable, but to find out the relative positions of $Zr(Ox)_4$, $Zr(Ox)_3(Dox)_1$, $Zr(Ox)_2(Dox)_2$, $Zr(Ox)_1(Dox)_3$, $Zr(Dox)_4$.

4.2.3.1.3 Mechanisms and pH

The experiment was designed to start with $Zr(Ox)_4$ in a triangular dodecahedron geometry, with a stepwise ligand replacement of Ox with Dox to give $Zr(Dox)_4$ *via* a dissociative pathway due to instability above 8-coordinate for an associative mechanism,[43] and inherent difficulty in modelling an interchange pathway. In the conclusions of Holland's $Zr(MeAHA)_4$ study, the conversion of $Zr(C_2O_4)_4$ to $Zr(MeAHA)_4$ prefers a basic rather than acidic dissociative pathway.[18,43]

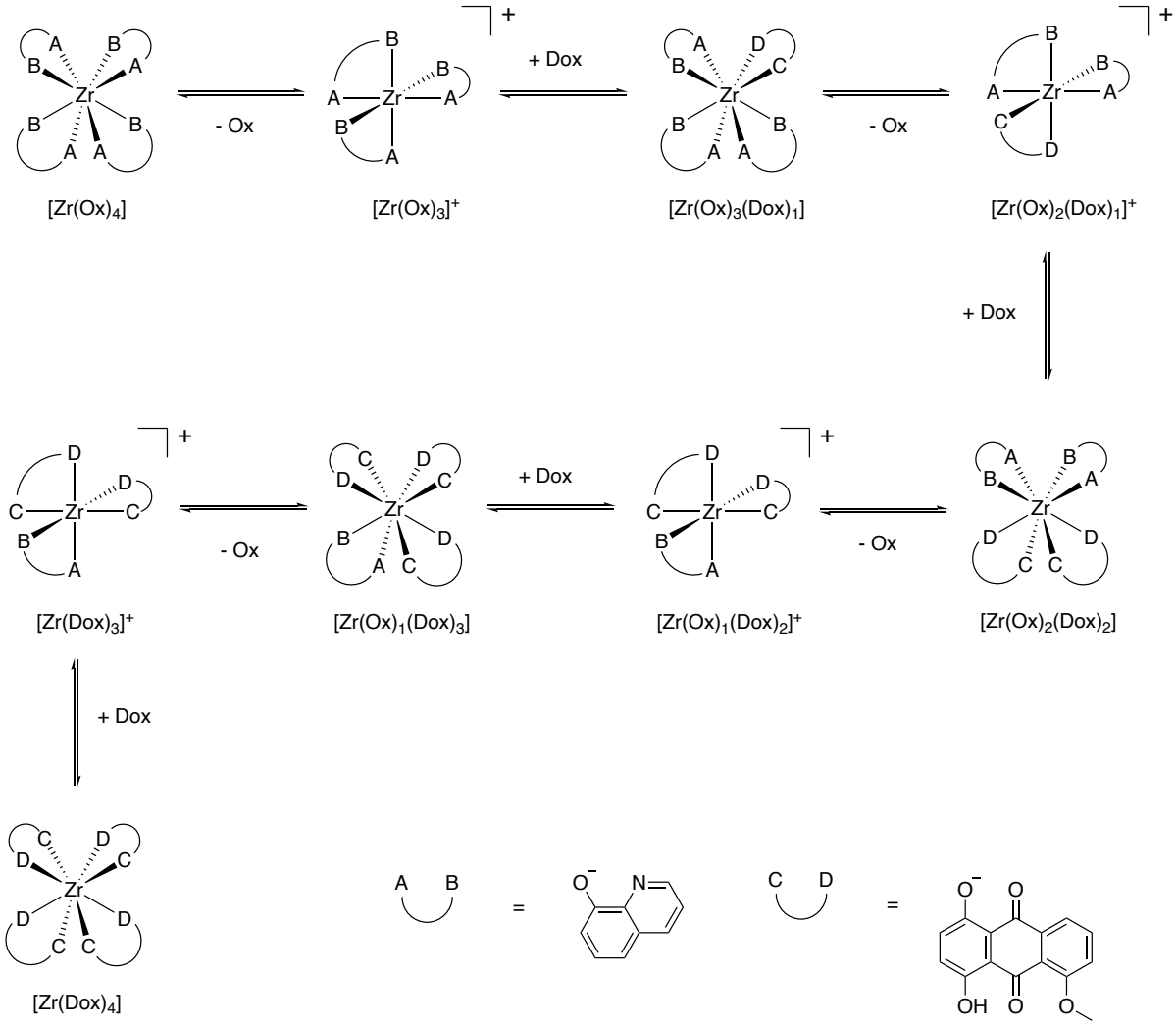
In the case of the liposome aqueous environment, the pH is expected to range from neutral to acidic (pH 5 within the liposome due to $(NH_4)_2SO_4$ buffer). Considering both of these factors, the DFT experiment was conducted under simulated basic, neutral, and acidic environments. As this entails dealing with different pH ranges, the dominant identity of each free ligand under simulated pH needed to be known. Therefore, the energies of proton transfer between HDox, Dox, HOx, and Ox was investigated, Scheme 4.1. Under simulated basic conditions, only Dox and Ox are available for complexation and existing in solution. Under neutral conditions, HOx and Dox will dominate, with excess HDox being consumed by free Ox (ΔG_{solv} = -27 kJ mol^{-1}). Therefore, it is unsurprising that with a free acid (simulated acidic environment), Ox has a greater stabilisation to be protonated than Dox (-67 vs. -40 kJ mol^{-1}). The conclusions of these calculations were also replicated in pKa prediction *via* Ji and co-workers' graph-convolutional neural network pKa predictor,[44] which provided pKa values of 7.8 for HOx, and 7.1 and 7.2 for HDox's two aldol groups. Under neutral conditions we would expect that any free Ox will be consumed to HOx by either the large excess of HDox or any H_3O^+, and under acidic conditions only HOx and HDox will be freely available in solution.



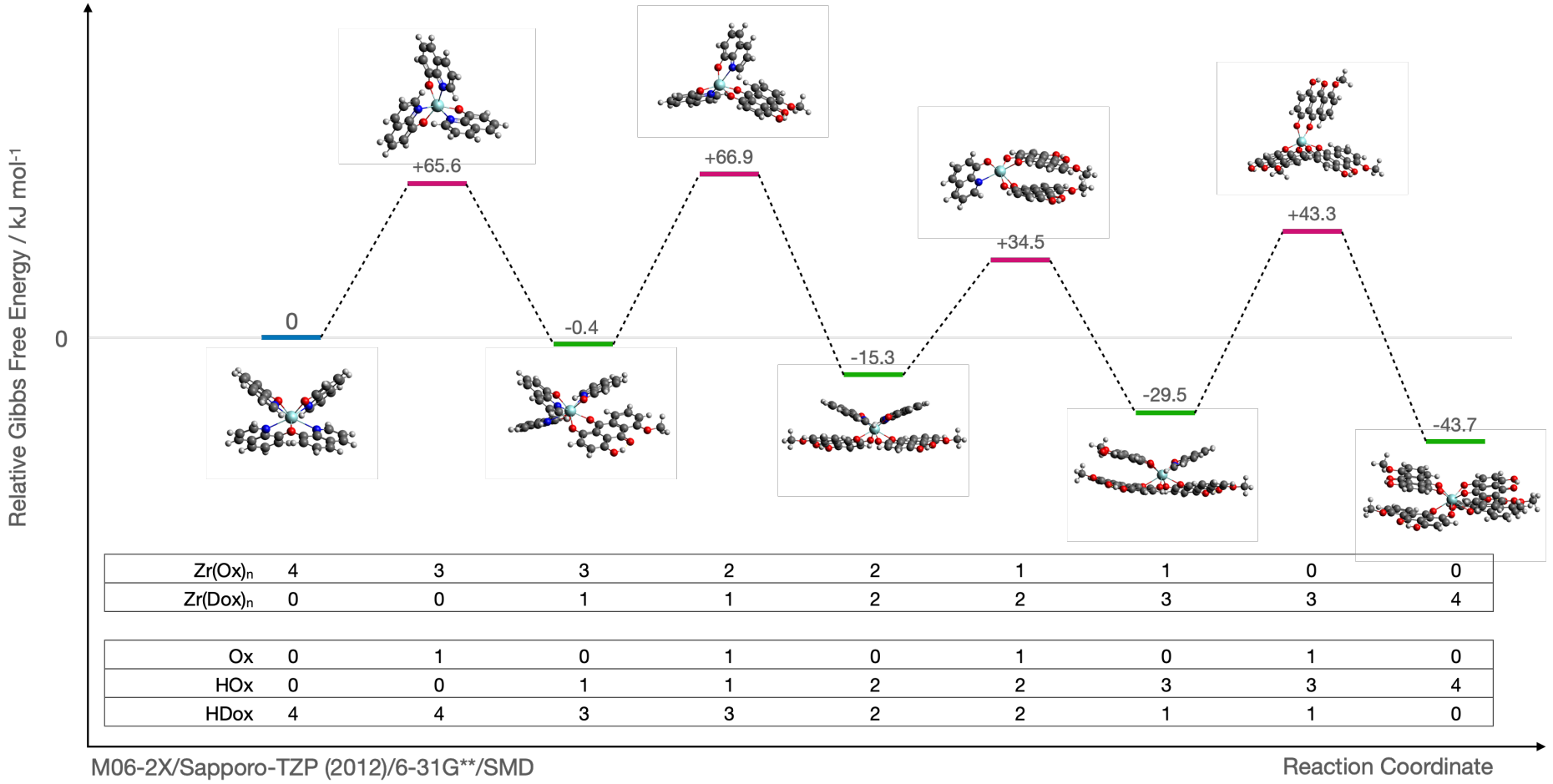
Scheme 4.1: DFT calculated protic equilibria of oxine (HOx) and the doxorubicin fragment (HDox), and their individual calculated ΔG_{solv} for their equilibria with H_3O^+. These calculations were used to aid in deciding reactants in the simulated conditions.

4.2.4 Reaction in a neutral environment

With the identity of the molecules available for the reaction, a reaction scheme for the neutral reaction was designed, Scheme 4.2. As per Scheme 4.1, Reaction 1, after every Ox dissociation, HDox will protonate Ox to HOx leaving a free Dox able to coordinate to the unsaturated zirconium species. Under this assortment of ligands and Zr(IV) species, Figure 4.8, every 6-coordinate zirconium complex arrangement was higher in energy than the starting complex $Zr(Ox)_4$: $[Zr(Ox)_3]^+$ +65.6 kJ mol^{-1}; $[Zr(Ox)_2(Dox)_1]^+$ +66.9 kJ mol^{-1}; $[Zr(Ox)_1(Dox)_2]^+$ +34.5 kJ mol^{-1}; $[Zr(Dox)_3]^+$ +43.3 kJ mol^{-1}. Once coordination was returned to 8 *via* Dox complexation, each successive substitution resulted in an increasingly stable complex – relative to $Zr(Ox)_4$: $Zr(Ox)_3(Dox)_1$ -0.4 kJ mol^{-1}; $Zr(Ox)_2(Dox)_2$ -15.3 kJ mol^{-1}; $Zr(Ox)_1(Dox)_3$ -29.5 kJ mol^{-1}; $Zr(Dox)_4$ -43.7 kJ mol^{-1}. Each substitution after $Zr(Ox)_3(Dox)_1$ contributes around 15 kJ mol^{-1} to stabilising the arrangement. Overall, the calculations suggest that under neutral conditions it is thermodynamically feasible for $Zr(oxinate)_4$ to fully react with doxorubicin and form $Zr(doxorubicin)_4$. It naturally then follows to ask what proportions these numbers translate to. As with finding the populations of the states in MRI in Chapter 1, the Boltzmann distribution is equally useful here in finding the distribution of Zr(IV) species within the liposome, Equation 4.5.



Scheme 4.2: Reaction scheme for the conversion of $Zr(Ox)_4$ to $Zr(Dox)_4$ via a dissociative pathway under either a simulated basic or neutral environment. In the neutral environment, Dox is generated via proton exchange between Ox and HDox as in Scheme 4.1, Reaction 1.



$Zr(Ox)_n$	4	3	3	2	2	1	1	0	0
$Zr(Dox)_n$	0	0	1	1	2	2	3	3	4

Ox	0	1	0	1	0	1	0	1	0
HOx	0	0	1	1	2	2	3	3	4
HDox	4	4	3	3	2	2	1	1	0

Figure 4.8: DFT calculated reaction coordinate of the relative solvated free energy in the conversion of $Zr(Ox)_4$ into $Zr(Dox)_4$ under simulated neutral conditions. The upper table inset shows the number of Ox or Dox ligands coordinated in the Zr(IV) complex for that arrangement, the lower inset shows the uncoordinated ligands Ox, HOx, and HDox accounting for stoichiometric equivalence between arrangements. Above each energy level is the relative change in Gibbs energy/kJ mol^{-1} to the starting arrangement of $Zr(Ox)_4$ and four HDox ligands in aqueous solution. All calculations at M06-2X/Sapporo-TZP-2012/6-31G(d,p)/SMD.

$$p_i = \frac{e^{-\frac{\Delta G_i}{RT}}}{\sum_j^n e^{-\frac{\Delta G_j}{RT}}}$$

Equation 4.5: Boltzmann probability distribution where p_i is the probability of state i, ΔG_i is the relative energy difference for state i to $Zr(Ox)_4$, T is 298.15 K, and n is the total number of states.

Table 4.3: Probability distribution of arrangements from Figure 4.8.

Complex	Probability
$Zr(Ox)_4$	2.16×10^{-8}
$[Zr(Ox)_3]^+$	7.01×10^{-20}
$Zr(Ox)_3(Dox)_1$	2.52×10^{-8}
$[Zr(Ox)_2(Dox)_1]^+$	4.13×10^{-20}
$Zr(Ox)_2(Dox)_2$	1.03×10^{-5}
$[Zr(Ox)_1(Dox)_2]^+$	1.97×10^{-14}
$Zr(Ox)_1(Dox)_3$	0.00318
$[Zr(Dox)_3]^+$	5.67×10^{-16}
$Zr(Dox)_4$	0.997

The results of the Boltzmann probability distribution, Table 4.3, suggest that in the simulated neutral conditions, and under thermodynamic control (where achieving all of the arrangements are kinetically feasible), there will be almost complete conversion of $Zr(Ox)_4$ to $Zr(Dox)_4$ (99.7%) with a minority of Zr(IV) complexed as $Zr(Ox)_1(Dox)_3$ (0.3%), with all of the other states not present in significant quantities.

The thermodynamics of a reaction only describes half of the situation, for true feasibility the activation energy of each step needs to be within the kinetic energy available to the molecules at experimental temperatures (298.15 K in the laboratory, or 310.15 K inside the body). For the reaction in the neutral environment dissociation consistently increased the energy of the system, raising the question whether an interchange mechanism would better describe the reaction in the neutral state – especially as the actual experiment operates with a 10^4 excess of doxorubicin. Locating the transition structures for these reactions would be beyond the affordable cost of a project of this size. However, for an idea of feasibility, the limit of activation energies for the reaction can be calculated and compared to the energies of the converged 6-coordinate dissociation product, *via* the Eyring-Polanyi equation – Equation 4.6.[45] At 298.15 K, using the 6-coordinate energy levels as proxy values for the activation energy of an interchange mechanism (largest forward step was +72.8 kJ mol^{-1}), the reactions would all

have half-lives of a magnitude below 10^1 seconds, Table 4.4. In reality, the interchange activation energies may not be at those levels, and assuming the dissociative mechanism is the dominant mechanism, activation energies at 298.15 K for either mechanism would need to remain below 100 kJ mol^{-1} (equivalent to a 10 hour half-life) to be feasible.

$$k = \frac{k_b T}{h}(c^0)^{1-m} e^{-\frac{\Delta G^{\ddagger}}{RT}}$$

Equation 4.6: Eyring-Polanyi equation from transition state theory for reaction rate,[45] where k_b is Boltzmann's constant, T is 298 K, h is Planck's constant, m is the number of reactants (2 for interchange), c^0 is the standard state concentration (1 M for these DFT calculations), and $\Delta G^{\ddagger}$ is the Gibbs activation energy.

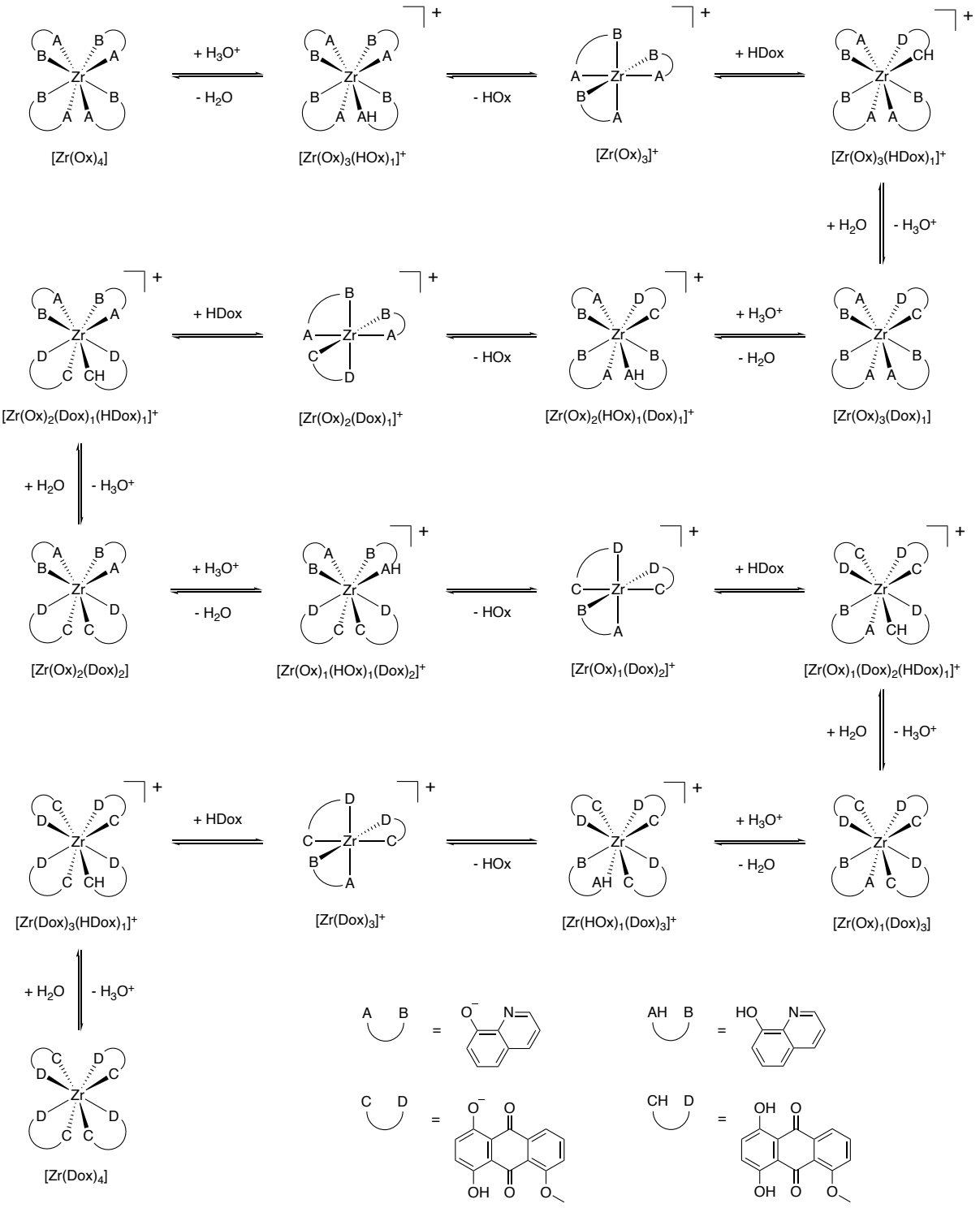
Table 4.4: Half-life values for reactions at different activation energies, as calculated via *Equation 4.6.*

$\Delta G^{\ddagger}$ kJ mol^{-1}	k/s^{-1}	$t_{1/2}$/s	$t_{1/2}$/minute
40	6.10×10^{5}	1.14×10^{-6}	1.89×10^{-8}
50	1.08×10^{4}	6.41×10^{-5}	1.07×10^{-6}
60	1.91×10^{2}	3.62×10^{-3}	6.04×10^{-5}
70	3.39	2.05×10^{-1}	3.41×10^{-3}
80	6.00×10^{-2}	11.6	1.93×10^{-1}
90	1.06×10^{-3}	6.53×10^{2}	1.09
100	1.88×10^{-5}	3.69×10^{4}	6.15×10^{2}
110	3.33×10^{-7}	2.08×10^{6}	3.47×10^{4}
120	5.89×10^{-9}	1.18×10^{8}	1.96×10^{6}

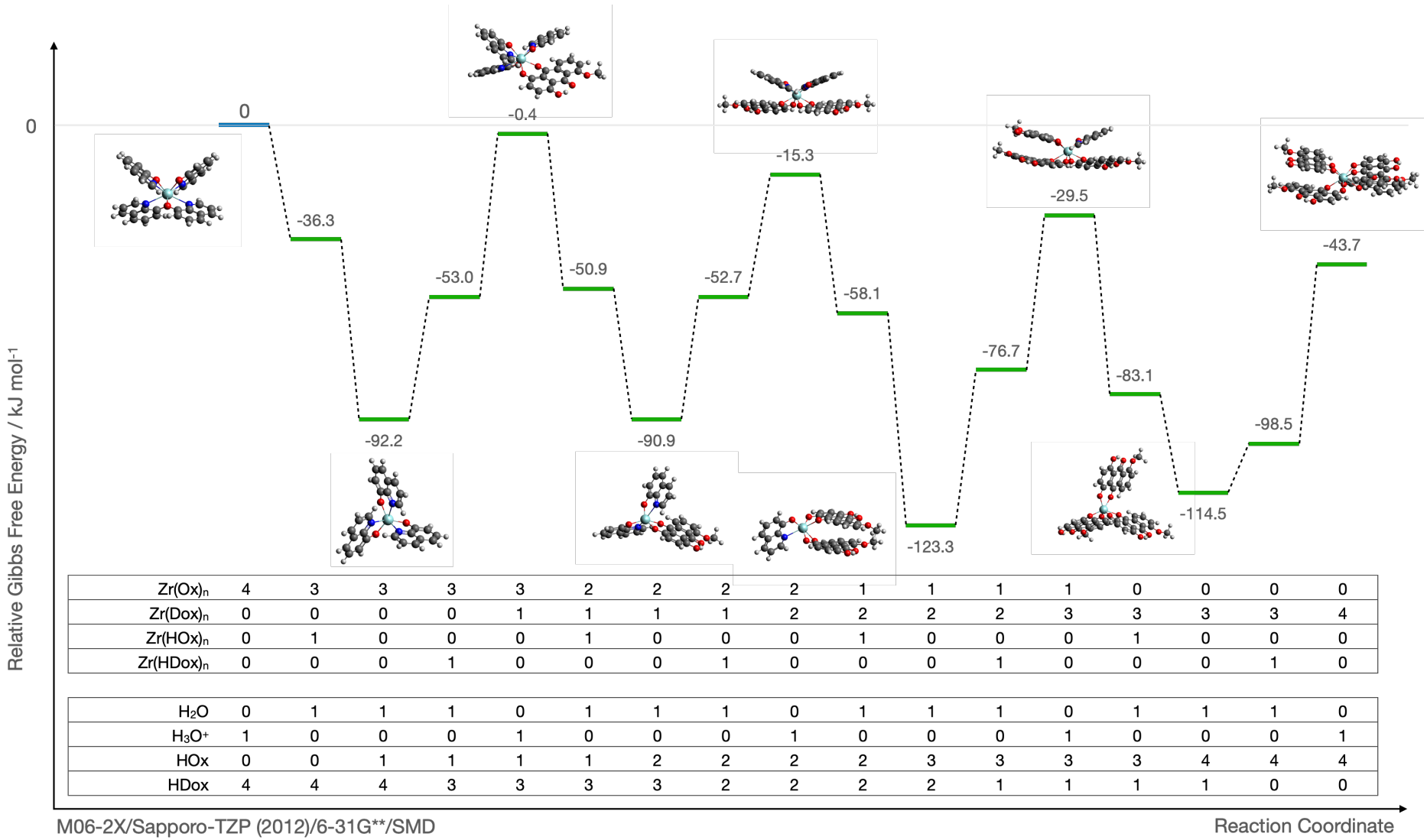
4.2.5 Reaction in an acidic environment

As the liposomes are formed with a pH 5 buffer, the acidic intraliposmal pH would be expected to be low enough to protonate both Ox and Dox. Therefore, the reaction mechanism was modified to accommodate free acid and only protonated ligands, Scheme 4.3. Under these conditions, the coordinated Ox is protonated by proton transfer from H_3O^+ and then decomplexes, allowing HDox to coordinate, which then transfers its proton to coordinated Ox *via* H_2O to begin the ligand substitution cycle again.

Geometry optimisations and frequency calculations were undertaken on the structures from Scheme 4.3 to provide the Gibbs free energy values in Figure 4.9. Interestingly, in this simulated acidic environment, the 6-coordinate species were, on the whole, more stable than the 8-coordinate species (-105.2 kJ mol^{-1} average vs. -45.8 kJ mol^{-1} average). These 8-coordinate species had the same stabilisation as in the neutral case relative to the $Zr(Ox)_4$ species, however once either of Ox or Dox were protonated, this led to a stabilisation of the system, with subsequent decomplexation leading to a further stabilisation. In this case, the $[Zr(Ox)_1(Dox)_2]^+$ complex provided the largest energy well with a value of -123.3 kJ mol^{-1} relative to $Zr(Ox)_4$ – this was 24.8 kJ mol^{-1} more stable than the most stable 8-coordinate species, $[Zr(Dox)_3(HDox)_1]^+$. This is shown in the probabilities from the Boltzmann distribution, Table 4.5. $[Zr(Ox)_1(Dox)_2]^+$ dominates at 97.2%, with $[Zr(Dox)_3]^+$ as a minor component at 2.8% (energy gap of 8.8 kJ mol^{-1}), all other states are essentially zero – emphasising the steep Boltzmann weights discussed in Section 4.2.3.1.2. Therefore, in the simulated acidic conditions modelled, $Zr(Ox)_4$ is thermodynamically unfavourable and would likely convert to $[Zr(Ox)_1(Dox)_2]^+$ and a minority as $[Zr(Dox)_3]^+$.



Scheme 4.3: Reaction scheme for the conversion of $Zr(Ox)_4$ to $Zr(Dox)_4$ via a dissociative pathway under an acidic environment.



$Zr(Ox)_n$	4	3	3	3	3	2	2	2	2	1	1	1	1	0	0	0	0
$Zr(Dox)_n$	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3	4
$Zr(HOx)_n$	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
$Zr(HDox)_n$	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0

H_2O	0	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	0
H_3O^+	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
HOx	0	0	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4
HDox	4	4	4	3	3	3	3	2	2	2	2	1	1	1	1	0	0

Figure 4.9: DFT calculated reaction coordinate of the relative solvated free energy in the conversion of $Zr(Ox)_4$ into $Zr(Dox)_4$ under simulated acidic conditions. The upper table inset shows the number of Ox, Dox, HOx, or HDox coordinated in the Zr(IV) complex for that arrangement, the lower inset shows the uncoordinated ligands oxine (HOx), or doxorubicin fragment (HDox), as well as free H_2O and H_3O^+ accounting for stoichiometric equivalence between arrangements. Alongside each energy level is the relative change in Gibbs energy/kJ mol^{-1} to the starting arrangement of $Zr(Ox)_4$ in aqueous solution. All calculations at M06-2X/Sapporo-TZP-2012/6-31G(d,p)/SMD.

Table 4.5: Probability distribution of arrangements from Figure 4.9, for $Zr(Ox)_4$, $Zr(Dox)_4$, and any arrangement with relative energy of -90 kJ mol^{-1} or below.

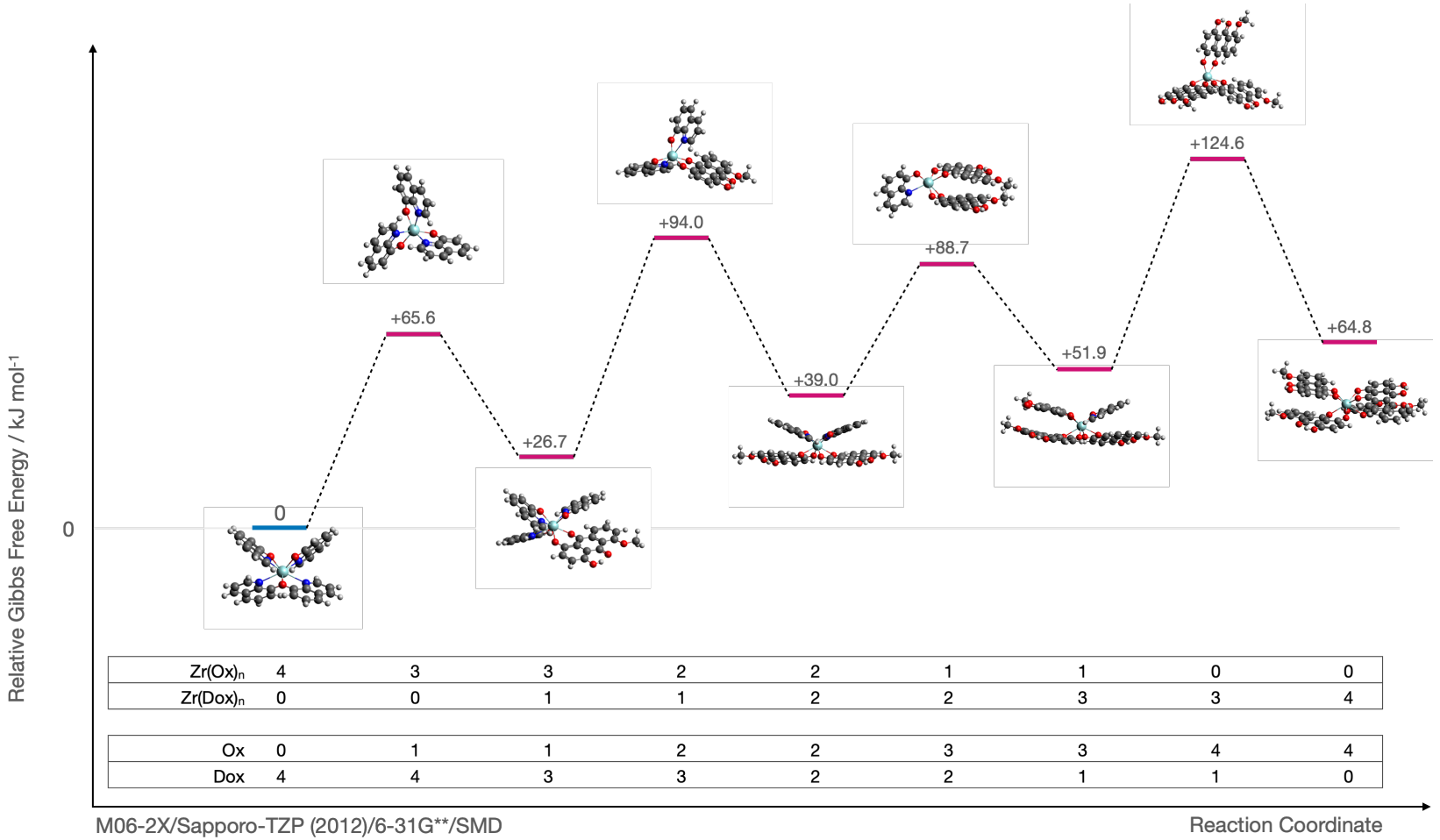
Complex	Probability
$Zr(Ox)_4$	2.46×10^{-22}
$[Zr(Ox)_3]^+$	3.46×10^{-6}
$[Zr(Ox)_2(Dox)_1]^+$	2.04×10^{-6}
$[Zr(Ox)_1(Dox)_2]^+$	0.972
$[Zr(Dox)_3]^+$	0.028
$[Zr(Dox)_3(HDox)_1]^+$	4.41×10^{-5}
$Zr(Dox)_4$	1.13×10^{-14}

4.2.6 Reaction in a basic environment

Finally, the basic scenario was calculated *via* the mechanism in Scheme 4.2, resulting in the energies in Figure 4.10. In a simulated basic environment only the deprotonated Ox and Dox ligands are available; the Zr(IV) complex progressively removes Ox and replaces with Dox. In this case, the $Zr(Ox)_4$ species dominates the reaction, and the progressive substitution of Ox with Dox is thermodynamically unfeasible (+26.7, +39.0, +51.9, +64.8 kJ mol^{-1} for each Dox substitution vs. $Zr(Ox)_4$). This deep energy well for $Zr(Ox)_4$ corresponds to a 99.998% distribution, with its nearest neighbour $Zr(Ox)_3(Dox)_1$ making up the rest at 0.002% (all others at a magnitude of 10^{-7} or below).

The preference for $ZrOx_4$ may not be surprising, considering Ox is more basic than Dox (pK_a prediction 7.8 vs 7.2;[44] and in Scheme 4.1, reaction 1, there is a -27 kJ mol^{-1} reduction in Gibbs energy for it to remove a proton from HDox). In the other two environments, the other molecules in the arrangement augment the energetics to favour substitution.

Across the three models it appears that the ligand substitution to form a desired Zr-doxorubicin complex is pH dependent, increasing in doxorubicin coordination as the pH falls to neutral, and then settling at $[Zr(Ox)_1(Dox)_2]^+$ at low pH.



$Zr(Ox)_n$	4	3	3	2	2	1	1	0	0
$Zr(Dox)_n$	0	0	1	1	2	2	3	3	4

Ox	0	1	1	2	2	3	3	4	4
Dox	4	4	3	3	2	2	1	1	0

Figure 4.10: DFT calculated reaction coordinate of the relative solvated free energy in the conversion of $Zr(Ox)_4$ into $Zr(Dox)_4$ under simulated basic conditions. The upper table inset shows the number of Ox or Dox ligands coordinated in the Zr(IV) complex for that arrangement, the lower inset shows the uncoordinated ligands Ox, and Dox accounting for stoichiometric equivalence between arrangements. Above each energy level is the relative change in Gibbs energy/kJ mol^{-1} to the starting arrangement of $Zr(Ox)_4$ and four Dox ligands in aqueous solution. All calculations at M06-2X/Sapporo-TZP-2012/6-31G(d,p)/SMD.

4.2.7 Isomerisation

To gauge the effect that isomerisation around the zirconium centre could have on the energy levels of the arrangements, seven isomers of $Zr(Ox)_2(Dox)_2$ were optimised. These investigated the effect on Gibbs energy of repositioning the Dox ligands in a triangular dodecahedron input geometry – skeletal diagrams of the input geometries are shown in Figure 4.11. The Boltzmann probability distribution of the seven isomers were calculated, Table 4.6, and showed a standard deviation of 10.6 kJ mol^{-1}. The probabilities in Table 4.6 suggest that each of the zirconium species will have non-negligible isomers that are within a small enough energy range to co-exist in solution – providing *computational* evidence that the conversion between isomers would indeed be possible and expected.[40–42] A 10.6 kJ mol^{-1} standard deviation is an acceptable size, as each state is within reach of the other (all probabilities are above 0.01), therefore the reaction coordinate will proceed in Figure 4.8, Figure 4.9, and Figure 4.10, even in the worst case scenario where the highest energy isomer was optimised by Gaussian.

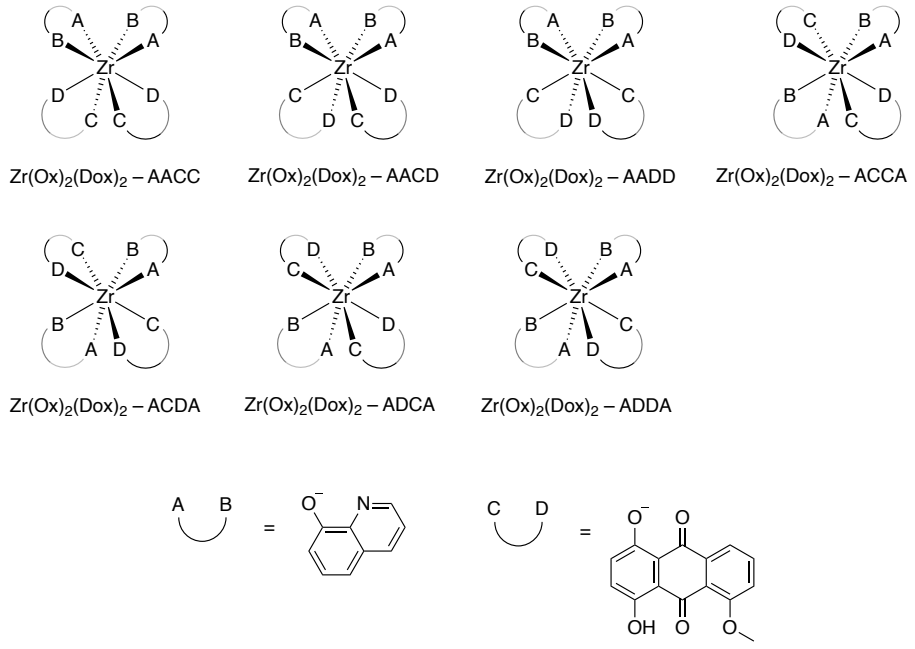


Figure 4.11: $Zr(Ox)_2(Dox)_2$ isomers to investigate the effect on Gibbs energy of repositioning the Dox ligands in a triangular dodecahedron input geometry, naming convention following the donor to Zr in the central vertical plane, starting with the top right, then top left, then bottom right and finally back left.

Table 4.6: Boltzmann probability distribution of the seven $Zr(Ox)_2(Dox)_2$ isomers analysed.

$Zr(Ox)_2(Dox)_2$ Isomer	Probability
AACC	0.54
AACD	0.14
AADD	0.04
ACCA	0.15
ACDA	0.03
ADCA	0.06
ADDA	0.05

4.2.8 Saturating the 6-coordinate zirconium complexes

In the simulated neutral and basic environments, the 6-coordinate Zr(IV) species were all higher in energy than the 8-coordinate species, Figure 4.8 and Figure 4.10. Continuous shape measure analysis was carried out on the four complexes, and apart from $[Zr(Dox)_3]^+$ (trigonal prism, 0.904), the other complexes had unfavourable CShM scores (> 3.00) for the standard 6-coordinate geometries (hexagon, pentagonal pyramid, octahedron, trigonal prism, or Johnson pentagonal pyramid (equal sides pentagonal pyramid). In an attempt to stabilise their energies, two water molecules were coordinated to the four complexes and geometry optimised, Figure 4.12.

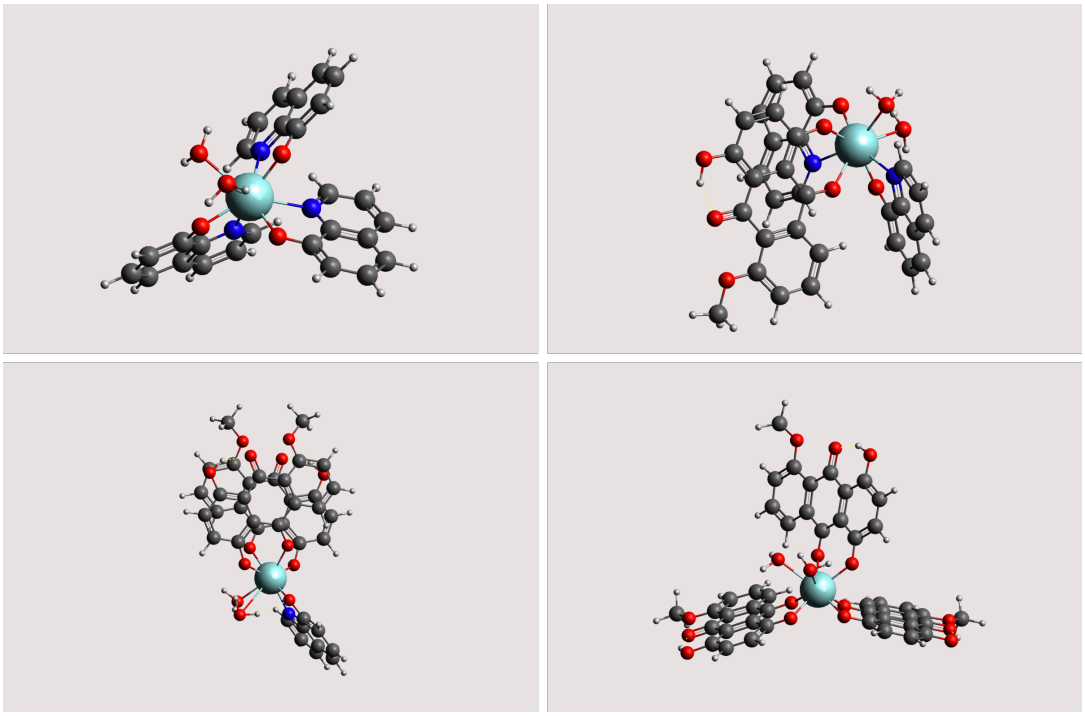


Figure 4.12: Optimised geometries of the previously 6-coordinate Zr(IV) species with two water molecules complexed. Top left: $[Zr(Ox)_3(H_2O)_2]^+$; top right: $[Zr(Ox)_2(Dox)_1(H_2O)_2]^+$; bottom left: $[Zr(Ox)_1(Dox)_2(H_2O)_2]^+$; bottom right: $[Zr(Dox)_3(H_2O)_2]^+$.

Table 4.7: CShM of the previously 6-coordinate Zr(IV) species with two water molecules complexed, calculated for square anti-prism, triangular dodecahedron, Johnson biaugmented trigonal prism, and Johnson snub diphenoid.

Complex	Square anti-prism	Triangular dodecahedron	Johnson biaugmented trigonal prism	Johnson snub disphenoid
$[Zr(Ox)_3(H_2O)_2]^+$	3.672	1.073	2.453	2.310
$[Zr(Ox)_2(Dox)_1(H_2O)_2]^+$	2.380	1.171	3.310	5.603
$[Zr(Ox)_1(Dox)_2(H_2O)_2]^+$	5.547	3.982	4.582	5.314
$[Zr(Dox)_3(H_2O)_2]^+$	1.084	2.179	2.726	4.945

These complexes optimised to far more regular geometries, Table 4.7 – $[Zr(Ox)_1(Dox)_2(H_2O)_2]^+$ only bound to one water and gave a CShM of 1.696 for capped octahedron. Nonetheless as seen in Figure 4.9, once explicit water is considered for the overall arrangement, the coordination of two water molecules leads to an overall destabilisation for all four complexes' arrangements, Table 4.8. This phenomena is also seen in Holland's study into $Zr(DFO)(H_2O)_x$ complexes,[46] where it was estimated *via* DFT calculation (at B3LYP/DGDZVP/IEFPCM) that $[Zr(DFO)(H_2O)_2]^+$ would have a smaller $\log(K_{ZrL})$ of 39.89 compared to 41.51 and 41.20 for $[Zr(DFO)(H_2O)]^+$ and $[Zr(DFO)]^+$. So, despite their unconventional geometry, the 6-coordinate Zr(IV) species are more stable as 6-coordinate than 8-coordinate with water.

Table 4.8: Table of the energy change between the arrangements without complexed water (but 2 water molecules within solution) and with two complexed water molecules for the 6-coordinate complexes.

Complex	Change in Gibbs energy upon hydration/kJ mol^{-1}
$[Zr(Ox)_3(H_2O)_2]^+$	+51.6
$[Zr(Ox)_2(Dox)_1(H_2O)_2]^+$	+37.0
$[Zr(Ox)_1(Dox)_2(H_2O)_2]^+$	+12.7
$[Zr(Dox)_3(H_2O)_2]^+$	+52.0

4.2.9 Increasing the temperature to 310.15 K

So far, all of the optimisations and frequency calculations were carried out at 298.15 K to match the introduction of $Zr(oxinate)_4$ into the liposome in the laboratory. Frequency calculations were undertaken with the keyword 'temperature=310.15' to match internal body temperature to investigate the difference in Boltzmann population distribution. For all of the optimisations (ligands and complexes), the frequency calculation showed that the optimised structure from 298.15 K was also a minimum on the PES for 310.15 K, implying that despite these calculations not being from reoptimised structures, the

structures will represent local (or perhaps global) minima on the PES. The change in Boltzmann population distribution for the anhydrous (6-coordinate Zr species included) and hydrated (6-coordinate species are brought to 8 *via* coordination of two water molecules) complexes is shown in Table 4.9 and Table 4.10. The calculations suggest that there would be almost no difference in distribution due to temperature whether inside the body or outside. For the neutral and basic environments, the changes were almost negligible for both the anhydrous and hydrated complexes. The acidic environment showed a swing of 6.2% from $[Zr(Ox)_1(Dox)_2]^+$ to $[Zr(Dox)_3]^+$ (2.8% to 9.0%) for the anhydrous complexes, and a swing of 4.6% from $[Zr(Ox)_1(Dox)_2]^+$ to $[Zr(Dox)_3(HDox)_1]^+$ (0.7% to 5.3%) for the hydrated complexes. Across the six scenarios, there was not a large effect from the 12 K increase of temperature between 298.15 K and 310.15 K.

Table 4.9: Change in Boltzmann distribution probabilities between 298.15 K and 310.15 K calculations for the anhydrous scenario; data shown for acidic, neutral, and basic environments. Complexes where the change in probability was below 0.001 for all three environments were omitted.

Zr Complex	Acidic	Neutral	Basic
$[Zr(Ox)_1(Dox)_2]^+$	-0.062	0.000	0.000
$Zr(Ox)_1(Dox)_3$	0.000	-0.001	0.000
$[Zr(Dox)_3]^+$	+0.062	0.000	0.000
$Zr(Dox)_4$	0.000	+0.001	0.000

Table 4.10: Change in Boltzmann distribution probabilities between 298.15 K and 310.15 K calculations for the hydrated scenario; data shown for acidic, neutral, and basic environments. Complexes where the change in probability was below 0.001 for all three environments were omitted.

Zr Complex	Acidic	Neutral	Basic
$[Zr(Ox)_1(Dox)_2]^+$	-0.046	0.000	0.000
$Zr(Ox)_1(Dox)_3$	0.000	-0.001	0.000
$[Zr(Dox)_3]^+$	0.000	0.000	0.000
$[Zr(Dox)_3(HDox)_1]^+$	+0.046	-	-
$Zr(Dox)_4$	0.000	+0.001	0.000

4.2.10 Limitations

As the DFT experiment can only model within reasonable constraints, there are certain aspects which deviate from reality. Most obvious in this case is the scale of the problem: in our experiment we had equal numbers of Ox ligands to Dox ligands, however in the liposome, $Zr(oxinate)_4$ is at an appropriate concentration for PET studies, and therefore, the corresponding doxorubicin available is 1000s of times more plentiful. The extreme excess of doxorubicin in the liposome will likely push the equilibria to

move away from $Zr(Ox)_4$ and towards the Dox containing Zr(IV) species. This couples with the next alteration to reality – oxine's solubility is pH dependent, and as such in the neutral and acidic environments, the reaction coordinate will tend towards the direction of $Zr(Dox)_4$ as more and more oxine is permanently removed from the system. Both of these effects help the kinetics of the thermodynamically feasible reaction.

Another consideration that needs to be made in translating the results of the *in silico* experiment to the laboratory is that the model does not factor in any of the hydrophobic neutral species ($Zr(Dox)_4$, $Zr(Ox)_1(Dox)_3$, $Zr(Ox)_2(Dox)_2$, $Zr(Ox)_3(Dox)_1$, $Zr(Ox)_4$) diffusing out of the liposome. The acidic nature of the liposome favours the formation of charged Zr species – the five species with the largest probabilities in the Boltzmann distribution are all charged, with $[Zr(Ox)_1(Dox)_2]^+$ dominating at (97.2%). This suggests that (as long as the charged species can be formed before neutral species are able to diffuse out), the pH 5 buffered liposome will most likely retain the zirconium complex.

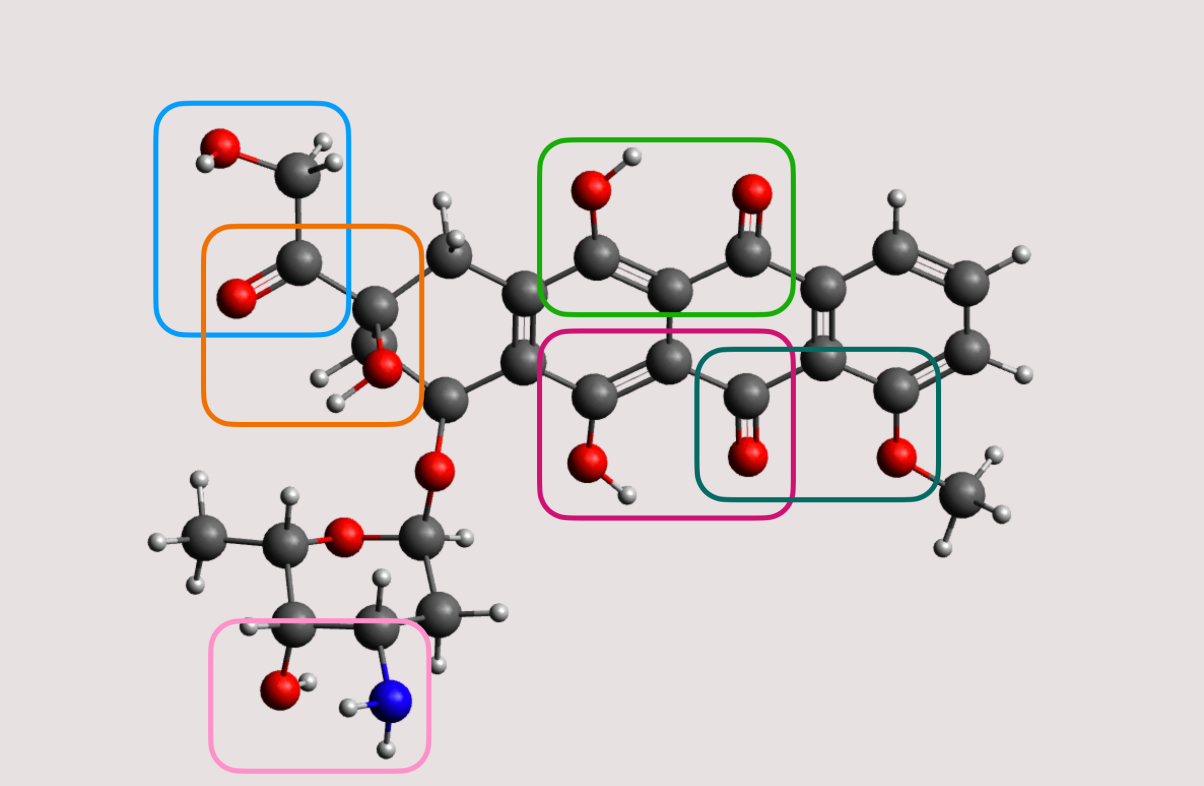


Figure 4.13: Possible coordinating groups on doxorubicin that could coordinate to zirconium; the aldol in the light green box (top, centre) was used in the DFT experiment.

Despite evidence suggesting the chosen aldol coordinating group is the one primarily seen in experiment,[37,38] there is still the possibility of donors from different parts of doxorubicin, Figure 4.13. Understanding the most energetically favourable donor could be accomplished by optimising $Zr(Dox)_4$

with each donor, then a Boltzmann population distribution will reveal their relative amounts. If any of them apart from the aldol chosen for the DFT study are energetically accessible, accurately modelling the system would then require an exponential increase in isomer consideration – mixed donors in complexes with multiple Dox ligands and flipping the Dox (as was considered in the DFT study) at each donor site. Evaluating the system with multiple coordinating groups quickly becomes unaffordable, and for this reason, other coordinating groups on doxorubicin were not considered.

In addition to this, a fragmented version of doxorubicin was used which limits the study as steric effects of the pyranose ring are not considered and may over-signify the steric effect of the non-coordinating aromatic ring on Dox. However, the decision to remove these non-aromatic parts of doxorubicin, whilst simply reducing the electron count, also reduces the amount of rotational freedom that the fragment has which would assist in reducing the complexity in the geometry optimisation.

4.3 Conclusion

The ligand substitution reaction between $Zr(oxinate)_4$ and doxorubicin in aqueous solution was modelled using DFT, replicating conditions within a liposome, as closely as possible. An appropriate basis set and functional was chosen *via* benchmark testing a selection of basis sets and functionals and their success in replicating the geometry of a crystal structure of $Zr(oxinate)_4$. The functional and basis set combination that was chosen was M06-2X/Sapporo-TZP-2012/6-31G(d,p) due to its combination of low 1D and 3D errors, relatively low computational cost, and suitability towards the desired SCRF method (SMD). The experiments were conducted in simulated acidic, neutral, and basic environments, with geometry optimisation and frequency calculations undertaken to provide Gibbs free energy values for each step of the ligand substitution between $Zr(Ox)_4$ and Dox (a fragment of doxorubicin) *via* a dissociative mechanism. For the neutral and acidic environments, ligand substitution of Dox for Ox in $Zr(Ox)_4$ was thermodynamically feasible, with Boltzmann distribution calculations suggesting that the neutral scenario would be dominated by $Zr(Dox)_4$ (99.7%), and the acidic dominated by $[Zr(Ox)_1(Dox)_2]^+$ (97.2%). Conversely, in the basic scenario, $Zr(Ox)_4$ had the lowest energy arrangement (100.0%), indicating that conversion to $Zr(Dox)_4$ is pH sensitive. In addition to this, the hydration of the 6-coordinate species from the dissociative mechanisms was explored. It was found that hydration raised the energy of the steps that were previously 6-coordinate, suggesting that the 6-coordinate zirconium complexes would likely remain 6-coordinate before complexation by Dox or HDox. Finally, the effect of temperature was investigated and found to have minimal effect on the distribution of species in all pH scenarios.

The experiments set out to model what would form when doxorubicin encountered $Zr(oxinate)_4$ inside a liposome; the calculations suggest that it would be thermodynamically feasible to form either $Zr(doxorubicinate)_4$ or $[Zr(oxinate)_1(doxorubicinate)_2]^+$ depending on the pH of the solution.

4.3.1 Future work

To further strengthen the utility of using *in silico* experiments to find the identity of zirconium species (whether in a liposome or in solution), it would be beneficial for spectroscopy studies to be carried out to find the identity of the zirconium species that occur between $Zr(oxinate)_4$ and doxorubicin. Furthermore, the species distribution could be determined *via* GC-MS or NMR. With respect to the DFT experiments themselves, adding more complete basis sets may have provided more accurate results. The comparison of Def2-TZVP with Sapporo-TZP-2012/6-31G(d,p), is slightly unfair, as geometry optimisations using the Def2-TZVP basis set will always require more computational time compared to optimisations using Sapporo-TZP-2012 for Zr, and the much smaller 6-31G(d,p) basis set for all other atoms. Consequently, it would have been a fairer test to benchmark a mixed Def2-TZVP/6-31G(d,p) basis set. Furthermore, for completeness, the Sapporo-TZP-2012/TZVP mixed basis set or even the Def2-TZVP/TZVP mixed basis set may have provided a more accurate Gibbs free energy value, whilst still retaining an appropriate computational cost.

The question then becomes, what is the level that the brief requires? In this case, the precedent set in the literature was to use the B3LYP functional, treating all atoms with the LANL2DZ basis set, and the IEFPCM solvation model. Therefore, a double zeta basis set (6-31G(d,p)) for the non-Zr atoms was also used, with an appreciation that the accuracy could suffer, but with the gain of faster computation. The overall HPC time on this project was just under 1000 hours on the highest-level resources available, and 1000 hours at the second highest (not including failed jobs that were deleted from the HPC log). Increasing the basis set size to being more comprehensive may have made the project unfeasible in the time available.

4.4 References

1 D. Navot, N. Laufer, J. Kopolovic, R. Rabinowitz, A. Birkenfeld, A. Lewin, M. Granat, E. J. Margalioth and J. G. Schenker, *N. Engl. J. Med.*, 1986, **314**, 806–811, DOI: 10.1056/NEJM198603273141302.

2 J. O'Shaughnessy, *Oncologist*, 2003, **8**, 1–2, DOI: 10.1634/theoncologist.8-suppl_2-1.

3 D. Bellarosa, A. Ciucci, A. Bullo, F. Nardelli, S. Manzini, C. A. Maggi and C. Goso, *J. Pharmacol. Exp. Ther.*, 2001, **296**, 276–283.

4 N. Ismaili, M. Amzerin and A. Flechon, *J. Hematol. Oncol.*, 2011, **4**, 1–11, DOI: 10.1186/1756-8722-4-35.

5 A. J. J. Wood, R. Alexanian and M. Dimopoulos, *N. Engl. J. Med.*, 1994, **330**, 484–489, DOI: 10.1056/NEJM199402173300709.

6 O. Tacar, P. Sriamornsak and C. R. Dass, *J. Pharm. Pharmacol.*, 2012, **65**, 157–170, DOI: 10.1111/j.2042-7158.2012.01567.x.

7 D. Wright, *NHS Chemotheraphy Protocol: Breast Cancer – Doxorubicin (21 day)*, Southampton, 2014.

8 S. Wang, E. A. Konorev, S. Kotamraju, J. Joseph, S. Kalivendi and B. Kalyanaraman, *J. Biol. Chem.*, 2004, **279**, 25535–25543, DOI: 10.1074/jbc.M400944200.

9 C. Carvalho, R. Santos, S. Cardoso, S. Correia, P. Oliveira, M. Santos and P. Moreira, *Curr. Med. Chem.*, 2009, **16**, 3267–3285, DOI: 10.2174/092986709788803312.

10 P. Liu, G. Chen and J. Zhang, *Molecules*, 2022, **27**, 1372, DOI: 10.3390/molecules27041372.

11 H. Hashizume, P. Baluk, S. Morikawa, J. W. McLean, G. Thurston, S. Roberge, R. K. Jain and D. M. McDonald, *Am. J. Pathol.*, 2000, **156**, 1363–1380, DOI: 10.1016/S0002-9440(10)65006-7.

12 F. Martin, A. Huang, B. Uziely, B. Kaufman and T. Safra, *Cancer Res.*, 1994, **54**, 987–992.

13 A. Gabizon, H. Shmeeda and Y. Barenholz, *Clin. Pharmacokinet.*, 2003, **42**, 419–436, DOI: 10.2165/00003088-200342050-00002.

14 M. E. R. O'Brien, N. Wigler, M. Inbar, R. Rosso, E. Grischke, A. Santoro, R. Catane, D. G. Kieback, P. Tomczak, S. P. Ackland, F. Orlandi, L. Mellars, L. Alland and C. Tendler, *Ann. Oncol.*, 2004, **15**, 440–449, DOI: 10.1093/annonc/mdh097.

15 P. Gawne, F. Man, J. Fonslet, R. Radia, J. Bordoloi, M. Cleveland, P. Jimenez-Royo, A. Gabizon, P. J. Blower, N. Long and R. T. M. de Rosales, *Dalton Trans.*, 2018, **47**, 9283–9293, DOI: 10.1039/C8DT00100F.

16 S. Edmonds, A. Volpe, H. Shmeeda, A. C. Parente-Pereira, R. Radia, J. Baguña-Torres, I. Szanda, G. W. Severin, L. Livieratos, P. J. Blower, J. Maher, G. O. Fruhwirth, A. Gabizon and R. T. M. de Rosales, *ACS Nano*, 2016, **10**, 10294–10307, DOI: 10.1021/acsnano.6b05935.

17 P. J. Gawne, F. Clarke, K. Turjeman, A. P. Cope, N. J. Long, Y. Barenholz, S. Y. A. Terry and

R. T. M. de Rosales, *Theranostics*, 2020, **10**, 3867–3879, DOI: 10.7150/thno.40403.

18 J. P. Holland and N. Vasdev, *Dalt. Trans.*, 2014, **43**, 9872–9884, DOI: 10.1039/C4DT00733F.

19 M. Bursch, J. Mewes, A. Hansen and S. Grimme, *Angew. Chem. Int. Ed.*, , DOI:10.1002/anie.202205735, DOI: 10.1002/anie.202205735.

20 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian, Inc., Wallingford, CT, 2016.

21 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396, DOI: 10.1021/jp810292n.

22 B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson and T. L. Windus, *J. Chem. Inf. Model.*, 2019, **59**, 4814–4820, DOI: 10.1021/acs.jcim.9b00725.

23 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652, DOI: 10.1063/1.464913.

24 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241, DOI: 10.1007/s00214-007-0310-x.

25 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615, DOI: 10.1039/b810189b.

26 Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656–5667, DOI: 10.1021/jp050536c.

27 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728, DOI: 10.1063/1.1674902.

28 P. J. Hay and W. R. Wadt, *J. Chem. Phys.*, 1985, **82**, 270–283, DOI: 10.1063/1.448799.

29 N. Godbout, D. R. Salahub, J. Andzelm and E. Wimmer, *Can. J. Chem.*, 1992, **70**, 560–571, DOI: 10.1139/v92-079.

30 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297, DOI: 10.1039/b508541a.

31 P. Kathirgamanathan, S. Surendrakumar, J. Antipan-Lara, S. Ravichandran, V. R. Reddy, S. Ganeshamurugan, M. Kumaraverl, V. Arkley, A. J. Blake and D. Bailey, *J. Mater. Chem.*, 2011, **21**, 1762–1771, DOI: 10.1039/c0jm02644a.

32 J. Kreutzer, P. Blaha and U. Schubert, *Comput. Theor. Chem.*, 2016, **1084**, 162–168, DOI: 10.1016/j.comptc.2016.03.030.

33 N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, 115, 2315–2372, DOI: 10.1080/00268976.2017.1333644.

34 J. Cirera, E. Ruiz and S. Alvarez, *Organometallics*, 2005, **24**, 1556–1562, DOI: 10.1021/om049150z.

35 S. Alvarez and J. Echeverría, *J. Phys. Org. Chem.*, 2010, **23**, 1080–1087, DOI: 10.1002/poc.1735.

36 J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc., Wallingford, CT, 3rd edn., 2015.

37 A. Kheirolomoom, L. M. Mahakian, C. Y. Lai, H. A. Lindfors, J. W. Seo, E. E. Paoli, K. D. Watson, E. M. Haynam, E. S. Ingham, L. Xing, R. H. Cheng, A. D. Borowsky, R. D. Cardiff and K. W. Ferrara, *Mol. Pharm.*, 2010, **7**, 1948–1958, DOI: 10.1021/mp100245u.

38 S. A. Abraham, K. Edwards, G. Karlsson, S. MacIntosh, L. D. Mayer, C. McKenzie and M. B. Bally, *Biochim. Biophys. Acta - Biomembr.*, 2002, **1565**, 41–54, DOI: 10.1016/S0005-2736(02)00507-2.

39 O. Gutten and L. Rulíšek, *Inorg. Chem.*, 2013, **52**, 10347–10355, DOI: 10.1021/ic401037x.

40 J. P. Holland, V. Divilov, N. H. Bander, P. M. Smith-Jones, S. M. Larson and J. S. Lewis, *J. Nucl. Med.*, 2010, **51**, 1293–1300, DOI: 10.2967/jnumed.110.076174.

41 U. Mazzi, W. C. Eckelman and W. A. Volkert, *Technetium and Other Radiometals in Chemistry and Medicine*, S.G.E., 2010.

42 F. Guérard, Y. S. Lee, R. Tripier, L. P. Szajek, J. R. Deschamps and M. W. Brechbiel, *Chem. Commun.*, 2013, **49**, 1002–1004, DOI: 10.1039/c2cc37549d.

43 W. B. Blumenthal, *J. Chem. Doc.*, 1969, **9**, 135–136, DOI: 10.1021/c160034a003.

44 X. Pan, H. Wang, C. Li, J. Z. H. Zhang and C. Ji, *J. Chem. Inf. Model.*, 2021, **61**, 3159–3165, DOI: 10.1021/acs.jcim.1c00075.

45 R. Ma, G. F. Schuette and L. J. Broadbelt, *Appl. Catal. A Gen.*, 2015, **492**, 151–159, DOI: 10.1016/j.apcata.2014.12.022.

46 J. P. Holland, *Inorg. Chem.*, 2020, **59**, 2070–2082, DOI: 10.1021/acs.inorgchem.9b03515.

Chapter 5

Can a DFT model make gallium chemists more efficient?

5.1 Introduction

Currently chemists researching mitochondrial function *via* gallium contrast agents will spend months to years in synthesising complexes without a rough estimate of the key performance indicator (KPI) for their application. This is an inefficient way for research to run – in most other industries prototyping would be a key stage in project development that is not possible in this field. Instead, chemists have to design their ligands and complexes based on intuition and hope that these designs will be sufficient to meet a non-trivial problem such as passing a gallium complex through a mitochondrial membrane. This understandably leads to years of work in developing complexes that do not meet their KPIs once evaluated in biological models.

5.1.1 Effect of anthracyclines on cardiomyocytes

As seen in Chapter 4, anthracyclines (like doxorubicin, Figure 5.1) are highly effective anti-cancer drugs used to treat a wide variety of cancers,[1–5] however, they have been shown to have high cardiotoxicity which can eventually result in heart failure.[6,7] The immediate threat of cancer is usually prioritised with the cardiotoxicity being secondary in this regard; the associated risk is managed by carefully controlling the dose.[8] The problem is compounded by an unpredictable onset of cardiovascular symptoms which can occur weeks to years after treatment with anthracyclines, by which time, significant damage to cardiac muscle cells (cardiomyocytes) may have occurred without the patient's or clinician's knowledge.

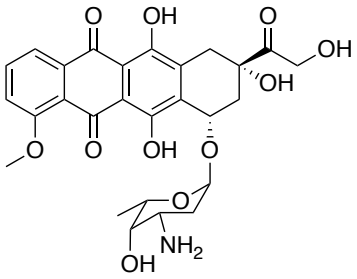


Figure 5.1: Skeletal diagram of doxorubicin.

5.1.2 Mitochondrial function

Mitochondria are membrane-bound organelles found in a large portion of cells in the body. The core function of mitochondria is to transform adenosine diphosphate (ADP) to adenosine triphosphate (ATP) *via* oxidative phosphorylation through a sequence of biochemical steps. The human heart generates 30 kilograms of ATP per day to meet its energy consumption needs, in maintaining blood pressure and basal metabolism, pumping over 100,000 times a day circulating around 5 litres of blood.[9,10] Mitochondria account for around 35% of cardiomyocyte volume, highlighting the massive energy demands the human heart places on the body.[11]

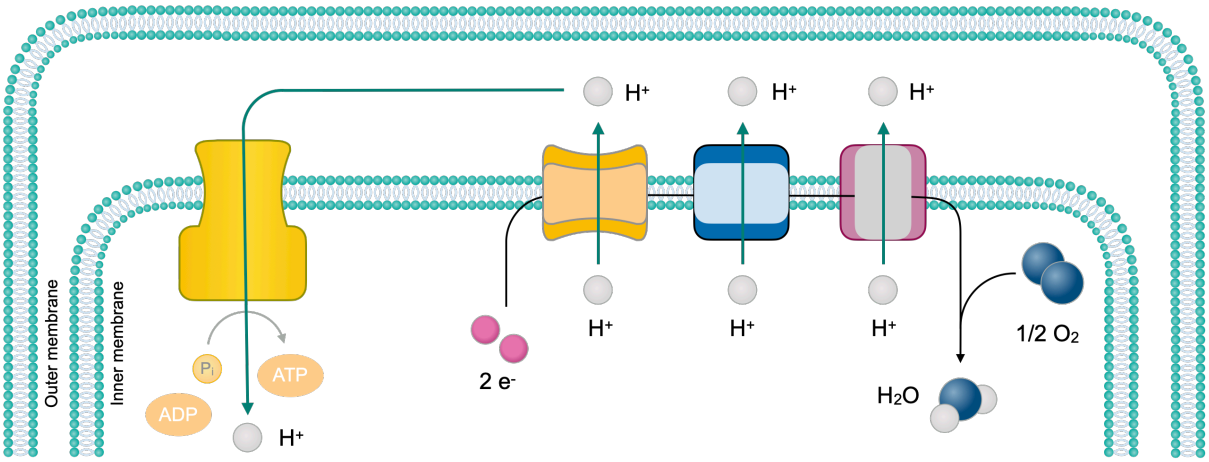


Figure 5.2: Schematic of the mitochondria, showing the electron transport chain facilitating the proton gradient required for the conversion of ADP to ATP.

Inside a mitochondrion, Figure 5.2, multiple processes are occurring with the end goal of converting ADP to ATP. The citric acid cycle produces electrons from glucose during glycolysis, which are given to the electron transport chain (ETC; a collection of protein complexes within the inner membrane), which *via* redox reactions drives a gradient of protons across the membrane as O_2 is reduced to H_2O. This enables ATP synthase to use the gradient to power oxidative phosphorylation of ADP with inorganic phosphate to produce ATP. Typically this electronic potential is -150 to -180 mV (negative within the inner membrane, positive outside).[12] With normal heart (and lung) function, cardiomyocytes will have a reliable supply of O_2 for the ETC to shuttle electrons to for its reduction to H_2O. Disruption to this supply affects the membrane potential, rising up to -200 mV, before fading to 0 mV, as the organelle stops working, ultimately leading to cell death.[13]

Once administered, doxorubicin tends to accumulate inside the mitochondria, due to its high binding affinity to cardiolipin – a phospholipid in the mitochondria's inner membrane. The heart is naturally a target for doxorubicin due to its high concentration of mitochondria within its cardiomyocytes. This binding disrupts the activity of the ETC, inserting itself into the redox pathway, and eventually causing the reduction of doxorubicin into a radical containing species. From here, it can give electrons to molecular oxygen creating reactive oxygen species. These are then able to cause peroxidation of lipids (leading to membrane rupture), and oxidative damage to DNA (cell DNA and mitochondrial DNA) leading to cell damage and death.[14–17]

5.1.3 Imaging mitochondrial function

Whilst the anthracycline induced disruption of the ETC is somewhat responsible for the cardiotoxicity, this effect can also be leveraged for imaging of the mitochondrial function. Every mitochondria-containing cell has two barriers before entry into the mitochondria, the cell membrane and the outer membrane of the mitochondria. Both of these are made up of phospholipid bilayers, consisting of two layers of phospholipids where the hydrophilic heads point towards the intracellular and extracellular space, and the hydrophobic tails point to each other. This means that molecules passively passing through the membrane need a certain degree of hydrophobicity to pass through the hydrophobic portion of the membrane, but not so much that they would rather stay there than keep passing through. As explained above, the inner membrane has a -150 to -180 mV membrane potential which attracts cations; the other membranes also have a similar but weaker membrane potential due to the distribution of endogenous cations being maintained to have an excess of K^+ cations inside the membrane and Ca^{2+} and Na^+ outside. Together these two effects can be exploited to create tracers that have a delocalised positive charge over their moderately lipophilic structures.

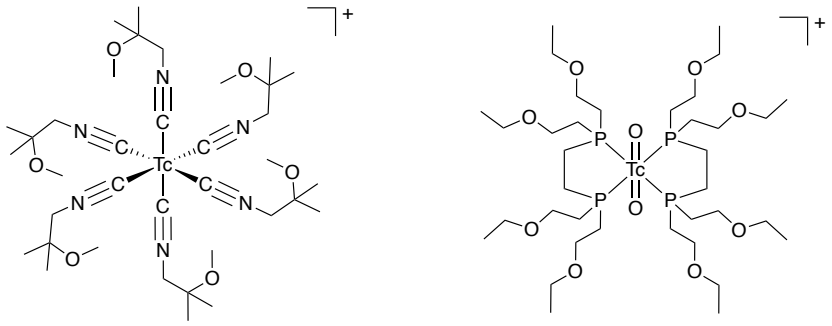
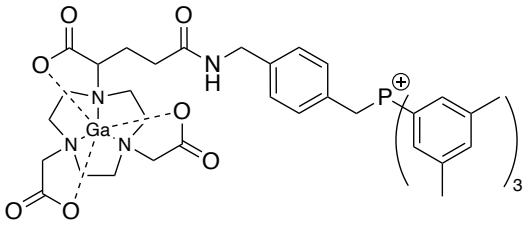


Figure 5.3: Technetium-based imaging agents for myocardial perfusion imaging.

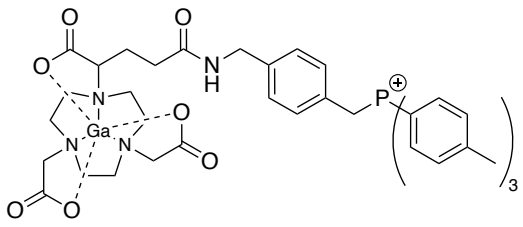
Mitochondrial and myocardial perfusion imaging was originally undertaken in the 1970s with the electron capture radioisotope thallium-201, but due to its long half-life (73 hours), low activity doses were given, resulting in poor resolution imaging.[18] This led to the use of technetium-99m, a SPECT radioisotope with a more clinically relevant half-life for myocardial perfusion imaging (6 hours), allowing for lower radiation exposure for the patient. Two successful clinical imaging agents, sestamibi and tetrofosmin, are shown in Figure 5.3. These imaging agents perform well as myocardial imaging agents as they are overall monocationic, yet due to their alkoxy containing ligands have a good balance of hydrophobic and hydrophilic character, and the symmetrical nature of the complexes ensure that the positive charge is localised in the centre. This is seen in [^{99m}Tc]Tc-sestamibi's log P value (with RCY of 98%) of 1.29.[19] These effects shield the metal core from hydrophilic interactions which would cause the complex to travel less effectively through the cell and organelle membranes.[20,21]

5.1.3.1 Gallium complexes

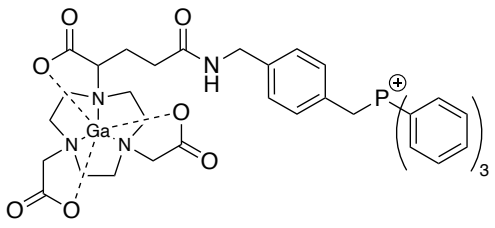
Gallium complexes have recently been researched to develop PET tracers capable of imaging the mitochondria and myocardial perfusion *via* the use of radioisotopes Ga-68 and Ga-67, leveraging the advantages of PET over SPECT seen in Chapter 1 with better resolution and sensitivity, and in the case of Ga-68, a shorter half-life (68 minutes) and therefore reduced radiation dose. In 2014, Sharma and co-workers developed Galmydar, Figure 5.4.[22] This complex was designed as a lipophilic monocationic agent, with aromatic arms functionalised with isopropoxy groups increasing the lipophilic character of the alkylamine chain skeleton. Ga-67 studies showed that axial *trans*-coordination of the phenolic binding groups improved cell penetration compared to a variant designed to be *cis*-coordinated, this was mirrored in the log P values with a value of 1.0 for the *trans* and 0.2 for the *cis*.[23] Electrostatic potential mapping highlighted that the distribution of positive charge for the *trans*-complex was uniformly distributed around the complex, whereas the *cis*-complex had a separation of charges, hypothesised to impede its cell penetration. The ability of Galmydar to image the effect of doxorubicin administration was shown in 2019 by Sivapackiam and co-workers.[24] They showed a 2-fold drop in retention of [^{68}Ga]Ga-Galmydar in myocardium with the administration of doxorubicin in rats, compared to the control group, *via* microPET-CT imaging and *ex vivo* biodistribution studies. The authors also showed that the effect was time- and dose-dependent – highlighting the need for dose control when used as a chemotherapeutic.



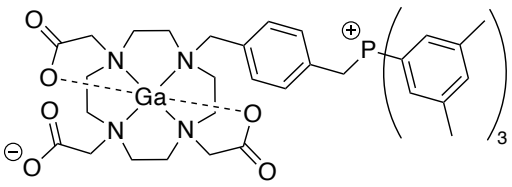
[Ga(NODAGA-xy-TXP)]$^+$



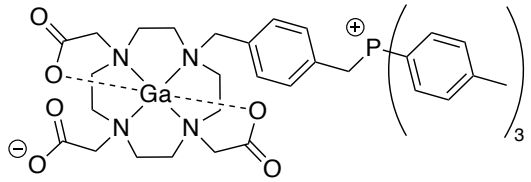
[Ga(NODAGA-xy-TTP)]$^+$



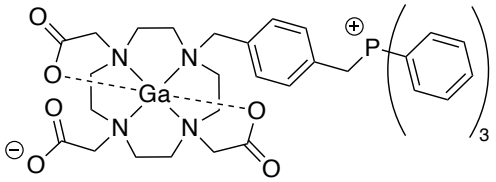
[Ga(NODAGA-xy-TPP)]$^+$



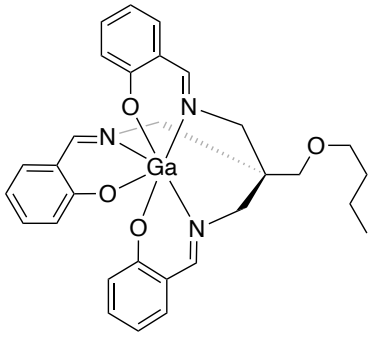
[Ga(DO3A-xy-TXP)]$^+$



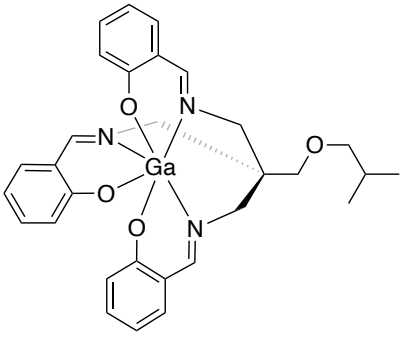
[Ga(DO3A-xy-TTP)]+



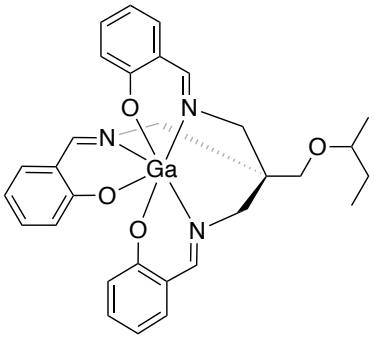
[Ga(DO3A-xy-TPP)]+



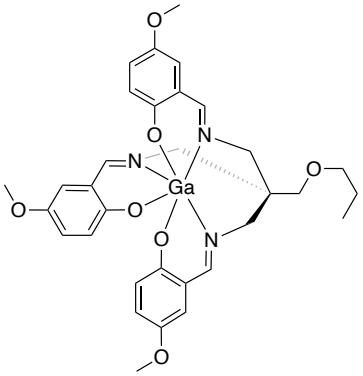
Ga((sal)$_3$tame-O-n-Bu)



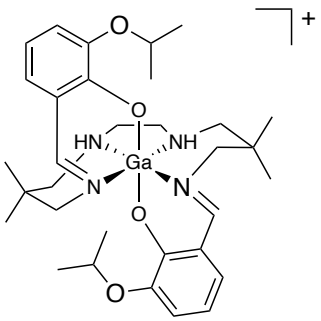
Ga((sal)$_3$tame-O-iso-Bu)



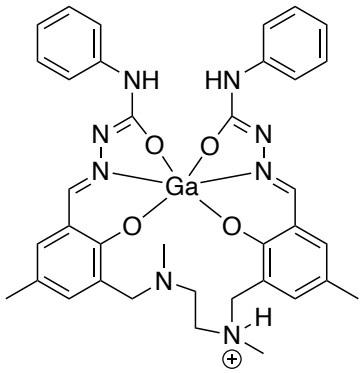
Ga((sal)$_3$tame-O-sec-Bu)



Ga((5-MeOsal)$_3$tame-O-n-Pr)



Galmydar



Ga(Hbsc)

Figure 5.4: Published gallium complexes with log P or log $D_{7.4}$ values.

Table 5.1: Published gallium complexes with log P or log $D_{7.4}$ values.

Complex	Log P or Log $D_{7.4}$	Value
$[Ga(NODAGA\text{-}xy\text{-}TXP)]^+$	Log $D_{7.4}$	-0.37
$[Ga(NODAGA\text{-}xy\text{-}TTP)]^+$	Log $D_{7.4}$	-1.00
$[Ga(NODAGA\text{-}xy\text{-}TPP)]^+$	Log $D_{7.4}$	-2.43
$[Ga(DO3A\text{-}xy\text{-}TXP)]^+$	Log $D_{7.4}$	-1.81
$[Ga(DO3A\text{-}xy\text{-}TTP)]^+$	Log $D_{7.4}$	-2.66
$[Ga(DO3A\text{-}xy\text{-}TPP)]^+$	Log $D_{7.4}$	-3.12
$Ga((sal)_3tame\text{-}O\text{-}n\text{-}Bu$	Log P	3.10
$Ga((sal)_3tame\text{-}O\text{-}iso\text{-}Bu$	Log P	3.10
$Ga((sal)_3tame\text{-}O\text{-}sec\text{-}Bu$	Log P	2.58
$Ga((5\text{-}MeOsal)_3tame\text{-}O\text{-}n\text{-}Pr$	Log P	2.54
Galmydar	Log P	1.00
Ga(Hbsc)	Log P	1.44

In 2018, Smith and co-workers developed a series of complexes around the DO3A scaffold, in a similar fashion to complexes published by Kardashinsky and co-workers in 2017, which similarly coupled arylphosphonium cations to chelated gallium.[25,26] Complexes reported by Smith and co-workers were specifically designed to investigate myocardial perfusion imaging, with tuneable lipophilicity from the alkyl functionalisation of the aryl groups on the phosphonium cation. These complexes demonstrated an ability to localise within the myocardium but showed fast washout in the Langendorff isolated perfused heart model (Figure 5.5), compared to [^{99m}Tc]Tc-sestamibi (Figure 5.6). The very low percentage of retention for the DO3A-xy-TXP complex indicated that it struggled to penetrate the cell and mitochondria membranes. The mechanism of retention for [^{99m}Tc]Tc-sestamibi was made clearer when the hearts were infused with the ionophore carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP, Figure 5.7). CCCP depolarises the inner membrane potential by shuttling protons back against the gradient set up by the ETC, therefore disrupting the retention of a cationic radiotracer.[27] Only [^{99m}Tc]Tc-sestamibi was affected by mitochondrial depolarisation; with the DO3A-xy-TXP complex showing the same trace as without CCCP. The CCCP result indicates that [^{99m}Tc]Tc-sestamibi's retention is dependent on the membrane potential, whereas $[^{68}Ga][Ga(DO3A\text{-}xy\text{-}TXP)]^+$ is not. This in turns suggests that [^{99m}Tc]Tc-sestamibi is able to penetrate the three membranes required to reach the inner mitochondria, and that $[^{68}Ga][Ga(DO3A\text{-}xy\text{-}TXP)]^+$ is not able to pass through the required membranes. This conclusion was supported by the log $D_{7.4}$ studies (log $D_{7.4}$ is the pH specified analogue of log P at pH 7.4) showing that $[^{68}Ga][Ga(DO3A\text{-}xy\text{-}TPP)]^+$, -TTP)]$^+$, and -TXP)]$^+$ had log $D_{7.4}$ values of -3.12 ± 0.17, -2.66 ± 0.09, and -1.81 ± 0.02 respectively (Table 5.1), indicating that the radiotracers were not lipophilic enough compared to [^{99m}Tc]Tc-sestamibi's log P of 1.29.

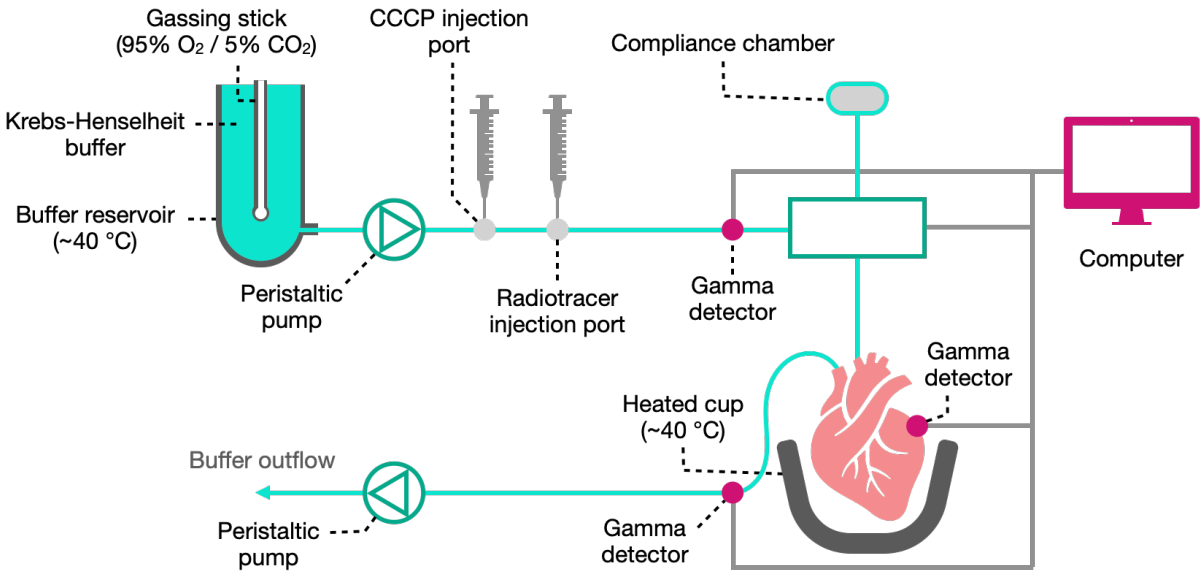


Figure 5.5: Schematic of the Langendorff isolated perfused heart model, showing the triple gamma detector system in pink. The system pumps oxygenated Krebs-Henseleit buffer (blue) at 40 °C through an isolated rat heart; gamma detection occurs before (arterial), at, and after the heart (venous) to monitor radiotracer retention.

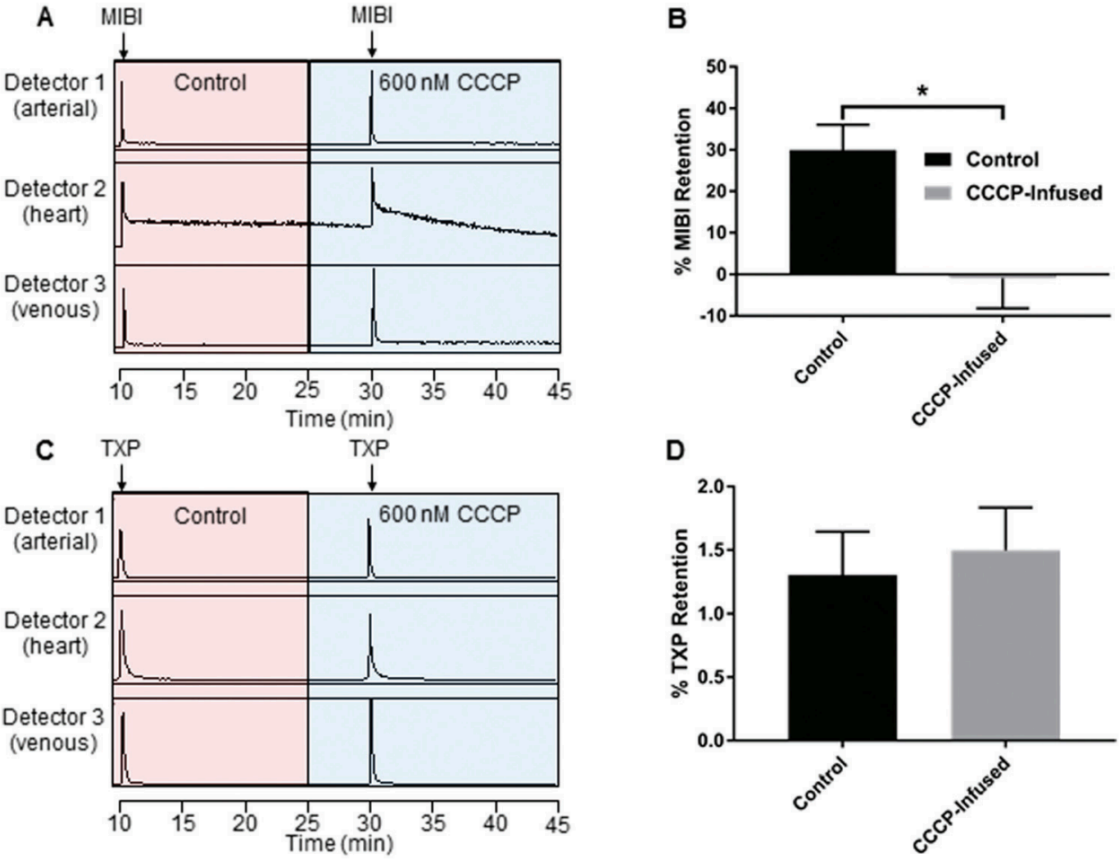


Figure 5.6: Langendorff model data for each of the gamma detectors positioned around the rat heart, and retention values inside the heart 15 minutes after injection of tracer without (control) and with 600 nM CCCP infusion. [^{99m}Tc]Tc-sestamibi (A and B) and [^{68}Ga][Ga(DO3A-xy-TXP)]$^{+}$ (C and D). [^{99m}Tc]Tc-sestamibi shows retention with very gradual washout in the heart – 30% retention after 15 minutes, and complete washout after injection into the CCCP infused heart. [^{68}Ga][Ga(DO3A-xy-TXP)]$^{+}$ shows only 1.3% retention in the control, and essentially the same level of retention after injection into the CCCP infused heart.[25] Data reproduced with permission from Smith and co-workers.[25]

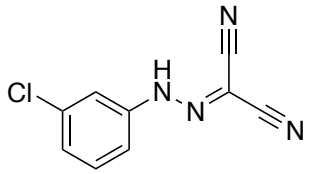


Figure 5.7: Carbonyl cyanide m-*chlorophenyl hydrazone (CCCP), an inhibitor of oxidative phosphorylation which disrupts the proton gradient in mitochondria.*

The design principles behind these tracers were improved upon by Osborne and co-workers, with a selection of NODAGA chelates, Figure 5.4.[28] These tracers used the NODAGA chelate, which contains one fewer carboxylate arm, and retained the arylphosphonium moiety. These complexes had improved log $D_{7.4}$ values over the DO3A complexes, however all three still preferentially dissolve in PBS buffer over octanol: −2.43 ± 0.07 for [^{68}Ga][Ga(NODAGA-xy-TPP)]$^+$, −1.00 ± 0.14 for [^{68}Ga][Ga(NODAGA-xy-TTP)]$^+$, and −0.37 ± 0.21 for [^{68}Ga][Ga(NODAGA-xy-TXP)]$^+$. [^{68}Ga][Ga(NODAGA-xy-TXP)]$^+$ showed the least hydrophilic measurement, and its retention was tested before and after depolarisation of the mitochondria membrane potential with CCCP. As with the DO3A complexes, the complex had very little retention in the cell (2.67 ± 0.83% after 15 minutes), and almost equivalent retention after depolarisation (2.19 ± 0.60%), indicating that the complexes struggle to penetrate the cell, and of the small proportion that enters, only a minority passes through the inner membrane. Compared to [^{99m}Tc]Tc-sestamibi, both are cationic, however it is the lipophilicity that appears to be the decisive factor for its retention in the mitochondria. Whilst a definitive log P or log $D_{7.4}$ value is not agreed upon in literature, a value between 0.5 and 1.2 is recommended (and more broadly between 1-3) to reduce washout.[29]

Several other gallium complexes have also been developed with reported log P values. In 1993, Green and co-workers developed four gallium complexes around a tripodal tris(salicylaldimine) scaffold, Figure 5.4. These complexes investigated the effect of different pendant arms and a methoxy group on the salicylaldimine, with log P values of 3.10, 3.10, 2.58, and 2.54 for the *n*-Bu, *iso*-Bu, *sec*-Bu, and *n*-Pr ^{67}Ga complexes respectively. More recently in 2021, Prado and co-workers developed a bis(semicarbazone) gallium complex (abbreviated as Ga(Hbsc)), Figure 5.4. This complex showed fast complexation (15 minutes, 20 °C) with high RCY of > 95% and provided a log P value of 1.40, ideal for myocardial perfusion imaging.

In the case of Smith and Osborne, both former researchers in the Long group, many months (to years) were spent developing and synthesising their ligands before testing. The main barrier for their success was the ability of the tracers to penetrate the cell and mitochondria membranes – something which could not be evaluated until full synthesis had been undertaken. As a proxy for this ability, log P (or log

$D_{7.4}$) can provide a handle on understanding whether a complex would be able to be lipophilic enough to pass through the membranes, and as such, a computational tool that can predict this months before final testing would allow researchers to more efficiently conduct their research and lead to the optimisation of their designs from years to weeks.

5.1.4 Computational efforts in finding log P

The partition coefficient is a measure of the ratio of a solute's concentration in organic solvent (in this case octanol) to its concentration in water. Essentially, it is like any other equilibrium constant and can be treated in the same way as the stability constant was treated in Chapter 3. In this case, the situation is less complicated due to only having one species to contend with, which does not change its state or composition between either side of the equilibrium. The partition coefficient can be defined in Equation 5.1 *via* the standard definition of Gibbs free energy at chemical equilibrium. In order to find the $\Delta G_{partition}$, the thermodynamic cycle can be defined, Figure 5.8, which involves determining the Gibbs energy of the optimised complex as an ideal gas, and then finding the ΔG_{solv} of dissolving in 1 M water (ΔG_{solv}^{w}), and the ΔG_{solv} of dissolving in 1 M octanol (ΔG_{solv}^{o}), at the same time taking into consideration the correction for a change in state (1 atm to 1 M; 1.98 kcal mol^{-1}). In practice, as the complex stays the same whether as a gas, in water, or in octanol, these extra considerations cancel out (Equation 5.2) and can be combined *via* a change of base in Equation 5.3.

This can be rearranged *via* change of base rules to the standard log P.

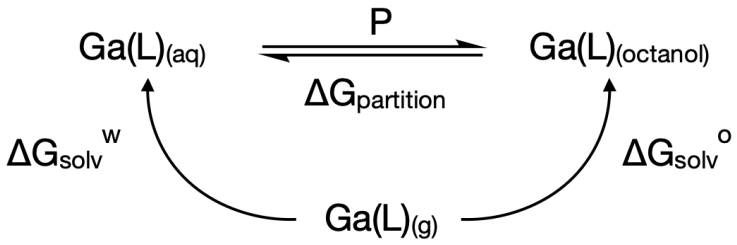


Figure 5.8: Thermodynamic cycle for the DFT determination of log P.

$$\Delta G_{partition} = -RT\ ln\ P$$

Equation 5.1: Gibbs energy of the partition equilibrium of a solute between two solvents.

$$\Delta G_{solv}{}^{w} = G\left(Ga(L)_{(aq)}\right) - G\left(Ga(L)_{(g)}\right) + 1.98\ kcal\ mol^{-1}$$
$$\Delta G_{solv}{}^{o} = G\left(Ga(L)_{(octanol)}\right) - G\left(Ga(L)_{(g)}\right) + 1.98\ kcal\ mol^{-1}$$
$$\Delta G_{partition}{}^{w\to o} = \Delta G_{solv}{}^{o} - \Delta G_{solv}{}^{w}$$
$$\Delta G_{partition}{}^{w\to o} = G\left(Ga(L)_{(octanol)}\right) - G\left(Ga(L)_{(aq)}\right)$$

Equation 5.2: Derivation of ΔG of partition between octanol and water.

$$\log P = -\frac{G\left(Ga(L)_{(octanol)}\right) - G\left(Ga(L)_{(aq)}\right)}{RT \ln 10}$$

Equation 5.3: Equation for log P via *the combination of Equation 5.1 and Equation 5.2, and change of base.*

There are multiple deviations from reality that may not be captured in a computational experiment. The first of these is that the partition of water and octanol in the two phases will not be completely distinct – the mole fraction of water in octanol has been reported as 0.270,[30] and octanol in water in 7×10^{-5} mole fractions.[31] This potentially becomes a problem as the introduction of a non-negligible amount of octanol mixing with the water will have physical properties that will be different from the pure water that we use to construct the SCRF cavity that our complexes will sit in. Additionally, there will be micro-heterogeneities within the saturated solvent (any more solvent will start to form bubbles of octanol), which makes proper sampling of the aqueous layer difficult.

This effect was investigated by Ouimet and Paluch in 2020.[31] The authors took three approaches to calculating log P. First, the untrained standard approach of a simple difference in Gibbs energies, second a trained approach where a coefficient was determined and placed on the energy difference, and third where two coefficients were determined (one for the water calculation, one for the octanol calculation). These trained approaches to determine coefficients aimed to capture the deviations from an untrained approach that would show in experimental log P values. Geometry optimisations were carried out at M06-2X/cc-pVDZ level, with single point energy calculations at M06-2X/6-31G(d) level with either SMD, SM8, or SM12 SCRF. The results of their calculations showed that the lowest root mean square error (RMSE; 0.440) was found in the untrained model with SM8, with an R^2 of 0.736. This would have ranked 4th in the SAMPL6 challenge. The other trained SM8 attempts had RMSEs between 0.64-1.02, and R^2 between 0.691-0.760; SM12 performed slightly worse with an average RMSE of 0.577 and R^2 of 0.575, and SMD RMSE of 0.731 and R^2 of 0.697. Overall, the authors concluded that the extra effort required to consider solvent mixing did not result in an equivalent gain in prediction accuracy. This was attributed to either the training dataset not matching the testing dataset (the SAMPL6 samples, Figure 5.9) closely enough; or that there is inaccuracy in log P experimental determination (they noted that of the fourteen molecules in their training dataset, DrugBank's log P references changed for eight of them,

and disappeared for three of them); or that SMD, SM8, and SM12 were already parameterised with log P data in their inception, and so the effect has already been partially accounted for.

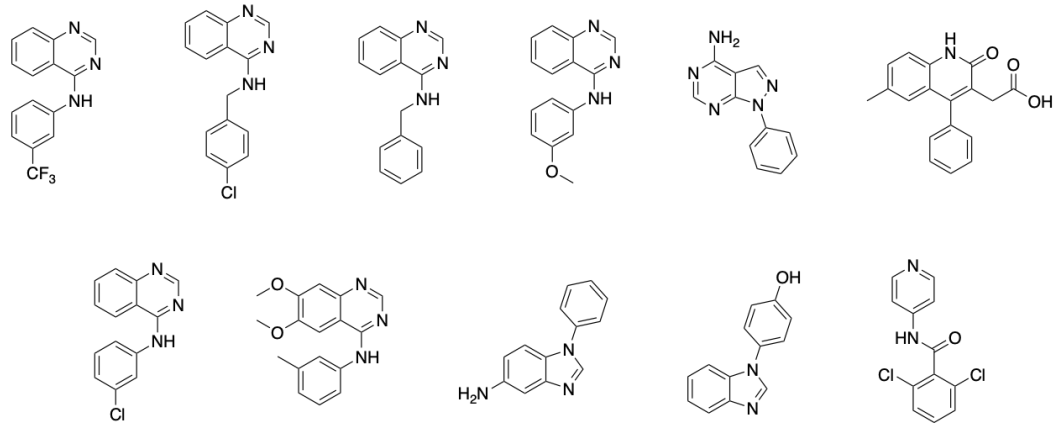


Figure 5.9: SAMPL6 molecules for log P determination.

Another possible complication is that the solute can act to alter the distribution – an effect which is incredibly difficult to account for in a computational experiment. It is possible that a molecule would be able to strongly bind solvent molecules from one phase and exist in that state in the other. Furthermore, as with each of the computational experiments in this wider work, isomerisation and conformation will play a role in providing a certain amount of unavoidable error due to the cost in sampling every single orientation of the ligand/complex.

There have been many attempts to provide a computational tool for the determination of log P, however only within the last few years have DFT methods shown to be able to suitably calculate log P values to an acceptable margin. This is shown in the SAMPL challenges discussed in Section 3.1.4. These challenges aim to benchmark the many different solutions computational chemistry researchers have designed to carry out property prediction – in the case of SAMPL6 (2018) and SAMPL7 (2020) log $D_{7.4}$ and P were assessed.

The results of the log P challenge in SAMPL6 were assessed in 2020 in the Journal of Computer-Aided Molecular Design by the organisers.[32] Of the 115 entries, quantum mechanical and empirical approaches performed better on average at predicting log P than molecular mechanics and mixed approaches. Empirical approaches take the structure of 1000s to 10,000s of molecules and their associated log P values, and then create a model that can predict log P from structure. Overall, the best submissions had RMSE values below 0.5, mean absolute error (MAE) below 0.4, and R^2 values slightly below 0.8. For context, over two thirds had RMSEs and MAEs above 1.0, and R^2 below 0.5; with the worst approaches showing RMSEs and MAEs above 4.0, and one third had R^2 values below 0.2. Two

of the best QM entries were using the COSMOtherm software with COSMO-RS as the solvation model (RMSE 0.38, R^2 0.77), or using the SMD model (RMSE 0.49, R^2 0.69).

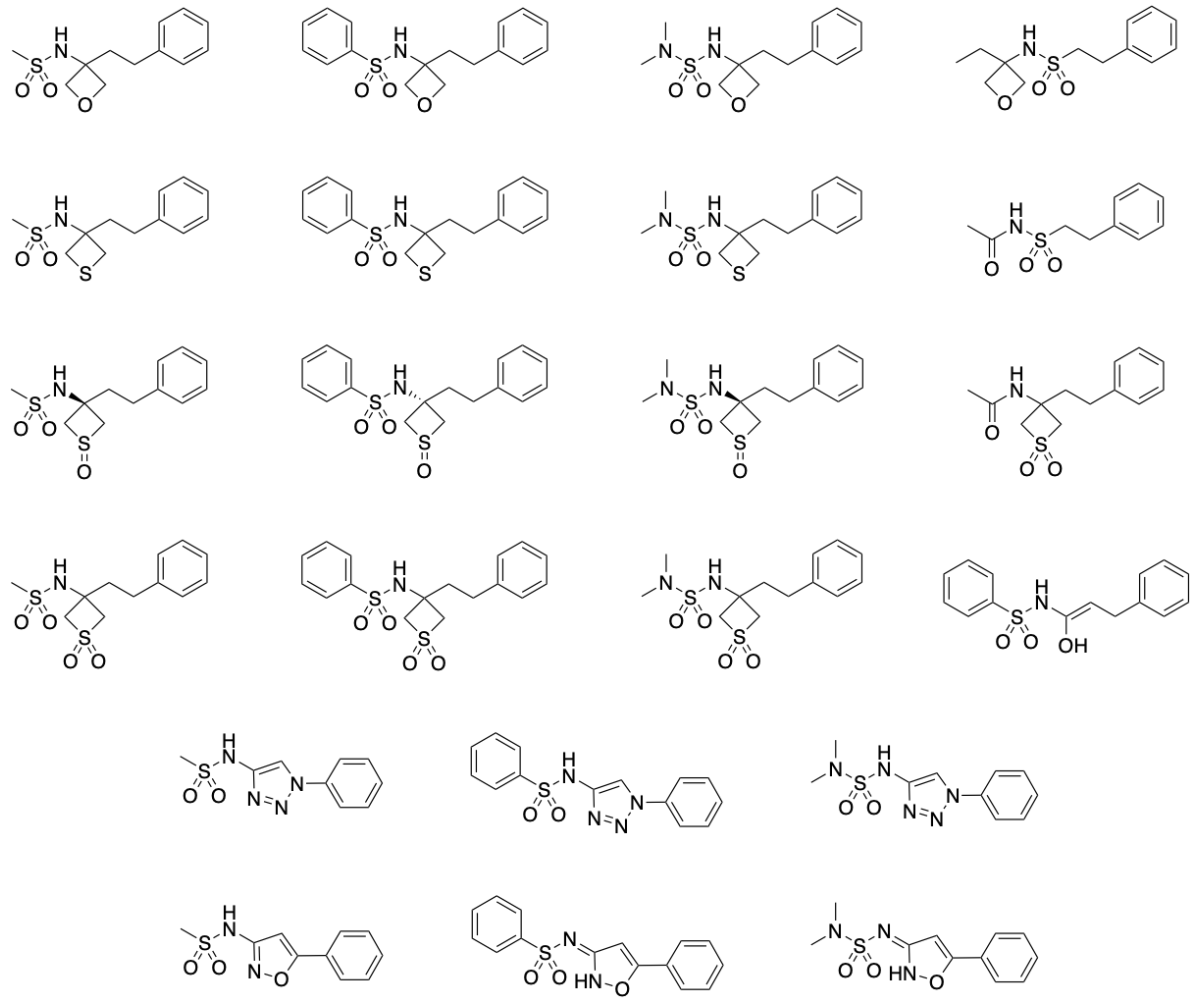


Figure 5.10: SAMPL7 molecules for log P determination.

SAMPL7's results were assessed by SAMPL organisers in 2021 – SAMPL7 sample molecules are shown in Figure 5.10.[33] Of the five most successful attempts in the log P challenge, three were empirical and two were QM. The empirical methods used various ML models – multi-linear regression trained on 60 training sulfonamides, and two neural network approaches one based on structure similarity on 233 training molecules, and the other on structure from 44,595 training molecules. These empirical methods performed better than QM methods in RMSE and MAE (0.4-0.8), though had lower R^2 values than SAMPL6 (0.4-0.5). Of the two QM methods, one used COSMOtherm software to find a collection of low energy conformers, then optimised at BP86/TZVP/COSMO, before SPE calculations at BP86/def2-TZVPD/COMSO-RS level; the other calculated structures with B3LYP/Def2-TZVP with an undefined solvation model in Gaussian 16. The COSMO method had better RMSE and MAE values (0.78 and 0.57 vs 1.55 and 1.34), yet roughly equal R^2 values (0.49 vs 0.52) to the Gaussian method.

The outcome of the SAMPL7 challenge indicates the power that machine learning models will have over the status quo QM approaches in the coming years, with greater access to more capable computing services and ever-increasing datasets. Neural networks will likely be able to 'understand' the problem faster and better than a first principles approach based on theory that has to be caveated with assumptions and approximations.

It is important to apply caution on the SAMPL6 and SAMPL7 results with respect to the use of DFT in this work – all of the samples tested in those challenges were much smaller molecules only containing C, H, N, O, S, Cl, F. The simpler the atom and molecule, the simpler it is to model with DFT. In this case, the gallium complexes used here are more difficult to optimise, and further from the molecules that were used to parameterise the SCRF models like IEFPCM, CPCM, and SMD used in the study. However, the standard that this work should be evaluated at will not be at the level of RMSE and MAE values seen in the SAMPL challenges – where RMSE is expected to be as close to 0.00 as possible, but that there will be a certain – but acceptable – level of unavoidable error. As such, the intention in this work is to test whether methods that are easily available for researchers with Gaussian software can perform calculations that can predict log P values within a *useful* accuracy (and in this specific use case, the accuracy is only required to be above or below 0).

5.1.5 Aim of this work

The aim of this work is to provide a general method for gallium chemists to run two DFT calculations (one in water, one in octanol), and decide whether their design meets their log P requirements. Considering the difficulty in this task for DFT (solvation models are not parameterised with examples such as large complexes we use here; the log P values that we are looking for are close to 0, which means that the ΔGs that we want to find are also small), a successful outcome would be a model that can say whether the complex will have a positive or negative log P. For an experimental log P of 1.00, this requires DFT to calculate a difference in ΔG_{solv} between octanol and water of only 5.7 kJ mol^{-1} (1.36 kcal mol^{-1}). Such a fine line between these states means that the value of log P is likely to be strongly dependent on how accurate the functionals and basis sets can calculate accurate energies, and most importantly how well the SCRF method can approximate the effect solvent has on the complex's electron density. Considering this, and the need of the synthetic researchers, providing a simple test that can dependably reveal whether the complex favours octanol or water before synthesis would be a success. As such, unlike in Chapter 3, this is (at first) a classification task rather than regression.

5.2 Results

In this analysis, the system that is being modelled is the determination of log P for gallium complexes. Usually this is measured *via* gamma counting the distribution of radioactive complex that is allowed to equilibrate within a mixture of n-octanol/water. The distribution of the gallium complexes follows the preference the complex has to dissolve in either octanol or water, which itself is described by its $\Delta G_{solvation}$. This change in Gibbs energy is possible to calculate in Gaussian. As well as log P, the partition coefficient can be determined at a chosen buffered pH such that the introduction of the solute does not significantly affect pH, typically set at pH 7.4 to mirror physiological pH.[28] For the purposes of this analysis, and in order to simplify the calculations, log $D_{7.4}$ and log P measurements will all be taken as one.

5.2.1 Computational methods

Density functional theory was used for all calculations with the Gaussian 16 program (version C.01).[34] Structures were geometry optimised without symmetry constraints, in aqueous solution using either IEFPCM, CPCM or SMD self-consistent reaction fields.[35–37] Each optimised structure was confirmed as a local minima on the PES *via* the presence of no imaginary frequencies. DFT calculations were undertaken with the B3LYP, M06-2X, or PW6B95D3 functional, and either 6-31G(d,p), TZVP, or def2-QZVPP basis sets.[38–42] The D3 version of Grimme's dispersion correction (*via* keyword empiricaldispersion=gd3) was applied where relevant.[43] In the cases where first order SCF failed to converge, quadratically convergent SCF was employed (*via* the scf = xqc keyword). Single point energy calculations were employed where relevant replacing the electronic energy component of the calculated Gibbs energy of optimised structures optimised at either PW6B95D3/TZVP/SMD or B3LYPD3/TZVP/SMD level. Default values for temperature and pressure were used (298.15 K, 1 atm). The default integration grid (UltraFine, 99 radial shells, 590 angular points per shell) was used throughout. The Orange 3 (version 3.35.0) data visualisation and analysis software was used to perform machine learning tasks *via* python 3.10.10.[44] Orange's neural network widget (a feed-forward network with back-propagation) was used with logistic activation and the Adam optimisation algorithm for optimisation. The neural network contained two hidden layers; the first layer had as many neurons as the dataset had features, the second had two-thirds of this amount +1. Data used in the neural network was pre-processed to have $\mu = 0$ and $\sigma = 1$. Orange's linear regression widget was also used with and without lasso regularisation (L1) (strength alpha set at 0.05).

5.2.2 Evaluating DFT model chemistries

As none of the published gallium complexes with log P (or log $D_{7.4}$) have published crystal structures, the usual method of finding the model chemistry which most accurately optimises the complex's bond lengths and angles cannot be used. Instead, the most successful of the set of complexes that have been

published in the group, $[Ga(NODAGA\text{-}xy\text{-}TXP)]^+$ (Figure 5.4), was tested with a range of model chemistries, with functional, basis set, and SCRF varied to find the model chemistry which could most accurately reproduce the experimental log P value, Table 5.2. For this analysis each model chemistry was used to optimise the complex in both water and octanol, with a minimum confirmed on the potential energy surface.

Table 5.2: Calculated log P results from the optimisations of $[Ga(NODAGA\text{-}xy\text{-}TXP)]^+$ in octanol and water. The method, basis set and SCRF method was varied, and compared to experimental log P.

Entry	Method	Basis Set	SCRF	Log P	$\lvert \text{Log P}_{\text{Actual}} - \text{Log P}_{\text{calc}} \rvert$
Experiment	-	-	-	-0.37	-
1	B3LYP	6-31G(d,p)	IEFPCM	-7.19	6.82
2	M06-2X	TZVP	IEFPCM	-5.73	5.36
3	PW6B95D3	TZVP	IEFPCM	-6.06	5.69
4	B3LYP	6-31G(d,p)	SMD	0.84	1.21
5	M06-2X	TZVP	SMD	0.26	0.63
6	PW6B95D3	TZVP	SMD	-0.47	0.10
7	B3LYP	6-31G(d,p)	CPCM	-3.58	3.21
8	M06-2X	TZVP	CPCM	-4.39	4.02
9	PW6B95D3	TZVP	CPCM	-4.33	3.96

The primary aim of the analysis in Table 5.2 was to explore if DFT was capable of correctly predicting which solvent the complex would prefer to be in. If this was possible, then the next aim was to identify which SCRF was best able to accurately replicate the experimental log P. Whilst it was expected that SMD would perform best at essentially predicting ΔG_{solv} (as that is its primary purpose),[45] it was surprising how far off IEFPCM and CPCM were from the experimental log P (MAE IEFPCM: 5.96; SMD: 0.65; CPCM: 3.73). Within the SMD results, the triple-zeta basis set TZVP outperformed the 6-31G(d,p) basis set, and the most capable PW6B95D3 functional essentially produced the correct log P (-0.47 vs -0.37). Whilst this represented only one result, so is not significant for all gallium complexes, PW6B95D3/TZVP/SMD appeared to be a good model for predicting log P and was chosen as the level at which geometry optimisations would be carried out to, before single point energy calculations would be used for other model chemistries (only method and basis set varied, SMD retained). Single point energy calculations on the already optimised structures were chosen to facilitate the timely completion of the project as the complexes are all quite large (117 atoms, 468 electrons for the NODAGA-xy-TXP gallium complex). This would otherwise result in lengthy optimisation times if optimising every single model chemistry or reducing the scope of the project with fewer model chemistries.

5.2.3 Success parameters

The brief for this project was to provide a predictive model that can, in the first instance, tell a gallium chemist that they need to improve the design of their ligands. This alone would prevent months of wasted synthesis time. As such, the bar for success is set at quite a modest height – at the bare minimum, the model should be able to determine whether a complex will prefer water or octanol. This means that the task is at its core, a classification task rather than a regression task. However, at its most ambitious, the model would be able to use regression to predict the actual value of log P. In this regard, the model will be judged on how far the calculated log P is from experimental. The calculated log P values will be correlated to experimental log P – as was done in Chapter 3. This will be judged on the predictive model's R^2 on the training dataset (as there are not enough gallium complexes with log P data to split into a testing dataset), and on the RMSE and MAE of the predictions to reality. In this instance, producing a model with RMSE and MAE scores with ± errors around 0.5 would be extremely successful, with usefulness decreasing as this error rises.

5.2.4 Simple linear regression

Alongside the three SMD model chemistries in Table 5.2, the D3 version of Grimme's dispersion correction was applied to M06-2X/TZVP and B3LYP/6-31G(d,p),[43] and the very large, quadruple-zeta basis set def2-QZVPP was also tested with PW6B95D3 to give six model chemistries, with Gibbs energy values determined on the PW6B95D3/TZVP optimised geometry. Dispersion corrections take the effect of instantaneous and fleeting dipole interactions (due to the non-fixed nature of electron density in a real molecule) into consideration, which can be important for precise energy calculation.[45] A larger, quadruple-zeta basis set was chosen to test whether the improvement in accuracy at the expense of large computational resources was worthwhile in this case.[46]

5.2.4.1 Criteria 1 – Positive or negative log P determination

In terms of the first criteria for a successful model, on the whole, the six models are able to determine whether the complexes analysed will have a positive or negative log P, Table 5.3 and Figure 5.11. $[Ga(NODAGA\text{-}xy\text{-}TTP)]^+$ was consistently calculated to be more stable in octanol than water, which may be as a result of $[Ga(NODAGA\text{-}xy\text{-}TTP]^+$'s experimental log P (-1.00) being close to 0, and therefore more difficult to predict for DFT. Nonetheless, from these results alone, the basic ability to understand whether a gallium complex will by hydrophobic enough to prefer dissolving in octanol *via* DFT has been validated. Even at the most basic level – B3LYP/6-31G(d,p), DFT can (mostly) qualitatively predict whether the log P will be negative or positive. Surpassing this level means that the models can now be evaluated quantitatively, producing predictive models which take calculated data to produce a log P prediction.

Table 5.3: Averaged calculated log P values for the six model chemistries tested, for each gallium complex. Model chemistries: PW6B95D3/TZVP, PW6B95D3/Def2QZVPP, M06-2XD3/TZVP, M06-2X/TZVP, B3LYPD3/6-31G(d,p), and B3LYP/6-31G(d,p), all with SMD for SCRF.

Complex	Experiment	Average log P
$[Ga(NODAGA\text{-}xy\text{-}TXP)]^+$	-0.37	-0.35
$[Ga(NODAGA\text{-}xy\text{-}TTP)]^+$	-1.00	1.96
$[Ga(NODAGA\text{-}xy\text{-}TPP)]^+$	-2.43	-1.46
$[Ga(DO3A\text{-}xy\text{-}TXP)]^+$	-1.81	-3.01
$[Ga(DO3A\text{-}xy\text{-}TTP)]^+$	-2.66	-4.03
$[Ga(DO3A\text{-}xy\text{-}TPP)]^+$	-3.12	-5.80
$Ga((sal)_3tame\text{-}O\text{-}n\text{-}Bu$	3.10	4.96
$Ga((sal)_3tame\text{-}O\text{-}iso\text{-}Bu$	3.10	4.34
$Ga((sal)_3tame\text{-}O\text{-}sec\text{-}Bu$	2.58	4.88
$Ga((5\text{-}MeOsal)_3tame\text{-}O\text{-}n\text{-}Pr$	2.54	2.88
Galmydar	1.00	5.84
Ga(Hbsc)	1.44	5.71

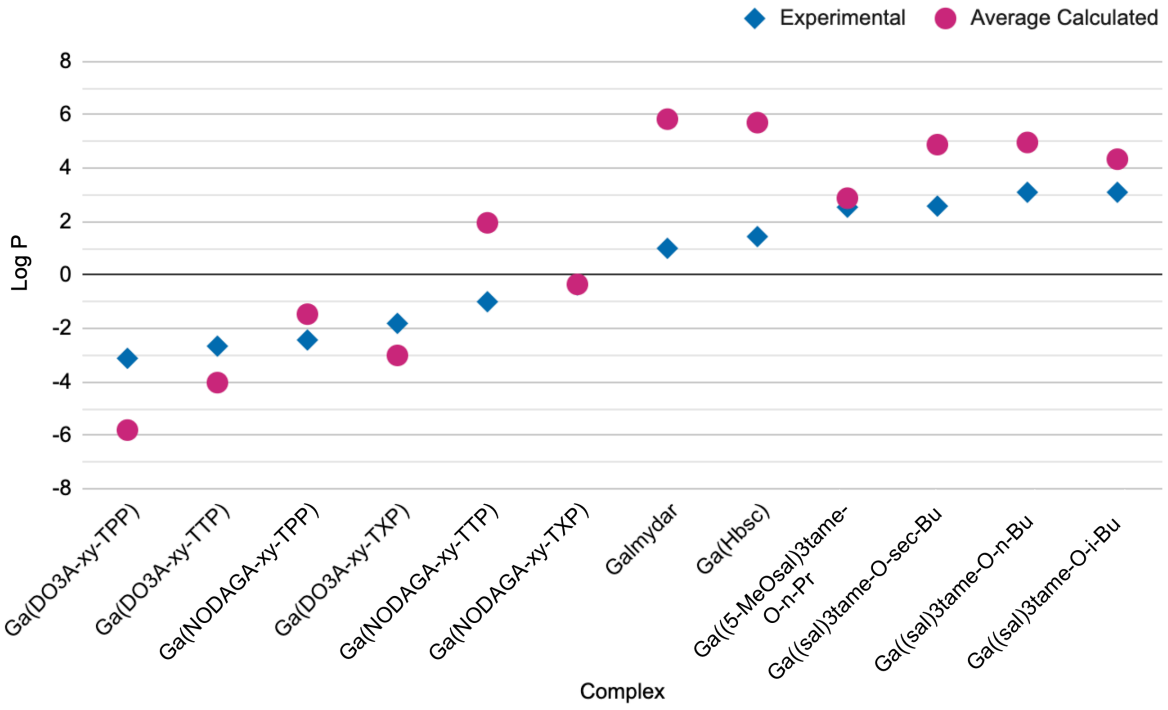


Figure 5.11: Average calculated log P plotted with experimental log P for each of the complexes, all model chemistries averaged.

5.2.4.2 Criteria 2 – Quantitative log P determination

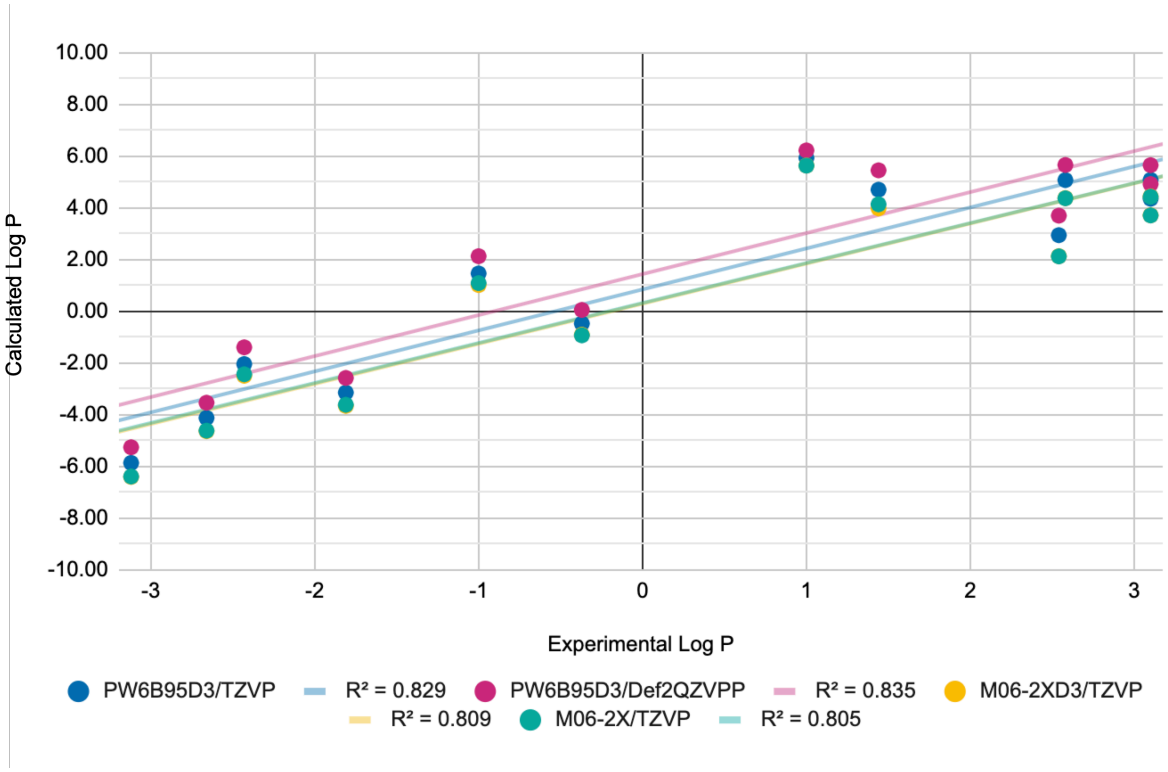


Figure 5.12: Calculated log P plotted against experimental log P for the analysed model chemistries (all with SMD). B3LYP models were omitted for clarity, B3LYP/6-31G(d,p) R^2 0.631, and B3LYPD3/6-31G(d,p) R^2 0.781.

As the qualitative criteria has been surpassed, the next test was for the method to have actual predictive power. This was determined *via* the R^2 scores when plotting calculated log P against experimental, Figure 5.12 (B3LYP models omitted for clarity, R^2 0.781 for B3LYPD3 and 0.631 for B3LYP). All models performed well with R^2 values between 0.631 to 0.835; these R^2 values fall within the top end of R^2 values in the SAMPL6 log P challenge – the four models in Figure 5.12 surpass that challenge's best result (R^2 of 0.77, for a COSMOtherm method). The predictions from the PW6B95D3/TZVP line gave an RMSE of 0.88 log units, and MAE of 1.03 log units, Table 5.4, Equation 5.4. Less powerful model chemistries also performed well e.g. M06-2X and M06-2XD3 provided very similar results, with near identical log P values, and R^2s for their linear trend lines. M06-2X and M06-2XD3 are typically less computationally expensive than PW6B95D3, so would represent a good cost-to-benefit ratio, especially for chemists without access to high performance computing. On the other hand, PW6B95D3/Def2-QZVPP only marginally outperformed the TZVP model (R^2 0.835 vs 0.829) despite the much larger basis set. This small improvement came at a considerable computational cost, as shown in Figure 5.13, which plots the averaged calculation time for the SPE calculations for $[Ga(DO3A\text{-}xy\text{-}TXP)]^+$ using a B3LYPD3/TZVP optimised geometry. The PW6B95D3/Def2-QZVPP calculation required over 36 times longer to complete than for PW6B95D3/TZVP, far outweighing its minor increase in accuracy. Unsurprisingly, the models' R^2 values increased in line with the functional's

quality as benchmarked by Mardirossian and co-workers,[47] with the larger basis set in Def2-QZVPP performing better than TZVP, as predicted by Bursch and co-workers.[46] The PW6 and M06 models all produced predictions with MAE below 1 and RMSE also around 1. The B3LYP models provided worse predictions, however B3LYPD3/6-31G(d,p) showed that with a suitable geometry the accuracy of its results was only slightly worse than far more expensive model chemistries. These error metrics suggest that these predictions are able to correspond to rule-of-thumb errors of around ±1 – as can be seen in the width of the 1σ prediction interval in Figure 5.14.

Table 5.4: Evaluation statistics of the predictive models from simple linear regression for each of the DFT model chemistries (all with SMD).

Model Chemistry	R^2	MAE	RMSE
PW6B95D3/TZVP	0.829	0.88	1.03
PW6B95D3/Def2QZVPP	0.835	0.88	1.01
M06-2XD3/TZVP	0.809	0.92	1.10
M06-2X/TZVP	0.805	0.94	1.12
B3LYPD3/6-31G(d,p)	0.781	1.04	1.20
B3LYP/6-31G(d,p)	0.631	1.57	1.74

$$Experimental \log P = \frac{Calculated \log P - 0.846}{1.587}$$

Equation 5.4: Equation for the linear trend line for PW6B95D3/TZVP/SMD calculated log P vs experimental log P.

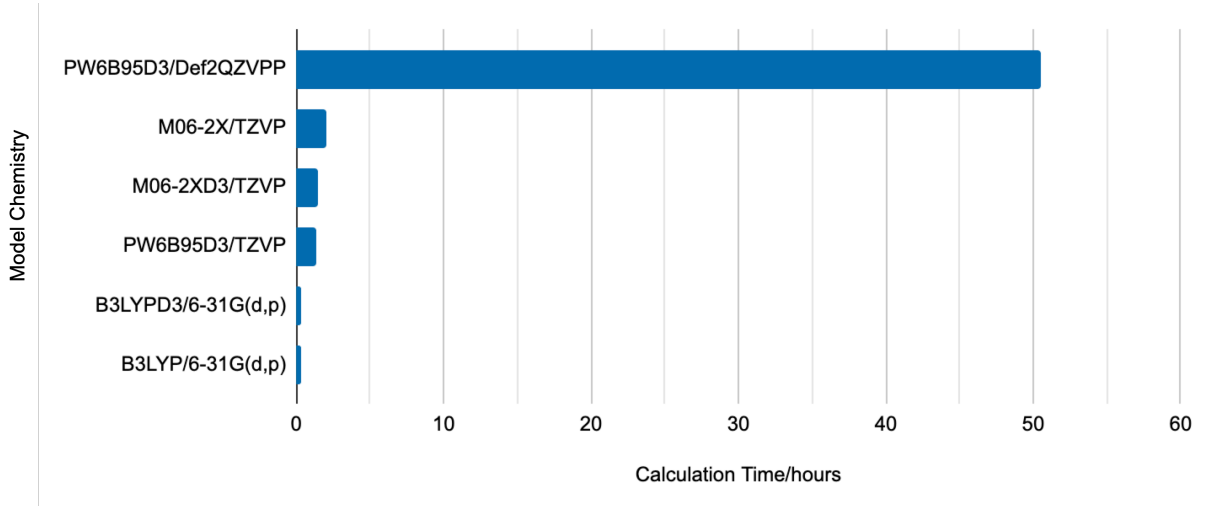


Figure 5.13: Calculation time for single point energy calculation of $[Ga(DO3A\text{-}xy\text{-}TXP)]^+$ from a B3LYPD3/TZVP optimised geometry (averaged value of water and octanol solution calculation; all with SMD).

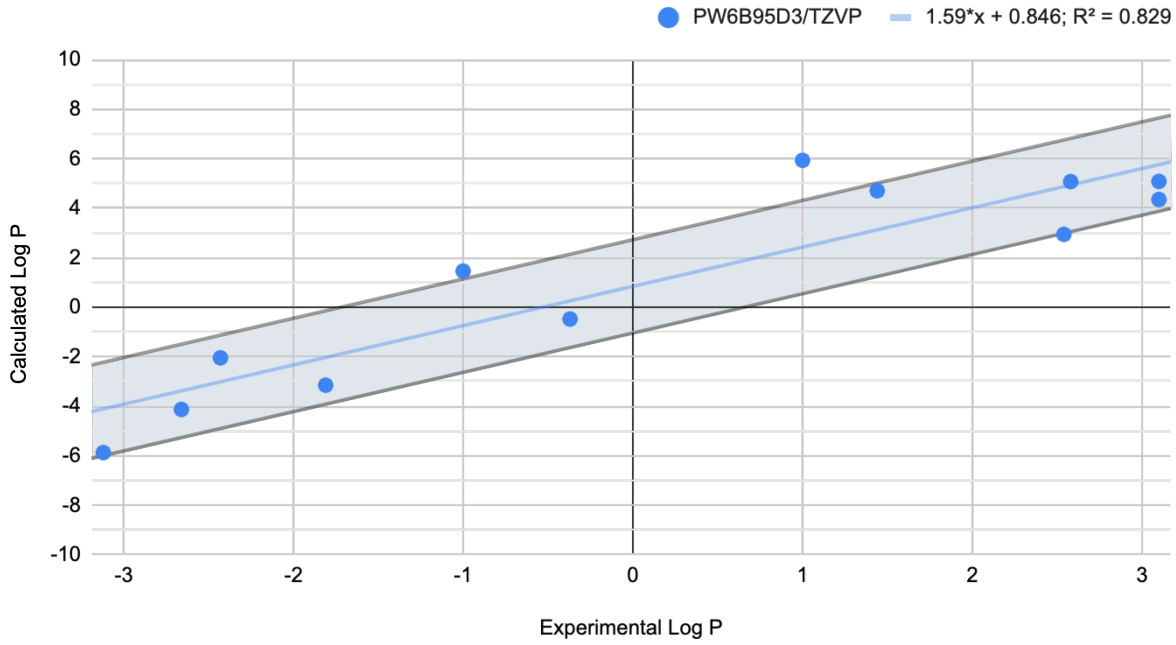


Figure 5.14: Plot of calculated vs experimental log P at PW6B95D3/TZVP/SMD level, fitted with a linear trend line and ± 1σ prediction interval.

5.2.5 Multivariate regression and neural network log P prediction

As in Chapter 3, machine learning methods were employed to investigate if a combination of model chemistries could potentially find a relationship that can more accurately model the true experimental value of log P. For these calculations, PW6B95D3/Def2-QZVPP was not included, as it would render the approach to be useless if most chemists could not generate the data required because PW6B95D3/Def2-QZVPP is too expensive.

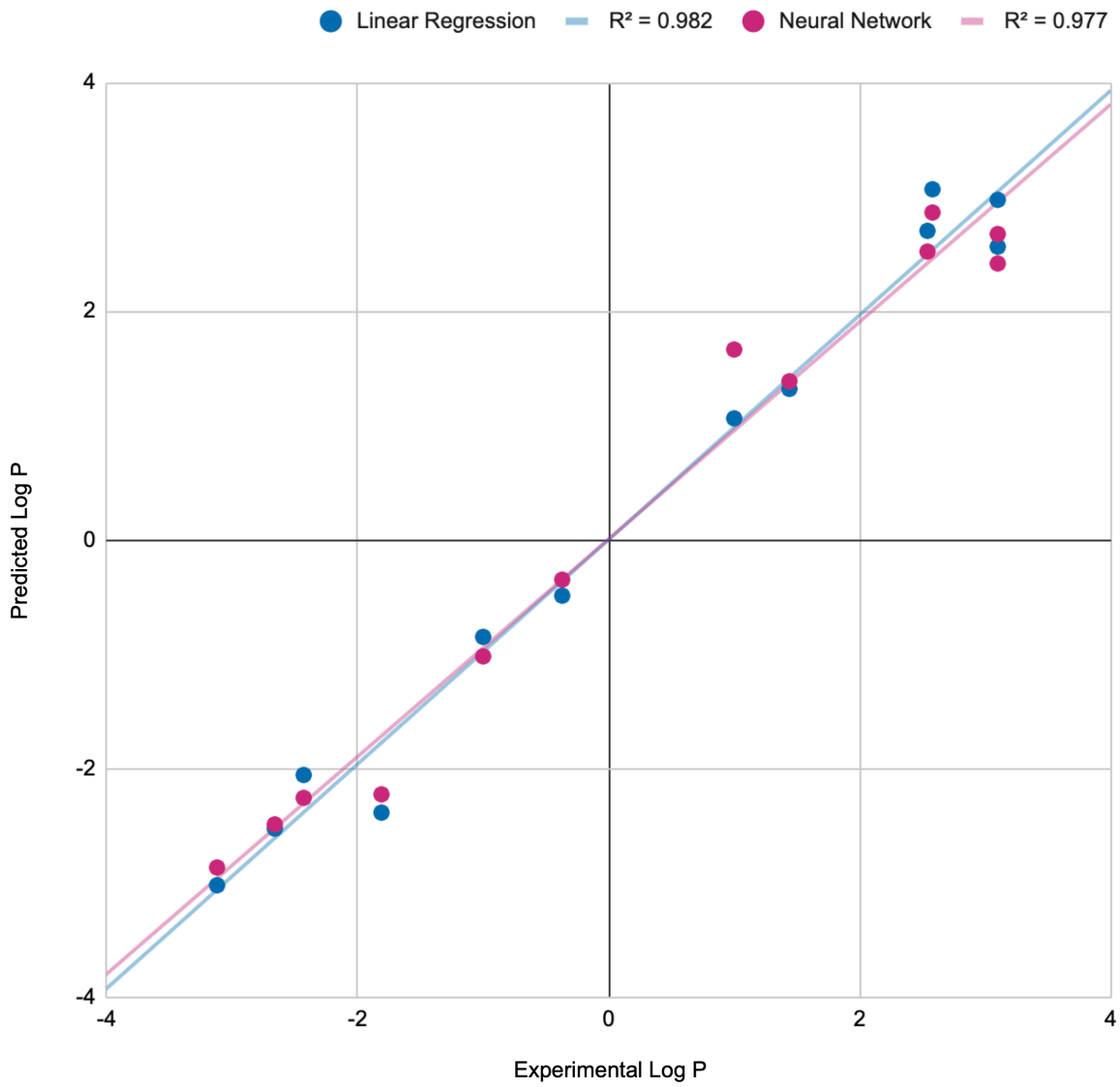


Figure 5.15: Plot of the predicted log P values vs experimental log P values for multivariate linear regression (no regularisation) and neural network (5,4 hidden layer neurons; no regularisation, max iterations 100,000) models.

Both multivariate linear regression and a neural network were able to provide models with excellent R^2 values on their predictions (Figure 5.15) 0.982 and 0.977 respectively. Unlike with the simple linear regressions, the models are much closer to having a calculated log P of 0, at an experimental log P of 0.

As there is not a lot of data available, these models have to be tested against the training dataset so it is unsurprising that the R^2, RMSE, and MAE would be good. These ML models are unlikely to be as robust against completely unseen testing data, and such high R^2 values are likely indicating that the trend is being over-fit to the training data. This would be improved with more complexes to train and test with. In an attempt to mitigate this, L1 regularisation (alpha = 0.05) was applied to the linear regression to drive the coefficients of the independent variables to 0, Table 5.5. This was able to reduce the dependence on the M06-2X data, whilst still maintaining an excellent R^2 (0.935).

Table 5.5: Statistics from the linear regression models, both with and without L1 regularisation, demonstrating the effect of regularisation to reduce the complexity of the required dataset whilst still providing acceptable R^2 values.

Statistic	Without regularisation	With L1 regularisation
intercept	-3.25257	-0.125596
PW6B95D3/TZVP coefficient	5.57835	0.535277
M06-2XD3/TZVP coefficient	12.1559	0
M06-2X/TZVP coefficient	-17.6471	0
B3LYPD3/6-31G(d,p) coefficient	0.204012	0.739232
B3LYP/6-31G(d,p) coefficient	0.171475	-0.752545
R^2	0.982	0.935

5.3 Discussion

Experimentally, log P is very unpredictable, where small impurities (including free gallium) can have a very large effect on the measured value. It is not abnormal to be unable to reproduce published syntheses across all branches of chemistry, and so the fear in producing an analysis such as this, is the true accuracy of the training data. For example, Galmydar and Ga(HBsc)'s experimental log P was a lot lower than DFT predicted, pointing at either DFT failure or a measurement failure. Whilst this specific use of DFT is pushing its accuracy limits, its reasonably good log P determination for the other complexes may also indicate some level of experimental measurement error.

In a similar regard, for all of the ionic complexes the counterions were not modelled. This was chosen to avoid the increased difficulty in sampling where the counterion would be once the complex is dissolved in solution, as with modelling explicit water molecules around the complex. The effect of the counterion is likely to contribute towards a deviation between the calculated (and therefore predicted) and experimental log P. Interestingly, the cationic complexes gave results which mostly matched the experimental values and did not largely deviate the simple linear regression. This indicates that for DFT prediction modelling, the gallium complex alone could potentially be sufficient. Equally, we did not factor in pH in these calculations, treating log $D_{7.4}$ and log P as one, to simplify the problem (important for its desired end users). To convert between log P and log $D_{7.4}$ would either require capturing pK_a data from the log $D_{7.4}$ complexes in the dataset (all from the Long group), or negotiating the challenges of DFT pK_a prediction – which would be another chapter within itself. With the pK_a data, log $D_{7.4}$ could be converted to log P *via* Equation 5.5:

$$\log D_{pH} = \log P - \log\left(1 + 10^{pK_a - pH}\right)$$

Equation 5.5: General equation to convert log P to log $D_{7.4}$ for acidic molecules.[33]

The confidence in these predictive models would be improved if there were more gallium complexes to build a larger training dataset – the increase in sample size would hopefully reduce variation, leading to higher R^2 values. This is especially important for the machine learning methods which are heavily influenced by the data that is fed into it (hence the large R^2 values, and possible over-fitting). The data that trained these models largely came from neutral complexes. If this model is used for future gallium complexes that are more suited to the myocardial perfusion imaging problem (i.e. cationic), it is possible that the confidence in the predictions will suffer.

5.4 Conclusions

The aim of this work was to provide a general method for gallium chemists to run two DFT calculations (one in water, one in octanol), and decide whether their design meets their log P requirements. With both simple linear regression and more complex machine learning methods (such as multivariate regression and neural networks), predictive models were able to be trained which had excellent R^2 values which surpassed the expectation set by the SAMPL6 and SAMPL7 challenges (R^2 values above 0.8 for all three models).

This is a very good achievement, especially considering the small window of difference between the energy states that DFT is being asked to perform at. For the simple linear regression PW6B95D3/Def2-QZVPP/SMD provided the largest R^2, and smallest MAE and RMSE – however this came at an exorbitant cost (over 36 times longer SPE calculation time than the next best model chemistry, PW6B95D3/TZVP/SMD). Because of this, PW6B95D3/TZVP/SMD was the recommended model chemistry to calculate log P values for the linear regression model's prediction (R^2 0.829, MAE 0.88, RMSE 1.03). For users with more constrained computational resources, M06-2X/TZVP/SMD (without dispersion effects) was able to produce a predictive model that had R^2, MAE, and RMSE within a similar level to the PW6B95D3 models (R^2 0.805, MAE 0.94, RMSE 1.12) – the M06-2X functional is typically faster in optimising geometries than PW6B95D3.

The predictive power of this simple model exceeds the level required to be useful for gallium chemists, reliably predicting log P values indicating preference for water solubility rather than octanol. This can be quantified by its MAE of 0.88, essentially indicating a ±1 window around a prediction – enough to communicate to a synthetic chemist that their design needs improvement or not. This result should be evaluated against what is currently available for gallium chemists to predict their log P. Currently they have not got anything, which results in lengthy synthesis with only intuition guiding them on the experimental log P value, thus, usability is more important than accuracy.

The machine learning methods, on the face of it, appear to have both. Due to the ease in carrying out SPE calculations once an acceptable geometry has been calculated, all of the model chemistries apart from PW6B95D3/Def2-QZVPP were used to train the models. This approach was chosen to let the artificial intelligence find if it could find the right combination of model chemistries to match reality most closely – in effect, the AI is tasked with taking each simple linear regression model and creating something that is better than the sum of its parts. For both multivariate regression and neural network, impressive R^2 values were achieved (0.982 (0.935 with regularisation) and 0.977) – indicating very good accuracy in its predictions. This, however, should be caveated due to the R^2 values being calculated from its training dataset – potentially leading to overfitting.

For a scenario as complex as solvation chemistry, where the current state-of-the-art computational chemistry (let alone at the level in this work) is unable to replicate reality reliably and completely, the future will likely involve moving towards bigger equalling better (more powerful computers allowing approximations to be eased – for instance, adding 10s of water molecules to the DFT calculation); or taking an ML empirical approach. The latter is already being used in the SAMPL challenges for small molecules.[32,33] This artificial intelligence approach mirrors how we use intuition ourselves – we look at a molecule and say that a certain functional group makes it more lipophilic, based off what we have seen before. Similarly, if a suitably powerful neural network is given a very large dataset of log P values for complexes across the periodic table (and given the structures), eventually the model would be trained to empirically 'understand' the problem by assigning weights to each minute aspect of a complex, to provide a predictor for log P. In the meantime, the situation could be improved with small models like the ones produced in this work being trained and released for others to use and improve with their own log P and log D data.

5.4.1 Future work

This work (and predictive model work in general) relies on the data that is fed into it. More data should be collected to strengthen the R^2 values from the simple linear regression. With the wealth of ligands available within the group – whether designed for gallium or not, it would be very feasible to complex all of these and determine their experimental log P values (most likely *via* NMR). With these in hand, geometry optimisation and frequency calculations could easily be run, increasing the size of the dataset, and providing a model that should have better predictive power.

Carrying out this work would also help to validate the extremely good machine learning models which were able to produce linear plots that had an intercept of 0, and a gradient that was very close to 1. The new data would be used as a testing dataset to evaluate whether the models can withstand being tested with unseen data. Regardless of the result, the data could be added to the training dataset to improve the weightings used in the linear regression and neural network.

As suggested in Section 5.3, a potential avenue of research is to standardise the data before it is fed into the models. This would require pK_a determination of the six NODAGA and DO3A complexes, and then calculating the log P value. Similarly reducing other constraints may lead to improved accuracy. With more time, more accurate sampling of the solute within the solvents would have been undertaken, with the effect of counterions and explicit solvent molecules being considered.

This project showed the potential effect that DFT log P determination could have on projects that depend on log P, such as in myocardial perfusion imaging. Proper prototyping and KPI prediction should be standard practice in projects that span months to years and can consume a lot of resources. In this specific regard, it could be beneficial to survey researchers for a list of complexes and log P (or log $D_{7.4}$) values that have already been collected in the past. Currently, there are no publicly available (and reliable) predictive tools available for bioinorganic researchers to predict log P (or log $D_{7.4}$) for complexes. The data may already be there – it just needs to be combined with DFT optimisation and processed.

5.5 References

1 D. Navot, N. Laufer, J. Kopolovic, R. Rabinowitz, A. Birkenfeld, A. Lewin, M. Granat, E. J. Margalioth and J. G. Schenker, *N. Engl. J. Med.*, 1986, **314**, 806–811, DOI: 10.1056/NEJM198603273141302.

2 J. O'Shaughnessy, *Oncologist*, 2003, **8**, 1–2, DOI: 10.1634/theoncologist.8-suppl_2-1.

3 D. Bellarosa, A. Ciucci, A. Bullo, F. Nardelli, S. Manzini, C. A. Maggi and C. Goso, *J. Pharmacol. Exp. Ther.*, 2001, **296**, 276–283.

4 N. Ismaili, M. Amzerin and A. Flechon, *J. Hematol. Oncol.*, 2011, **4**, 1–11, DOI: 10.1186/1756-8722-4-35.

5 A. J. J. Wood, R. Alexanian and M. Dimopoulos, *N. Engl. J. Med.*, 1994, **330**, 484–489, DOI: 10.1056/NEJM199402173300709.

6 F. A. van Nimwegen, M. Schaapveld, C. P. M. Janus, A. D. G. Krol, E. J. Petersen, J. M. M. Raemaekers, W. E. M. Kok, B. M. P. Aleman and F. E. van Leeuwen, *JAMA Intern. Med.*, 2015, **175**, 1007, DOI: 10.1001/jamainternmed.2015.1180.

7 S. Myrehaug, M. Pintilie, R. Tsang, R. Mackenzie, M. Crump, Z. Chen, A. Sun and D. C. Hodgson, *Leuk. Lymphoma*, 2008, **49**, 1486–1493, DOI: 10.1080/10428190802140873.

8 D. Cardinale, F. Iacopo and C. M. Cipolla, *Front. Cardiovasc. Med.*, 2020, **7**, 3–5, DOI: 10.3389/fcvm.2020.00026.

9 X. Wang, X. Zhang, D. Wu, Z. Huang, T. Hou, C. Jian, P. Yu, F. Lu, R. Zhang, T. Sun, J. Li, W. Qi, Y. Wang, F. Gao and H. Cheng, *Elife*, 2017, **6**, 1–20, DOI: 10.7554/eLife.23908.

10 G. W. Dorn, *Biochim. Biophys. Acta - Mol. Cell Res.*, 2013, **1833**, 233–241, DOI: 10.1016/j.bbamcr.2012.03.008.

11 E. Page and L. P. McCallister, *Am. J. Cardiol.*, 1973, **31**, 172–181, DOI: 10.1016/0002-9149(73)91030-8.

12 A. Rovini, K. Heslop, E. G. Hunt, M. E. Morris, D. Fang, M. Gooz, A. A. Gerencser and E. N. Maldonado, *FASEB J.*, 2021, **35**, 1–14, DOI: 10.1096/fj.202001693R.

13 B. Kadenbach, R. Ramzan, R. Moosdorf and S. Vogt, *Mitochondrion*, 2011, **11**, 700–706, DOI: 10.1016/j.mito.2011.06.001.

14 N. Wenningmann, M. Knapp, A. Ande, T. R. Vaidya and S. Ait-Oudhia, *Mol. Pharmacol.*, 2019, **96**, 219–232, DOI: 10.1124/mol.119.115725.

15 K. Renu, A. V.G., T. P. P.B. and S. Arunachalam, *Eur. J. Pharmacol.*, 2018, **818**, 241–253, DOI: 10.1016/j.ejphar.2017.10.043.

16 G. Takemura and H. Fujiwara, *Prog. Cardiovasc. Dis.*, 2007, **49**, 330–352, DOI: 10.1016/j.pcad.2006.10.002.

17 U. Schlattner, M. Tokarska-Schlattner and T. Wallimann, *Biochim. Biophys. Acta - Mol. Basis Dis.*, 2006, **1762**, 164–180, DOI: 10.1016/j.bbadis.2005.09.004.

18 R. A. Pagnanelli and D. A. Basso, *J. Nucl. Med. Technol.*, 2010, **38**, 1–3, DOI: 10.2967/jnmt.109.068593.

19 Y.-S. Kim, Z. He, W.-Y. Hsieh and S. Liu, *Bioconjug. Chem.*, 2007, **18**, 929–936, DOI: 10.1021/bc0603182.

20 A. G. Jones, M. J. Abrams, A. Davison, J. W. Brodack, A. K. Toothaker, S. J. Adelstein and A. I. Kassis, *Int. J. Nucl. Med. Biol.*, 1984, **11**, 225–234, DOI: 10.1016/0047-0740(84)90004-4.

21 J. D. Kelly, A. M. Forster, B. Higley, C. M. Archer, F. S. Booker, L. R. Canning, K. Wai Chiu, B. Edwards, H. K. Gill, M. McPartlin, K. R. Nagle, I. A. Latham, R. D. Pickett, A. E. Storey and P. M. Webbon, *J. Nucl. Med.*, 1993, **34**, 222–227.

22 V. Sharma, J. Sivapackiam, S. E. Harpstrite, J. L. Prior, H. Gu, N. P. Rath and D. Piwnica-Worms, *PLoS One*, 2014, **9**, e109361, DOI: 10.1371/journal.pone.0109361.

23 J. Sivapackiam, S. E. Harpstrite, N. P. Rath and V. Sharma, *Medchemcomm*, 2017, **8**, 158–161, DOI: 10.1039/C6MD00474A.

24 J. Sivapackiam, S. Kabra, S. Speidel, M. Sharma, R. Laforest, A. Salter, M. P. Rettig and V. Sharma, *PLoS One*, 2019, **14**, e0215579, DOI: 10.1371/journal.pone.0215579.

25 A. J. Smith, P. J. Gawne, M. T. Ma, P. J. Blower, R. Southworth and N. J. Long, *Dalton Trans.*, 2018, **47**, 15448–15457, DOI: 10.1039/C8DT02966K.

26 M. Kardashinsky, N. Lengkeek and L. M. Rendina, *J. Label. Compd. Radiopharm.*, 2017, **60**, 4–11, DOI: 10.1002/jlcr.3448.

27 Y. Q. Zhang, X. Shen, X. L. Xiao, M. Y. Liu, S. L. Li, J. Yan, J. Jin, J. L. Gao, C. L. Zhen, N. Hu, X. Z. Zhang, Y. Tai, L. S. Zhang, Y. L. Bai and D. L. Dong, *Br. J. Pharmacol.*, 2016, 3145–3158, DOI: 10.1111/bph.13578.

28 B. E. Osborne, T. T. C. Yue, E. C. T. Waters, F. Baark, R. Southworth and N. J. Long, *Dalton Trans.*, 2021, **50**, 14695–14705, DOI: 10.1039/d1dt02224e.

29 S. Liu, *Dalton Trans.*, 2007, 1183, DOI: 10.1039/b618406e.

30 B. E. Lang, *J. Chem. Eng. Data*, 2012, **57**, 2221–2226, DOI: 10.1021/je3001427.

31 J. A. Ouimet and A. S. Paluch, *J. Comput. Aided. Mol. Des.*, 2020, **34**, 575–588, DOI: 10.1007/s10822-020-00293-2.

32 M. Işık, T. D. Bergazin, T. Fox, A. Rizzi, J. D. Chodera and D. L. Mobley, *J. Comput. Aided. Mol. Des.*, 2020, **34**, 335–370, DOI: 10.1007/s10822-020-00295-0.

33 T. D. Bergazin, N. Tielker, Y. Zhang, J. Mao, M. R. Gunner, K. Francisco, C. Ballatore, S. M. Kast and D. L. Mobley, *Evaluation of log P, pK a, and log D predictions from the SAMPL7 blind challenge*, Springer International Publishing, 2021, vol. 35, DOI: 10.1007/s10822-021-00397-3.

34 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov,

J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian, Inc., Wallingford, CT, 2016.

35 S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129, DOI: 10.1016/0301-0104(81)85090-2.

36 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396, DOI: 10.1021/jp810292n.

37 M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669–681, DOI: 10.1002/jcc.10189.

38 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652, DOI: 10.1063/1.464913.

39 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241, DOI: 10.1007/s00214-007-0310-x.

40 Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656–5667, DOI: 10.1021/jp050536c.

41 R. Ditchfield, W. J. Hehre and J. A. Pople, *J. Chem. Phys.*, 1971, **54**, 724–728, DOI: 10.1063/1.1674902.

42 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297, DOI: 10.1039/b508541a.

43 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, , DOI:10.1063/1.3382344, DOI: 10.1063/1.3382344.

44 J. Demšar, T. Curk, A. Erjavec, C. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik and B. Zupan, *J. Mach. Learn. Res.*, 2013, **14**, 2349–2353.

45 J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc., Wallingford, CT, 3rd edn., 2015.

46 M. Bursch, J. Mewes, A. Hansen and S. Grimme, *Angew. Chem. Int. Ed.*, , DOI:10.1002/anie.202205735, DOI: 10.1002/anie.202205735.

47 N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, 115, 2315–2372, DOI: 10.1080/00268976.2017.1333644.

Chapter 6

Conclusions and Future work

6.1 Conclusions

The work in this thesis spans multiple applications of molecular imaging – from the synthesis of manganese-based contrast agents (MnBCAs) for PET-MRI, to developing methods that can both explore the chemotherapeutic application of doxorubicin, and aid in investigating its negative effects on mitochondria in the heart. In the first half of this thesis, novel ligand families for MnBCAs were investigated, and computational methods were developed to predict the key performance indicators (KPIs) that MnBCAs are evaluated against. The second half of the thesis described the DFT modelling of potential Zr-doxorubicin complexes inside a simulated liposomal formulation such as DOXIL, and then developing a predictive tool for gallium chemists to calculate the partition coefficient of their complex – a KPI required for cardiac mitochondrial function imaging.

The first results chapter of this work focused on investigating multiple ligand families for potential MnBCAs. Sulfonamides and amides were explored due to their ability to provide an X donor that is also able to form part of the ligand skeleton – which is advantageous for providing an overall monoanionic complex. Ligand designs were based around coupling the sulfonamide or amide bond *via* reacting an amine bridge with either a sulfonyl chloride, sulfonyl fluoride, acyl chloride, or carboxylic acid. Repeated attempts at amine protection/deprotection chemistry yielded three useful amine bridges. The synthesis of sulfonyl chlorides proved to be challenging, potentially due to the phenanthroline/pyridyl moiety enabling decomposition of the sulfonyl chloride before coupling could occur. SuFEx chemistry was also explored as an alternative route to building the sulfonamide ligands, however, like with the sulfonyl chloride route, attempts ended without success. As with the sulfonamides, macrocyclic amide formation met difficulty. Amide coupling *via* carboxylic acid and acyl chloride coupling methods were unsuccessful. A final attempt of closing the macrocycle *via* a double *N*-alkylation ended with aziridine formation instead. Acyclic phenanthroline amides proved to be more synthetically achievable, with complexation conditions eventually found to favour organic solvents with organic bases able to deprotonate the amide and complex to Mn^{2+}. Picolyl amides were explored to improve on the poor solubility observed with acyclic phenanthroline amide complexation. **PicAm-1** was synthesised in 43% overall yield. This ligand complexed well in anhydrous conditions with KHMDS and $MnCl_2$ in THF. ^{1}H NMRD profiles at 25 °C and 37 °C were recorded, which revealed two dispersions (at both high and low fields) suggesting that the coordination of the complex was not what was expected. This, coupled with the large r_1 at 0.01 MHz indicating a q of 3, suggested that the amido donor (as the highest pK_a donor) and pyridyl donor may have decomplexed in solution.

Complementary to the synthetic MnBCA work, data analysis was carried out investigating the KPIs for synthetic chemists in the MnBCA field. Chief amongst these is the stability constant – crucial for understanding the ability of the MnBCA to remain complexed in solution. Geometry optimisation and

frequency calculations were undertaken on a collection of the most studied MnBCAs and their corresponding ligands with Gaussian 16.[1] A range of functional, basis set, and solvation model combinations were tested, with permutations designed to range from low computational cost to best-practice for accurate results. The DFT models were trained on a selection of 10 MnBCAs to expose the models to a variety of MnBCA ligand types. The resultant thermodynamic values from the training optimisations were used to create calculated $\log(K_{MnL})$ values, which were correlated to experimental values *via* linear least squares regression. From all permutations, the three models with the highest correlation (*via* R^2) were taken forward to validate against a testing dataset to test prediction accuracy, and each model's ability to differentiate isomers. Against the testing dataset, the three models gave predictions with mean absolute errors (MAEs) of 2.23, 1.84, 1.79 log units respectively, and root mean square errors (RMSEs) of 4.91, 3.72, 3.92 log units respectively; indicating model 2 ('with water × IEFPCM(SMD) × PW6B95D3/TZVP') performed best against the testing dataset. These data showed that whilst the DFT methods investigated could not outperform the empirical approach of Uzal-Varela (MAE of 0.99 log units) which was trained on 150 complexes,[2] its prediction accuracy of ± ~2 log units is still within a useful range for synthetic chemists – especially when designing novel ligands that are not part of the original Uzal-Varela analysis. Alongside this, the models were tested to differentiate between isomers – which produce identical predictions in the Uzal-Varela method. All three models predicted the relative positions of the tested CDTA, and PC2A isomers, however struggled with the relative positions for Mn^{2+} complexes of 1,3-CBuDTA and 1,4-CBuDTA.

Additionally, the bond lengths and geometries from DFT geometry optimisations (M06-2X/TZVP/IEFPCM) on 47 MnBCAs were analysed for any correlation to MnBCA KPIs. Overall, both continuous shape measurements (a measure of deviation from perfect geometry) and bond length distributions did not correlate significantly to the stability constant, water exchange rate, or hydration number. The large dataset from the geometry analysis was expanded to include their corresponding ligands to repeat the primary analysis from the MnBCA DFT work. With an expanded dataset, the R^2 value for the 'with water' × IEFPCM × M06-2X/TZVP model's predictions increased from 0.142 to 0.199 (Figure 6.1) – suggesting improved results would be obtained for the best three models above with more data in the training dataset. The average residual value between prediction and experimental was 0.00 log units, but with σ of 5.76 log units, indicating that there was not an inherent error in the method, but rather that the MnBCA dataset has a large degree of variance when analysed *via* DFT. This hampered efforts to provide useful predictive tools – as seen in the wide 95% confidence interval in Figure 6.1.

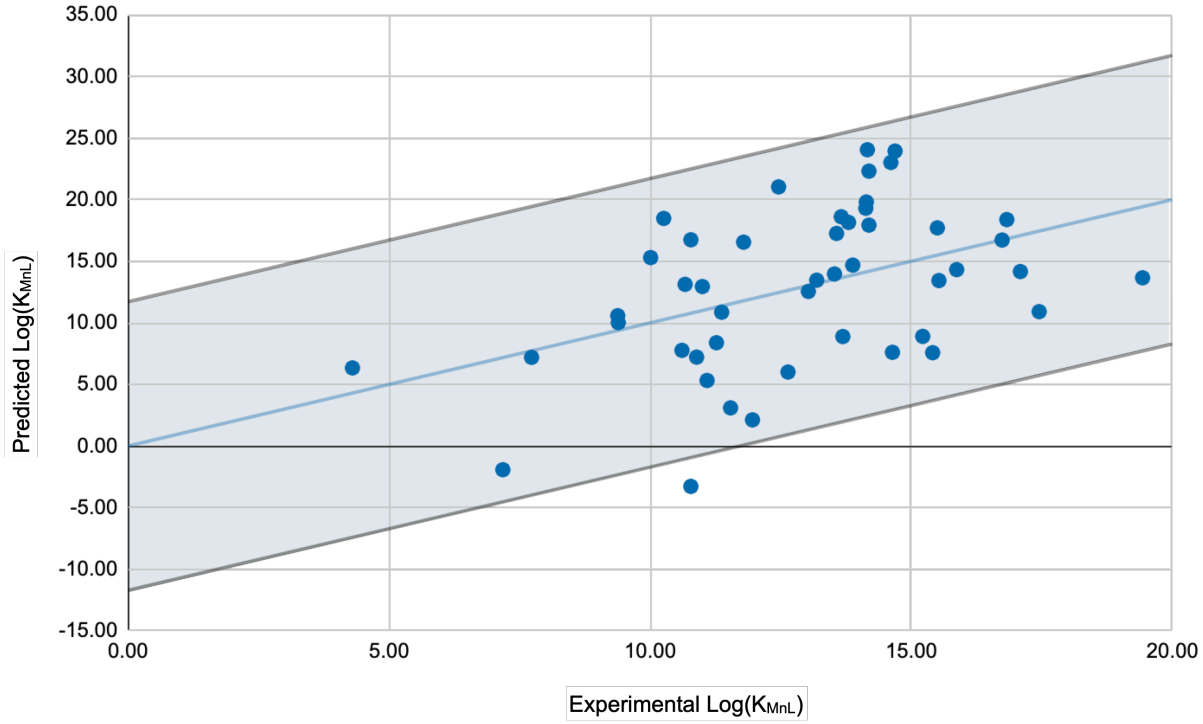


Figure 6.1: Predicted vs experimental log(K_{MnL}) values from the 'with water × M06-2X/TZVP × IEFPCM' model on 47 MnBCAs in Chapter 3, plotted with a 95% confidence interval; R^2: 0.199.

Finally, machine learning (ML) methods were used to find correlations between the data generated *via* DFT optimisations and the stability constant. ML models were able to produce R^2 values of their predictions on unpruned datasets that outperformed simple linear regression – SVR with the 'DFT energies' dataset produced an R^2 of 0.371. Dataset cleaning (removing MnBCAs that were not clinically viable) greatly improved the predictive power of the multivariate linear regression and neural network models. The linear regression model using 'handpicked' variables (those that most contributed to improving the R^2 of prediction (CShM, coordination number, geometry type, molecular weight, the six DFT features (thermal and Gibbs energies of ligand and complex, calculated log(K_{MnL}), and the ΔG used in the log(K_{MnL}) calculation), *q*, the four DFT charge calculations, and average Mn-carboxylate, Mn-donor, Mn-water, Mn-3° amine bond lengths)) predicted log(K_{MnL}) with an R^2 of 0.879, MAE of 0.50 log units, and RMSE of 0.52 log units, far surpassing the required level of accuracy needed to guide a decision on a proposed MnBCA's stability constant.

The other half of the work in this thesis concerned the identity of the resultant zirconium complex inside a doxorubicin containing liposome and predicting the partition coefficient (log P) of gallium complexes primarily used to image cardiac mitochondria (investigating the cardiotoxic effect of anthracyclines such as doxorubicin). For the liposomal modelling work, a proposed dissociative ligand substitution was modelled between $Zr(oxinate)_4$ and a fragment of doxorubicin in aqueous solution *via* DFT calculations in Gaussian 16. The M06-2X functional with Sapporo-TZP-2012 basis set for Zr,

6-31G(d,p) for all other atoms, and SMD solvation model was chosen after testing a range of functionals and basis sets for accuracy and computational cost in replicating the geometry of the crystal structure of $Zr(oxinate)_4$. The experiments simulated acidic, neutral, and basic environments, and found that the ligand substitution of oxine for the doxorubicin fragment was thermodynamically feasible. Boltzmann distribution calculations indicated that 99.7% of zirconium would be complexed as $Zr(Dox)_4$ in neutral conditions, and 97.2% as $[Zr(oxinate)(Dox)_2]^+$ in acidic conditions. Simulated basic conditions showed that essentially all zirconium would stay as $Zr(oxinate)_4$ (100.0%), indicating a pH sensitivity to the reaction. Alongside this, the effect of hydrating the 6-coordinate zirconium complexes after dissociation was investigated and showed greater stability at 6-coordinate than with two complexed water molecules. Increasing the modelled temperature from 298.15 K to 310.15 K was also explored and was found to affect the distribution of zirconium complexes only marginally under all simulated conditions. Given the likely acidic pH within the doxorubicin containing liposome, under the modelled conditions, DFT calculations suggest that the conversion of $Zr(oxinate)_4$ to $[Zr(oxinate)(doxorubicinate)_2]^+$ would occur.

Finally, Gaussian 16 was used to calculate the log P values of gallium complexes with applications in myocardial perfusion imaging *via* DFT calculations. Simple linear regression of calculated log P with experimental log P was able to provide an excellent correlation (R^2 of 0.829, MAE of 0.88 log units, RMSE of 1.03 log units, Figure 6.2) using the PW6B95D3 functional, TZVP basis set, and SMD solvation model, surpassing expectations from the SAMPL6 and SAMPL7 challenges.[3,4]

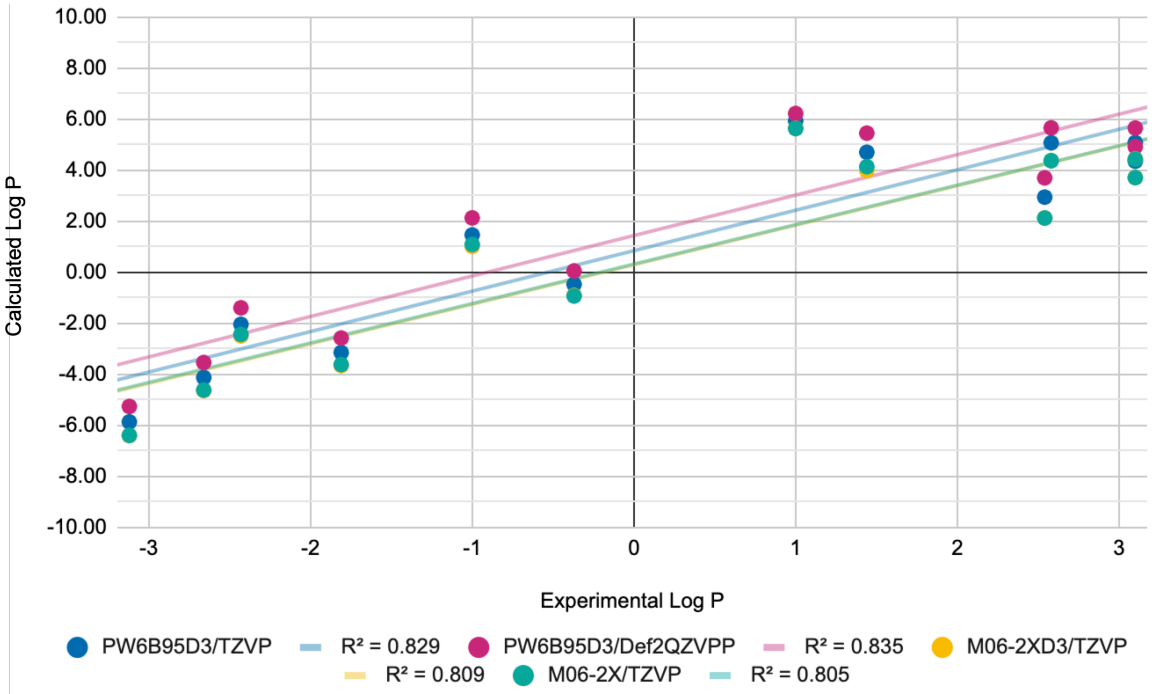


Figure 6.2: Calculated log P of gallium complexes in Chapter 5 plotted against experimental log P for the analysed model chemistries (all with SMD).

In addition to the high R^2 of the computationally expensive PW6B95D3 functional, both M06-2X and M06-2X with the D3 version of Grimme's dispersion correction were able to provide good predictions at lower computational cost (R^2 of 0.805 and 0.809 respectively). These simple models are able to provide predictions at accuracies exceeding the required level to evaluate whether a gallium complex will be lipophilic enough to be a successful myocardial perfusion imaging agent. The ML methods further improved on this result, with both multivariate regression and a neural network providing even better correlation between predicted and experimental log P values (0.982 (0.935 with regularisation) and 0.977 respectively), using calculated log P values from PW6B95D3/TZVP, M06-2X/TZVP (with and without D3), and B3LYP/6-31G(d,p) (with and without D3) as inputs for the models. The ability to predict log P in literature has so far largely been restricted to SAMPL challenges on small molecules, which largely focus on accuracy over utility. For the wider field, this has led to improving solvation models with COSMO-RS and SMD, or using machine learning methods to deconstruct a molecule into parts readable by the network which (through back-propagation) are then linked to contributions to log P.[55] These approaches either use non-standard software packages such as COSMOtherm, or require large datasets to build a neural network model. The models developed in this thesis provide myocardial perfusion imaging chemists with the means to prototype their designs *via* two DFT calculations with readily available software, improving the efficiency of their design processes.

6.2 Future work

Developing a reliable synthetic route to form sulfonamides for pyridyl/phenanthroline-based ligands would be beneficial for MnBCAs due to their rigidity, lipophilicity, and inclusion of an X donor as part of the ligand skeleton. Multiple general method routes were attempted during this work; none were ultimately successful, potentially due to decomposition to a chloride aided by the nitrogen on the aryl ring. Inspiration could be taken from the synthesis of PyFluor, in which 2-mercaptopyridine is oxidised with NaOCl (as with our attempted methods), and then trapped with fluoride from sat. KHF_2 in MeCN.[8] With a phenanthroline sulfonyl fluoride in hand, further investigation of the unsuccessful SuFEx reactions with PyFluor would be merited, initially evaluating the reaction with primary amines, before moving to a range of protected secondary amines. As for the amide-based ligands, synthesising **MnPicAm-1**, and **MnPicAm-2** in greater quantities for further characterisation and evaluation of their suitability as MnBCAs would be desired. However, the initial results of the ^{1}H NMRD profile of **MnPicAm-1** suggest that the Mn^{2+} complex may not be stable enough to be appropriate as MnBCAs – indicating a potential mismatch between Mn^{2+} and amido-donors not seen for Mn^{3+} complexes. As such, further development of $q = 2$ Mn^{3+} complexes is warranted, as despite the lower number of unpaired electrons, having a larger hydration number and stability, as well as simpler complexation conditions could make them ideal candidates for MnBCAs. Lastly, increasing lipophilicity is typically considered

for rational design of MnBCA ligands. However, as of yet, there has not been a study understanding the effect of log P on biodistribution, or elimination profile for MnBCAs. This potential study could be multifaceted, experimentally measuring log P values, and testing the biodistribution of the most widely studied MnBCAs, whilst also using the methodology developed in Chapter 5 to predict the log P of MnBCAs before synthesis.

Due to the low predictive power of the models from the MnBCA KPI prediction *via* DFT studies, efforts should be made to improve their accuracy. As the solvation model is likely to be the most influential in the Gibbs free energy calculations, evaluating the COSMO-RS solvation model *via* COSMOtherm software may lead to improvements in the R^2 values of the models. Additionally, more accurately sampling the true environment of the MnBCA in solution (i.e. including more explicit water molecules into the geometry optimisations; identifying and optimising conformers and isomers; consideration of different protonation states) could provide more accurate Gibbs values, however these additions would likely come at a prohibitive cost. Therefore, smaller scale evaluation of the effect of each change would be required to only focus on the changes with most effect on improving the accuracy of stability constant prediction. The success of the ML models with M06-2X/TZVP/IEFPCM may be improved further by instead using the functional, basis set, and solvation model combinations for the most successful model when calculating $\log(K_{MnL})$ in Chapter 3 (PW6B95D3/TZVP/IEFPCM(SMD) with two explicit solvent water molecules). As with all ML methods, providing more useful data to train with is likely to improve the ability of the model to describe reality, as the effect of outliers in the dataset should reduce. Ultimately, once reliable prediction of the stability constant is achieved, these methods should be able to be translated to accurately predicting r_1 once more precise r_1 data is experimentally measured.

For the zirconium DFT study, evaluating the conclusions experimentally should be the primary next step to understand the utility of the method. This would require replicating the conditions within the liposome (i.e. pH and reactant concentrations), and spectroscopically determining species distribution *via* GC-MS or NMR. These results would help to understand how impactful each of the approximations made in the study were towards the final result. Additionally, the DFT method itself could be improved with wider functional and basis set testing. As the geometries of each of the complexes have been optimised at a suitable level, low-cost SPE calculations (as were used in Chapter 5) could be used to rapidly test more complete basis sets and more powerful functionals.[9]

The log P prediction of gallium complexes used in myocardial perfusion imaging surpassed expectations. To further strengthen these predictions, more gallium complexes should be synthesised, have their log P experimentally determined, and then DFT geometry optimised. Expanding the number of ligands used to train the models (both the simple linear regressions and the ML models) should improve accuracy. Alongside this, increasing the number of gallium complexes allows for the potential

of investigating different aspects of the model. For instance, with the current data, the model assumes a linear relationship between what is calculated with DFT and the experimental log P value. Including very lipophilic and very lipophobic complexes will test how well the relationship holds up at the extremes, therefore serving to evaluate the robustness of the model.

6.3 Final remarks

Overall, this thesis has detailed efforts to improve the design process of molecular imaging agents. The field largely lacks validation of design principles – in part due to how difficult and time-consuming synthesis is. There are some good examples where analytics or computational chemistry in literature has helped to improve the understanding of the best ligand design considerations, for example Uzal-Varela and co-workers' publications on GBCA and MnBCA's stability constants, or Holland's work in predicting the stability constants of zirconium radiotracers.[2,6,7] However, there remains a large unmet need across the field – as seen in the gallium work – for analytics that could help improve efficiency and drive 'rational design'. To this end, a greater adoption of statistical literacy and prototyping (for example *via* quantum mechanical DFT methods, or empirical approaches such as using ML methods) would serve to advance our understanding of the underlying chemistry, reduce wasted effort, and hopefully improve the rate of innovation in medical imaging.

6.4 References

1 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian, Inc., Wallingford, CT, 2016.

2 R. Uzal-Varela, F. Pérez-Fernández, L. Valencia, A. Rodríguez-Rodríguez, C. Platas-Iglesias, P. Caravan and D. Esteban-Gómez, *Inorg. Chem.*, 2022, **61**, 14173–14186, DOI: 10.1021/acs.inorgchem.2c02364.

3 M. Işık, T. D. Bergazin, T. Fox, A. Rizzi, J. D. Chodera and D. L. Mobley, *J. Comput. Aided. Mol. Des.*, 2020, **34**, 335–370, DOI: 10.1007/s10822-020-00295-0.

4 T. D. Bergazin, N. Tielker, Y. Zhang, J. Mao, M. R. Gunner, K. Francisco, C. Ballatore, S. M. Kast and D. L. Mobley, *J. Comput. Aided. Mol. Des.*, 2021, **35**, 771–802, DOI: 10.1007/s10822-021-00397-3.

5 D. Roy and C. Patel, *Molecules*, 2023, **28**, 801, DOI: 10.3390/molecules28020801.

6 R. Uzal-Varela, A. Rodríguez-Rodríguez, H. Wang, D. Esteban-Gómez, I. Brandariz, E. M. Gale, P. Caravan and C. Platas-Iglesias, *Coord. Chem. Rev.*, 2022, **467**, 214606, DOI: 10.1016/j.ccr.2022.214606.

7 J. P. Holland, *Inorg. Chem.*, 2020, **59**, 2070–2082, DOI: 10.1021/acs.inorgchem.9b03515.

8 M. K. Nielsen, C. R. Ugaz, W. Li and A. G. Doyle, *J. Am. Chem. Soc.*, 2015, **137**, 9571–9574, DOI: 10.1021/jacs.5b06307.

9 M. Bursch, J. Mewes, A. Hansen and S. Grimme, *Angew. Chem. Int. Ed.*, 2022, **61**, e202205735, DOI: 10.1002/anie.202205735.

Chapter 7

Experimental Procedures

7.1 General considerations

All reactions were conducted in air, at atmospheric pressure, and with stirring unless otherwise stated; anhydrous solvents were used in all reactions under a nitrogen environment. Reactions under a nitrogen environment were carried out in a nitrogen-filled glovebox or with standard Schlenk techniques. All reagents and solvents were obtained from commercial sources (Sigma Aldrich, Fluorochem, Alfa Aesar, VWR, and Tokyo Chemical Industry) and used without further purification, unless otherwise stated. Anhydrous solvents (hexane, toluene, DCM, THF) were sources from MBraun MB-SPS 800 solvent towers and stored over 3 Å molecular sieves and degassed for 1 hour. Unless otherwise stated, flash column chromatography was performed automatically using a Biotage Isolera Four unit and 5 g, 10 g, 25 g, 50 g, or 100 g Biotage® Sfär Duo Flash Cartridge. Reversed-phase flash column chromatography was also performed with the Biotage Isolera Four unit, and a 6 g or 12 g Biotage® Sfär C18 (HP-Sphere C18 / SNAP ULTRA C18) Flash Cartridge. Mobile phase A contained H_2O with 0.1% TFA, and mobile phase B contained MeCN with 0.1% TFA. Mobile phase C contained NH_4OAc buffer (pH 7.4), and mobile phase D contained MeCN with 10% NH_4OAc buffer (pH 7.4)). Thin-layer chromatography was performed on silica-coated aluminium sheets.

^{1}H, $^{13}C\{^{1}H\}$, $^{31}P\{^{1}H\}$, $^{19}F\{^{1}H\}$, COSY, HSQC, HMBC NMR spectra were recorded on a Bruker Av400 spectrometer at 298 K unless otherwise stated. Chemical shifts are reported in ppm and referenced to residual signals from the solvent for NMR spectra. Where required, $^{13}C\{^{1}H\}$ NMR spectra were assigned with the help of 2D correlation spectroscopy, such as HMQC and HMBC, or with DEPT 45, 90, or 135 spectra. NMR spectra were processed and analysed using MestReNova software. Signal multiplicities used when analysing are as follows: s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet), dd (doublet of doublets), dt (doublet of triplets), ddd (doublet of doublets of doublets), ddt (doublet of doublets of triplets), qd (quartet of doublets), br (broad).

High resolution electrospray ionisation mass spectrometry (ESI-HRMS) was performed at Imperial College London on a Waters LCT Premier mass spectrometer by Dr Lisa Haigh and Malgorzata Puchnarewicz of the mass spectrometry service. MALDI ToF mass spectrometry was performed on Waters/Micromass MALDI micro MX spectrometer with MassLynx mass spectrometry software using positive ionisation.

Single crystal X-ray diffraction (XRD) data was collected and analysed by Dr Andrew White at Imperial College London.

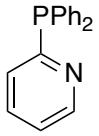
^{1}H NMRD profiles between 0.01 and 10 MHz, at 25 °C and 37 °C, were measured with a bench-top fast field cycling relaxometer fitted with a VTC90 temperature control unit (Stelar SMARtracer™). The

concentration of manganese complexes was measured *via* ^{1}H NMR measurements with the well-established Evans method.[1]

7.2 Synthetic chemistry

7.2.1 Phosphasalens

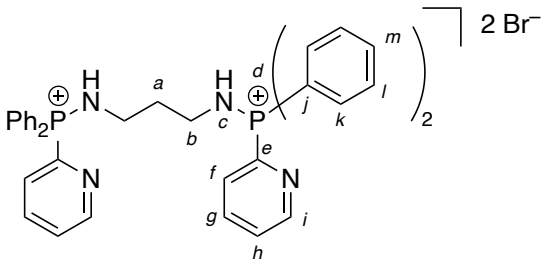
Diphenyl-2-pyridyl phosphine (2.1)[2]



2-Bromopyridine (1.45 mL, 15.2 mmol) was dissolved in diethyl ether (anhydrous, 45 mL) and cooled to -78 °C, under nitrogen. ^{n}BuLi (6.08 mL, 2.5 M in hexanes, 15.2 mmol) was added dropwise at -78 °C. The dark red solution was stirred at room temperature for 30 minutes. PPh_2Cl (2.73 mL, 15.2 mmol) was added *via* syringe at -78 °C, the resultant light brown solution was stirred overnight at 20 °C, to form a clay-coloured suspension. Solvent was removed *via* vacuum, leaving an orange solid coating the walls of the Schlenk flask. The product was extracted *via* hexane (1 × 30 mL, 1 × 10 mL, 1 × 5 mL). Solvent was removed *via* vacuum, the yellow solid was used without further purification, (2.43 g, 9.21 mmol, 30%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 8.72 (d, $^{3}J_{HH}$ = 3.6 Hz, 1H), 7.55 (t, $^{3}J_{HH}$ = 6.0 Hz, 1H), 7.45–7.30 (m, 10H), 7.17 (t, $^{3}J_{HH}$ = 6.2 Hz, 1H), 7.08 (d, $^{3}J_{HH}$ = 7.6 Hz, 1H).
^{31}P{^{1}H} NMR (121.5 MHz, $CDCl_3$, δ, ppm): −4.03 (s).

Phosphasalen 1 Salt (PS1-salt)

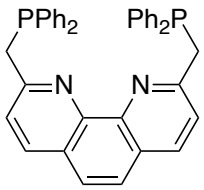


2.1 (33.0 mL, 0.129 M in DCM, 4.27 mmol) was cooled to -78 °C. Br_2 (200 μL, 3.88 mmol) was added dropwise at -78 °C, this dark red/brown solution was stirred at 20 °C for 2 hours, under nitrogen. At

-78 °C, tributylamine (924 µL, 3.88 mmol), then 1,3-diaminopropane (162 µL, 1.94 mmol) were added, and the resultant dark brown solution was stirred at 20 °C for 18 hours. Solvent was removed *via* vacuum, to afford a dark brown gummy, gelatinous solid. THF (anhydrous, 15 mL) was added, and sonicated, to give a light brown suspension. This was stirred overnight to afford a tan suspension. The suspension was filtered, washed with THF (5 × 5 mL), and dried under vacuum. Light beige/off-white powder collected (1.07 g, 1.40 mmol, 72%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 8.83 (d, $^{3}J_{HH}$ = 4.2 Hz, 2H, $C^{i}H$), 8.60 (dd, $^{3}J_{HH}$ = 7.6 Hz, $^{3}J_{HH}$ = 6.3 Hz , 2H, $C^{f/g}H$), 8.10 (m, 2H, $C^{f/g}H$), 7.92–7.80 (m, 10H, $C^{k/l}H$, and NH or $C^{h}H$), 7.76-7.69 (m, 4H, $C^{m}H$), 7.68-7.55 (m, 10H, $C^{k/l}H$, and $C^{h}H$ or NH), 3.22 (m, 4H, $C^{b}H$), 1.98 (m, 2H, $C^{a}H$).
^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 151.3 (d, $^{3}J_{CP}$ = 21.2 Hz, C^{j}), 147.5 (d, $^{1}J_{CP}$ = 536 Hz, C^{e}) 138.1 (d, $^{3}J_{CP}$ = 10.1 Hz, C^{g}), 134.8 (d, $^{4}J_{CP}$ = 3.0 Hz, C^{m}), 134.0 (d, $^{3}J_{CP}$ = 11.1 Hz, C^{l}), 131.2 (d, $^{2}J_{CP}$ = 23.2 Hz, C^{f}), 129.8 (d, $^{2}J_{CP}$ = 14.1 Hz, C^{k}), 127.5 (d, $^{4}J_{CP}$ = 3.5 Hz, C^{h}), 120.7 (d, $^{1}J_{CP}$ = 103.3 Hz, C^{j}), 40.1 (s, C^{b}), 33.1 (s, C^{a}).
^{31}P{^{1}H} NMR (121.5 MHz, $CDCl_3$, δ, ppm): 30.8 (s).
HR-MS (ES^+ ToF, g mol^{-1}): Found: 677.1605; Calc'd: 677.1599, $C_{37}H_{36}N_4BrP_2$ [M-Br].

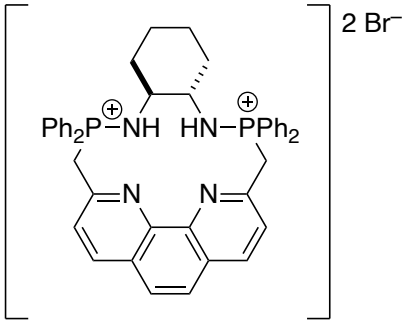
2,9-Bis((diphenylphosphino)methyl)-1,10 phenanthroline (2.2)[3]



In a Schlenk flask, neocuproine (0.95 g) was dissolved in THF (anhydrous, 35 mL) and cooled to -40 °C, under nitrogen. LDA solution (18.7 mL, 1 M in THF and hexanes) was added *via* syringe at -40 °C, the resultant dark red/black solution stirred for 1 hour at 20 °C. PPh_2Cl (1.64 mL) was added *via* syringe at 0 °C and stirred for 18 hours at room temperature. The resultant dark red/black solution was quenched with degassed, distilled water (2 mL) and stirred for 1 hour. Upon addition an orange solid precipitated. Solvent was removed *via* vacuum and external trap. DCM (anhydrous, 30 mL) was added to dissolve the yellow and white residue coating the Schlenk walls, which was then washed with degassed, distilled water (2 x 10 mL). The orange organic layer was then dried with $MgSO_4$ and cannula filtered. Solvent was removed *via* vacuum and external trap, creating a sticky orange foam as DCM evaporated. Upon evaporating to dryness *via* Schlenk line, an orange flake foam was collected, and dried and used without further purification (2.41 g, mmol, 92% crude).

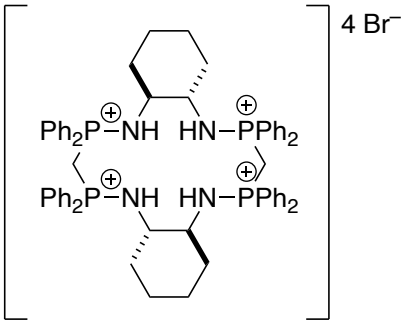
$^{31}P\{^{1}H\}$ NMR (121.5 MHz, $CDCl_3$, δ, ppm): -10.23 (s).

Phosphasalen 2 Salt (PS2-salt) – attempted synthesis



2.2 (0.40 g, 0.69 mmol) was dissolved in DCM (anhydrous, 20 mL) and was cooled to -78 °C, under nitrogen. Br_2 (71 µL, 1.39 mmol) was added dropwise at -78 °C, this dark red/brown solution was stirred at 20 °C for 2 hours. At -78 °C, tributylamine (333 µL, 1.39 mmol), then (±)-*trans*-1,2-diaminocyclohexane (83 µL, 0.69 mmol) were added, and the resultant dark red solution was stirred at 20 °C for 16 hours. Solvent was removed via vacuum, to afford a dark brown foam solid. THF (anhydrous, 15 mL) was added and the Schlenk flask was sonicated. This afforded a very thick, tan suspension. This suspension was gravity filtered through a G4 sintered funnel and washed with THF (7 × 5 mL). The resultant tan powder was transferred to a vial and dried *via* vacuum on the Schlenk line for 1 hour. A tan powder was collected (crude 0.35 g, 0.41 mmol, 59%). Purification *via* manual flash column chromatography on basified alumina (*via* NEt_3 (0.5 mL); 0 → 2.5% EtOH in DCM) was carried out, however fractions did not show evidence of product when characterised *via* HR-MS.

Phosphasalen 1 Salt (PS3-salt) – attempted synthesis



In a Schlenk flask, DPPM (293 mg, 0.76 mmol) was added to DCM (anhydrous, 20 mL) and cooled to -78 °C, under nitrogen. Br_2 (78 µL, 1.52 mmol) was added dropwise at -78 °C and stirred at 20 °C for

2 hours. A white precipitate was formed upon addition. At -78 °C, tributylamine (365 µL, 1.52 mmol), then (±)-*trans*-1,2-diaminocyclohexane (91 µL, 0.76 mmol) were added, and the resultant lime solution was stirred at 20 °C for 16 hours. Solvent was removed *via* vacuum, to afford a yellow flake foam solid. THF (anhydrous, 15 mL) was added, and the Schlenk flask was sonicated to remove solids from the walls. This THF suspension was transferred to a centrifuge tube and centrifuged at 4500 rpm for 5 minutes. The supernatant was pipetted off and the wet solid was dried *in vacuo*. Off-white solid collected (crude 179 mg, 0.14 mmol, 36%).

HR-MS (ES^+ ToF, g mol^{-1}): Found: 248.1093; Calc'd: 248.1098, $C_{62}H_{68}N_4P_4$ [M-4Br] or $C_{31}H_{34}N_2P_2$ [half of the macrocycle]; Found: 495.2109; Calc'd: 495.2114, $C_{62}H_{66}N_4P_4$ [M-2H-4Br] or $C_{31}H_{33}N_2P_2$ [half of the macrocycle, singly deprotonated].

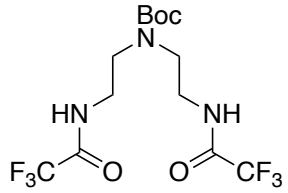
7.2.1.1 Phosphasalen hydrolysis experiment

6,6'-((1,2-phenylenebis(azaneylylidene))bis(diphenyl-λ^5-phosphaneylylidene))bis(2,4-di-*tert*-butylphenol) (**CB134**) (20 mg, 22.5 µmol) was added to water (2 mL) and stirred for 24 hours. $^{31}P\{^1H\}$ NMR was carried out showing evidence of iminophosphorane hydrolysis.

$^{31}P\{^1H\}$ NMR (121.5 MHz, $CDCl_3$, δ, ppm): 38.4 (s, **CB134**), 25.5 (s, expected phosphine oxide peak).

7.2.2 Sulfonamides

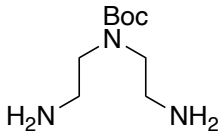
T*ert*-butyl bis(2-(2,2,2-trifluoroacetamido)ethyl)carbamate (2.10)[4]



Diethylenetriamine (1.40 mL, 10.0 mmol) and DCM (20 mL) were combined and cooled to 0 °C. Ethyl trifluoroacetate (2.62 mL, 22.0 mmol) was dissolved in DCM (10 mL) and added to the previous solution at 0 °C. This was stirred for 30 minutes at 0 °C. DCM (5 mL) was added, and the reaction was warmed to 20 °C and stirred for 1 hour. Boc_2O (4.80 g, 22.0 mmol) and NEt_3 (3.07 mL, 22.0 mmol) in DCM (10 mL) were added dropwise and stirred overnight at 20 °C. This colourless solution was washed with sat. $NaHCO_3$ (2 × 20 mL), distilled water (2 × 20 mL), dried *via* Na_2SO_4, and then concentrated *via* rotary evaporator to one third of the volume. An equal volume of hexane was added, and the solution was stored overnight at -20 °C, to provide white crystals. These were collected, washed with hexane and water, and dried under vacuum to provide a white powder (2.06 g, 4.86 mmol, 49%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 3.15 (m, 8H, CH_2), 1.67 (m, 4H, CH_2), 1.37 (m, 9H, CH_3).
HR-MS (ES^+ ToF, g mol^{-1}): Found: 446.1488; Calc'd: 446.1490, $C_{15}H_{23}N_3O_4F_6Na$ [M+Na].

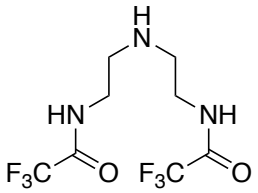
Tert-butyl bis(2-aminoethyl)carbamate (2.9-Boc)[4]



2.10 (2.06 g, 4.86 mmol) was dissolved in MeOH (50 mL). Aqueous NaOH (3 M, 42 mL), was added and the solution was stirred at 20 °C for 6 hours. The solution was concentrated *via* rotary evaporator to remove most of the MeOH, and then extracted with DCM (3 × 40 mL), dried *via* Na_2SO_4, and then concentrated *via* rotary evaporator to provide a colourless oil (0.91 g, 3.91 mmol, 39 % overall).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 3.24 (d, $^3J_{HH}$ = 20.4 Hz, 4H), 2.67 (t, $^3J_{HH}$ = 6.8 Hz, 4H), 1.64 (m, $^3J_{HH}$ = 6.9 Hz, 4H), 1.44 (s, 9H), 1.20 (s, 4H).
^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 155.9, 79.5, 44.4, 43.9, 39.6, 39.5, 32.7, 32.1, 28.6.

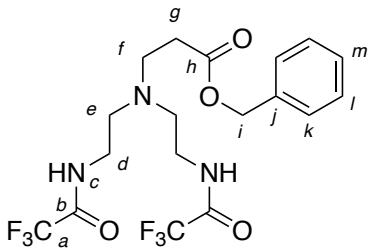
***N*,*N*'-(azanediylbis(ethane-2,1-diyl))bis(2,2,2-trifluoroacetamide) (2.11)[4]**



Diethylenetriamine (1.08 mL, 10.0 mmol) in DCM (10 mL) was cooled to 0 °C. Ethyl trifluoroacetate (2.62 mL, 22.0 mmol) was dissolved in DCM (10 mL) and added to the previous solution at 0 °C and stirred for 30 minutes at 0 °C. DCM (5 mL) was added, and the reaction was warmed to 20 °C and stirred for 1 hour. TLC showed consumption of diethylenetriamine. This colourless solution was then concentrated *via* rotary evaporator to provide a white solid (2.95 g, 10.0 mmol, 100%).

^{1}H NMR (400 MHz, DMSO-d_6, δ, ppm): 9.36 (s, 2H), 3.40 (br s, 1H), 3.26 (t, $^3J_{HH}$ = 6.5 Hz, 4H), 2.69 (t, $^3J_{HH}$ = 6.5 Hz, 4H).

Benzyl 3-(bis(2-(2,2,2-trifluoroacetamido)ethyl)amino)propanoate (2.12)



2.11 (295 mg, 1.00 mmol), benzyl acrylate (146 µL, 1.00 mmol), and water (1 mL) were combined and sonicated at 20 °C for 2 hours. This biphasic mixture was extracted with EtOAc (2 × 5 mL), dried with Na_2SO_4, and solvent removed *via* vacuum. Crude oil was then purified *via* automated column chromatography (0-100% EtOAc in hexane, product R_f 0.35 in 1:1 EtOAc/hexane TLC), to provide a colourless oil (160 mg, 0.35 mmol, 35%).

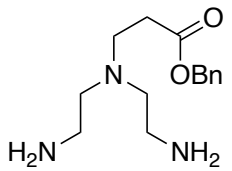
^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 7.42 – 7.30 (m, 5H, Ph), 7.25 (br s, 2H, N^{c}H), 5.13 (s, 2H, C^{i}H), 3.39 (q, $^3J_{HH}$ = 5.7 Hz, 4H, C^{d}H), 2.74 (t, $^3J_{HH}$ = 6.5, 2H, C^{g}H), 2.61 (t, $^3J_{HH}$ = 5.7 Hz, 4H, C^{e}H), 2.50 (t, $^3J_{HH}$ = 6.5 Hz, 2H, C^{f}H).

$^{13}C\{^1H\}$ NMR (101 MHz, $CDCl_3$, δ, ppm): 173.5 (C^h), 157.5 (q, $^2J_{CF}$ = 37.7 Hz, C^b), 116.0 (q, $^1J_{CF}$ = 287.5 Hz, C^a), 135.3 (C^j), 128.8 ($C^{k/l/m}$), 128.8 ($C^{k/l/m}$), 128.7 ($C^{k/l/m}$), 67.1 (C^i), 52.3 (C^e), 48.2 (C^g), 37.2 (C^d), 32.5 (C^f).

$^{19}F\{^1H\}$ NMR (377 MHz, $CDCl_3$, δ, ppm): -76.0.

HR-MS (ES^+ ToF, g mol^{-1}): Found: 458.1520; Calc'd: 458.1515, $C_{18}H_{21}F_6N_3O_4$ [M+H].

Benzyl 3-(bis(2-aminoethyl)amino)propanoate (2.6-Bn) – attempted synthesis 1



MeOH (3.6 mL), H_2O (1.4 mL), and K_2CO_3 (500 mg) were combined and added to **2.12** (51 mg, 0.14 mmol). This was stirred for 2 hours at 20 °C, then the pH was adjusted to pH 8 *via* the drop-wise addition of HCl (12 M). The residue was dissolved in H_2O (15 mL) and extracted with $CHCl_3$ (3 × 10 mL), then washed with H_2O (10 mL). The organic phase was dried with Na_2SO_4, and solvent removed *via* vacuum, to provide a colourless oil which was identified as benzyl alcohol.

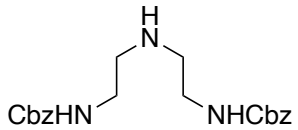
Benzyl 3-(bis(2-aminoethyl)amino)propanoate (2.6-Bn) – attempted synthesis 2

MeOH (3.6 mL), H_2O (1.4 mL), and K_2CO_3 (5 mg) were combined and added to **2.12** (80 mg, 0.22 mmol). This was stirred for 48 hours at 20 °C, pH was monitored and kept at pH 10 *via* addition of K_2CO_3. This solution was concentrated *via* vacuum, to provide a colourless oil which was identified as benzyl alcohol.

Benzyl 3-(bis(2-aminoethyl)amino)propanoate (2.6-Bn) – attempted synthesis 3

2.12 (181 mg, 0.50 mmol) was dissolved in ethanol (10 mL), and $NaBH_4$ (189 mg, 5.00 mmol) was added. This was stirred at 20 °C for 1 hour, and then heated to 80 °C for 30 minutes. This clear solution was then concentrated *via* vacuum to produce a white solid. Aqueous $NaHCO_3$ (10 mL, 1 M) was added, and then extracted with $CHCl_3$ (3 × 10 mL), dried *via* Na_2SO_4, and then concentrated *via* vacuum to produce a colourless oil which was identified as benzyl alcohol.

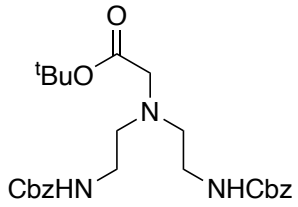
Dibenzyl (azanediylbis(ethane-2,1-diyl))dicarbamate (2.13)[5]



CbzOSu (4.98 g, 20.0 mmol) in DCM (30 mL) was cooled to -40 °C, and added to a -40 °C solution of diethylentriamine (1.11 mL, 10.3 mmol) in DCM (30 mL), and stirred for 16 hours at 20 °C. This was then concentrated *via* vacuum, DCM (30 mL) was added and washed with sat. $NaHCO_3$, dried *via* Na_2SO_4, filtered, and then concentrated *via* vacuum to provide a colourless oil. This crystallised overnight, to provide a dry white powder (2.34 g, 6.29 mmol, 63%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 7.40 – 7.27 (m, 10H), 5.18 (s, 1H), 5.08 (s, 3H), 3.26 (t, $^3J_{HH}$ = 5.9 Hz, 4H), 2.85 – 2.61 (m, 4H), 1.56 (s, 3H),

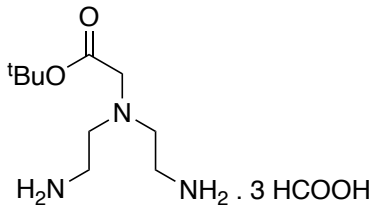
T*ert*-butyl bis(2-(((benzyloxy)carbonyl)amino)ethyl)glycinate (2.14)[6]



2.13 (2.34 g, 6.29 mmol) in MeCN (anhydrous, 40 mL) was cooled to 0 °C, under a nitrogen atmosphere. DIPEA (1.83 mL, 10.5 mmol) in DCM (anhydrous, 30 mL), and *tert*-butyl bromoacetate (1.22 mL, 8.39 mmol) were sequentially added at 0 °C. This solution was stirred for 16 hours at 20 °C. This was then concentrated *via* vacuum, dissolved in DCM (60 mL), washed with sat. $NaHCO_3$ (2 × 30 mL), dried *via* Na_2SO_4, filtered, and then concentrated *via* vacuum to provide a clear yellow oil. This was purified *via* automated chromatography in hexane/EtOAc to provide a colourless liquid (2.38 g, 4.91 mmol, 78%, R_f 0.49 in 1:1 hexane/EtOAc).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 7.36 – 7.27 (m, 10H), 5.49 (s, 2H), 5.06 (s, 4H), 3.22 (d, $^3J_{HH}$ = 6.1 Hz, 4H), 2.73 (t, $^3J_{HH}$ = 5.8 Hz, 4H), 1.64 (m, 2H), 1.43 (s, 9H).
^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 171.3, 156.8, 136.9, 128.6, 128.2, 128.1, 81.7, 66.7, 56.2, 54.4, 39.3, 28.2.
HR-MS (ES^+ ToF, g mol^{-1}): Found: 486.2593; Calc'd: 486.2604; $C_{26}H_{36}N_3O_6$ [M+H].

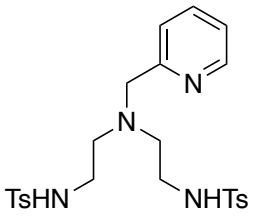
Tert-butyl bis(2-aminoethyl)glycinate formate (2.7-^{t}Bu)[6]



2.14 (2.30 g, 4.74 mmol) in 4% formic acid/MeOH (50 mL) was added to a suspension of 10% Pd/C (474 mg) in 4% formic acid/MeOH (50 mL) under a nitrogen atmosphere. This was stirred for 2.5 hours at 20 °C. The suspension was filtered *via* celite pad, concentrated *via* vacuum, then MeOH (10 mL) was added and concentrated *via* vacuum twice. The resultant golden oil was dissolved in a 20% water/MeCN solution (5 mL) and eluted through a Sep-Pak Plus tC18 cartridge. This was then concentrated *via* vacuum, to provide a golden oil (1.60 g, 4.51 mmol, 95%).

^{1}H NMR (400 MHz, D_2O) δ 8.35 (s, 3H), 3.40 (s, 2H), 3.05 (dd, $^{3}J_{HH}$ = 6.6, 5.2 Hz, 4H), 2.91 (dd, $^{3}J_{HH}$ = 6.6, 5.2 Hz, 4H), 1.44 (s, 9H).
HR-MS (ES^{+} ToF, g mol^{-1}): Found: 218.1859; Calc'd: 218.1869, $C_{10}H_{24}N_3O_2$ [M+H].

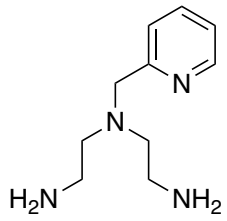
***N*,*N*'-(((pyridin-2-ylmethyl)azanediyl)bis(ethane-2,1-diyl))bis(4-methylbenzenesulfonamide) (2.15)[7]**



2-(Aminomethyl)pyridine (665 µL, 6.4 mmol) and tosyl aziridine (3.00 g, 14.2 mmol) were dissolved in EtOH (anhydrous, 45 mL) under nitrogen and stirred at 85 °C for 6 hours to provide a dark solution. This solution was concentrated *via* vacuum and purified *via* automated column chromatography (hexane → EtOAc) to provide a colourless oil which solidified into pink flakes at -20 °C, (2.02 g, 4.0 mmol, 62%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 7.84 (d, $^{3}J_{HH}$ = 8.0 Hz, 2H), 7.35 (d, $^{3}J_{HH}$ = 7.9 Hz, 2H), 2.45 (s, 4H), 2.37 (s, 4H).

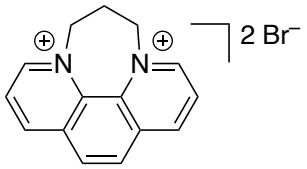
N^1-(2-aminoethyl)-N^1-(pyridin-2-ylmethyl)ethane-1,2-diamine (2.8)[7]



2.15 was dissolved in sulfuric acid (95%, 7 mL) under nitrogen. This green solution was stirred at 90 °C for 48 hours, then cooled to 0 °C and H_2O (14 mL) was slowly added dropwise. NaOH (sat.) was added until pH 14, then the tan-yellow solution was concentrated *via* vacuum. DCM (20 mL) was added to this beige residue, to provide a white suspension in a beige liquid. This was filtered, dried *via* Na_2SO_4, and then concentrated *via* vacuum, to provide a yellow oil (146 mg, 0.76 mmol, 86%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 8.52 (dd, $^3J_{HH}$ = 5.1, 1.8 Hz, 1H), 7.65 (td, $^3J_{HH}$ = 7.7, 1.8 Hz, 1H), 7.42 (d, $^3J_{HH}$ = 7.8 Hz, 1H), 7.15 (dd, $^3J_{HH}$ = 7.5, 5.0 Hz, 1H), 3.77 (s, 2H), 2.78 (t, $^3J_{HH}$ = 6.0 Hz, 4H), 2.62 (t, $^3J_{HH}$ = 6.0 Hz, 4H), 2.47 (br s, 5H).

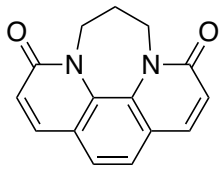
6,7-Dihydro-5*H*-[1,4]diazepino[1,2,3,4-*lmn*][1,10]phenanthroline-4,8-diium bromide (2.16)[8]



Phenanthroline (3.00 g, 16.65 mmol) was dissolved in chlorobenzene (25 mL), and heated to 70 °C. 1,3-Dibromopropane (8.45 mL, 83.24 mmol) was added dropwise at 70 °C. The temperature was raised to 120 °C and stirred for four hours, to provide an orange suspension. This suspension was filtered, washed with petroleum ether, and dried *via* vacuum to provide a yellow powder (6.36 g, 16.65 mmol, 100%).

^{1}H NMR (400 MHz, D_2O, δ, ppm): 9.65 (dd, $^3J_{HH}$ = 5.8, 1.4 Hz, 2H), 9.44 (dd, $^3J_{HH}$ = 8.5, 1.4 Hz, 2H), 8.58 (s, 2H), 8.53 (dd, $^3J_{HH}$ = 8.5, 5.8 Hz, 2H), 5.13 (t, $^3J_{HH}$ = 7.0 Hz, 5H), 3.41 (p, $^3J_{HH}$ = 7.0 Hz, 2H).

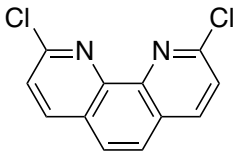
6,7-Dihydro-5*H*-[1,4]diazepino[1,2,3,4-*lmn*][1,10]phenanthroline-3,9-dione (2.17)[8]



2.16 (6.10 g, 16.0 mmol) was sonicated in ^{t}BuOH (100 mL) for 15 minutes. Potassium *tert*-butoxide (7.18 g, 64.0 mmol) was added slowly over 5 minutes at 40 °C and stirred for four hours, to provide an orange suspension. This suspension was filtered, dissolved in water, then extracted via $CHCl_3$ (3 × 60 mL). The combined organic layers were dried *via* Na_2SO_4, and then concentrated *via* vacuum to provide a light brown powder (1.36 g, 5.4 mmol, 32%). The filtrate was concentrated *via* vacuum and then purified *via* automated column chromatography (0 → 100% MeOH in DCM; R_f = 0.63 in 10% MeOH in DCM) to also provide a light brown powder which was combined with the original crop (2.63 g, 10.4 mmol, 62%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 7.71 (d, $^3J_{HH}$ = 9.5 Hz, 2H), 7.36 (s, 2H), 6.79 (d, $^3J_{HH}$ = 9.5 Hz, 2H), 4.31 (s, 4H), 2.45 (p, $^3J_{HH}$ = 6.5 Hz, 2H).

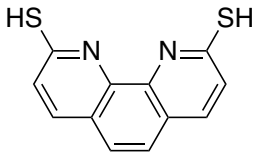
2,9-Dichloro-1,10-phenanthroline (2.18)[8]



In a 3-neck round bottom flask, **2.17** (2.00 g, 7.93 mmol) was suspended in $POCl_3$ (25 mL). PCl_5 (3.30 g, 15.86 mmol) was added, then the headspace was flushed with nitrogen for 10 minutes. This suspension was then heated at 110 °C for 16 hours. Once complete, the reaction was cooled to 20 °C and excess $POCl_3$ was distilled off. Unreacted PCl_5 and remaining $POCl_3$ were then quenched via dropwise addition of approximately 100 mL of ice-water until effervescence stopped. The pH was adjusted to pH 8 *via* the addition of sat. Na_2CO_3 (approximately 600 mL). The resulting brown suspension was filtered and dried *via* vacuum to provide a brown solid, which was then purified *via* automated column chromatography (0 → 100% DCM in Hexane), to provide a yellow solid (1.58 g, 6.34 mmol, 81%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 8.22 (d, $^3J_{HH}$ = 8.4 Hz, 2H), 7.83 (s, 2H), 7.66 (d, $^3J_{HH}$ = 8.4 Hz, 2H).

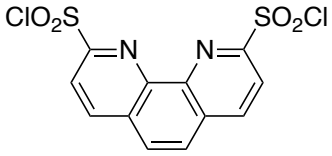
1,10-Phenanthroline-2,9-dithiol (2.19)[9]



2.18 (1.50 g, 6.02 mmol), $NaSH.xH_2O$ (2.63 g, 175% w/w), and DMF (25 mL) were combined and heated under reflux at 135 °C for 4 hours, to provide a dark green solution. Toluene (25 mL) was added, and the solution was concentrated *via* vacuum. H_2O (25 mL) and AcOH (25 mL, 2M) were sequentially added with stirring. This suspension was then filtered to provide a yellow brown solid, which was then purified *via* automated column chromatography (0 → 100% DCM in Hexane), to provide an orange solid (1.23 g, 5.02 mmol, 84%).

^{1}H NMR (400 MHz, TFA-*d*, δ, ppm) 8.58 (d, $^3J_{HH}$ = 8.9 Hz, 2H), 8.09 (s, 2H), 7.94 (d, $^3J_{HH}$ = 8.9 Hz, 2H).

1,10-Phenanthroline-2,9-disulfonyl dichloride (2.3) – attempted synthesis 1



This attempt was an adaptation of the sulfonamide coupling reaction published by Bahrami and co-workers, in which the sulfonyl chloride is generated *in-situ* and coupled with an amine.[10]

2.19 (0.200 g, 0.819 mmol) was suspended in MeCN (10 mL), then H_2O_2 (30% w/w, 501 µL, 4.91 mmol) was added which caused the yellow-orange suspension to turn red-orange. $SOCl_2$ (119 µL, 1.64 mmol) was added dropwise, and stirred for 5 minutes. This caused the suspension to aggregate to an orange gum. The gum was broken into smaller parts and (**2.9-Boc**) (0.189 g, 0.819 mmol) in pyridine (1 mL) was added dropwise, causing a lot of fuming. The gum turned into a green-brown sticky solid and stirred for 15 minutes, resulting in a milky white suspension with a green-brown residue on the bottom of the flask. This suspension was partitioned between DCM and water. The organic phase was

collected and purified *via* automated column chromatography (0 → 100% MeOH in DCM), to provide a cream yellow solid which was identified as **2.18**.

1,10-Phenanthroline-2,9-disulfonyl dichloride (2.3) – attempted synthesis 2

This attempt was an adaptation of a sulfonyl chloride synthesis published by Lewis and co-workers.[11]

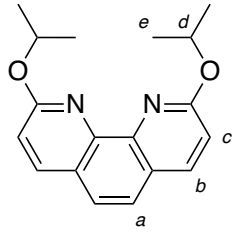
2.19 (294 mg, 1.2 mmol) and DCM (15 mL) were combined at 0 °C. Water (5.6 mL) and HCl (12 M, 1.6 mL) were added and stirred vigorously at 0 °C. NaOCl (10-15% Cl_2 content, 14.1 mL) was added drop wise at 0 °C and stirred for 1 hour. This was then extracted with DCM, washed with Na_2SO_3, dried *via* Na_2SO_4, and then concentrated *via* rotary evaporator to provide a yellow oil (crude 368 mg). This was then purified *via* automated column chromatography (100% DCM) to provide a white solid (125 mg) which was identified as **2.18**.

1,10-Phenanthroline-2,9-disulfonyl dichloride (2.3) – attempted synthesis 3

This attempt was an adaptation of a sulfonyl chloride synthesis published by Madabhushi and co-workers.[12]

2.19 (244 mg, 1.00 mmol), Oxone® (1.54 g, 5.00 mmol), KCl (149 mg, 2.00 mmol), and water (10 mL) were combined and stirred at 20 °C for 20 minutes. **2.19** did not dissolve well after 15 minutes of stirring, therefore HCl (3 mL, 12 M) was added to aid dissolution and also provide a chloride source. This suspension was stirred for a further 15 minutes, then extracted with $CHCl_3$ (3 × 10 mL). This created a solid interface between the phases, and the solid interface and organic layers were collected, dried *via* Na_2SO_4, and then concentrated *via* vacuum. This oil was then purified *via* automated column chromatography (hexane/DCM/MeOH) to provide a cream yellow solid (47 mg) which was identified as **2.18**.

2,9-diisopropoxy-1,10-phenanthroline (2.20)

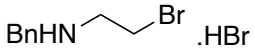


This attempt was an adaptation of the one-pot sulfonamide coupling reaction published by Flegeau and co-workers, in which a sulfinic acid is generated *in-situ*, oxidised, and then coupled with an amine.[13]

2,9-Diiodo-1,10-phenanthroline (173 mg, 0.40 mmol), DABSO (116 mg, 0.48 mmol), palladium(II) acetate (5.0 mg, 20.0 μmol, 5 mol %), and CataCXium A (11.0 mg, 31.6 μmol, 7.5 mol %), were combined under a nitrogen atmosphere. Triethylamine (anhydrous, 336 μL, 2.40 mmol), and 2-propanol (anhydrous, 1.5 mL) were added *via* syringe, and then the suspension was stirred at 75 °C for 16 hours. This solution was then cooled to room temperature, and 2-propanol (anhydrous, 8.5 mL) was added *via* syringe. (**2.9-Boc**) (81 mg, 0.40 mmol) was dissolved in 2-propanol (anhydrous, 5 mL), and added *via* syringe. This solution was then cooled to 0 °C, and NaOCl (10-15% available Cl_2, 805 μL, 1.20 mmol) was added dropwise (1 drop every 10 seconds) with stirring. This solution was stirred at 0 °C for 60 minutes, and then at 20 °C for 90 minutes, to provide a grey-brown suspension. This was poured into saturated sodium thiosulfate solution (20 mL) and extracted with EtOAc (3 × 50 mL), washed with brine (75 mL), dried *via* Na_2SO_4, and concentrated *via* vacuum to provide a brown solid. This was purified *via* automated column chromatography (DCM → MeOH) to provide an orange solid which was identified as 2,9-diisopropoxy-1,10-phenanthroline.

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 8.04 (d, $^3J_{HH}$ = 8.7 Hz, 2H, C^bH), 7.56 (s, 2H, C^aH), 7.00 (d, $^3J_{HH}$ = 8.7 Hz, 2H, C^cH), 5.76 (hept, $^3J_{HH}$ = 6.2 Hz, 2H, C^dH), 1.52 (d, $^3J_{HH}$ = 6.2 Hz, 12H, C^eH).
MS (ES^+ ToF, g mol^{-1}): Found: 297.2; Calc'd: 297.16, $C_{18}H_{21}N_2O_2$ [M+H].

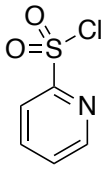
***N*-benzyl-2-bromoethan-1-amine hydrobromide (2.22)[14]**



N-benzylethanolamine (1.88 mL, 13.2 mmol) was dissolved in chloroform (anhydrous, 30 mL) under a nitrogen atmosphere. $SOBr_2$ (2.05 mL, 26.45 mmol) was added, and the orange solution was stirred at 65 °C for 2 hours. This orange liquid darkened and thickened, and as such the solution was stirred very rapidly. The solution was cooled to 20 °C, then Et_2O (20 mL) and EtOH (4 mL) were added causing an immediate orange precipitation. This was filtered to provide a white microcrystalline solid (3.62 g, 12.3 mmol, 93%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 9.63 (s, 2H), 7.69 – 7.56 (m, 2H), 7.49 – 7.38 (m, 3H), 4.21 (s, 2H), 3.73 (t, $^3J_{HH}$ = 7.1 Hz, 2H), 3.26 (t, $^3J_{HH}$ = 7.1 Hz, 2H).
^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 130.7, 130.3, 129.6, 129.1, 51.3, 47.4, 24.4.

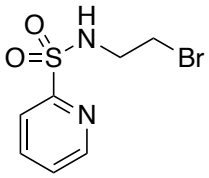
Pyridine-2-sulfonyl chloride (2.4) – attempted synthesis



This attempt was an adaptation of a sulfonyl chloride synthesis published by Madabhushi and co-workers.[12]

Oxone® (4.15 g, 13.5 mmol) and KCl (0.40 g, 5.4 mmol) were dissolved in H_2O (20 mL). 2-Mercaptopyridine (0.60 g, 5.40 mmol) was added and vigorously stirred at 20 °C for 20 minutes. This was then extracted with EtOAc (3 × 10 mL), dried *via* Na_2SO_4, then concentrated *via* vacuum, and purified *via* automated column chromatography (DCM → MeOH), however ^{1}H NMR showed no evidence of product peaks in any of the fractions.

***N*-(2-bromoethyl)pyridine-2-sulfonamide (2.21) – attempted synthesis 1**



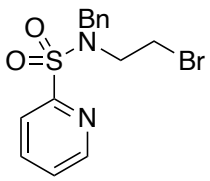
This attempt was an adaptation of a low-temperature sulfonyl chloride synthesis and trapping with amine published by Wright and co-workers.[15]

2-Mercaptopyridine (556 mg, 5.00 mmol) in DCM (25 mL) and HCl (1 M, 25 mL) was cooled to -5 °C and stirred for 10 minutes. To this, NaOCl (10%, 15 mL) was added dropwise and stirred vigorously for 15 minutes. This biphasic mixture was then poured into a separating funnel charged with ice-water, and the organic phase was collected into a conical flask at -78 °C. To this colourless solution, 2-bromoethylamine hydrobromide (1.23 g, 6.00 mmol) was added at -78 °C, and stirred at 0 °C for 30 minutes. This reaction was washed with H_3PO_4 (1 M, 20 mL), H_2O (20 mL), brine (20 mL), dried *via* Na_2SO_4, and concentrated *via* vacuum to provide a pale yellow-white oil (249 mg). However, ^{1}H NMR showed no evidence of expected product peaks in the crude.

_N_-(2-bromoethyl)pyridine-2-sulfonamide (2.21) – attempted synthesis 2

2-Mercaptopyridine (111 mg, 1.00 mmol) in DCM (10 mL) and HCl (1 M, 10 mL) was cooled to -10 °C and stirred for 10 minutes. To this, NaOCl (11%, 2.65 mL, 3.30 mmol) was added dropwise and stirred vigorously for 15 minutes. This biphasic mixture was then poured into a separating funnel charged with ice-water, and the organic phase was collected into a conical flask at -78 °C. To this colourless solution, 2-bromoethylamine hydrobromide (560 mg, 2.50 mmol) was added at -78 °C, and stirred at 0 °C for 30 minutes. This reaction was washed with HCl (1 M, 10 mL), H_2O (10 mL), brine (10 mL), dried *via* Na_2SO_4, and concentrated *via* vacuum to provide a pale yellow-white oil (249 mg). However, ^{1}H NMR showed no evidence of expected product peaks in the crude.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 1



PyFluor (81 mg, 0.5 mmol), **2.22** (295 mg, 1.0 mmol), and water (3 mL) were combined in a vial and stirred for three days. This provided a thin skin of microcrystalline solid floating on the water with PyFluor at the bottom. This was extracted with EtOAc (3 × 3 mL), dried *via* Na_2SO_4, and concentrated *via* vacuum to provide a white film (91 mg). However, ^{1}H NMR showed no evidence of expected product peaks in the crude, and almost exclusively PyFluor.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 2

This attempt was an adaptation of a sulfonamide synthesis published by Berg and co-workers.[16]

PyFluor (81 mg, 0.5 mmol), **2.22** (295 mg, 1.0 mmol), and THF (3 mL) were combined in a vial and stirred for three days. This provided a white suspension; the solids were filtered and found to be **2.22** *via* ^{1}H NMR. The filtrate was concentrated *via* vacuum and analysed *via* ^{1}H NMR. This showed no evidence of expected product peaks in the crude, and almost exclusively PyFluor.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 3

This attempt was an adaptation of a sulfonamide synthesis published by Brouwer and co-workers.[17]

PyFluor (81 mg, 0.5 mmol), **2.22** (295 mg, 1.0 mmol), and DCM (3 mL) were combined in a vial and stirred for three days. This provided a white suspension; the solids were filtered and found to be **2.22** *via* ^{1}H NMR. The filtrate was concentrated *via* vacuum and analysed *via* ^{1}H NMR. This showed no evidence of expected product peaks in the crude, and almost exclusively PyFluor.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 4

This attempt was an adaptation of a sulfonamide synthesis published by Bogolubsky and co-workers.[18]

PyFluor (81 mg, 0.5 mmol), **2.22** (148 mg, 0.5 mmol), NEt_3 (147 µL, 1.05 mmol) and MeCN (3 mL) were combined in a vial and stirred for three days. The resultant yellow-orange suspension was sonicated at 80 °C for four hours, filtered, and concentrated *via* vacuum to provide a brown oil (74 mg). However, ^{1}H NMR showed no evidence of expected product peaks in the crude, and almost exclusively PyFluor.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 5

This attempt was an adaptation of a sulfonamide synthesis published by Zelli and co-workers.[19]

DMF (5 mL) was partially evaporated under rotary evaporator to remove trace amounts of dimethylamine. PyFluor (161 mg, 1.00 mmol), **2.22** (148 mg, 0.50 mmol), DBU (262 µL, 1.75 mmol), and DMF (3.5 mL) were combined and stirred at 80 °C for 6 hours. This brown solution was then cooled to room temperature and concentrated *via* vacuum to provide an orange-brown liquid (584 mg). However, ^{1}H NMR showed no evidence of expected product peaks in the crude.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 6

PyFluor (81 mg, 0.5 mmol), **2.22** (148 mg, 0.5 mmol), Cs_2CO_3 (505 mg, 1.55 mmol) and MeCN (3 mL) were combined in a vial and stirred for 2 hours. This provided a white suspension which was concentrated *via* vacuum, dissolved in $CHCl_3$ (15 mL), washed with H_2O (2 × 15 mL), dried *via* Na_2SO_4, filtered, and concentrated *via* vacuum. However, ^{1}H NMR showed no evidence of expected product peaks in the crude and was almost exclusively *N*-benzyl-2-bromoethan-1-amine.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 7

This attempt was an adaptation of a sulfonamide synthesis published by Rojas and co-workers.[20]

PyFluor (81 mg, 0.5 mmol), **2.22** (142 mg, 0.48 mmol), NEt_3 (140 μL, 1.00 mmol) and MeCN (anhydrous, 1.40 mL) were combined in a vial under nitrogen and stirred at 60 °C for 1 hour. An instant white precipitate formed upon addition of NEt_3, a crystalline solid precipitated out after cooling to 20 °C. H_2O (20 mL) was added and extracted with DCM (3 × 20 mL). These organic phases were dried *via* Na_2SO_4 and concentrated *via* vacuum to provide a light-yellow oil (134 mg). However, ^1H and $^{19}F\{^1H\}$ NMR showed no evidence of expected product peaks in the crude, and almost exclusively PyFluor.

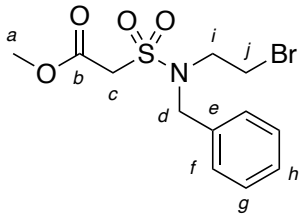
N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 8

PyFluor (81 mg, 0.5 mmol), **2.22** (148 mg, 0.5 mmol), NEt_3 (140 μL, 1.00 mmol) and MeCN-d_3 (1.40 mL) were combined in a vial under nitrogen and stirred at 60 °C for 1 hour. An instant white precipitate formed upon addition of NEt_3, a crystalline solid precipitated out after cooling to 20 °C. This suspension was filtered at 60 °C, and ^1H and $^{19}F\{^1H\}$ NMR on the resultant filtrate showed no evidence of expected product peaks in the crude, and almost exclusively PyFluor.

N-benzyl-N-(2-bromoethyl)pyridine-2-sulfonamide (2.21-Bn) – attempted synthesis 9

PyFluor (81 mg, 0.5 mmol), **2.22** (148 mg, 0.5 mmol), NEt_3 (140 μL, 1.00 mmol) and MeCN (anhydrous, 2 mL) were combined and sealed in a microwave vial under nitrogen and stirred at 150 °C for 10 minutes under microwave irradiation. This light-brown solution was concentrated *via* vacuum, partitioned between water (5 mL) and EtOAc (5 mL), and extracted with EtOAc (3 × 5 mL), concentrated *via* vacuum to provide a brown oil (103 mg). However, ^1H and $^{19}F\{^1H\}$ NMR showed no evidence of expected product peaks in the crude, and mostly PyFluor and degradation.

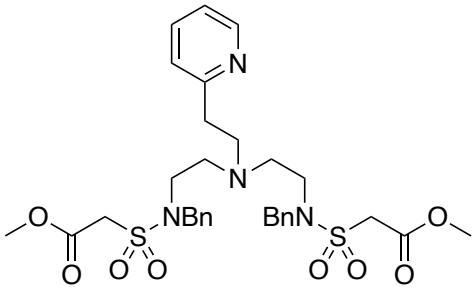
Methyl 2-(N-benzyl-N-(2-bromoethyl)sulfamoyl)acetate (2.23)



2.22 (118 mg, 0.40 mmol) was suspended in MeCN (anhydrous, 2 mL) under nitrogen. Triethylamine (anhydrous, 117 µL, 0.84 mmol), and methyl 2-(chlorosulfonyl)-acetate (46 µL, 0.40 mmol) were sequentially added causing the suspension to turn slightly white/yellow. This was stirred at 20 °C for 2 hours, then concentrated *via* vacuum, and partitioned between EtOAc (10 mL) and H_2O (10 mL), and then extracted with EtOAc (3 × 10 mL). Organic phases were combined and concentrated *via* vacuum to provide a colourless oil. This was purified *via* automated column chromatography (hexane → DCM) and concentrated *via* vacuum to provide a colourless liquid (51 mg, 0.15 mmol, 36%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 7.42 – 7.30 (m, 5H, Ph), 4.49 (s, 2H, C^dH), 4.05 (s, 2H, C^cH), 3.82 (s, 3H, C^aH), 3.61 (t, $^3J_{HH}$ = 7.6 Hz, 2H, C^iH), 3.19 (t, $^3J_{HH}$ = 7.1 Hz, 2H, C^jH).
^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 163.9 (C^b), 135.7 (C^e), 129.2 ($C^{f/g/h}$), 128.7 ($C^{f/g/h}$), 128.6 ($C^{f/g/h}$), 56.5 (C^c), 53.4 (C^d), 53.3 (C^a), 50.3 (C^i), 29.0 (C^j).
HR-MS (ES^+ ToF, g mol^{-1}): Found: 350.0060; Calc'd: 350.0056, $C_{12}H_{17}NO_4SBr$ [M+H].

Dimethyl 2,2'-((benzyl(2-((2-(benzyl(hydrosulfonyl)amino)ethyl)(2-(pyridin-2-yl)ethyl)amino)-ethyl)amino)sulfonyl)diacetate (2.24) – attempted synthesis

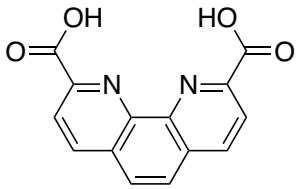


2.23 (250 mg, 0.71 mmol) was dissolved in DCM (5 mL) and transferred to a 3-neck round-bottomed flask, then concentrated *via* in-line trap. This was then dissolved in MeCN (anhydrous, 10 mL) under nitrogen atmosphere. K_2CO_3 (493 mg, 3.57 mmol), KI (57 mg, 0.34 mmol) and 2-pyridylethylamine (41 µL, 0.34 mmol) were sequentially added. This suspension was stirred for 3 days at 85 °C and

monitored every 24 hours *via* TLC and LC-MS. However, LC-MS showed no product formation throughout the reaction.

7.2.3 Amides

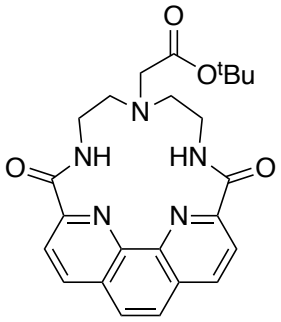
1,10-Phenanthroline-2,9-dicarboxylic acid (2.25)[21]



SeO_2 was dissolved in dioxane (25 mL) and water (0.7 mL) and stirred for 5 minutes at 110 °C. A suspension of neocuproine (1.00 g, 4.8 mmol) in dioxane (25 mL) was added dropwise over 10 minutes and stirred for a further 30 minutes at 110 °C, to provide a dark brown-purple suspension. This suspension was then decanted to a new flask whilst still hot and left overnight to cool. The resultant beige solid was filtered and washed with Et_2O (3 × 15 mL) and dried in a vacuum desiccator to afford a pale brown solid (2,9-dicarbaldehyde-1,10-phenanthroline, crude 1.13 g) which was used without further purification. 2,9-Dicarbaldehyde-1,10-phenanthroline (1.13 g, 4.8 mmol) was added to nitric acid (65%, 25 mL) and heated to reflux for 3 hours. This orange solution was then poured whilst hot onto ice (~20 mL). Yellow needle crystals crystallised overnight. These were isolated and dried *via* vacuum desiccator to provide yellow needle crystals (1.29 g, 4.8 mmol, 100%), and used without further purification.

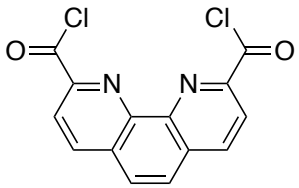
^{1}H NMR (400 MHz, DMSO-d_6, δ, ppm): 8.75 (d, $^3J_{HH}$ = 8.3 Hz, 2H), 8.43 (d, $^3J_{HH}$ = 8.3 Hz, 2H), 8.23 (s, 2H).

T*ert*-butyl 2-(2,10-dioxo-3,6,9-triaza-1(2,9)-phenanthrolinacyclodecaphane-6-yl)acetate (PhenDA-1-^{t}Bu) – attempted synthesis 1



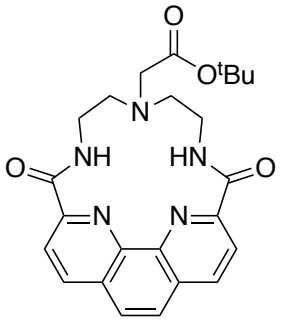
Under nitrogen, DIPEA (435 µL, 2.5 mmol) was added to **2.7-^{t}Bu** (178 mg, 0.5 mmol) in MeCN (anhydrous, 30 mL) and stirred for 20 minutes at 20 °C. HBTU (569 mg, 1.5 mmol) and DIPEA (435 µL, 2.5 mmol) were added to **2.25** (153 mg, 0.5 mmol) in MeCN (anhydrous, 30 mL) and stirred for 20 minutes at 20 °C also under nitrogen. The amine suspension was transferred *via* cannula into the solution containing **2.25**. This solution was stirred at 20 °C for 3 days, after which LC-MS analysis did not show evidence of product formation. The cloudy solution was heated to 40 °C and stirred for a further 16 hours, concentrated *via* vacuum, and purified *via* automated column chromatography (DCM → MeOH). However, HR-MS analysis showed no evidence of expected product peaks in any of the fractions.

1,10-Phenanthroline-2,9-dicarbonyl dichloride (2.26)[22]



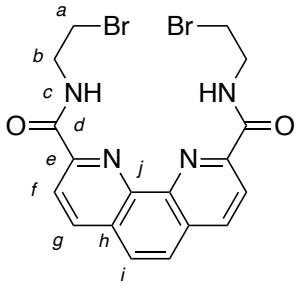
2.25 (536 mg, 2.00 mmol) was added to DCM (anhydrous, 50 mL) under nitrogen and heated at 40 °C for 10 minutes. Oxalyl chloride (687 µL, 8.00 mmol) and DMF (anhydrous, 1 drop) were added, producing effervescence. This suspension was stirred at 40 °C for 24 hours, concentrated *via* vacuum to provide a yellow-white solid. This was used without further purification.

T*ert*-butyl 2-(2,10-dioxo-3,6,9-triaza-1(2,9)-phenanthrolinacyclodecaphane-6-yl)acetate (PhenDA-1-^{t}Bu) – attempted synthesis 2



To a solution of **2.26** (153 mg, 0.50 mmol) in DCM (anhydrous, 30 mL), DIPEA (261 µL, 1.50 mmol) was added under nitrogen, and cooled to 0 °C. Separately under a nitrogen atmosphere, **2.7-^{t}Bu** (178 mg, 0.50 mmol) and DIPEA (435 µL, 2.50 mmol) were dissolved in DCM (anhydrous, 30 mL), and cooled to 0 °C. The amine solution was added dropwise to the acyl chloride solution at 0 °C. This white suspension was then stirred at 20 °C for 3 days. This suspension was washed with sat. $NaHCO_3$, dried *via* Na_2SO_4, and then concentrated *via* rotary evaporator. This solid was then purified *via* automated column chromatography, however no fractions contained the correct equivalents of phenanthroline and *tert*-butyl ^{1}H NMR peaks.

N^{2},N^{9}-bis(2-bromoethyl)-1,10-phenanthroline-2,9-dicarboxamide (2.27)



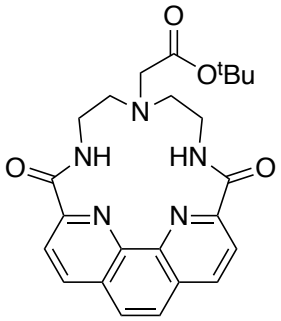
2.26 (0.61 g, 2.00 mmol) was dissolved in $CHCl_3$ (anhydrous, 30 mL) under nitrogen and stirred at 0 °C for 20 minutes. Separately 2-bromoethylamine hydrobromide (1.23 g, 6.00 mmol) and NEt_3 (1.95 mL, 14.00 mmol) were dissolved in $CHCl_3$ (anhydrous, 30 mL) under nitrogen and stirred at 0 °C for 20 minutes. The amine solution was added *via* syringe drop-wise to the acyl chloride at 0 °C. The slightly yellow, almost colourless solution was stirred for 16 hours gradually warming to 20 °C, then washed with H_2O (2 × 20 mL), dried *via* Na_2SO_4, and concentrated *via* vacuum. This was purified *via* automated column chromatography (DCM → MeOH), to provide a cream solid (443 mg, 0.93 mmol, 46%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 9.07 (br t, $^{3}J_{HH}$ = 6.1, 2H, $N^{c}H$), 8.59 (d, $^{3}J_{HH}$ = 8.2 Hz, 2H, $C^{f}H$), 8.47 (d, $^{3}J_{HH}$ = 8.3 Hz, 2H, $C^{g}H$), 7.95 (s, 2H, $C^{i}H$), 4.06 (q, $^{3}J_{HH}$ = 6.1 Hz, 4H, $C^{b}H$), 3.73 (t, $^{3}J_{HH}$ = 6.0 Hz, 4H, $C^{a}H$).

^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 164.5 (C^{d}), 149.5 (C^{e}), 144.4 (C^{j}), 138.0 (C^{g}), 130.8 (C^{h}), 128.0 (C^{i}), 121.7 (C^{f}), 41.4 (C^{b}), 32.9 (C^{a}).

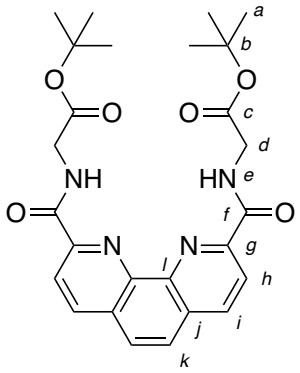
HR-MS (ES^{+} ToF, g mol^{-1}): Found: 478.9728; Calc'd: 478.9718, $C_{18}H_{17}N_4O_2Br$ [M+H].

T*ert*-butyl 2-(2,10-dioxo-3,6,9-triaza-1(2,9)-phenanthrolinacyclodecaphane-6-yl)acetate (PhenDA-1-^{t}Bu) – attempted synthesis 3



Under a nitrogen atmosphere, *tert*-butyl 2-aminoacetate (57 µL, 0.42 mmol), K_2CO_3 (576 mg, 4.17 mmol), and MeCN (anhydrous, 20 mL) were combined and stirred at 85 °C for 1 hour. To this suspension **2.27** (200 mg, 0.42 mmol) was added, stirred at 85 °C for 3 days, and monitored *via* LC-MS. This suspension was filtered and concentrated *via* vacuum to provide a yellow solid (102 mg). LC-MS analysis after 24 and 48 hours showed the majority of the reactants were converted to a peak which matches the molecular weight of (1,10-phenanthroline-2,9-diyl)bis(aziridin-1-ylmethanone)+H (**2.28**+H) at 319.12.

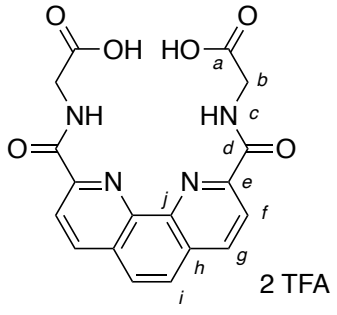
Di-*tert*-butyl-2,2'-((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))diacetate (PhenDA-2-^{t}Bu)



2.26 (0.61 g, 2.00 mmol) was dissolved in $CHCl_3$ (anhydrous, 30 mL) under nitrogen and stirred at 0 °C for 20 minutes. Separately *tert*-butyl 2-aminoacetate (820 µL, 6.00 mmol) and NEt_3 (1.95 mL, 14.00 mmol) were dissolved in $CHCl_3$ (anhydrous, 30 mL) under nitrogen and stirred at 0 °C for 20 minutes. The amine solution was slowly added *via* cannula to the acyl chloride at 0 °C. The slightly yellow, almost colourless solution was stirred for 16 hours gradually warming to 20 °C, to provide a colourless solution which was washed with H_2O (2 × 20 mL), dried *via* Na_2SO_4, and concentrated *via* vacuum. This was purified *via* automated column chromatography (DCM → MeOH), to provide a white solid (0.71 g, 1.43 mmol, 71%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 9.39 (t, $^{3}J_{HH}$ = 5.6 Hz, 2H, N^{e}H), 8.74 – 8.29 (m, 4H, C$^{h/i}$H), 7.89 (s, 2H, C^{k}H), 4.30 (d, $^{3}J_{HH}$ = 5.6 Hz, 4H, C^{d}H), 1.53 (s, 18H, C^{a}H).
^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 169.3 (C^{f}), 164.6 (C^{c}), 149.2 (C^{g}), 144.1 (C^{l}), 137.6 (C^{i}), 130.4 (C^{j}), 127.7 (C^{k}), 121.2 (C^{h}), 82.2 (C^{b}), 42.5 (C^{d}), 28.3 (C^{a}).
HR-MS (ES^{+} ToF, g mol^{-1}): Found: 495.2247; Calc'd: 495.2238, $C_{26}H_{31}N_4O_6$ [M+H].

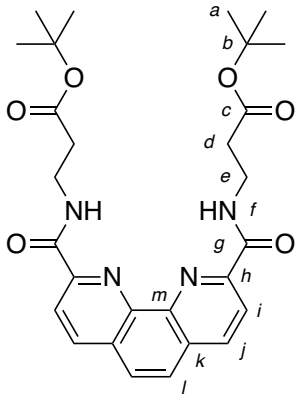
2,2'-((1,10-Phenanthroline-2,9-dicarbonyl)bis(azanediyl))diacetic acid trifluoroacetate (PhenDA-2)



PhenDA-2-^{t}Bu (0.40 g, 0.81 mmol) was dissolved in a mixture of DCM (4 mL) and TFA (4 mL). This solution was stirred at 20 °C for 16 hours and concentrated *via* vacuum to provide a cream solid (0.36 mg, 0.58 mmol, 72%).

^{1}H NMR (400 MHz, 333 K, DMSO-d_6, δ, ppm): 9.49 (t, $^{3}J_{HH}$ = 6.0 Hz, 2H, N^{c}H), 8.74 (d, $^{3}J_{HH}$ = 8.3 Hz, 2H), 8.45 (d, $^{3}J_{HH}$ = 8.2 Hz, 2H), 8.18 (s, 2H, C^{i}H), 4.17 (d, $^{3}J_{HH}$ = 6.0 Hz, 4H, C^{b}H).
^{13}C{^{1}H} NMR (101 MHz, 333 K, DMSO-d_6, δ, ppm): 170.7 (C^{d}), 163.9 (C^{a}), 148.8 (C^{e}), 143.4 (C^{j}), 138.0 (C^{g}), 130.1 (C^{h}), 127.7 (C^{i}), 120.7 (C^{f}), 41.2 (C^{b}).
^{19}F{^{1}H} NMR (471 MHz, 333 K, DMSO-d_6, δ, ppm) -74.7.
HR-MS (ES^{+} ToF, g mol^{-1}): Found: 383.0993; Calc'd: 383.0986, $C_{18}H_{15}N_4O_6$ [M+H].

Di-*tert*-butyl 3,3'-((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))dipropionate (PhenDA-3-^{t}Bu)

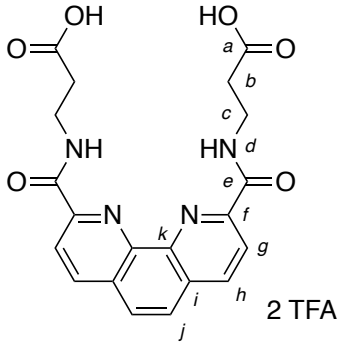


2.26 (0.61 g, 2.00 mmol) was dissolved in $CHCl_3$ (anhydrous, 30 mL) under nitrogen and stirred at 0 °C for 20 minutes. Separately *tert*-butyl 3-aminopropanoate (1.09 g, 6.00 mmol) and NEt_3 (1.95 mL, 14.00 mmol) were dissolved in $CHCl_3$ (anhydrous, 30 mL) under nitrogen and stirred at 0 °C for 20 minutes. The amine solution was added *via* syringe drop-wise to the acyl chloride at 0 °C. The slightly yellow, almost colourless solution was stirred for 16 hours gradually warming to 20 °C, to provide a colourless solution which was washed with H_2O (2 × 20 mL), dried *via* Na_2SO_4, and concentrated *via* vacuum. This was purified *via* automated column chromatography (DCM → MeOH), to provide a white solid (0.69 g, 1.31 mmol, 66%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 9.14 (m, 2H, N^{f}H), 8.55 (d, $^{3}J_{HH}$ = 8.3 Hz, 2H, C^{i}H), 8.40 (d, $^{3}J_{HH}$ = 8.3 Hz, 2H, C^{j}H), 7.89 (s, 2H, C^{l}H), 3.88 (m, 4H, C^{e}H), 2.71 (m, 4H, C^{d}H) , 1.41 (s, 18H, C^{a}H).

$^{13}C\{^1H\}$ NMR (101 MHz, $CDCl_3$, δ, ppm): 171.7 (C^c), 164.5 (C^g), 149.7 (C^h), 144.3 (C^m), 137.6 (C^j), 130.4 (C^k), 127.7 (C^l), 121.4 (C^i), 80.8, C^aH, 35.6 (C^d), 35.5 (C^e), 28.2 (C^a).
HR-MS (ES^+ ToF, g mol^{-1}): Found: 523.2563; Calc'd: 523.2557, $C_{28}H_{35}N_4O_6$ [M+H].

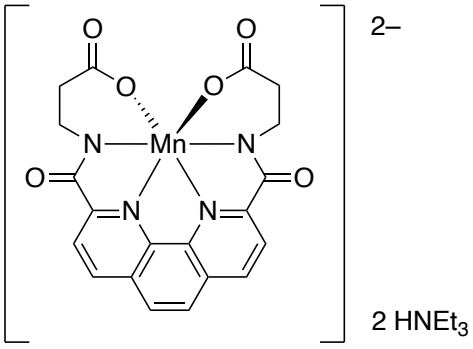
3,3'-((1,10-Phenanthroline-2,9-dicarbonyl)bis(azanediyl))dipropionic acid trifluoroacetate (PhenDA-3)



PhenDA-3-^{t}Bu (125 mg, 0.24 mmol) was dissolved in a mixture of DCM (4 mL) and TFA (4 mL). This solution was stirred at 20 °C for 16 hours and concentrated *via* vacuum to provide a cream solid (0.15 g, 0.24 mmol, 98%).

^1H NMR (400 MHz, DMSO-d_6, δ, ppm): 9.38 (t, $^3J_{HH}$ = 6.0 Hz, 2H, N^dH), 8.74 (d, $^3J_{HH}$ = 8.3 Hz, 2H, C^gH), 8.44 (d, $^3J_{HH}$ = 8.3 Hz, 2H, C^hH), 8.18 (s, 2H, C^jH), 3.83 – 3.63 (m, 4H, C^cH), 2.71 (t, $^3J_{HH}$ = 7.0 Hz, 4H, C^bH).
$^{13}C\{^1H\}$ NMR (101 MHz, DMSO-d_6, δ, ppm): 173.3 (C^a), 163.4 (C^e), 149.2 (C^f), 143.4 (C^k), 138.2 (C^g), 130.2 (C^i), 127.8 (C^j), 120.8 (C^h), 35.0 (C^c), 33.7 (C^b).
$^{19}F\{^1H\}$ NMR (471 MHz, DMSO-d_6, δ, ppm) -75.1.
HR-MS (ES^+ ToF, g mol^{-1}): Found: 411.1291; Calc'd: 411.1305, $C_{20}H_{19}N_4O_6$ [M+H].

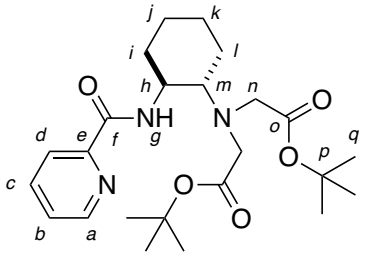
$(HNEt_3)_2$[Mn(PhenDA-3)] (MnPhenDA-3)



PhenDA-3 (50 mg, 0.078 mmol) and $Mn(OAc)_2.4H_2O$ (21 mg, 0.078 mmol) were suspended in MeCN (5 mL) and stirred for 1 hour at 60 °C. Triethylamine (65 µL, 0.470 mmol) was added and stirred for 2.5 hours at 60 °C, to provide a cream suspension. This was concentrated *via* vacuum, and the residue was suspended in cold EtOH (-78 °C), filtered, and washed with cold EtOH. This provided a yellow solid (24 mg, 0.036 mmol, 46%).

MS ($MALDI^+$ ToF, g mol^{-1}): Found: 464.8; Calc'd: 464.1, $C_{20}H_{14}MnN_4O_6$ [M+3H].

Di-*tert*-butyl 2,2'-(((1*S*,2*S*)-2-(picolinamido)cyclohexyl)azanediyl)diacetate (PicAm-1-^{t}Bu)

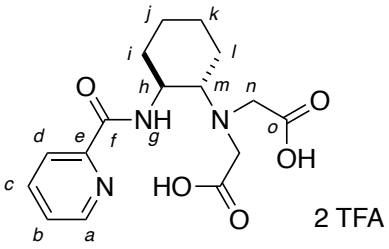


(±)-*Trans*-1,2-diaminocyclohexane (360 µL, 3.00 mmol) was dissolved in THF (anhydrous, 50 mL) and stirred at 20 °C for 1 hour under nitrogen. Separately picolinic acid (369 mg, 3.00 mmol), N,N-carbonyldiimidazole (486 mg, 3.00 mmol), and THF (anhydrous, 10 mL) were combined under nitrogen and stirred at 50 °C for 90 minutes. THF (anhydrous, 40 mL) was added to the picolinic acid solution, which was then added dropwise *via* cannula to the (±)-*trans*-1,2-diaminocyclohexane solution and stirred at 20 °C for 1 hour. This was then concentrated *via* vacuum to provide a golden oil, which was dissolved in MeCN (anhydrous, 30 mL) under nitrogen atmosphere, and transferred *via* cannula to a 2-neck round bottomed flask with K_2CO_3 (3.73 g, 27.00 mmol) under nitrogen. This flask was cooled to 0 °C, and *tert*-butyl bromoacetate (1.88 mL, 12.8 mmol) was added dropwise *via* syringe. This white

suspension was stirred at 60 °C for 2 days, to provide a creamy yellow suspension. This cream suspension was G4 filtered to provide a golden liquid, which was concentrated *via* vacuum to provide a dark golden, viscous oil, which was purified *via* automated column chromatography (hexane → EtOAc), and concentrated *via* vacuum, to provide a yellow-white solid (580 mg, 1.30 mmol, 43%).

^{1}H NMR (400 MHz, $CDCl_3$, δ, ppm): 9.02 (d, $^{3}J_{HH}$ = 5.9 Hz, 1H, $N^{g}H$), 8.63 – 8.50 (m, 1H, $C^{c}H$), 8.16 (dt, $^{3}J_{HH}$ = 7.8, 1.1 Hz, 1H, $C^{a}H$), 7.80 (td, $^{3}J_{HH}$ = 7.7, 1.8 Hz, 1H, $C^{d}H$), 7.37 (ddd, $^{3}J_{HH}$ = 7.6, 4.7, 1.2 Hz, 1H, $C^{b}H$), 3.76 (dp, $^{3}J_{HH}$ = 10.4, 5.0 Hz, 1H, $C^{h}H$), 3.52 – 3.26 (m, 4H, $C^{n}H$), 2.75 (td, $^{3}J_{HH}$ = 11.1, 3.3 Hz, 1H, $C^{m}H$), 2.53 (m, 1H, $C^{i}H$), 2.05 (m, 1H, $C^{l}H$), 1.80 (m, 1H, $C^{i}H$), 1.69 (m, 1H, $C^{l}H$), 1.40 (s, 18H, $C^{q}H$), 1.31 – 1.07 (m, 4H, $C^{j/k}H$).
^{13}C{^{1}H} NMR (101 MHz, $CDCl_3$, δ, ppm): 171.9 (C^{o}), 164.9 (C^{f}), 151.1 (C^{e}), 148.3 (C^{c}), 137.1 (C^{d}), 125.7 (C^{b}), 122.2 (C^{a}), 80.9 (C^{p}), 63.9 (C^{m}), 52.4 (C^{n}), 51.4 (C^{h}), 32.9 (C^{i}), 28.2 (C^{q}), 27.3 (C^{l}), 25.9 (C^{j}), 24.8 (C^{k}).
HR-MS (ES^{+} ToF, g mol^{-1}): Found: 448.2820; Calc'd: 448.2811, $C_{24}H_{38}N_3O_5$ [M+H].

2,2'-(((±)-Trans-2-(picolinamido)cyclohexyl)azanediyl)diacetic acid trifluoroacetate (PicAm-1)

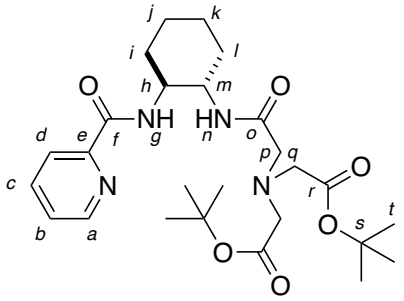


PicAm-1-^{t}Bu (400 mg, 0.89 mmol) was dissolved in a mixture of DCM (5 mL) and TFA (5 mL). This solution was stirred at 20 °C for 16 hours and concentrated *via* vacuum (repeated multiple times after diluting in DCM) to provide an orange oil. This was lyophilised to provide a light-yellow powder (507 mg, 0.89 mmol, 100%).

^{1}H NMR (400 MHz, D_2O, δ, ppm): 8.83 (dd, $^{3}J_{HH}$ = 5.7, 1.5 Hz, 1H, $C^{a}H$), 8.65 (td, $^{3}J_{HH}$ = 8.0, 1.5 Hz, 1H, $C^{b}H$), 8.50 (d, $^{3}J_{HH}$ = 8.0 Hz, 1H, $C^{d}H$), 8.20 – 8.08 (m, 1H, $C^{c}H$), 4.36 (td, $^{3}J_{HH}$ = 11.2, 4.4 Hz, 1H, $C^{h}H$), 4.11 (s, 4H, $C^{n}H$), 3.38 (td, $^{3}J_{HH}$ = 11.7, 3.5 Hz, 1H, $C^{m}H$), 2.24 – 2.12 (m, 1H, CyH), 2.07 (dd, $^{3}J_{HH}$ = 13.0, 3.9 Hz, 1H, CyH), 1.87 (d, $^{3}J_{HH}$ = 11.9 Hz, 1H, CyH), 1.76 (d, $^{3}J_{HH}$ = 12.2 Hz, 1H, CyH), 1.56 (qd, $^{3}J_{HH}$ = 12.3, 3.5 Hz, 2H, CyH), 1.48 – 1.22 (m, 2H, CyH).

$^{13}C\{^{1}H\}$ NMR (101 MHz, D_2O, δ, ppm): 169.1 (C^n), 162.7 (q, $^2J_{CF}$ = 36.0 Hz, TFA salt), 161.7 (C^e), 147.2 (C^b), 143.5 (C^a), 142.0 (C^f), 129.8 (C^c), 125.2 (C^d), 116.2 (q, $^1J_{CF}$ = 291.5 Hz, TFA salt), 68.6 (C^l), 52.9 (C^m), 49.0 (C^g), 30.6 (CyH), 23.6 (CyH), 23.4 (CyH), 23.3 (CyH).
HR-MS (ES^+ ToF, g mol^{-1}): Found: 336.1563; Calc'd: 336.1559, $C_{16}H_{22}N_3O_5$ [M+H].

Di-*tert*-butyl-2,2'-((2-oxo-2-(((1*S*,2*S*)-2-(picolinamido)cyclohexyl)amino)ethyl)azanediyl)-diacetate (PicAm-2-^{t}Bu)

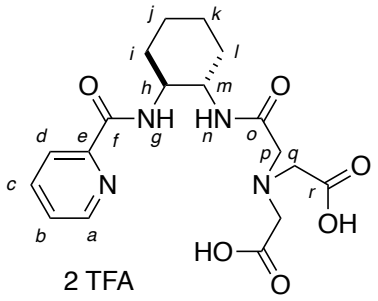


(±)-*Trans*-1,2-diaminocyclohexane (300 µL, 2.50 mmol) was dissolved in THF (anhydrous, 50 mL) and stirred at 20 °C for 1 hour under nitrogen. Separately picolinic acid (308 mg, 2.50 mmol), N,N-carbonyldiimidazole (405 mg, 2.50 mmol), and THF (anhydrous, 10 mL) were combined under nitrogen and stirred at 50 °C for 90 minutes. THF (anhydrous, 40 mL) was added to the picolinic acid solution, which was then added dropwise *via* cannula to the (±)-*trans*-1,2-diaminocyclohexane solution and stirred at 20 °C for 1 hour. This provided a white suspension, which was concentrated *via* vacuum to provide a very light-yellow oil. This was placed under nitrogen again (~10 quick backfills with vacuum and nitrogen) and dissolved in DCM (anhydrous, 20 mL). Simultaneously 2-(bis(2-(*tert*-butoxy)-2-oxoethyl)amino)acetic acid (758 mg, 2.50 mmol), N,N-carbonyldiimidazole (405 mg, 2.50 mmol), and THF (anhydrous, 10 mL) were combined under nitrogen and stirred at 50 °C for 1 hour. The DCM solution was added dropwise *via* cannula to the new N,N-carbonyldiimidazole solution and stirred at 20 °C for 1 hour, to provide a lightly golden solution. Sat. K_2CO_3 (40 mL) was added, extracted with DCM (3 × 20 mL), the organic phases were combined, dried *via* Na_2SO_4, filtered, and then concentrated *via* vacuum. This was purified *via* automated column chromatography (DCM → 10% MeOH in DCM) and concentrated *via* vacuum to provide a clear oil (837 mg, 1.66 mmol, 66%).

^1H NMR (400 MHz, $CDCl_3$, δ, ppm): 8.55 (d, $^3J_{HH}$ = 4.8 Hz, 1H, Py), 8.19 (d, $^3J_{HH}$ = 7.1 Hz, 1H, NH), 8.12 (d, $^3J_{HH}$ = 7.8 Hz, 1H, Py), 8.06 (d, $^3J_{HH}$ = 7.8 Hz, 1H, NH), 7.78 (td, $^3J_{HH}$ = 7.7, 1.6 Hz, 1H, Py), 7.37 (dd, $^3J_{HH}$ = 7.6, 4.8 Hz, 1H, Py), 3.88 (d, $^3J_{HH}$ = 7.7 Hz, 2H, $C^{h/m}H$), 3.29 (s, 4H, C^qH), 3.23 (d, $^3J_{HH}$ = 4.3 Hz, 2H, C^pH), 2.19 (d, $^3J_{HH}$ = 10.0 Hz, 1H, CyH), 2.00 (d, $^3J_{HH}$ = 11.6 Hz, 1H, CyH), 1.80 (dd, $^3J_{HH}$ = 15.6, 6.7 Hz, 3H, CyH), 1.45 (s, 3H, CyH), 1.40 (s, 18H, C^tH).

$^{13}C\{^{1}H\}$ NMR (101 MHz, $CDCl_3$, δ, ppm): 171.5 (C^o), 170.7 (C^r), 164.4 (C^f), 150.1 (C^e), 148.5 (Py), 137.2 (Py), 126.0 (Py), 122.1 (Py), 81.6 (C^s), 58.9 (C^p), 56.6 (C^q), 54.0 (C^h), 52.3 (C^m), 32.6 (CyH), 32.4 (CyH), 28.2 (C^tH), 25.0 (CyH), 24.9 (CyH).
HR-MS (ES^+ ToF, g mol^{-1}): Found: 505.3027; Calc'd: 505.3021, $C_{26}H_{41}N_4O_6$ [M+H].

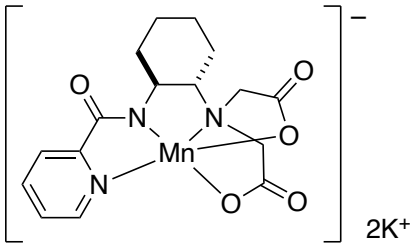
2,2'-((2-Oxo-2-(((±)-trans-2-(picolinamido)cyclohexyl)amino)ethyl)azanediyl)diacetic acid trifluoroacetate (PicAm-2)



PicAm-2-^{t}Bu (500 mg, 0.99 mmol) was dissolved in a mixture of DCM (5 mL) and TFA (5 mL). This solution was stirred at 20 °C for 16 hours and concentrated *via* vacuum (repeated multiple times after diluting in DCM) to provide a brown oil (crude 966 mg). This was lyophilised to provide white solid, which melted upon standing into a light-coloured oil. This was purified *via* reversed-phase automated column chromatography (**A** → **B**) to provide a white solid (337 mg, 0.54 mmol, 55%).

^{1}H NMR (400 MHz, D_2O, δ, ppm): 8.78 (d, $^3J_{HH}$ = 5.4 Hz, 1H, C^aH), 8.54 – 8.42 (m, 1H, C^cH), 8.25 (d, $^3J_{HH}$ = 8.0 Hz, 1H, C^dH), 8.05 – 7.96 (m, 1H, C^bH), 4.08 – 3.64 (m, 8H, $C^{h/m/p/q/}H$), 2.12 – 1.25 (m, 8H, CyH).
$^{13}C\{^{1}H\}$ NMR (101 MHz, D_2O, δ, ppm): 169.5 (C^r), 165.9 (C^o), 161.9 (C^f), 144.9 ($C^{b/c}$), 144.8 (C^a), 144.3 (C^e), 128.9 ($C^{b/c}$), 124.1 (C^h), 55.9 (C^p), 55.9 (C^q), 54.5 (C^h), 53.0 (C^m), 30.9 ($C^{i/l}$), 24.0 ($C^{j/k}$).
HR-MS (ES^+ ToF, g mol^{-1}): Found: 393.1759; Calc'd: 393.1774, $C_{18}H_{25}N_4O_6$ [M+H].

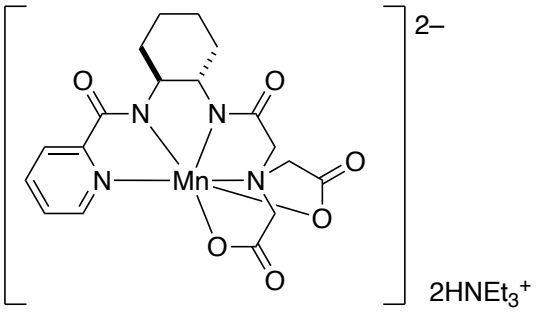
K_2[Mn(PicAm-1)] (MnPicAm-1)



Under a nitrogen atmosphere, **PicAm-1** (60 mg, 106 μmol), and KHMDS (106 mg, 532 μmol) were suspended in THF (anhydrous, 5 mL), and stirred at 20 °C for 2 hours. $MnCl_2$ (13.4 mg, 106 μmol) was added, and stirred at 20 °C for 16 hours. This provided a cloudy brown suspension, which was concentrated *via* vacuum, and purified *via* automated reversed-phase column chromatography (**C → D**) and lyophilised, to provide a yellow-white hygroscopic solid (84 mg, >100% due to absorbed water).

r_1 (0.01 MHz, 298 K, mM^{-1} s^{-1}): 18.95

$(HNEt_3)_2$[Mn(PicAm-2)] (MnPicAm-2)



PicAm-2 (50 mg, 81 μmol) and $Mn(OAc)_2.6H_2O$ (22 mg, 89 μmol) were suspended in MeCN (5 mL) and stirred at 60 °C for 1 hour. Then triethylamine (67 μL, 81 μmol) was added, and stirred for a further 2 hours. This suspension was concentrated *via* vacuum to provide a yellow film. Cold EtOH (-78 °C) was added, however a suspension did not form, and G4 filtration did not collect anything. This solution was concentrated *via* vacuum, dissolved in H_2O, and lyophilised, to produce a yellow-white hygroscopic solid (69 mg, >100 % due to absorbed water).

MS (ES^+ ToF, g mol^{-1}): Found: 446.1; Calc'd: 446.1, $C_{16}H_{22}N_3O_5$ [M+3H].

7.3 References

1 D. M. Corsi, C. Platas-Iglesias, H. van Bekkum and J. A. Peters, *Magn. Reson. Chem.*, 2001, **39**, 723–726, DOI: 10.1002/mrc.922.

2 A. Brück and K. Ruhland, *Organometallics*, 2009, **28**, 6383–6401, DOI: 10.1021/om900324a.

3 T. Takeshita, K. Sato and Y. Nakajima, *Dalton Trans.*, 2018, **47**, 17004–17010, DOI: 10.1039/C8DT04168G.

4 A. Blum, J. Böttcher, B. Sammet, T. Luksch, A. Heine, G. Klebe and W. E. Diederich, *Bioorganic Med. Chem.*, 2008, **16**, 8574–8586, DOI: 10.1016/j.bmc.2008.08.012.

5 M. Adamczyk, J. R. Fishpaugh and K. J. Heuser, *Org. Prep. Proced. Int.*, 1998, **30**, 339–348, DOI: 10.1080/00304949809355294.

6 C. Giordano, C. Del Bianco, S. Faini, A. Napoli, G. Sindona and L. Cellai, *Synthesis (Stuttg).*, 2004, **2004**, 1835–1843, DOI: 10.1055/s-2004-829143.

7 M. E. G. Skinner, Y. Li and P. Mountford, *Inorg. Chem.*, 2002, **41**, 1110–1119, DOI: 10.1021/ic011043z.

8 H. C. Guo, R. H. Zheng and H. J. Jiang, *Org. Prep. Proced. Int.*, 2012, **44**, 392–396, DOI: 10.1080/00304948.2012.697745.

9 A. P. Krapcho, S. Sparapani, A. Leenstra and J. D. Seitz, *Tetrahedron Lett.*, 2009, **50**, 3195–3197, DOI: 10.1016/j.tetlet.2009.01.138.

10 K. Bahrami, M. M. Khodaei and M. Soheilizad, *J. Org. Chem.*, 2009, **74**, 9287–9291, DOI: 10.1021/jo901924m.

11 EU Pat., EP19880300343 19880115, 1987, 16.

12 S. Madabhushi, R. Jillella, V. Sriramoju and R. Singh, *Green Chem.*, 2014, **16**, 3125–3131, DOI: 10.1039/C4GC00246F.

13 E. Flegeau, J. Harrison and M. Willis, *Synlett*, 2015, **27**, 101–105, DOI: 10.1055/s-0035-1560578.

14 T. Takeda and M. Terada, *J. Am. Chem. Soc.*, 2013, **135**, 15306–15309, DOI: 10.1021/ja408296h.

15 S. W. Wright and K. N. Hallstrom, *J. Org. Chem.*, 2006, **71**, 1080–1084, DOI: 10.1021/jo052164+.

16 S. Berg, M. Bergh, S. Hellberg, K. Högdin, Y. Lo-Alfredsson, P. Söderman, S. Von Berg, T. Weigelt, M. Ormö, Y. Xue, J. Tucker, J. Neelissen, E. Jerning, Y. Nilsson and R. Bhat, *J. Med. Chem.*, 2012, **55**, 9107–9119, DOI: 10.1021/jm201724m.

17 A. J. Brouwer, T. Ceylan, T. van der Linden and R. M. J. Liskamp, *Tetrahedron Lett.*, 2009, **50**, 3391–3393, DOI: 10.1016/j.tetlet.2009.02.130.

18 A. V. Bogolubsky, Y. S. Moroz, P. K. Mykhailiuk, S. E. Pipko, A. I. Konovets, I. V. Sadkova and A. Tolmachev, *ACS Comb. Sci.*, 2014, **16**, 192–197, DOI: 10.1021/co400164z.

19 R. Zelli, S. Tommasone, P. Dumy, A. Marra and A. Dondoni, *European J. Org. Chem.*, 2016, **2016**, 5102–5116, DOI: 10.1002/ejoc.201600732.

20 J. J. Rojas, R. A. Croft, A. J. Sterling, E. L. Briggs, D. Antermite, D. C. Schmitt, L. Blagojevic, P. Haycock, A. J. P. White, F. Duarte, C. Choi, J. J. Mousseau and J. A. Bull, *Nat. Chem.*, 2022, **14**, 160–169, DOI: 10.1038/s41557-021-00856-2.

21 C. L. Xiao, C. Z. Wang, L. Y. Yuan, B. Li, H. He, S. Wang, Y. L. Zhao, Z. F. Chai and W. Q. Shi, *Inorg. Chem.*, 2014, **53**, 1712–1720, DOI: 10.1021/ic402784c.

22 A. M. Craciun, A. Rotaru, C. Cojocaru, I. I. Mangalagiu and R. Danac, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, 2021, **249**, 119318, DOI: 10.1016/j.saa.2020.119318.

Chapter 8

Appendix

8.1 Compound characterisation data

Phosphasalen 1 Salt (PS1-salt)

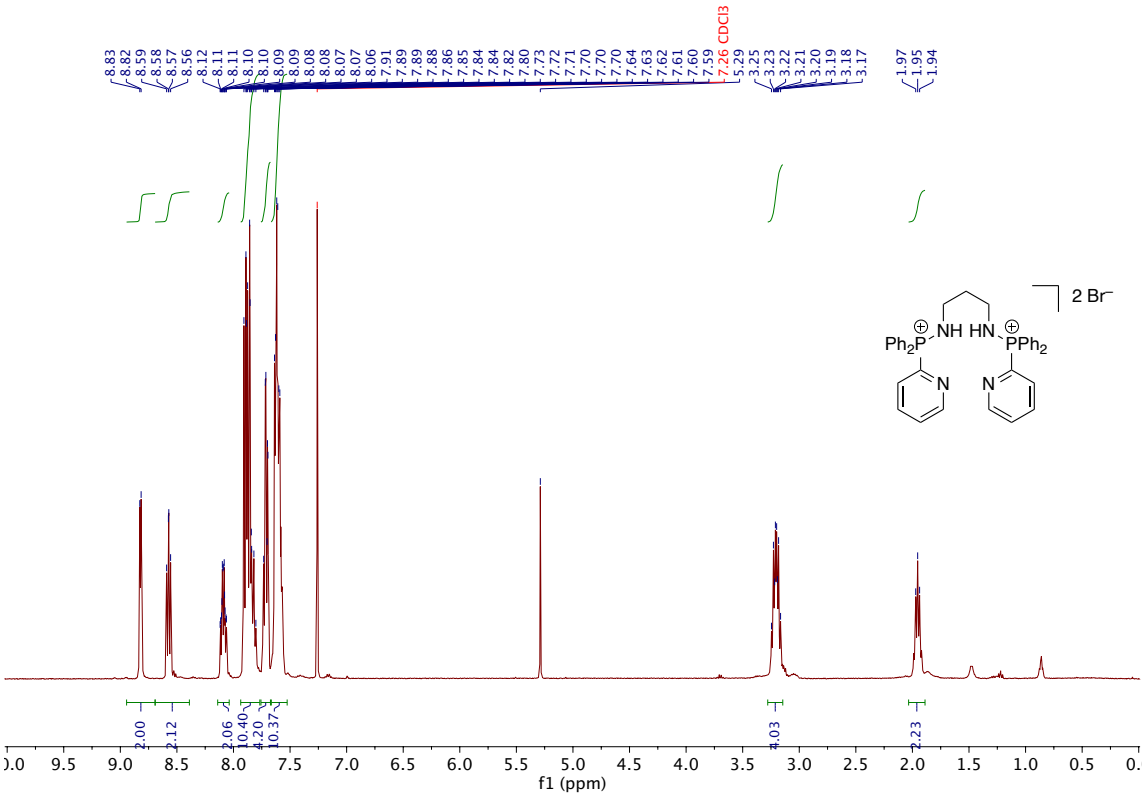


*Figure 8.1: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **PS1-salt**.*

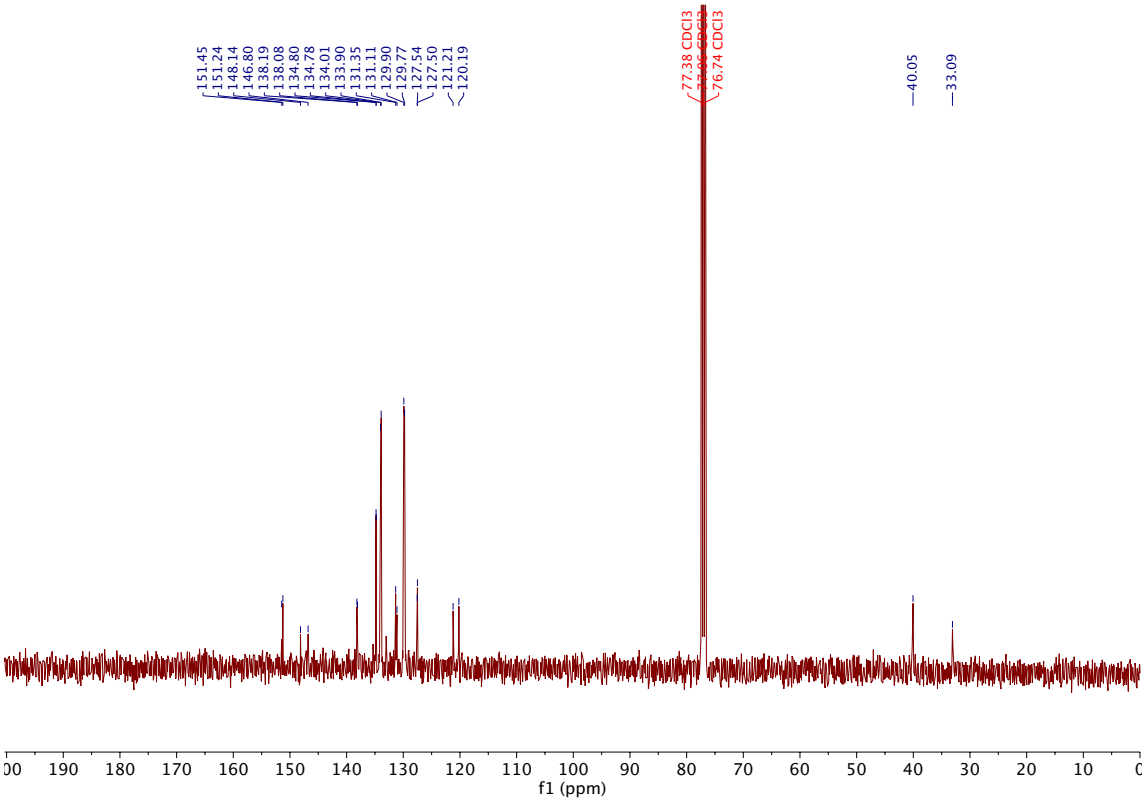


*Figure 8.2: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, $CDCl_3$) of **PS1-salt**.*



*Figure 8.3: HMQC NMR spectrum of **PS1-salt**.*

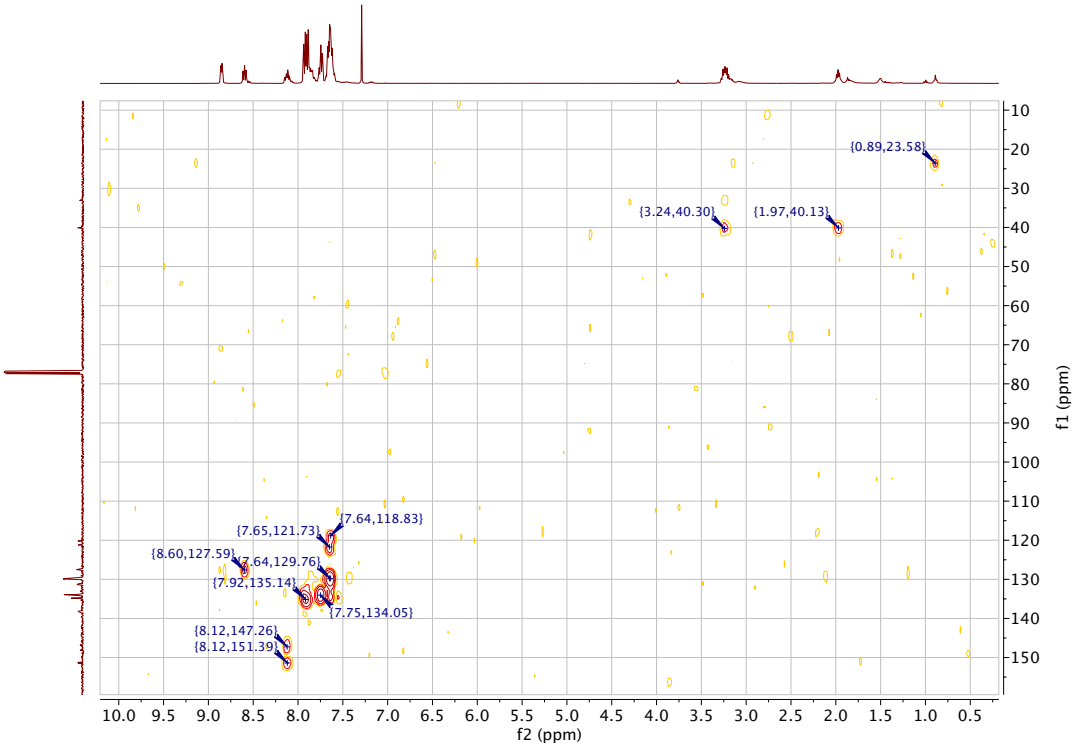


*Figure 8.4: HMBC NMR spectrum of **PS1-salt**.*

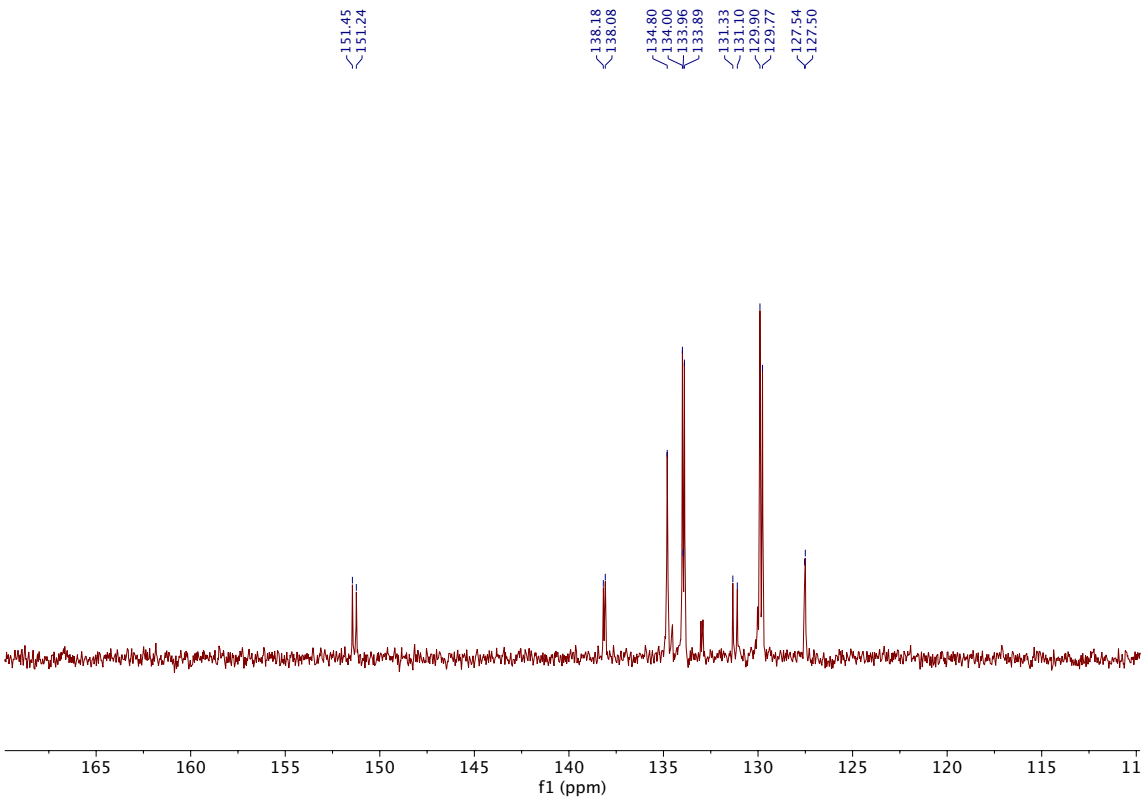


*Figure 8.5: DEPT-90 NMR spectrum of **PS1-salt**.*

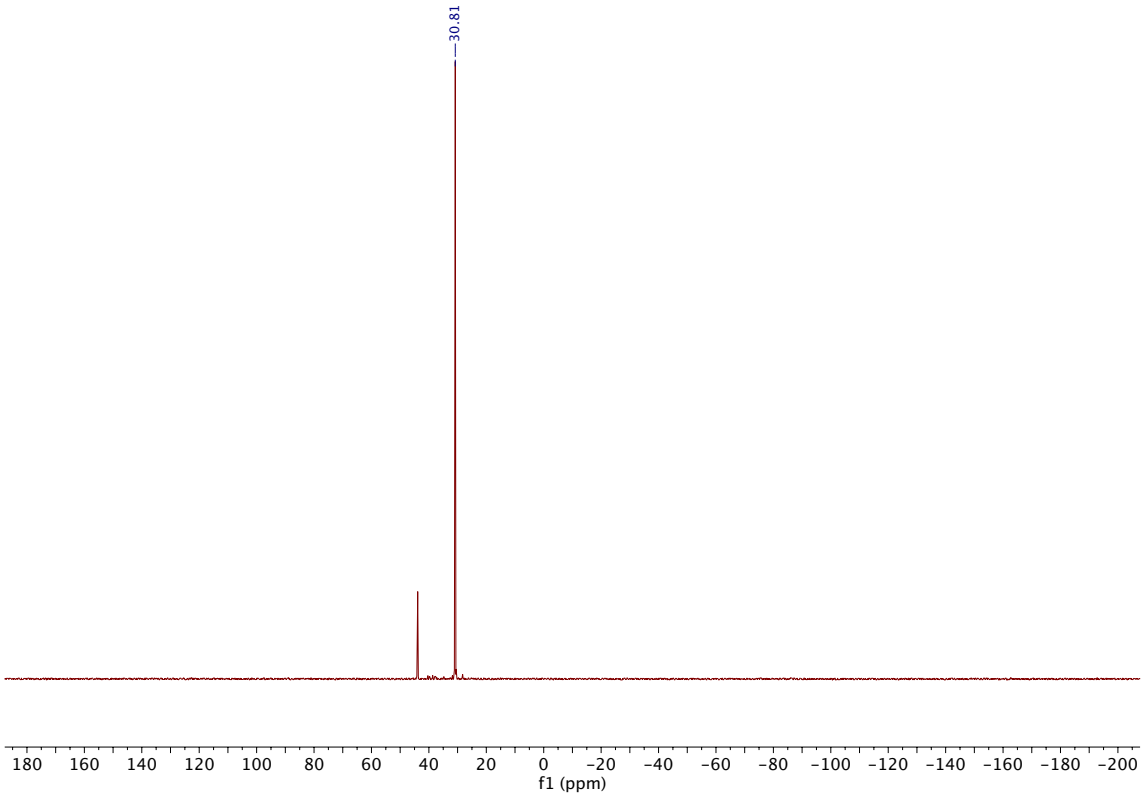


*Figure 8.6: $^{31}P\{^{1}H\}$ NMR spectrum (121.5 MHz, $CDCl_3$) of **PS1-salt**.*

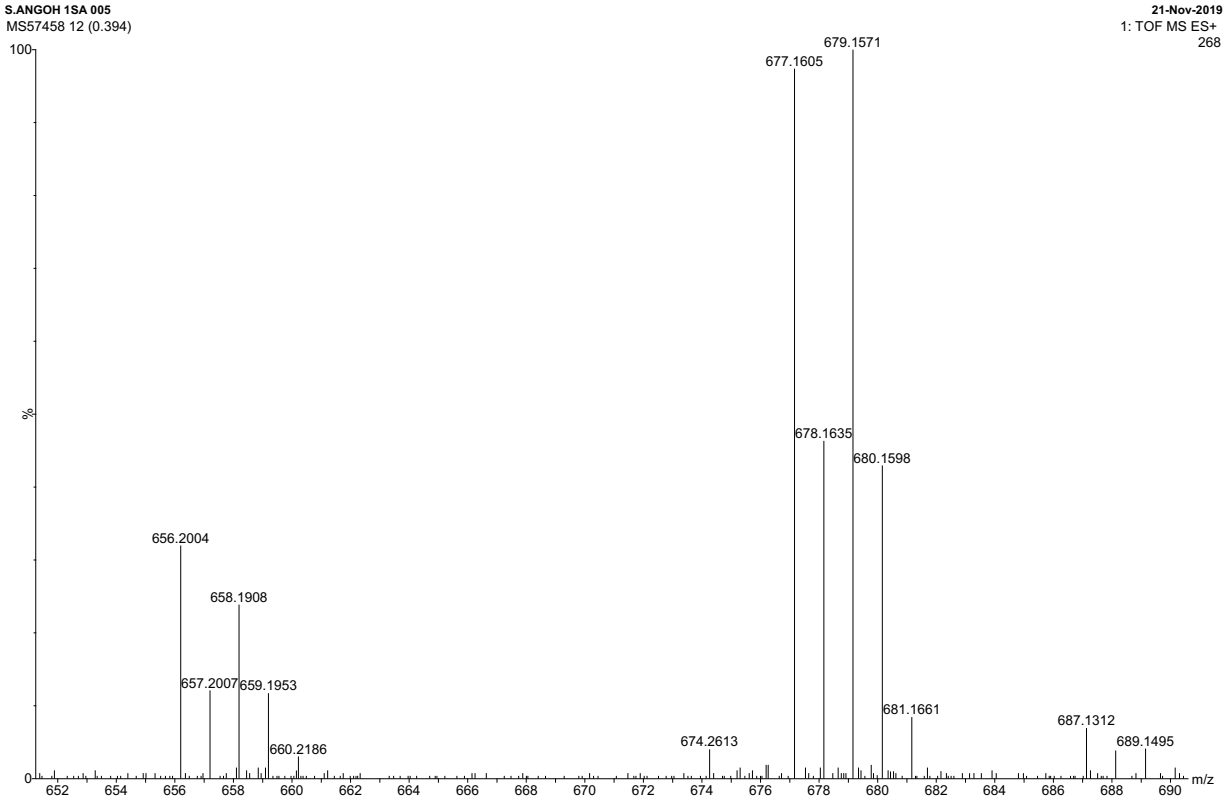


*Figure 8.7: HR-MS (ES^+ ToF) of **PS1-salt**.*

Phosphasalen 3 Salt (PS3-salt)

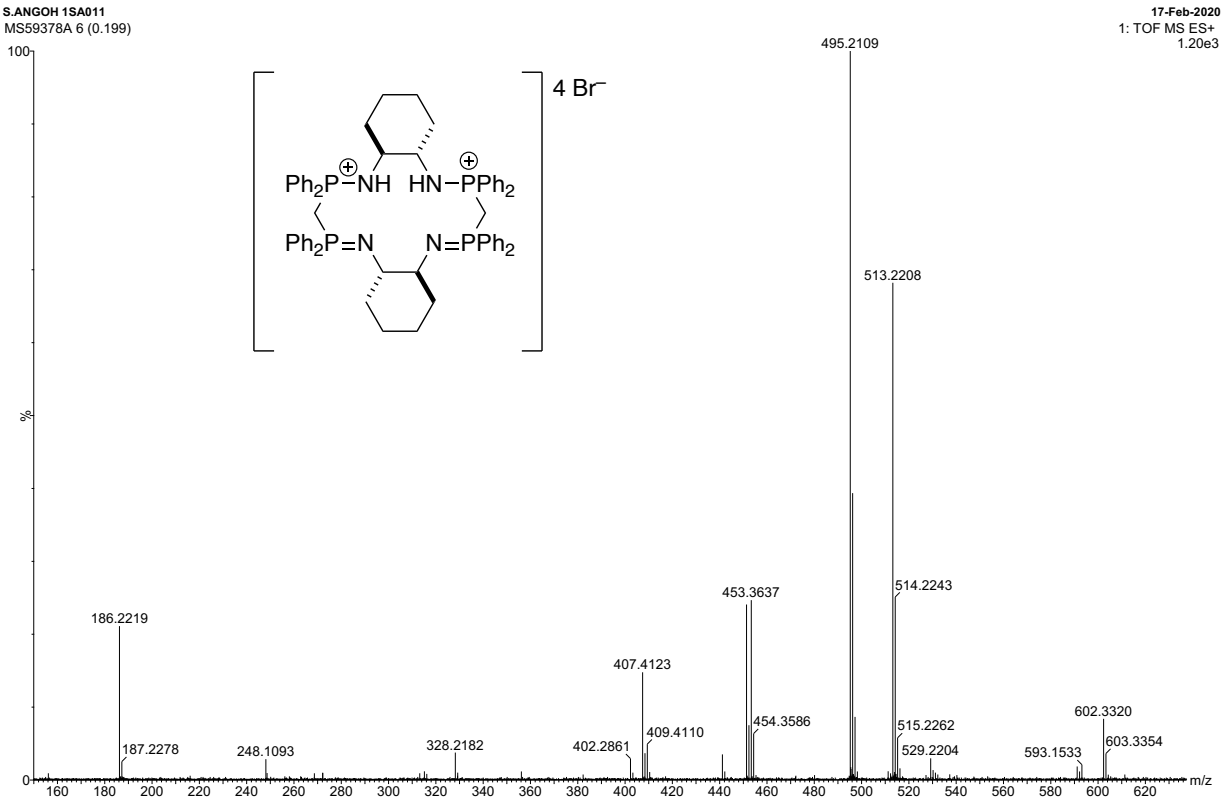


*Figure 8.8: HR-MS (ES⁺ ToF) of **PS3-salt**.*

Benzyl 3-(bis(2-(2,2,2-trifluoroacetamido)ethyl)amino)propanoate, (2.12)

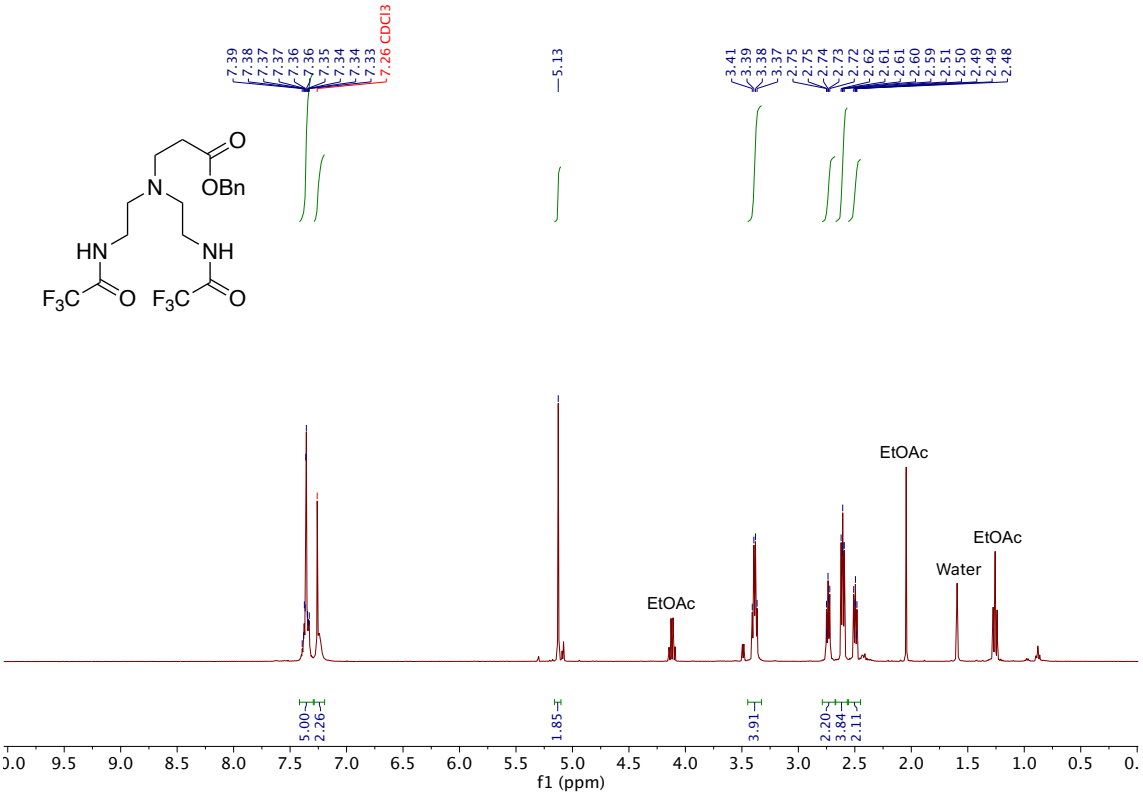


*Figure 8.9: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **2.12**.*

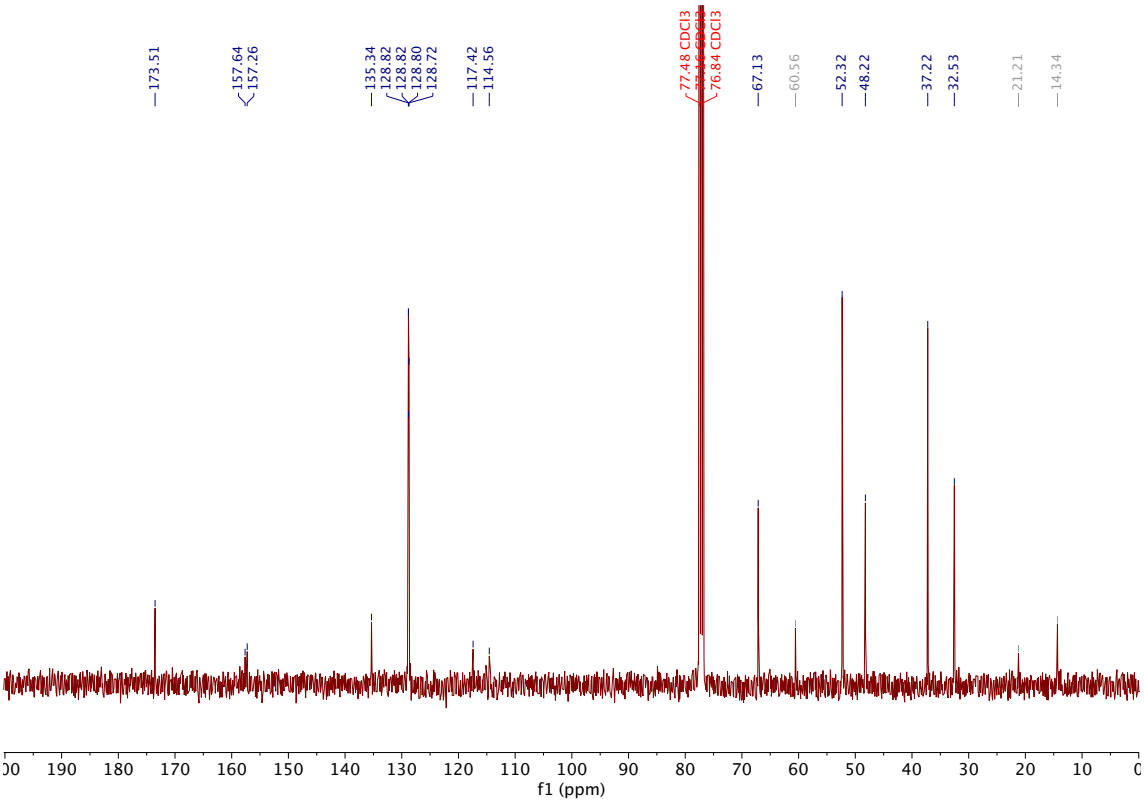


*Figure 8.10: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, $CDCl_3$) of **2.12**.*



*Figure 8.11: HMQC NMR spectrum of **2.12**.*

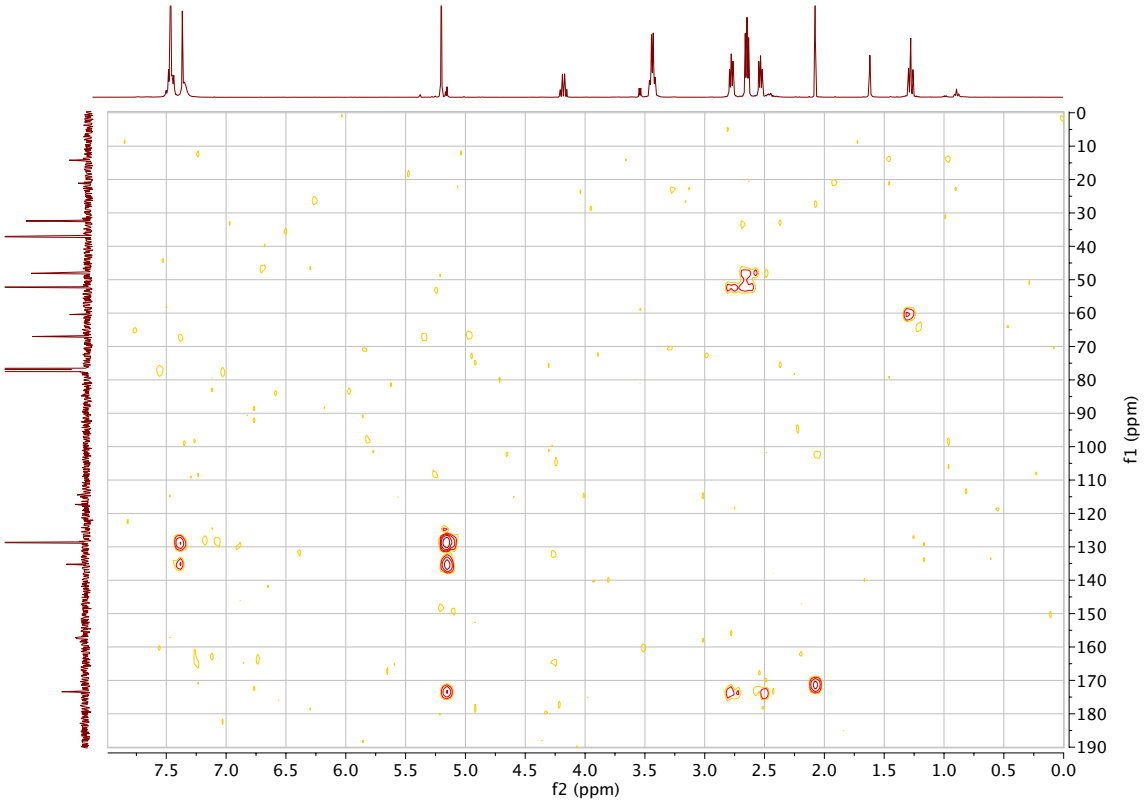


*Figure 8.12: HMBC NMR spectrum of **2.12**.*

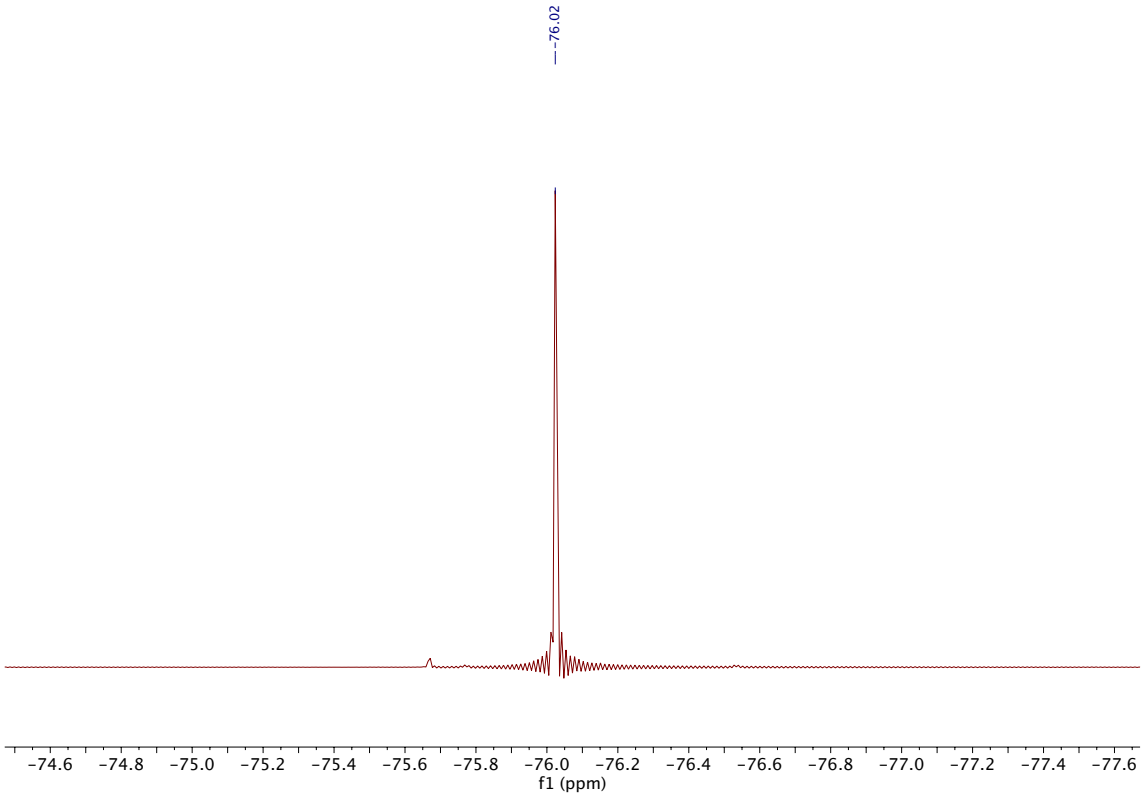


*Figure 8.13: $^{19}F\{^{1}H\}$ NMR spectrum (377 MHz, $CDCl_3$) of **2.12**.*

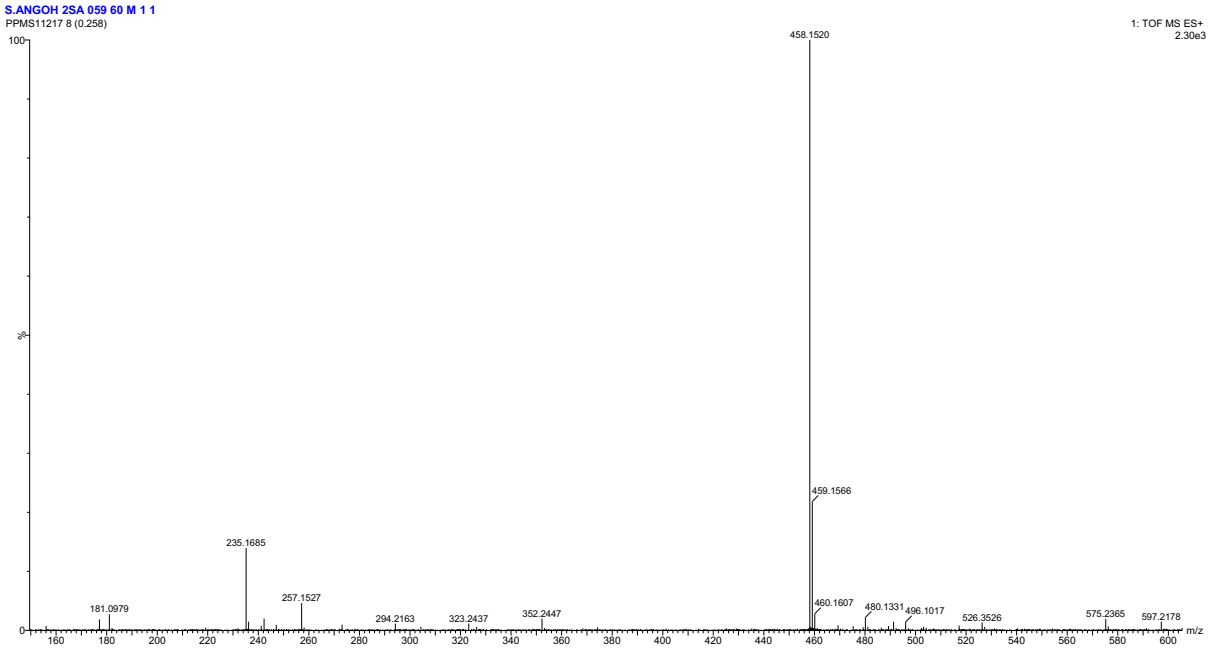


*Figure 8.14: HR-MS (ES^+ ToF) of **2.12**.*

2,9-diisopropoxy-1,10-phenanthroline (2.20)

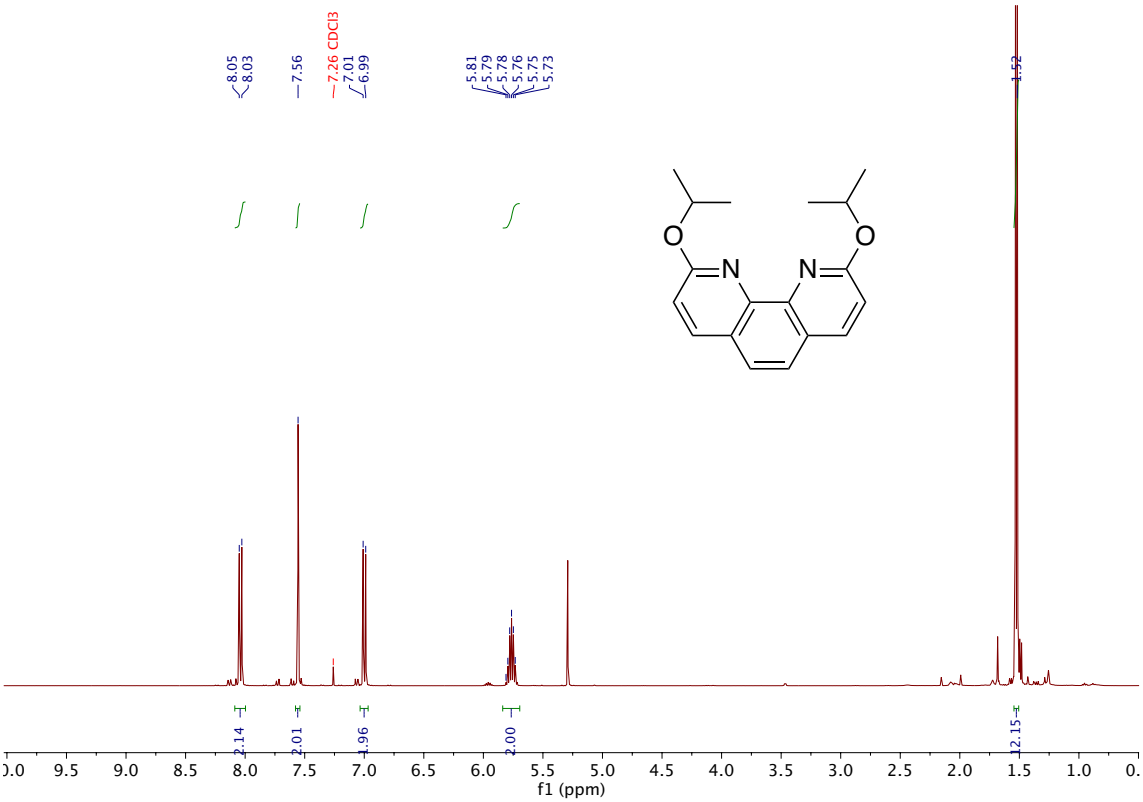


*Figure 8.15: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **2.20**.*

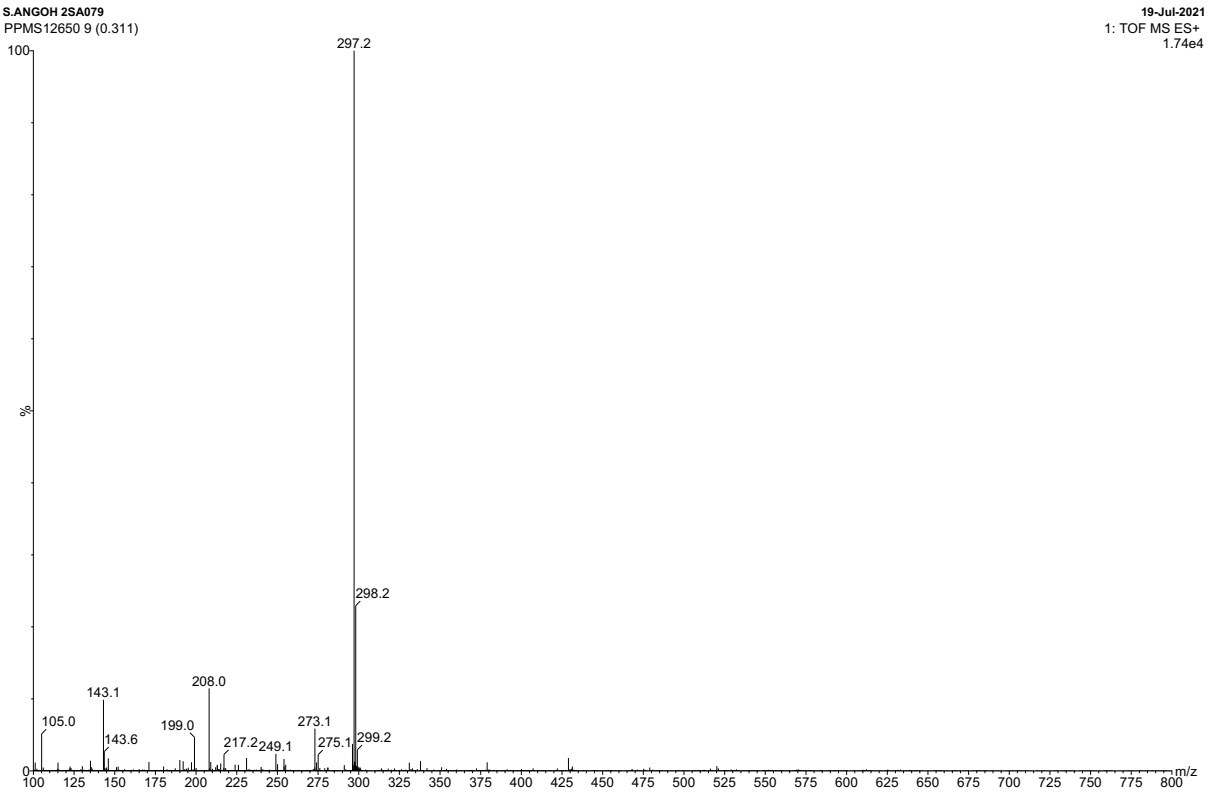


*Figure 8.16: HR-MS (ES^{+} ToF) of **2.20**.*

Methyl 2-(N-benzyl-N-(2-bromoethyl)sulfamoyl)acetate (2.23)

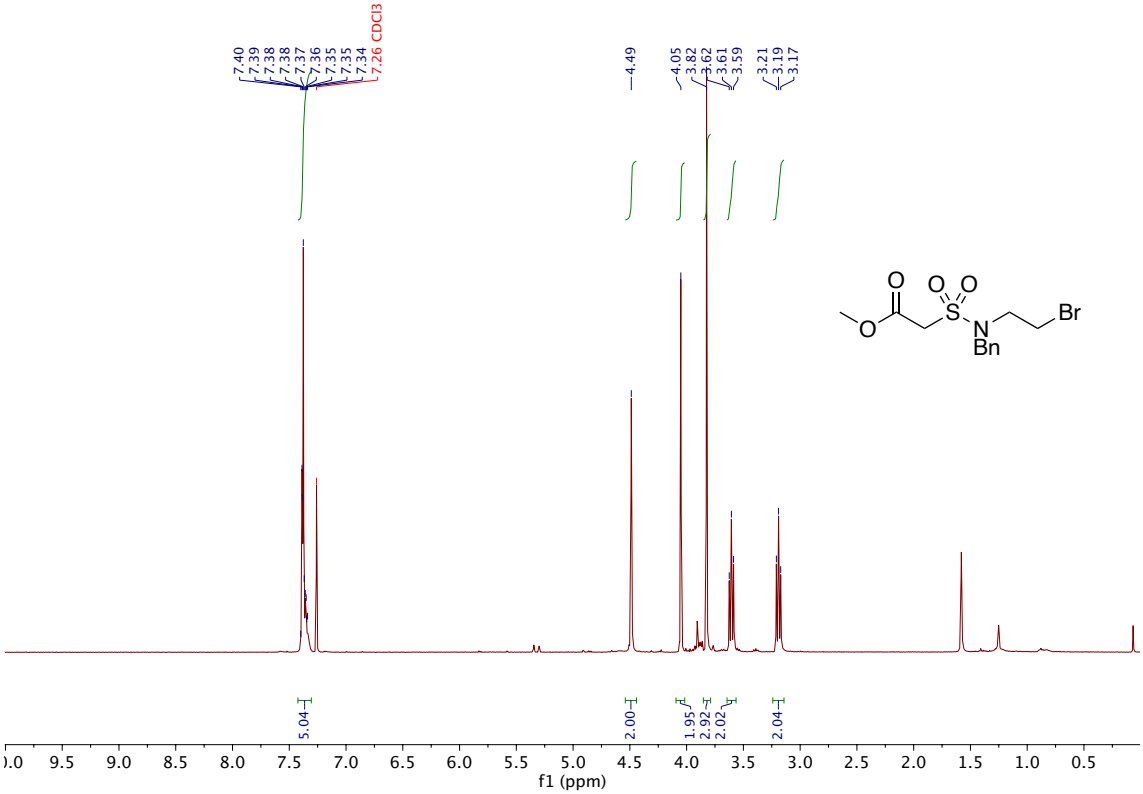


*Figure 8.17: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **2.23**.*

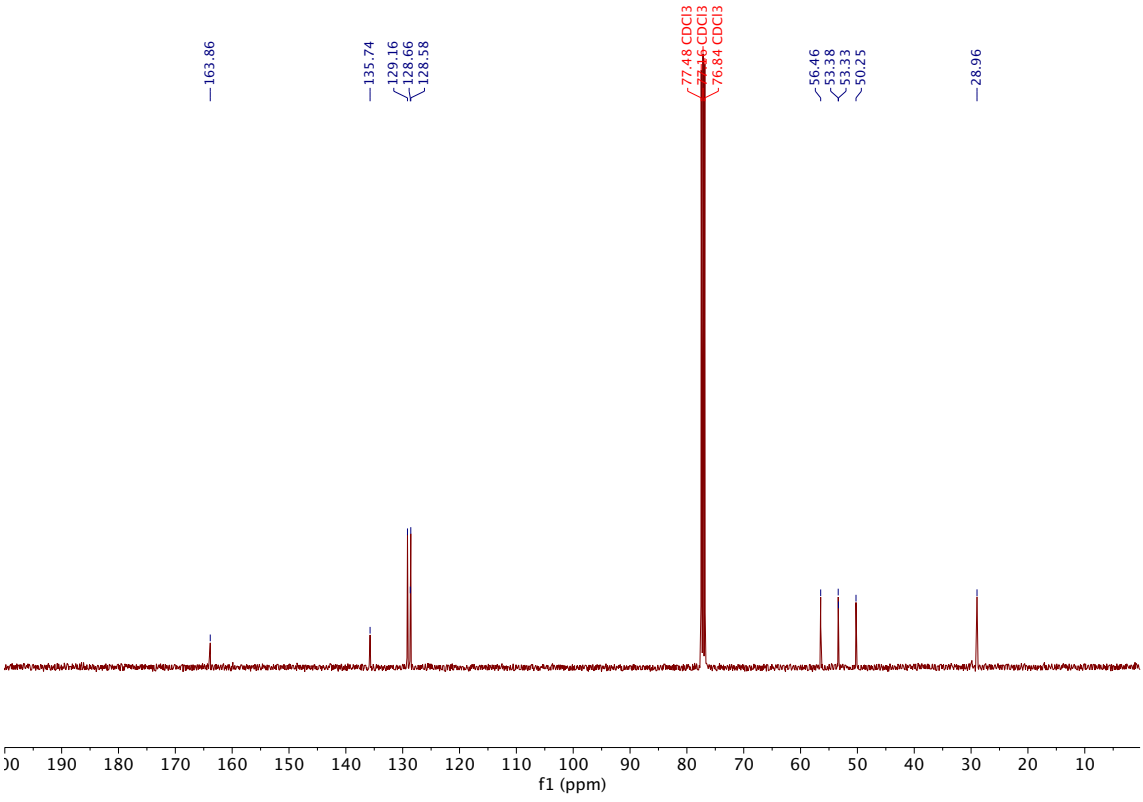


*Figure 8.18: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, $CDCl_3$) of **2.23**.*

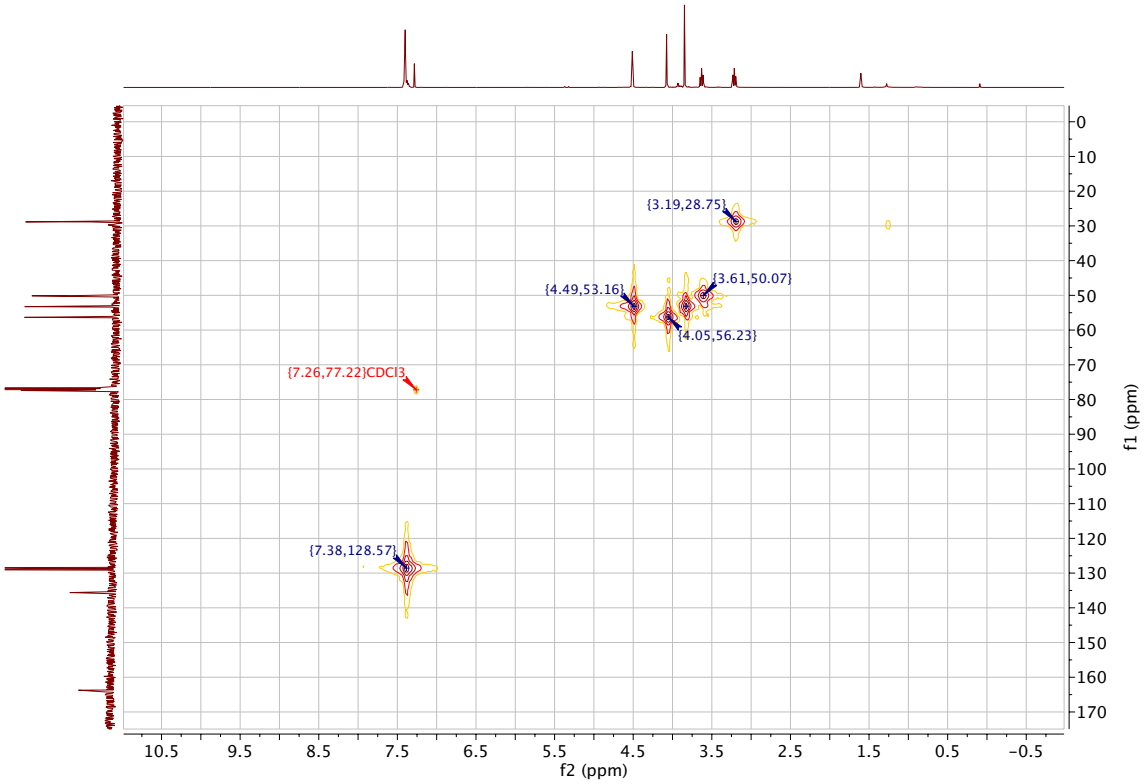


*Figure 8.19: HMQC NMR spectrum of **2.23**.*

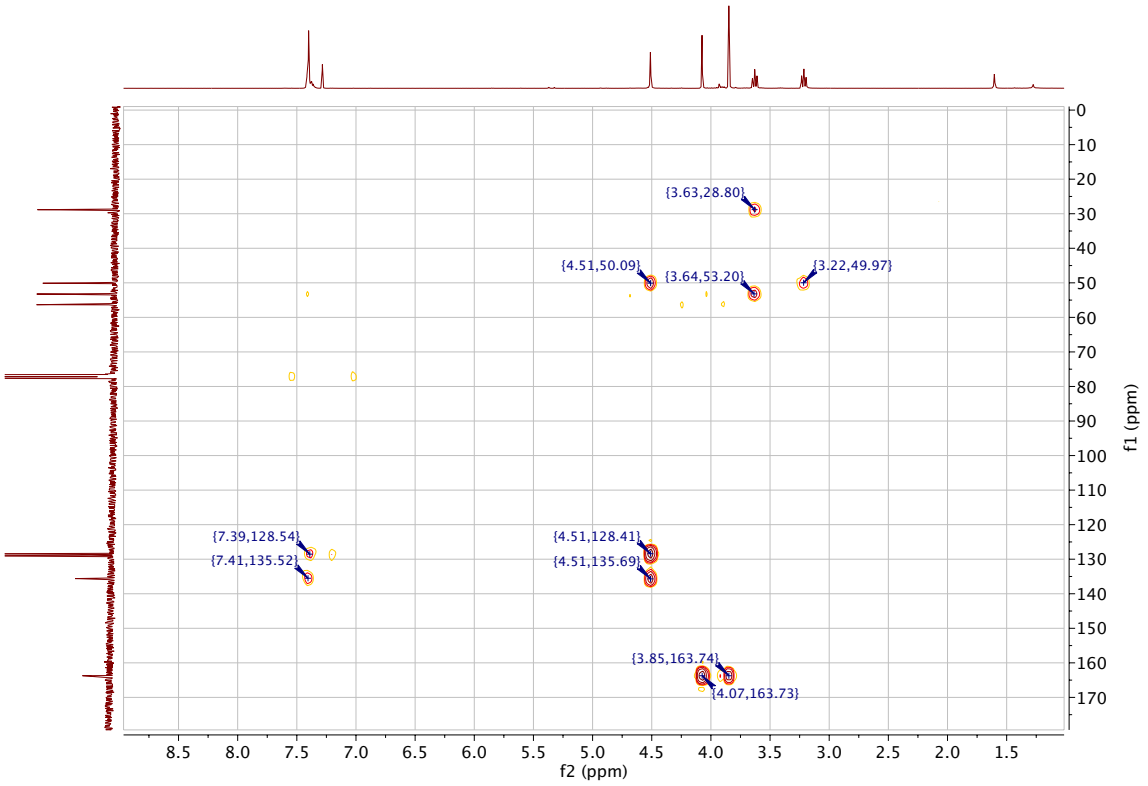


*Figure 8.20: HMBC NMR spectrum of **2.23**.*

N^2,N^9-bis(2-bromoethyl)-1,10-phenanthroline-2,9-dicarboxamide (2.27)

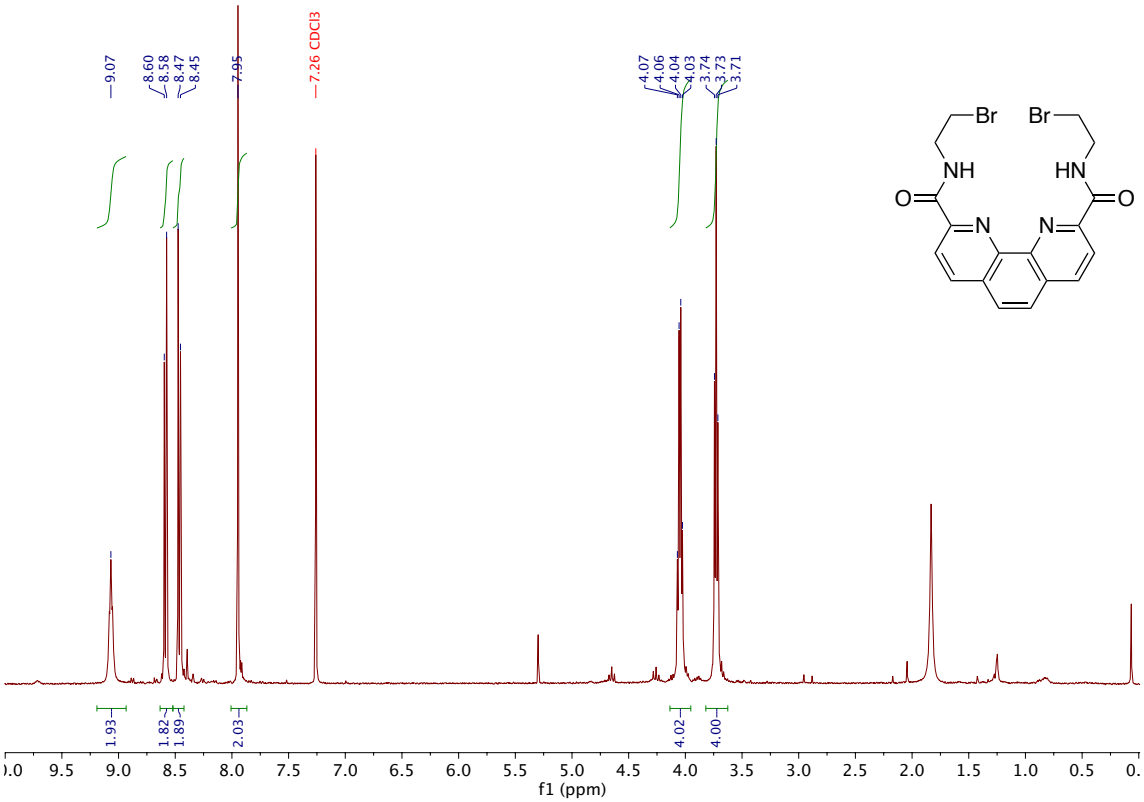


*Figure 8.21: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **2.27**.*

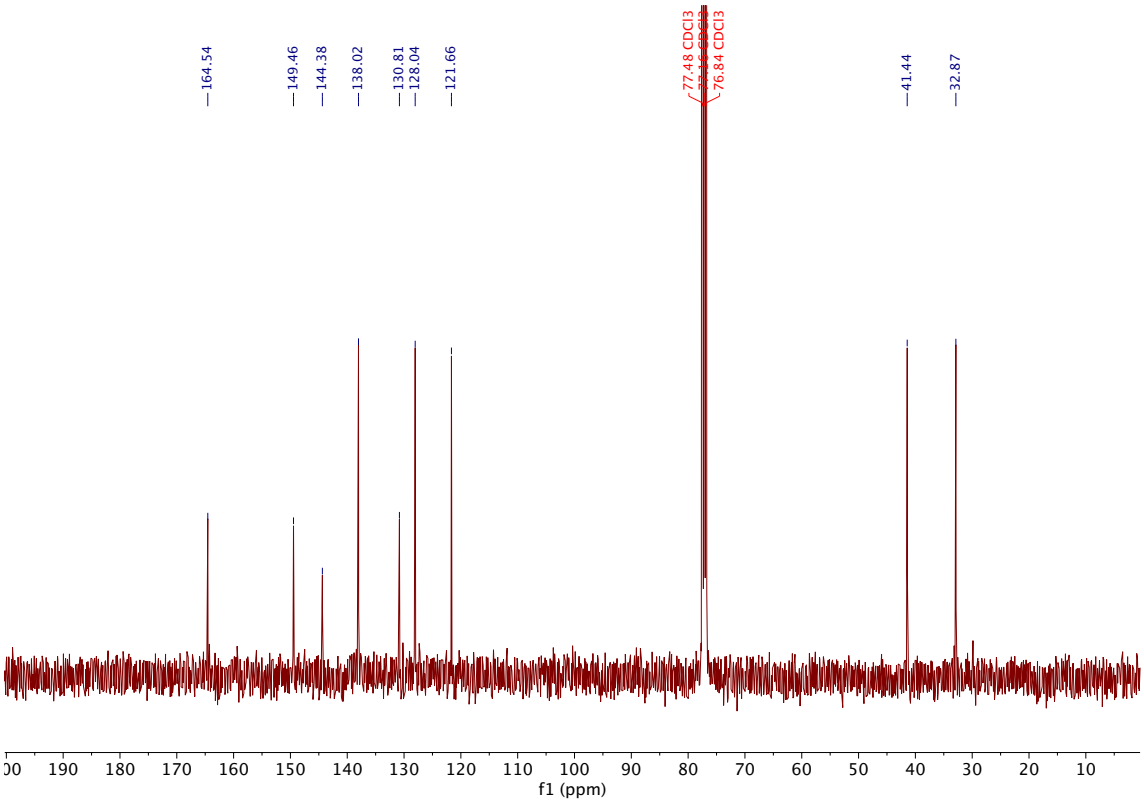


*Figure 8.22: $^{13}C\{^1H\}$ NMR spectrum (101 MHz, $CDCl_3$) of **2.27**.*

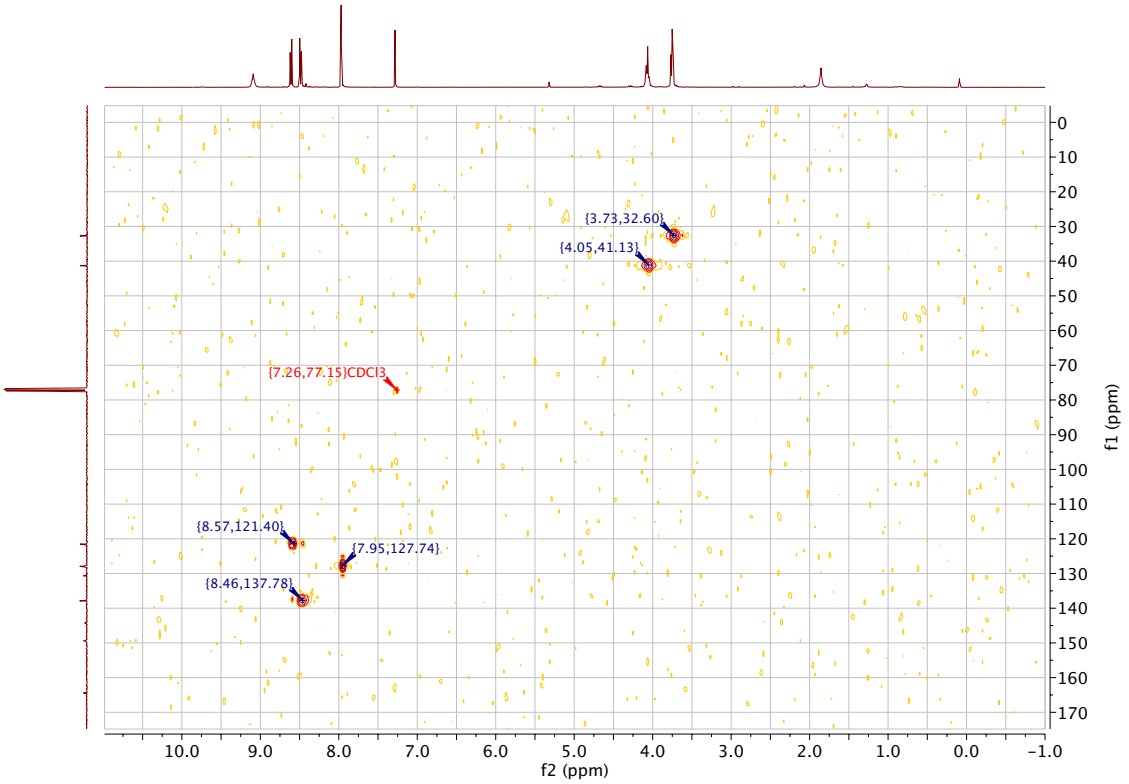


*Figure 8.23: HMQC NMR spectrum of **2.27**.*

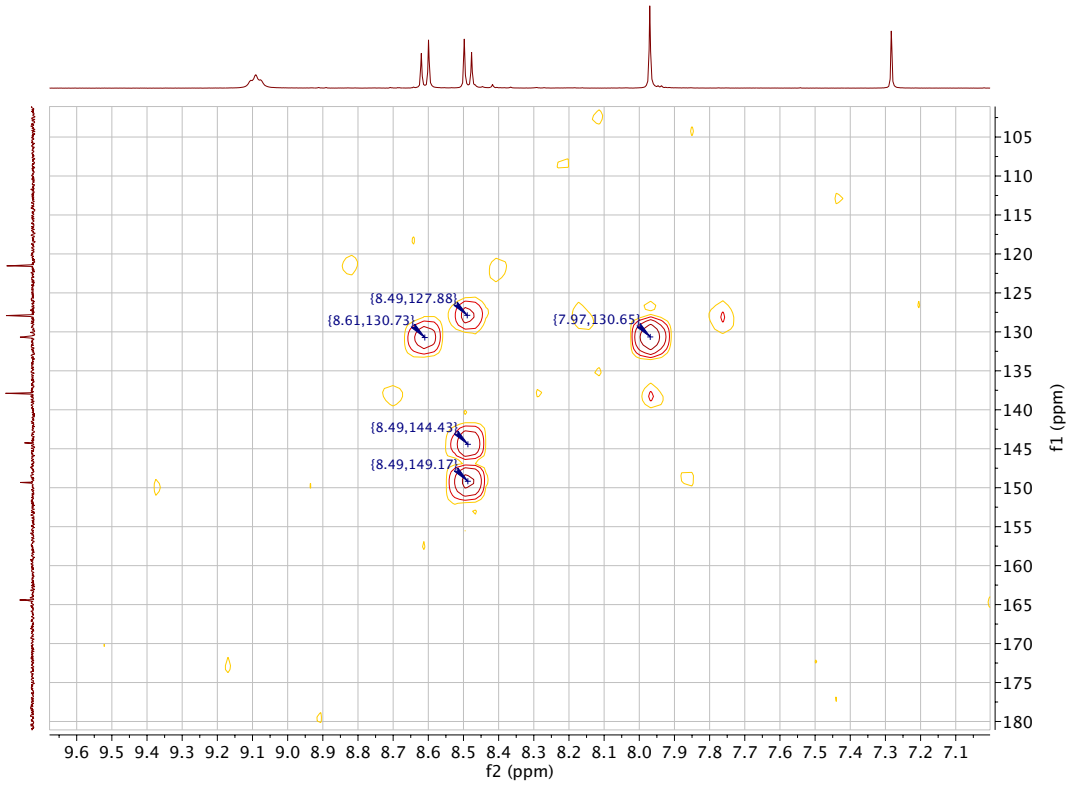


*Figure 8.24: HMBC NMR spectrum of **2.27**.*

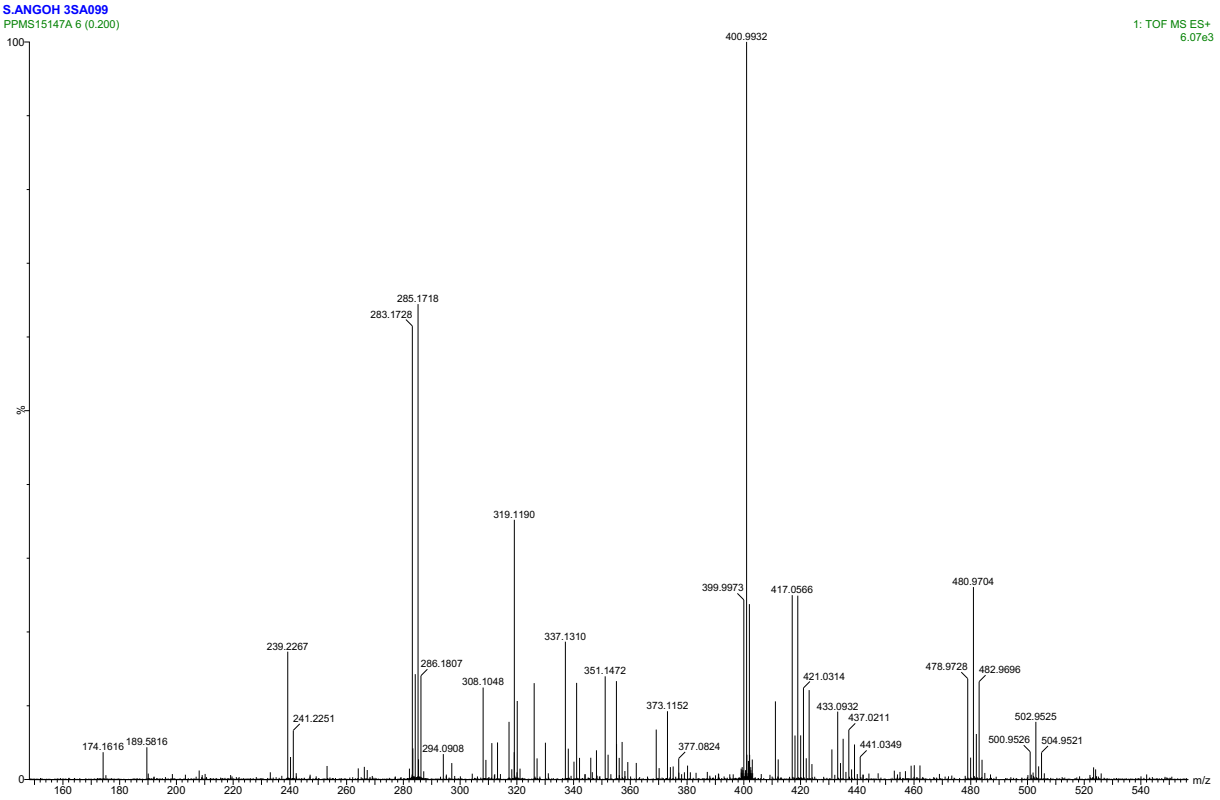


*Figure 8.25: HR-MS (ES$^+$ ToF) of **2.27**.*

Di-*tert*-butyl 2,2'-((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))diacetate (PhenDA-2-^{t}Bu)

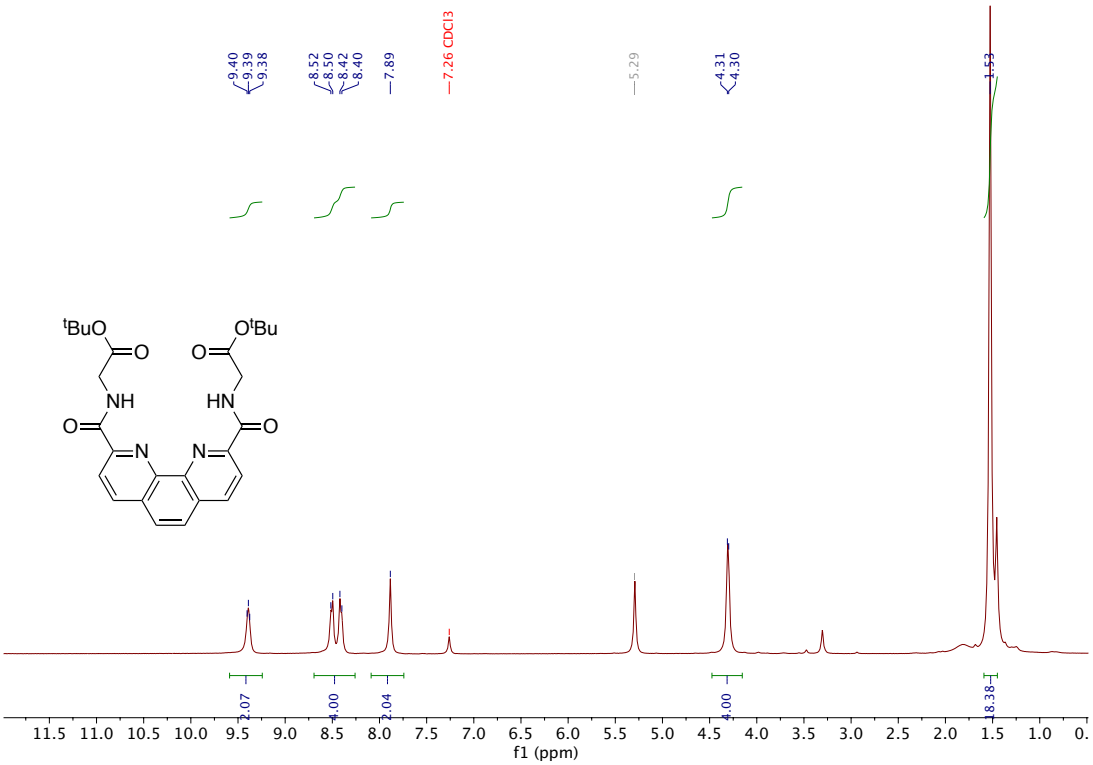


*Figure 8.26: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **PhenDA-2-^{t}Bu**.*

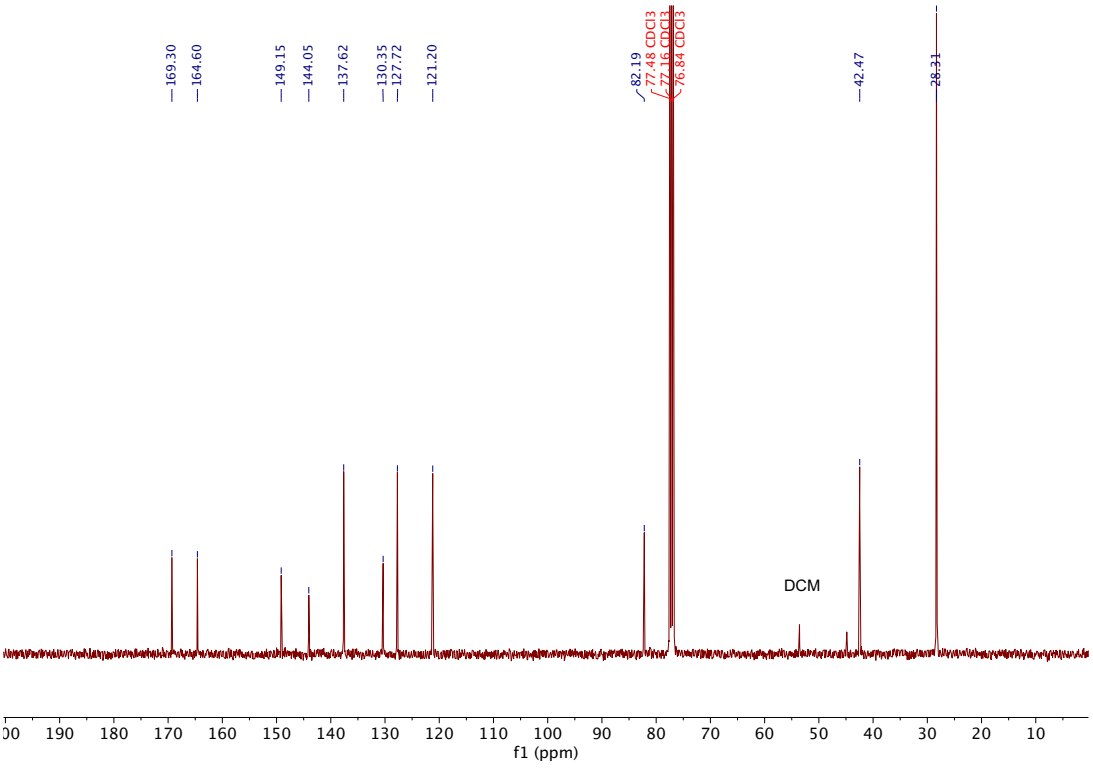


*Figure 8.27: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, $CDCl_3$) of **PhenDA-2-^{t}Bu**.*

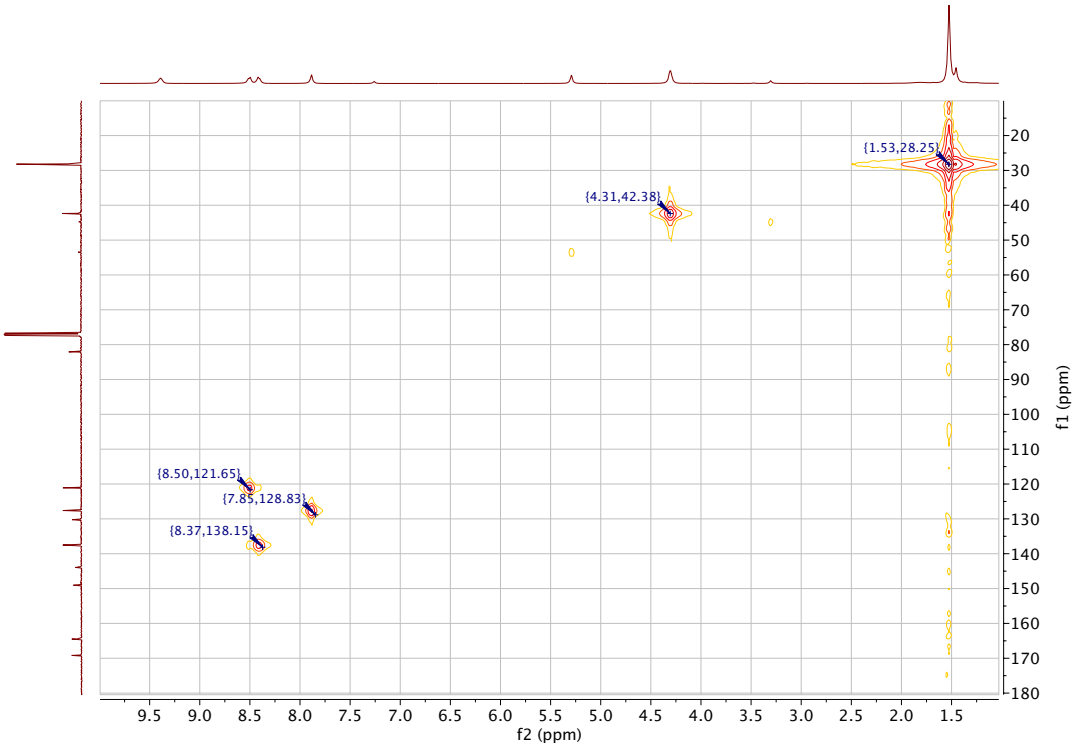


Figure 8.28: HMQC NMR spectrum of ***PhenDA-2-^{t}Bu***.

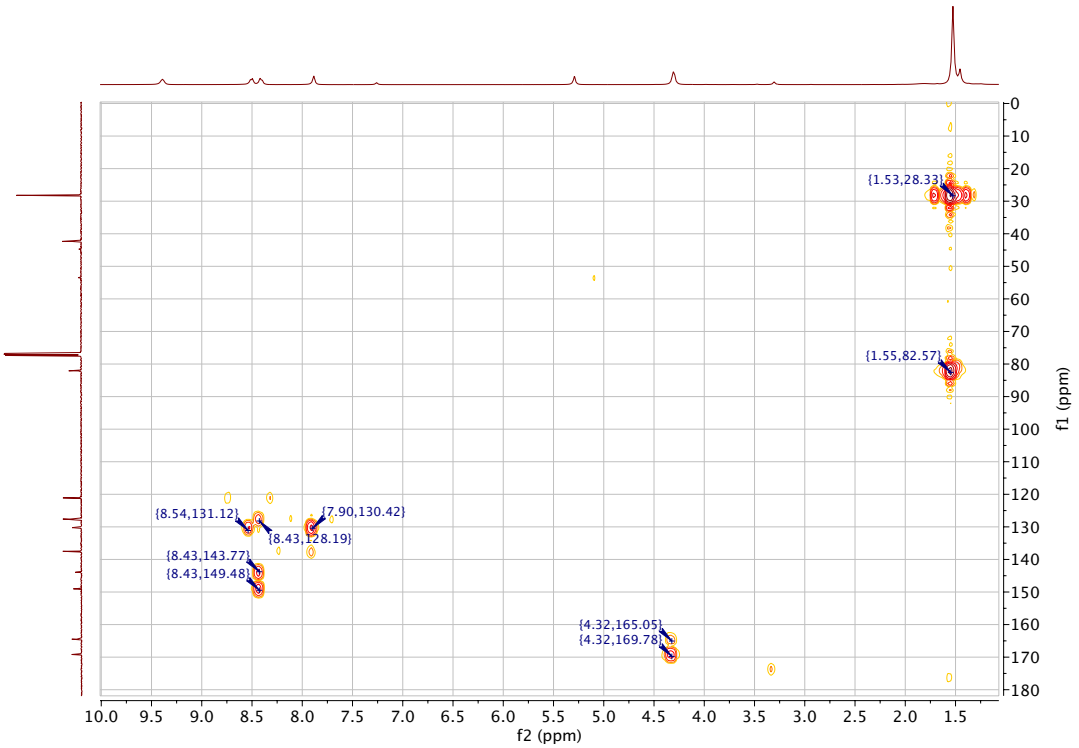


Figure 8.29: HMBC NMR spectrum of ***PhenDA-2-^{t}Bu***.

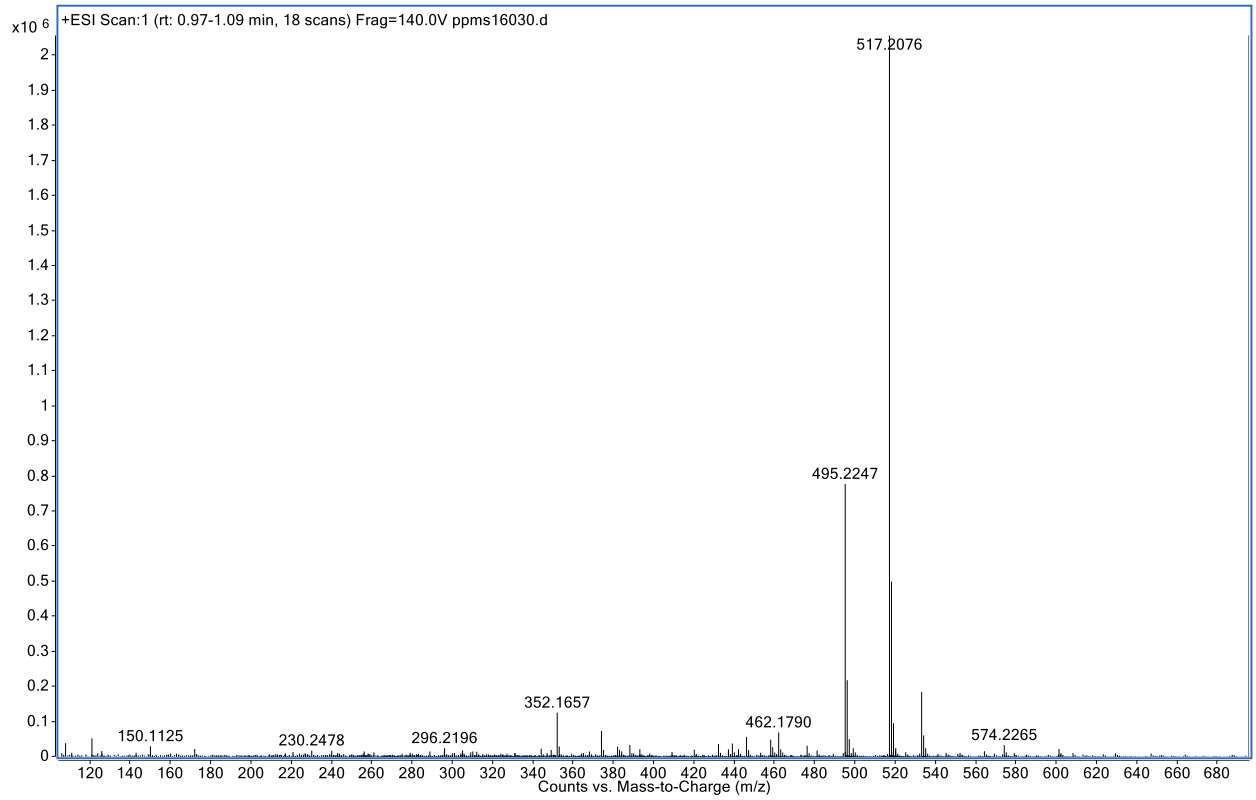


Figure 8.30: HR-MS (ES$^+$ ToF) of ***PhenDA-2-^{t}Bu****.*

2,2'-((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))diacetic acid trifluoroacetate (PhenDA-2)

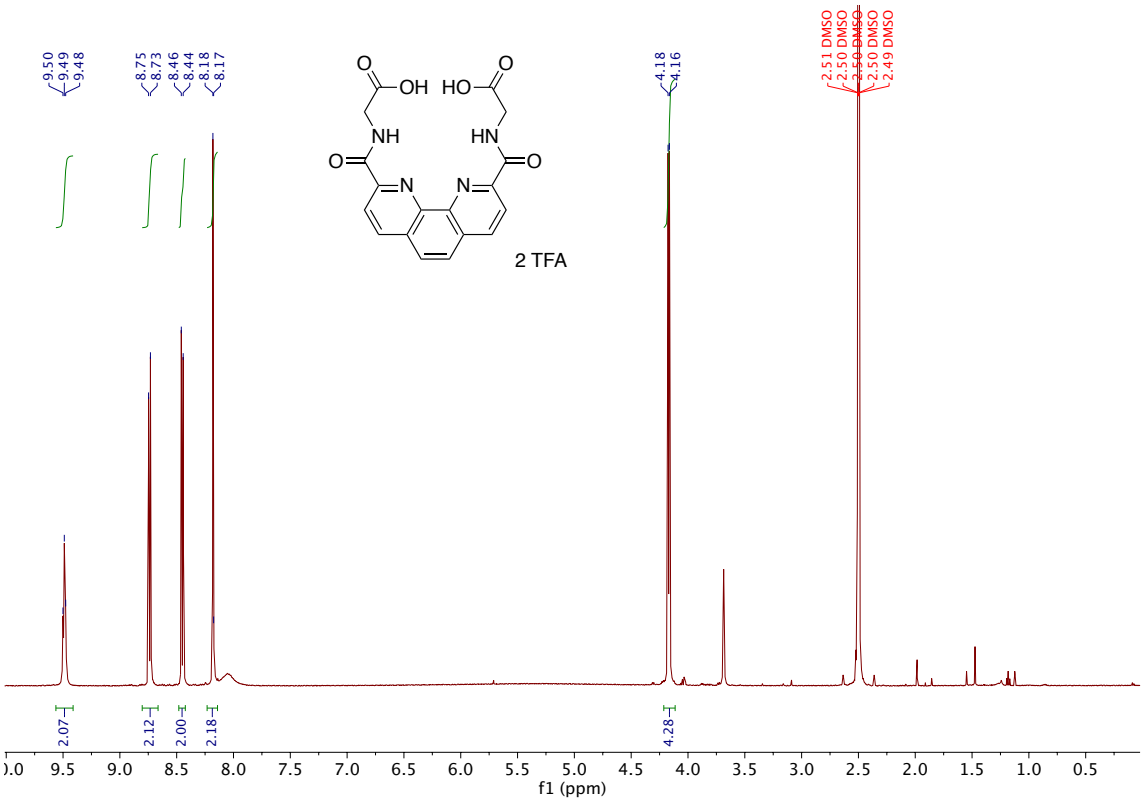


*Figure 8.31: ^{1}H NMR spectrum (400 MHz, 333 K, DMSO-d_6) of **PhenDA-2**.*

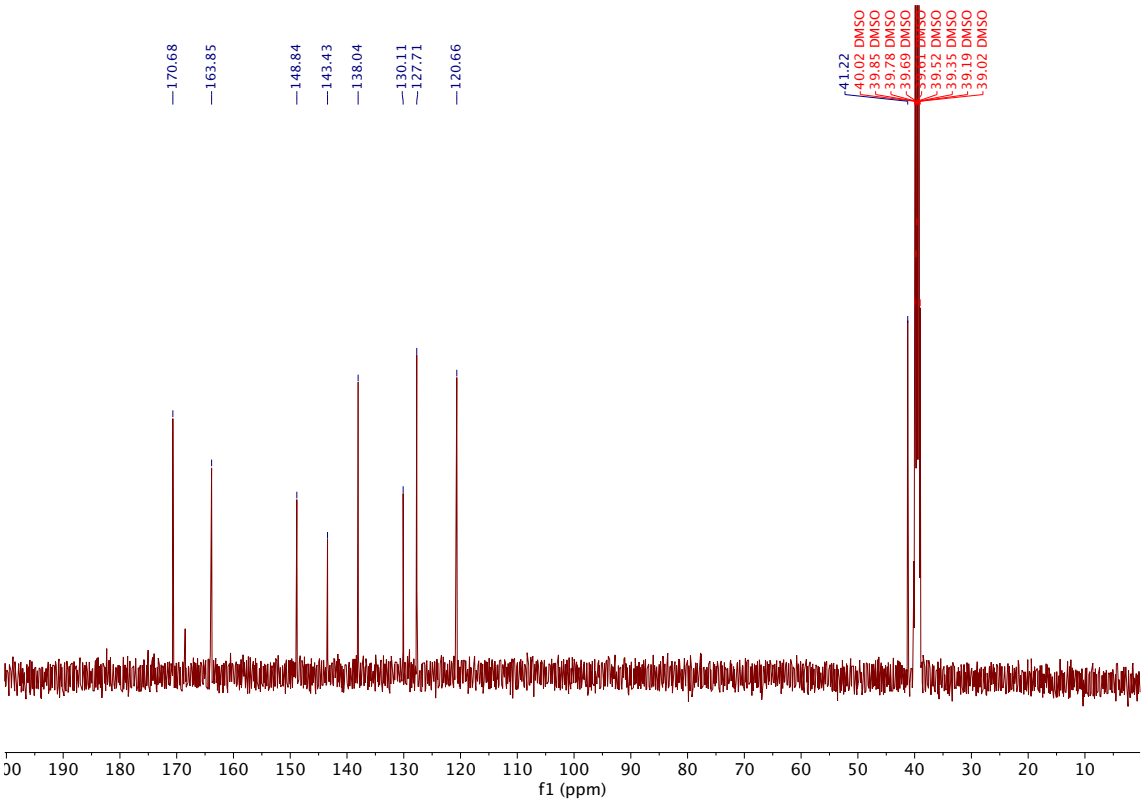


*Figure 8.32: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, 333 K, DMSO-d_6) of **PhenDA-2**.*

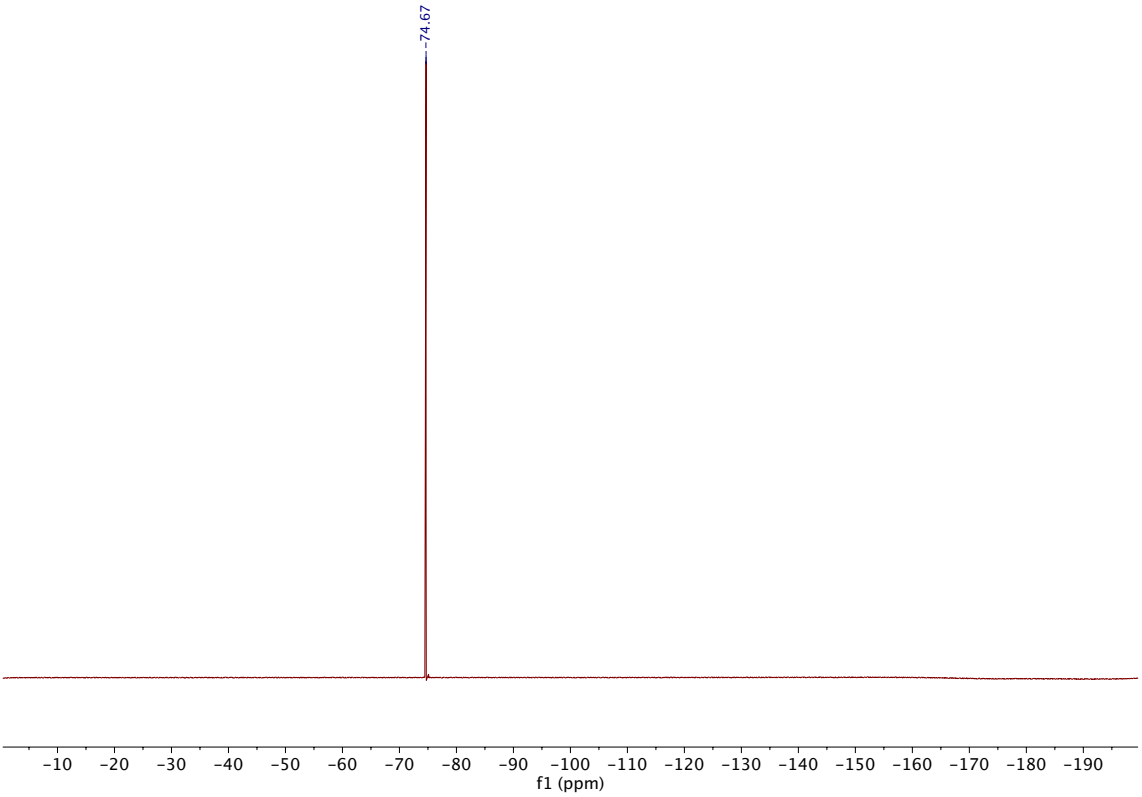


*Figure 8.33: $^{19}F\{^{1}H\}$ NMR spectrum (471 MHz, 333 K, DMSO-d_6) of **PhenDA-2**.*

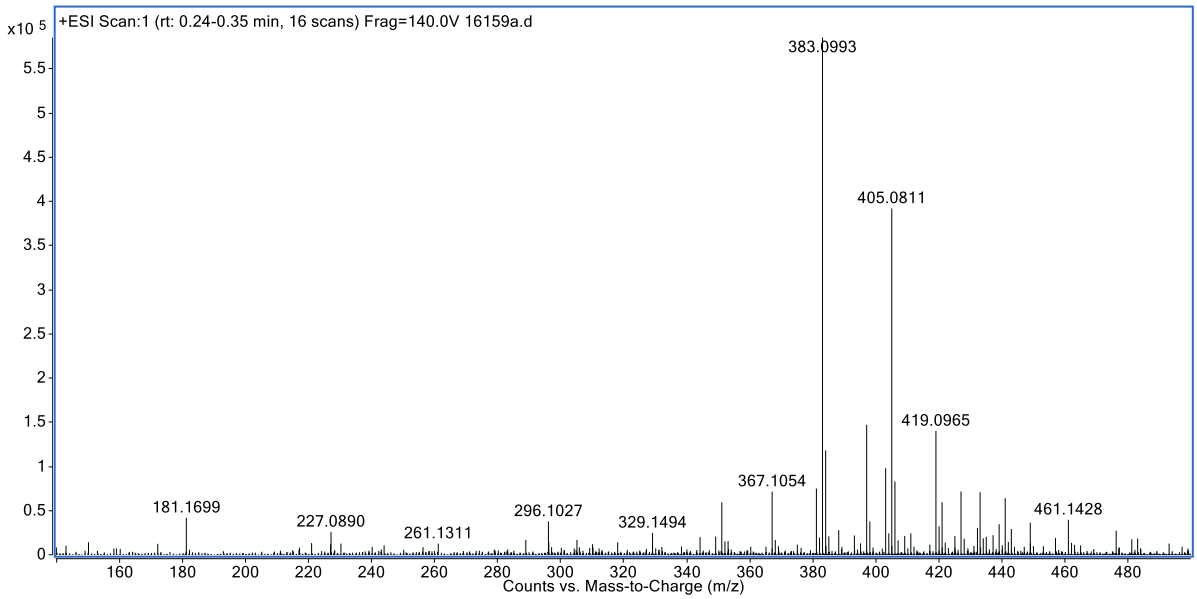


*Figure 8.34: HR-MS (ES$^+$ ToF) of **PhenDA-2**.*

Di-*tert*-butyl 3,3'-((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))dipropionate (PhenDA-3-^{t}Bu.

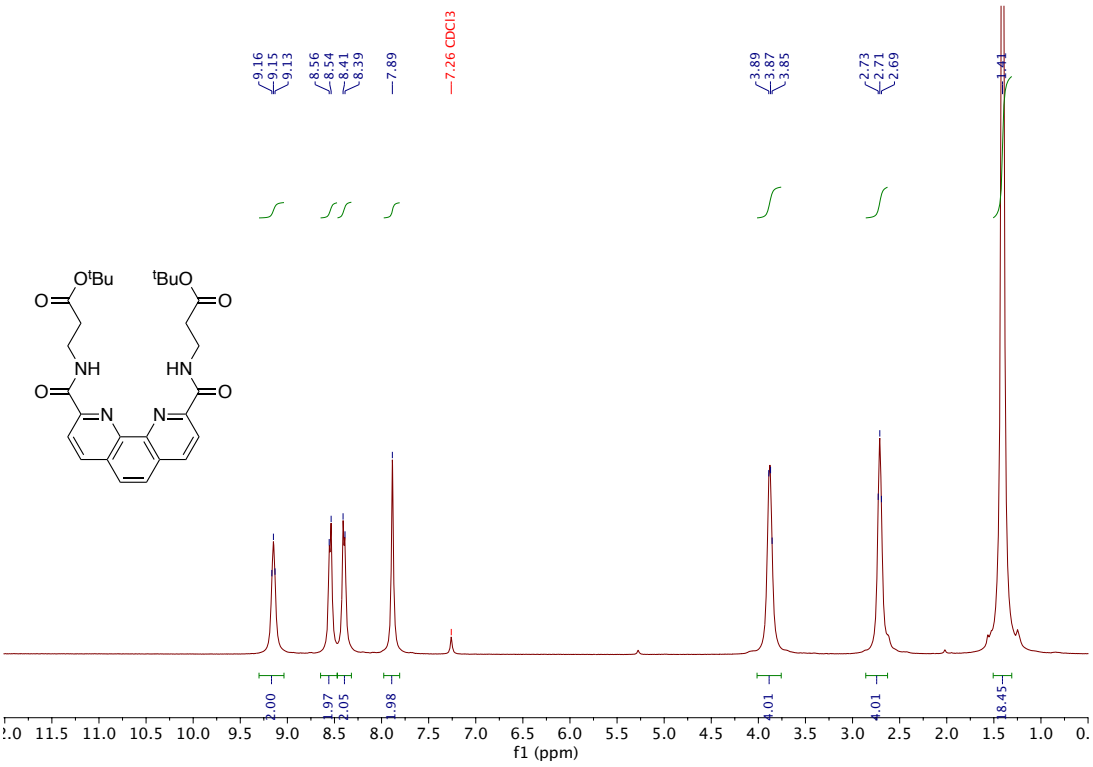


*Figure 8.35: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **PhenDA-3-^{t}Bu**.*

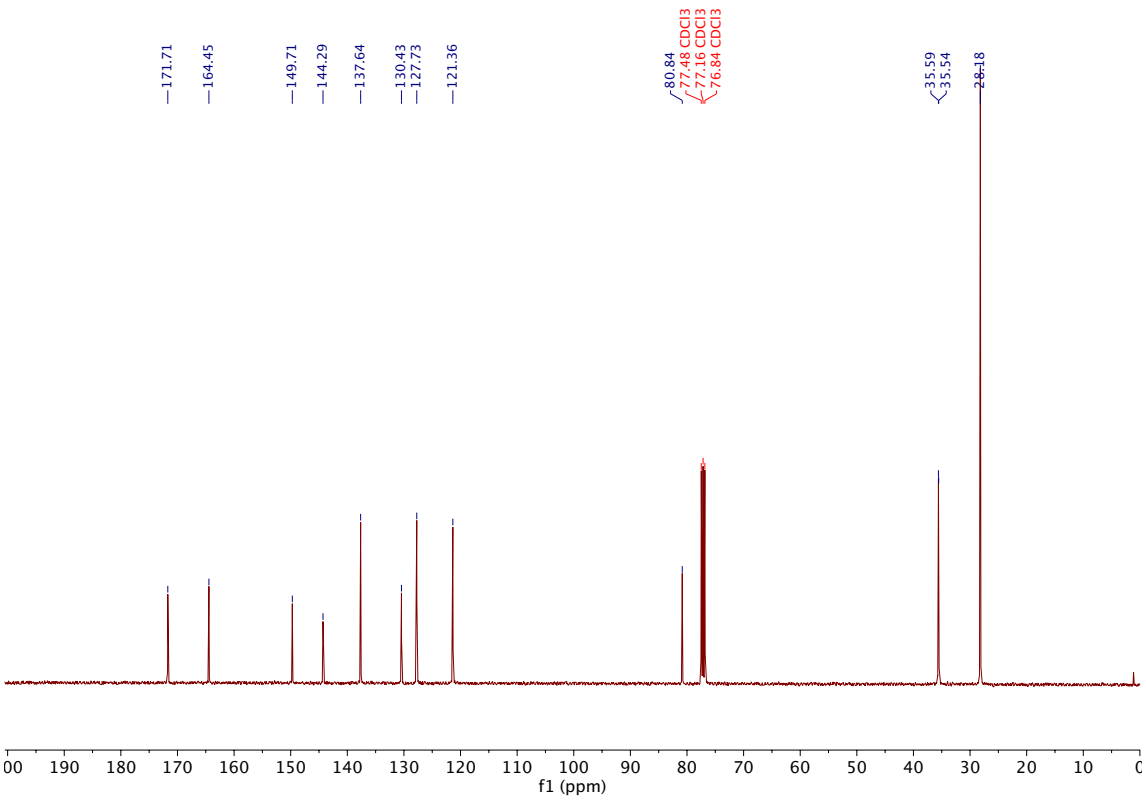


*Figure 8.36: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, $CDCl_3$) of **PhenDA-3-^{t}Bu**.*

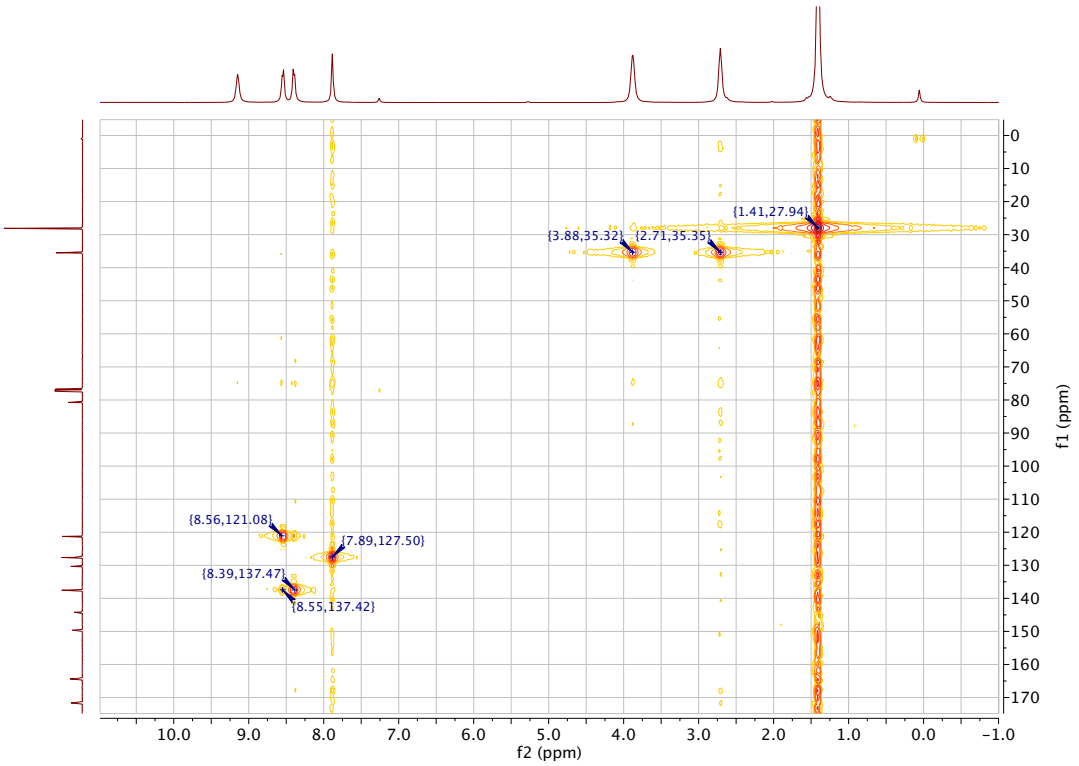


*Figure 8.37: HMQC NMR spectrum of **PhenDA-3-^{t}Bu**.*

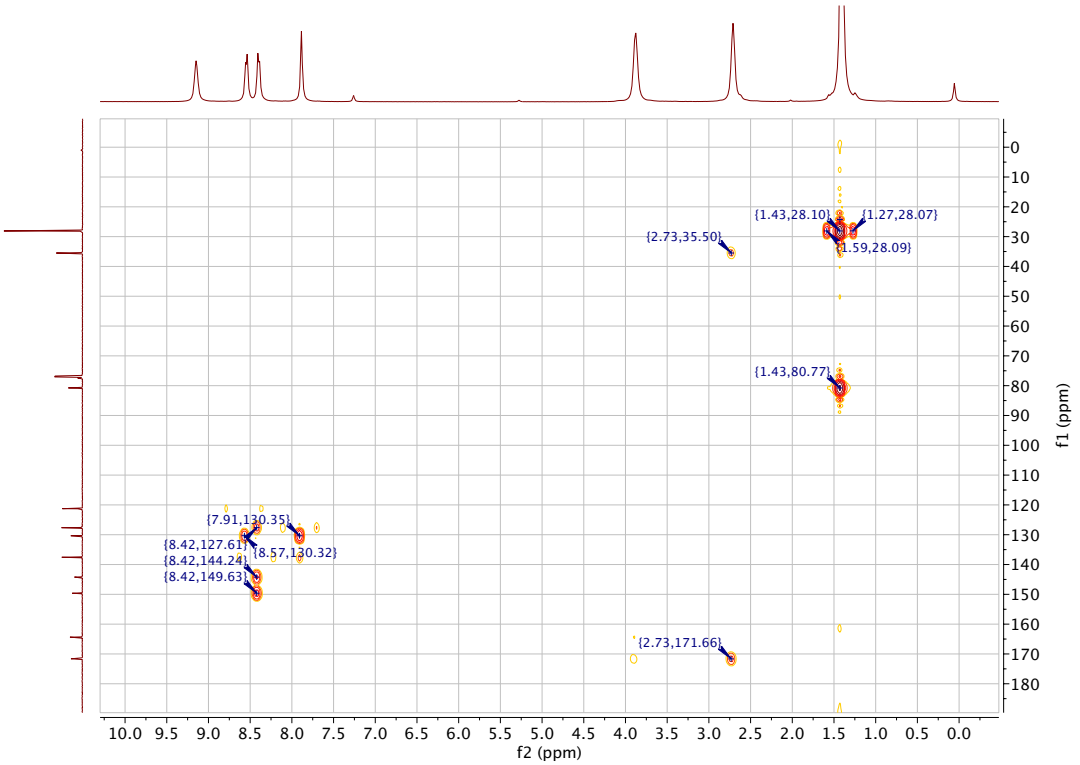


*Figure 8.38: HMBC NMR spectrum of **PhenDA-3-^{t}Bu**.*

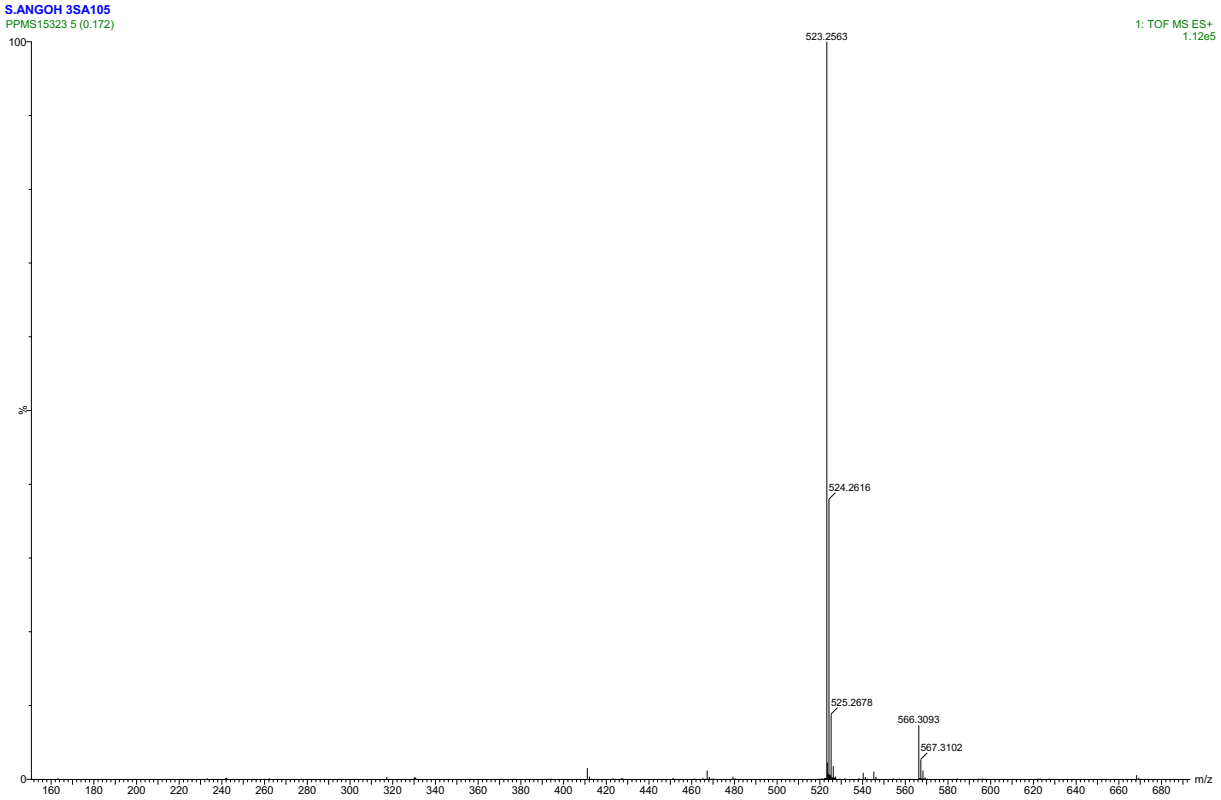


*Figure 8.39: HR-MS (ES$^+$ ToF) of **PhenDA-3-^{t}Bu**.*

3,3'-((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))dipropionic acid trifluoroacetate (PhenDA-3)

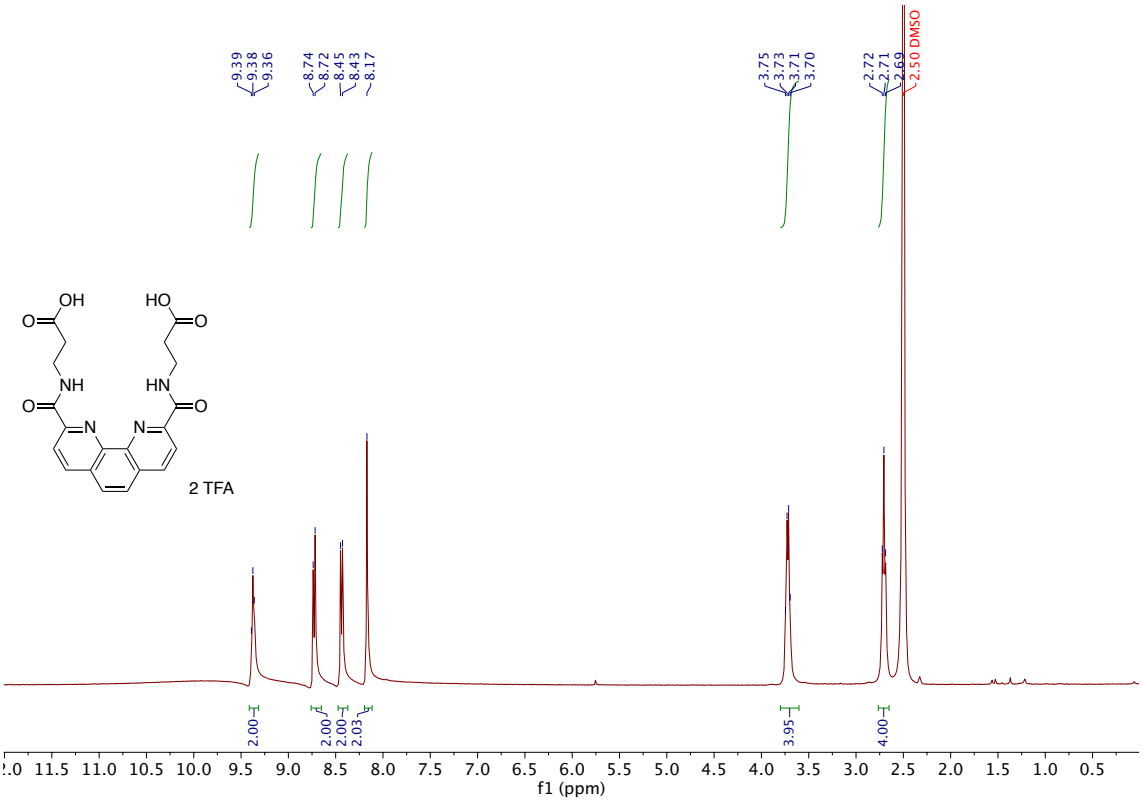


*Figure 8.40: ^{1}H NMR spectrum (400 MHz, DMSO-d_6) of **PhenDA-3**.*

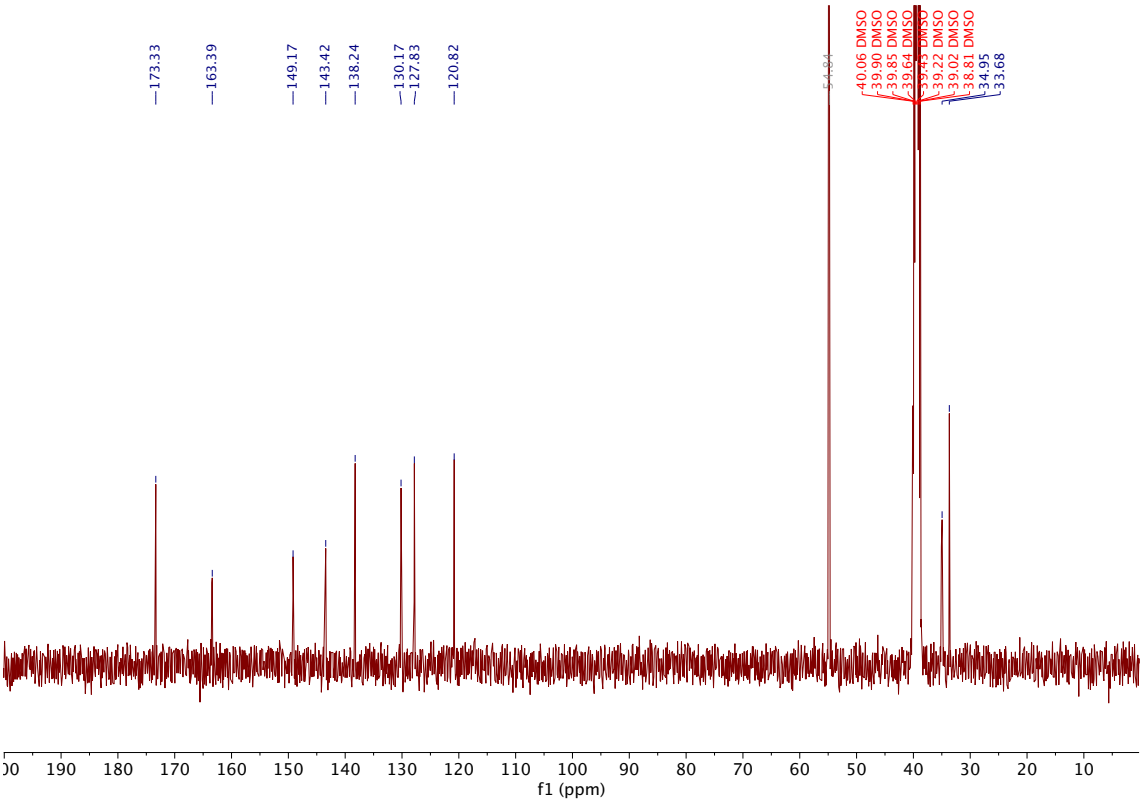


*Figure 8.41: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, DMSO-d_6) of **PhenDA-3**.*

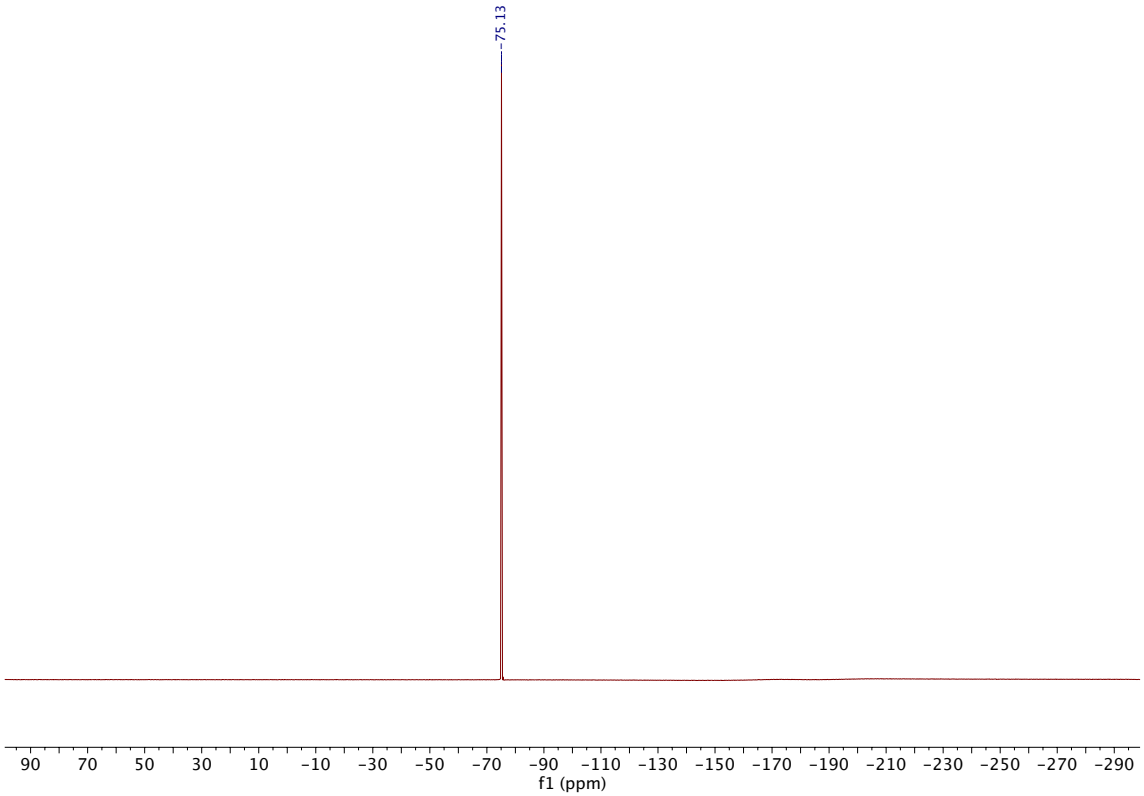


*Figure 8.42: $^{19}F\{^{1}H\}$ NMR spectrum (471 MHz, DMSO-d_6) of **PhenDA-3**.*

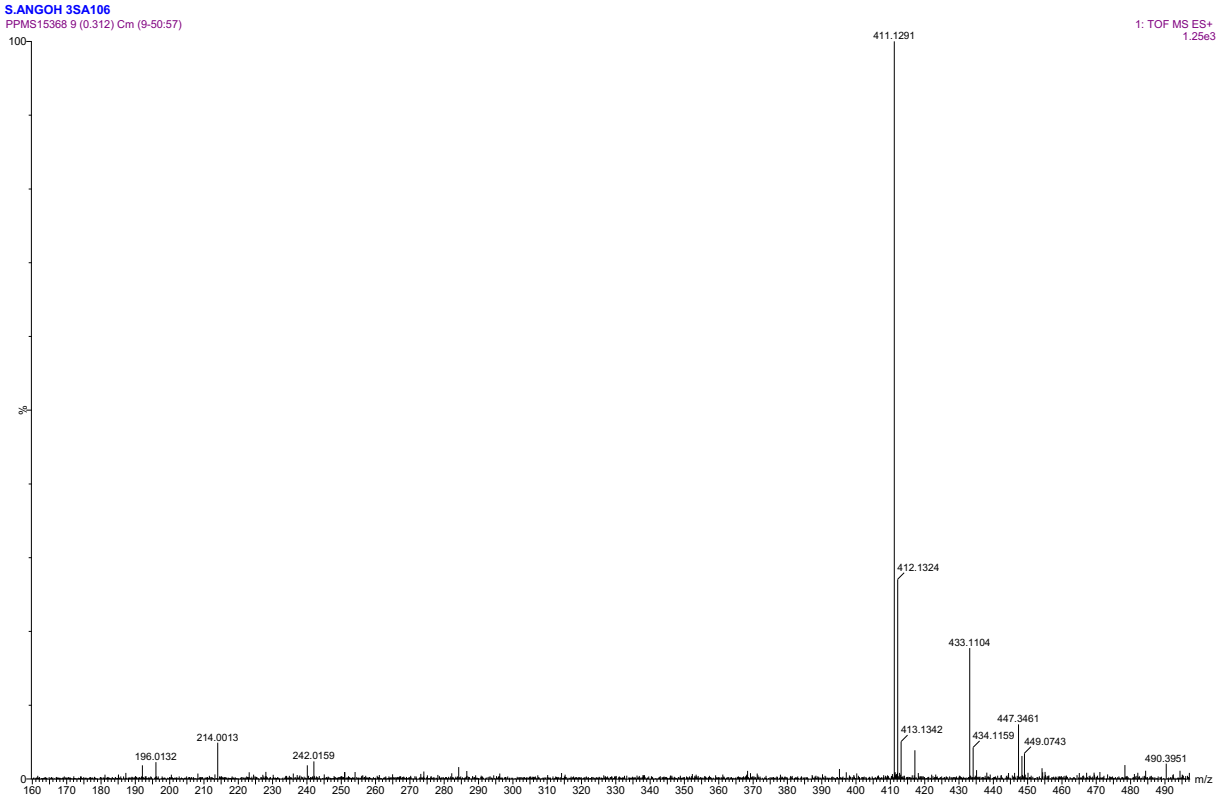


*Figure 8.43: HR-MS (ES$^+$ ToF) of **PhenDA-3**.*

MnPhenDA-3

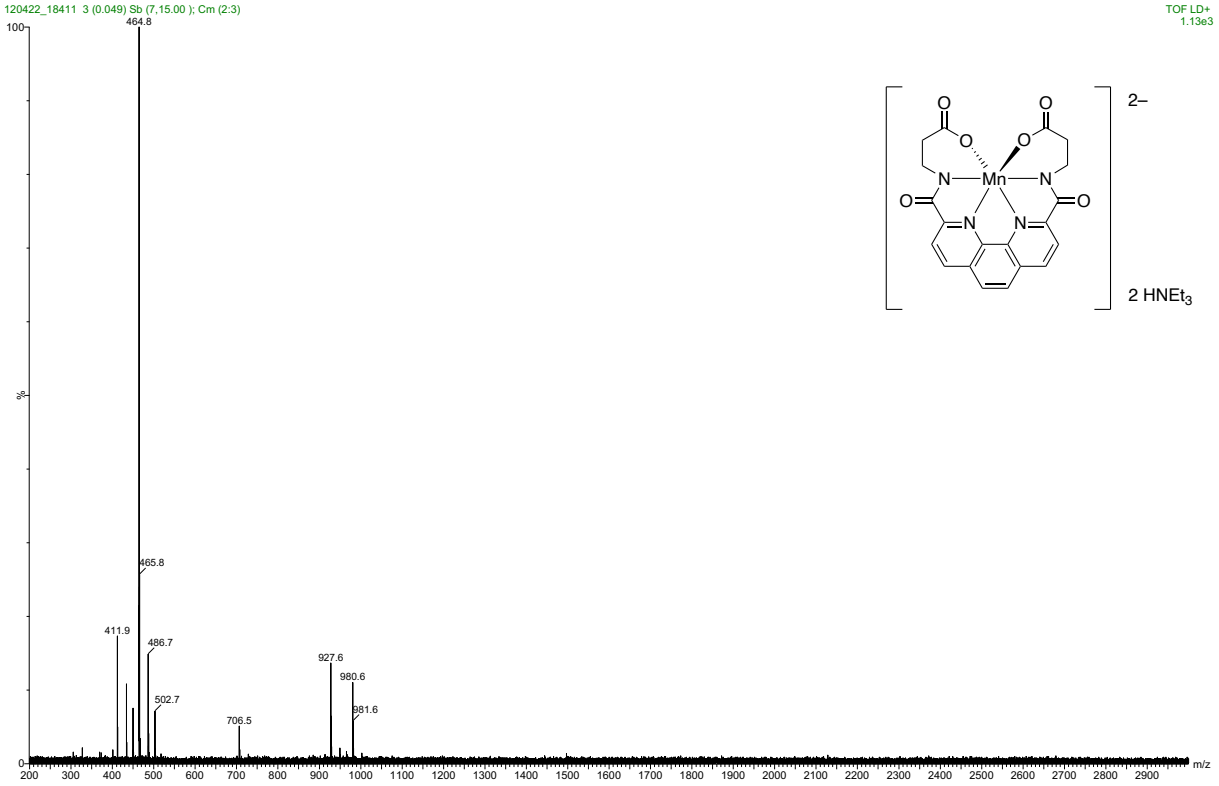


*Figure 8.44: HR-MS (ES⁺ ToF) of **MnPhenDA-3**.*

Di-*tert*-butyl 2,2'-(((1*S*,2*S*)-2-(picolinamido)cyclohexyl)azanediyl)diacetate (PicAm-1-^{t}Bu)

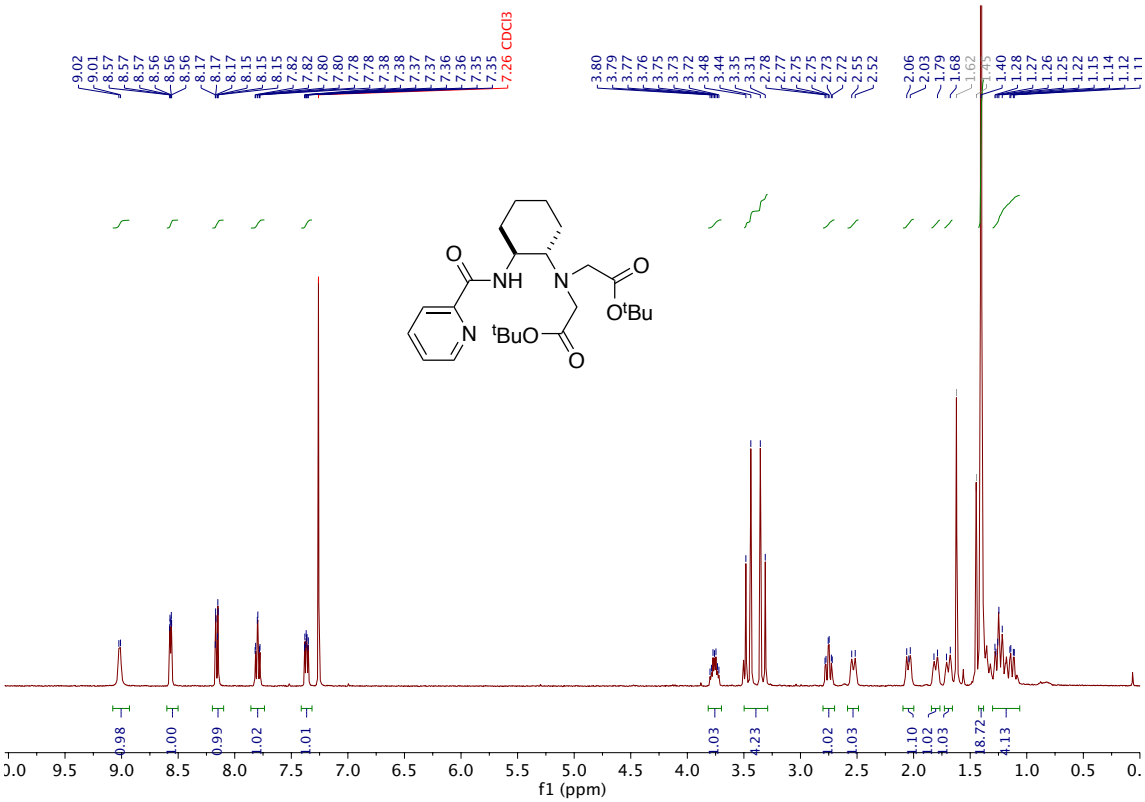


*Figure 8.45: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of **PicAm-1-^{t}Bu**.*

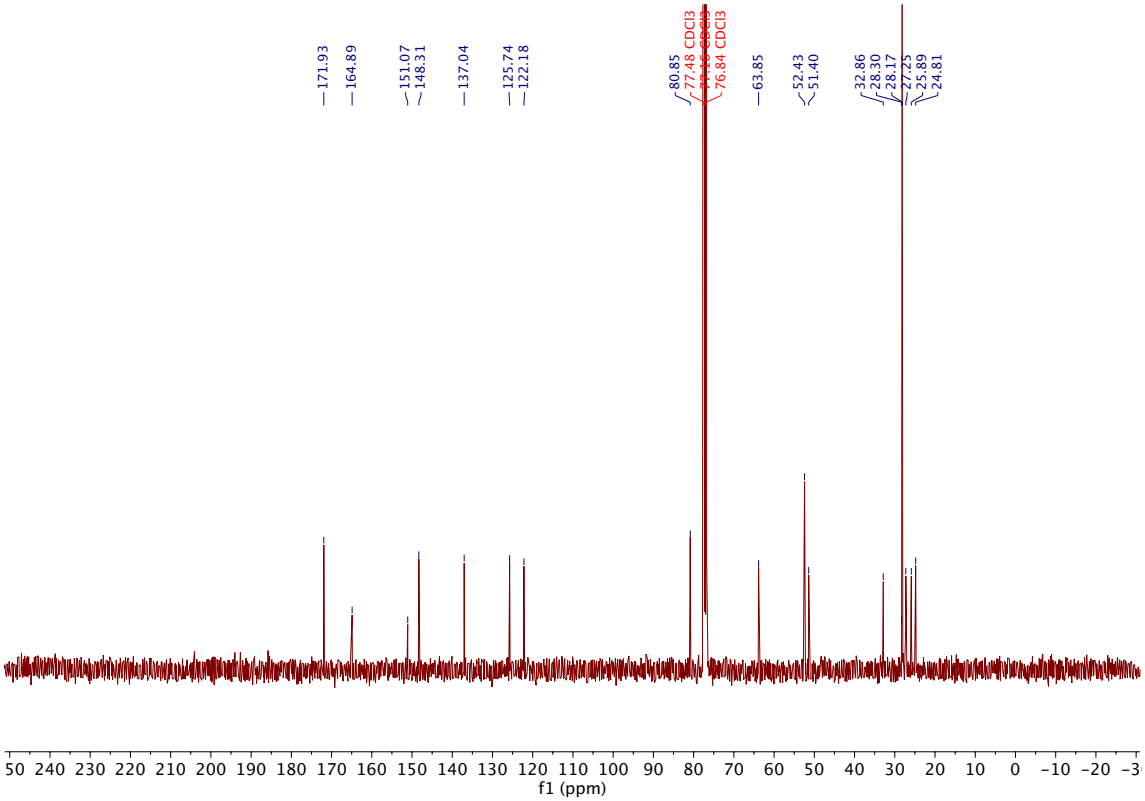


*Figure 8.46: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, $CDCl_3$) of **PicAm-1-^{t}Bu**.*

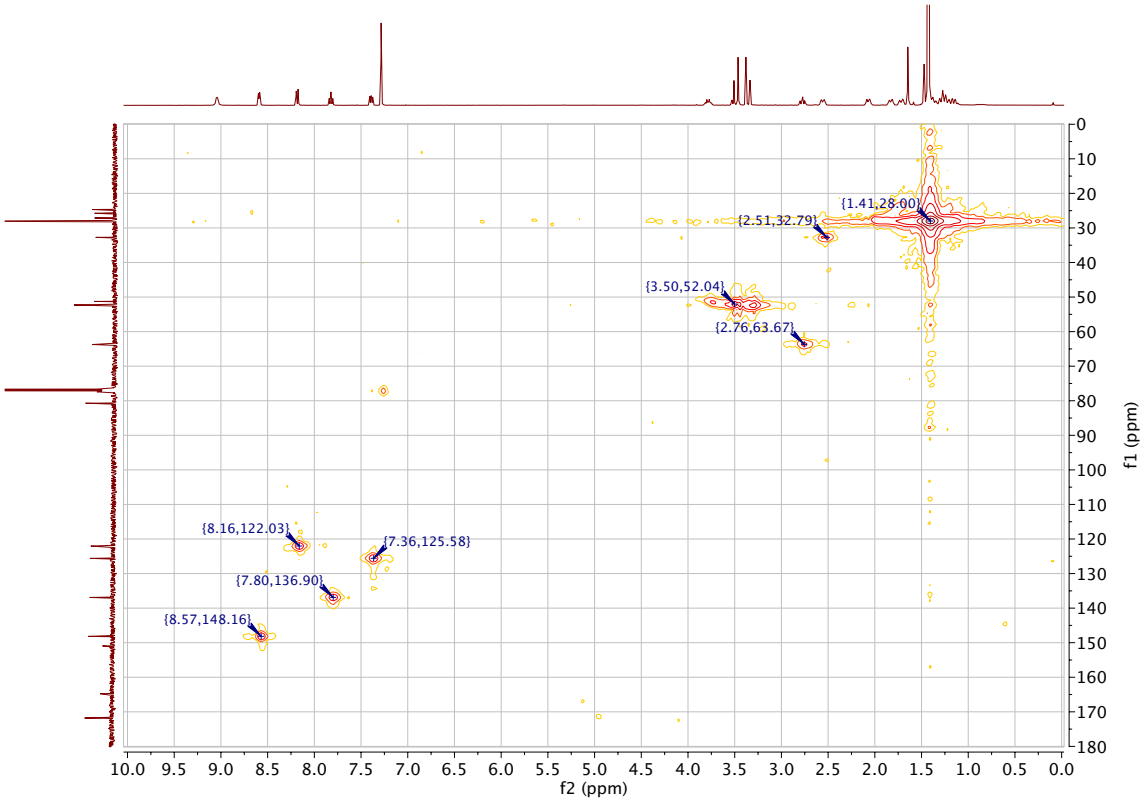


*Figure 8.47: HMQC NMR spectrum of **PicAm-1-^{t}Bu**.*

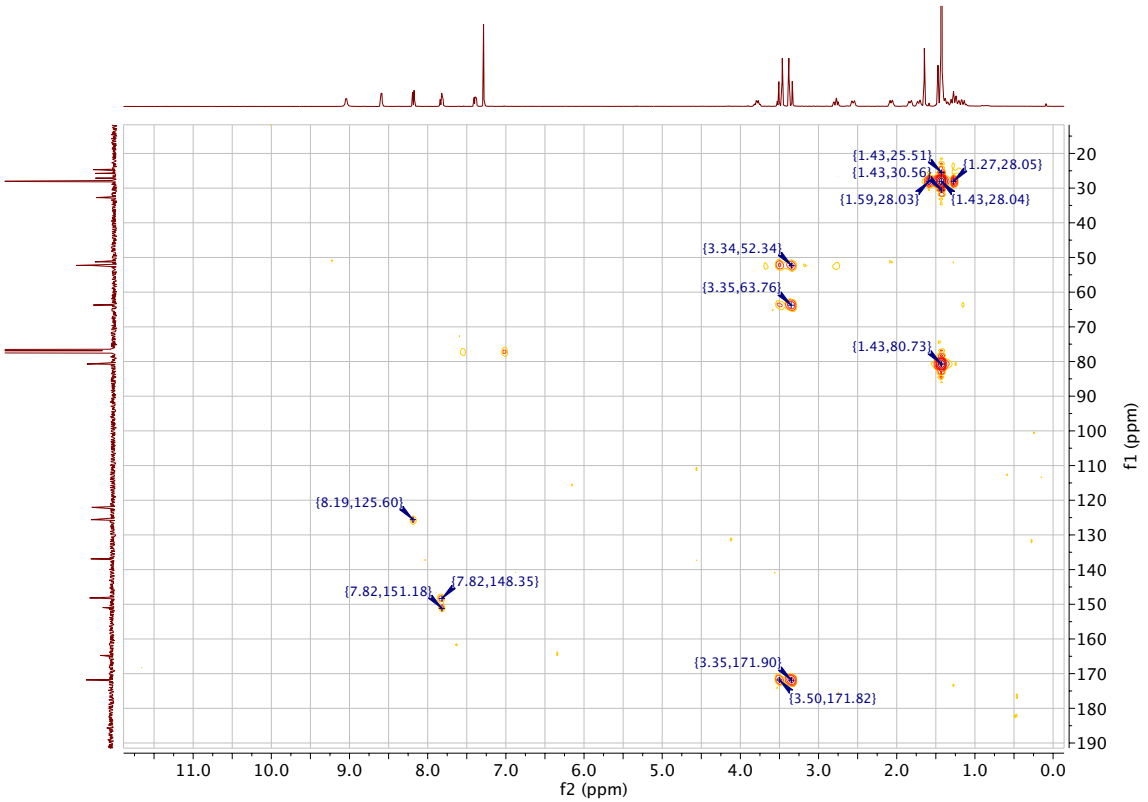


*Figure 8.48: HMBC NMR spectrum of **PicAm-1-^{t}Bu**.*

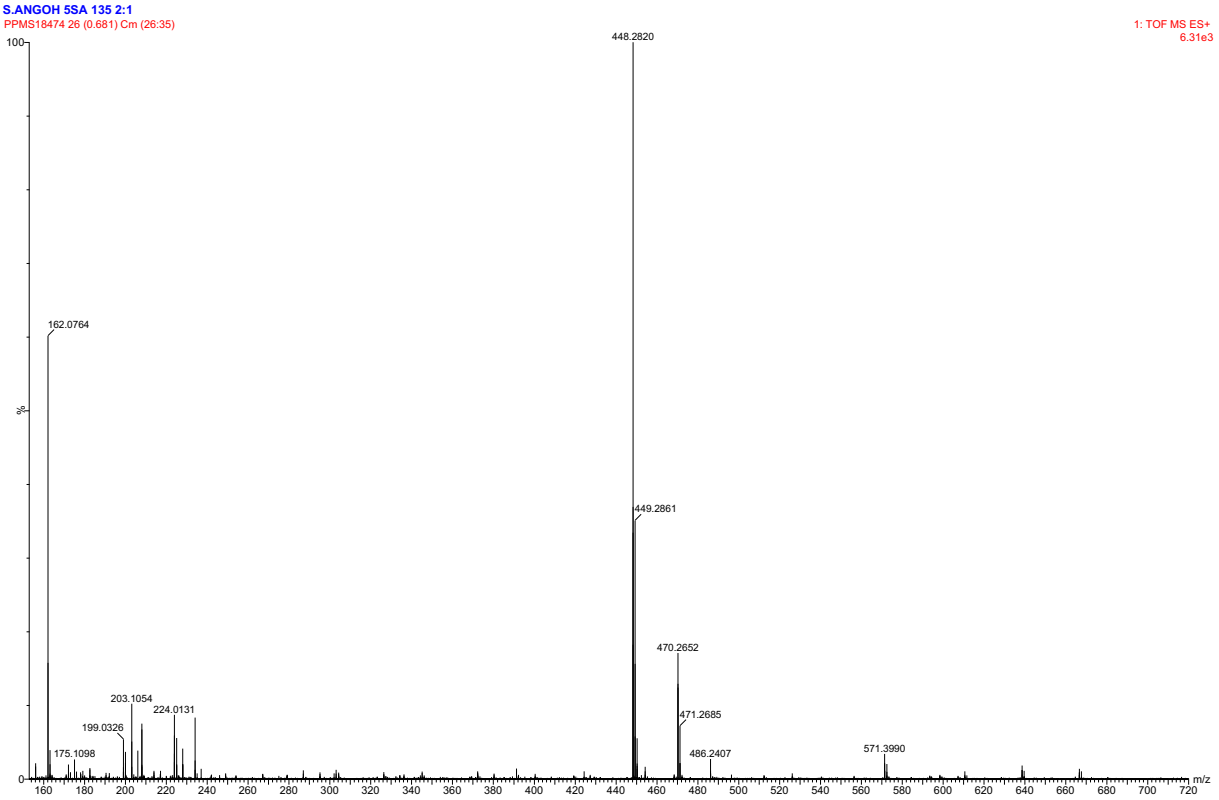


*Figure 8.49: HR-MS (ES$^+$ ToF) of **PicAm-1-^{t}Bu**.*

2,2'-(((±)-trans-2-(picolinamido)cyclohexyl)azanediyl)diacetic acid trifluoroacetate (PicAm-1)

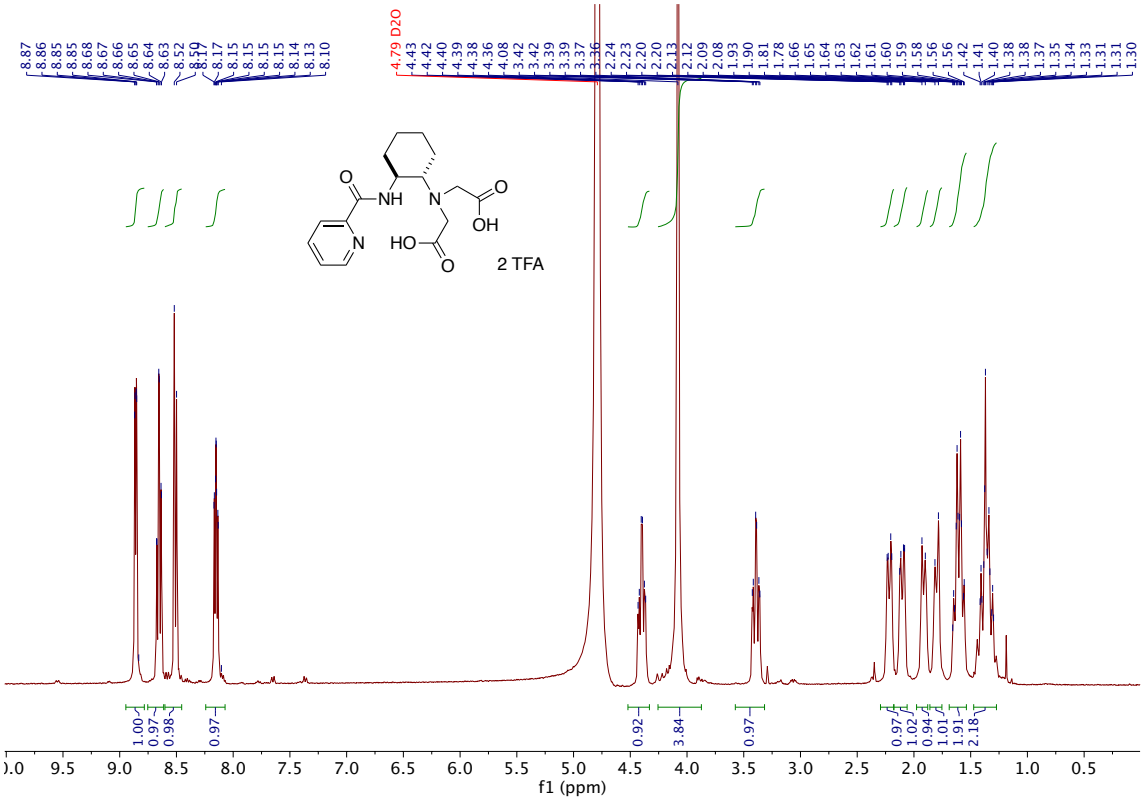


*Figure 8.50: ^{1}H NMR spectrum (400 MHz, D_2O) of **PicAm-1**.*

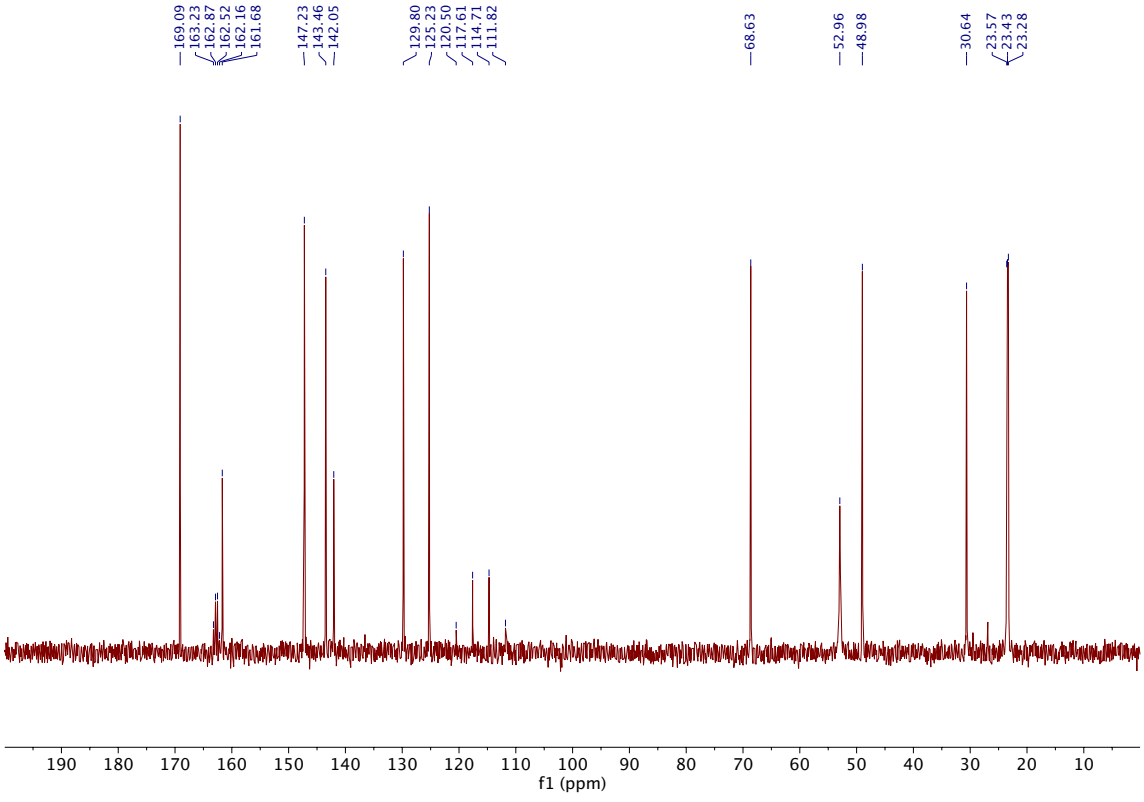


*Figure 8.51: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, D_2O) of **PicAm-1**.*

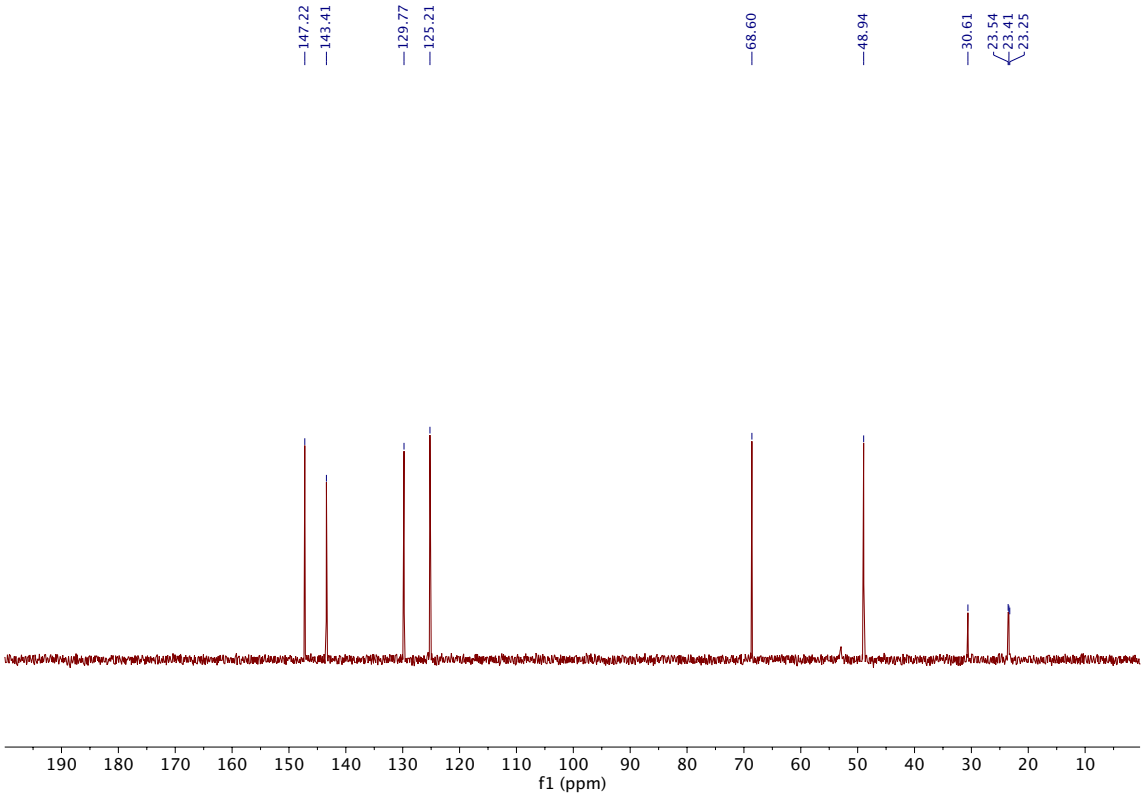


*Figure 8.52: DEPT-90 NMR spectrum of **PicAm-1**.*

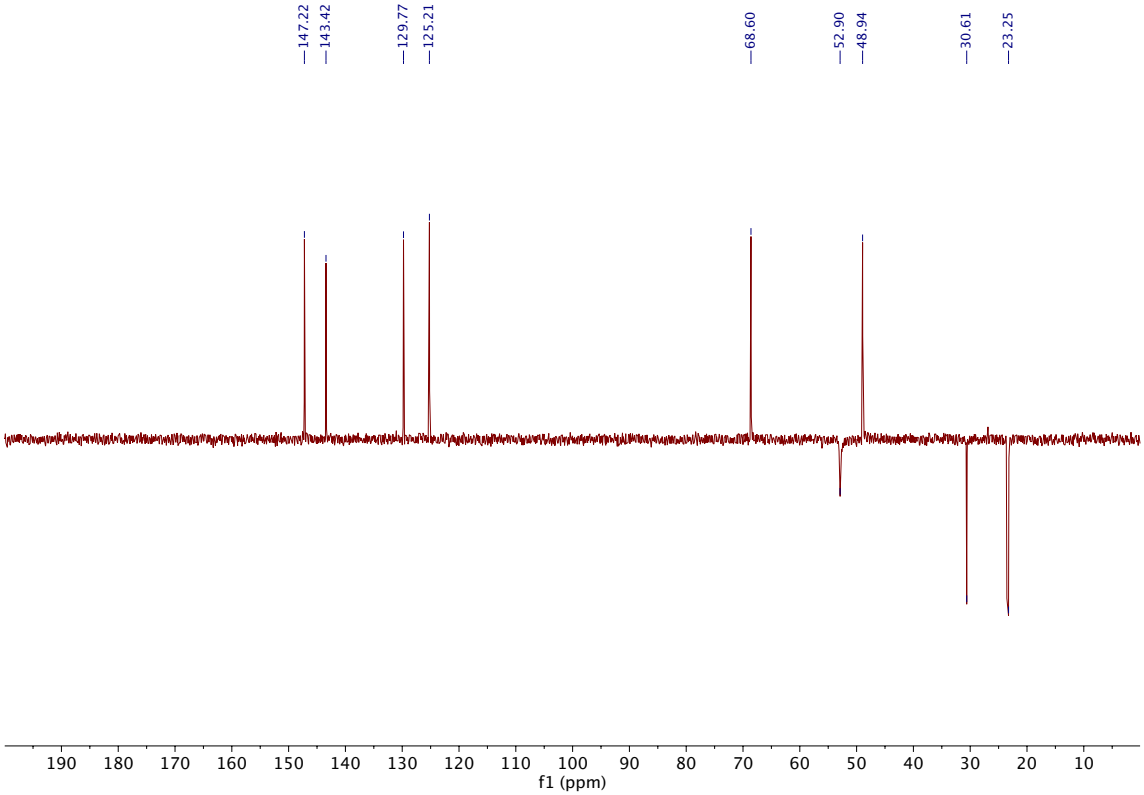


*Figure 8.53: DEPT-135 NMR spectrum of **PicAm-1**.*

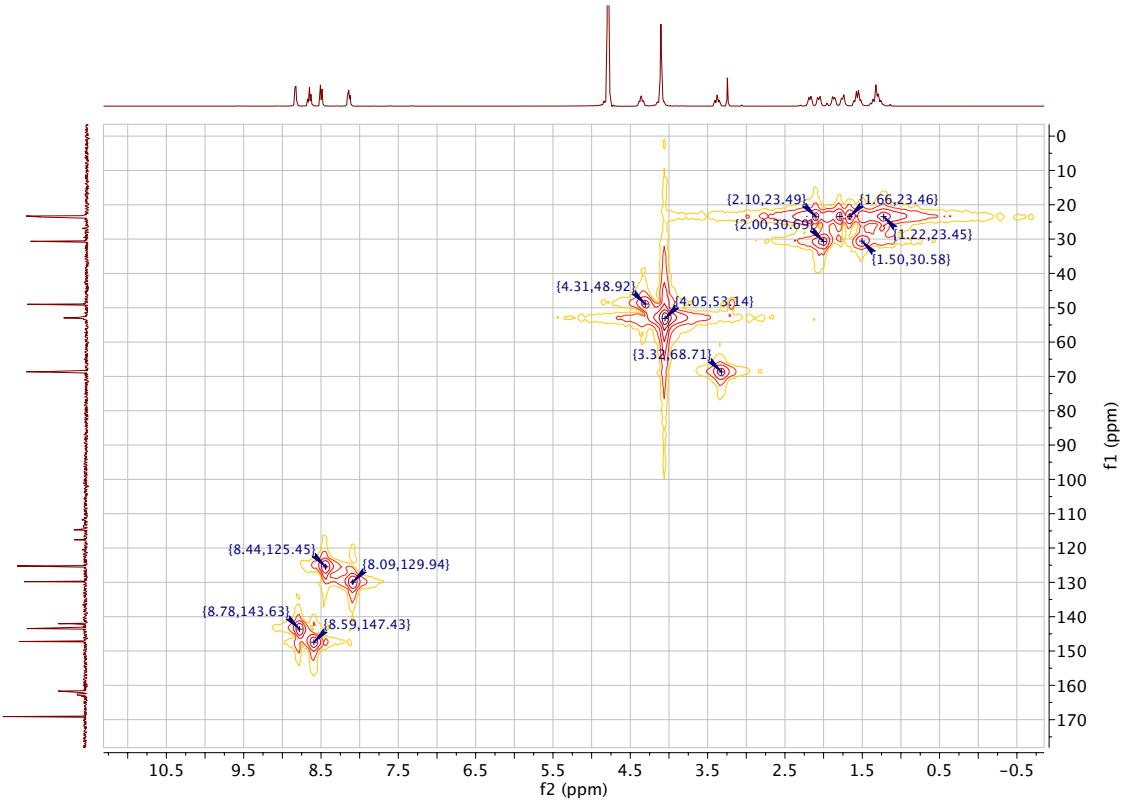


*Figure 8.54: HMQC NMR spectrum of **PicAm-1**.*

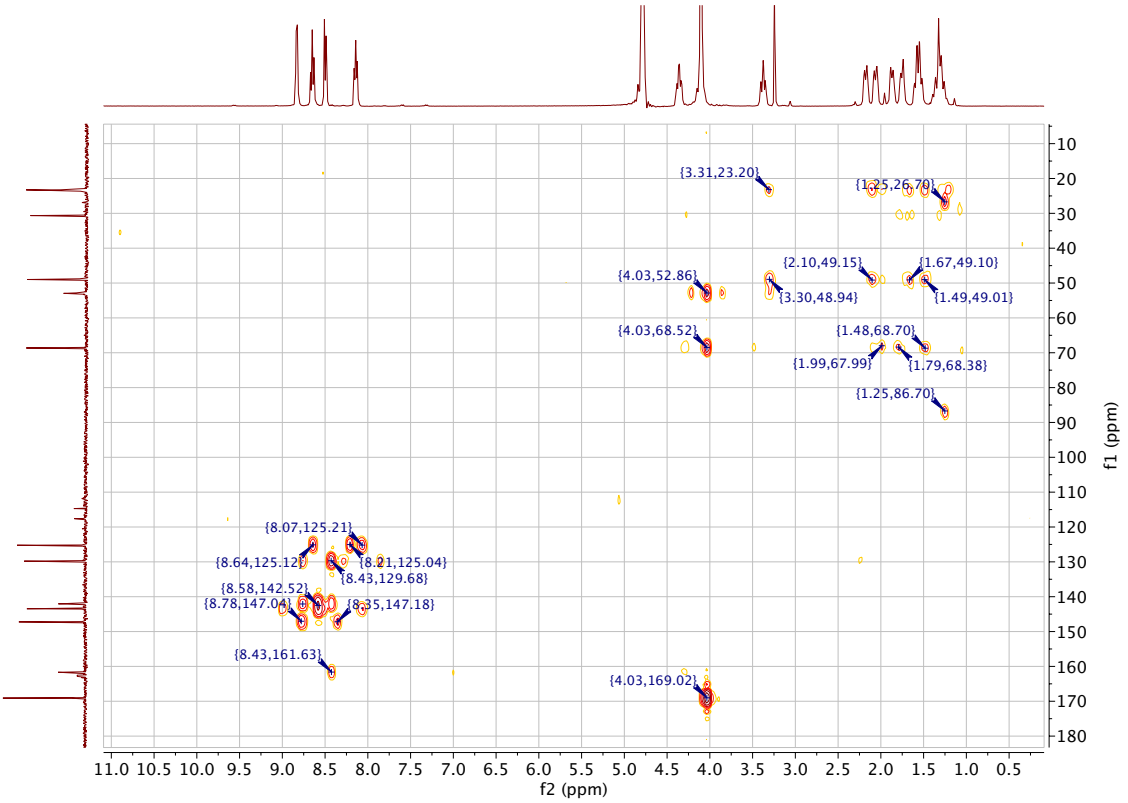


*Figure 8.55: HMBC NMR spectrum of **PicAm-1**.*

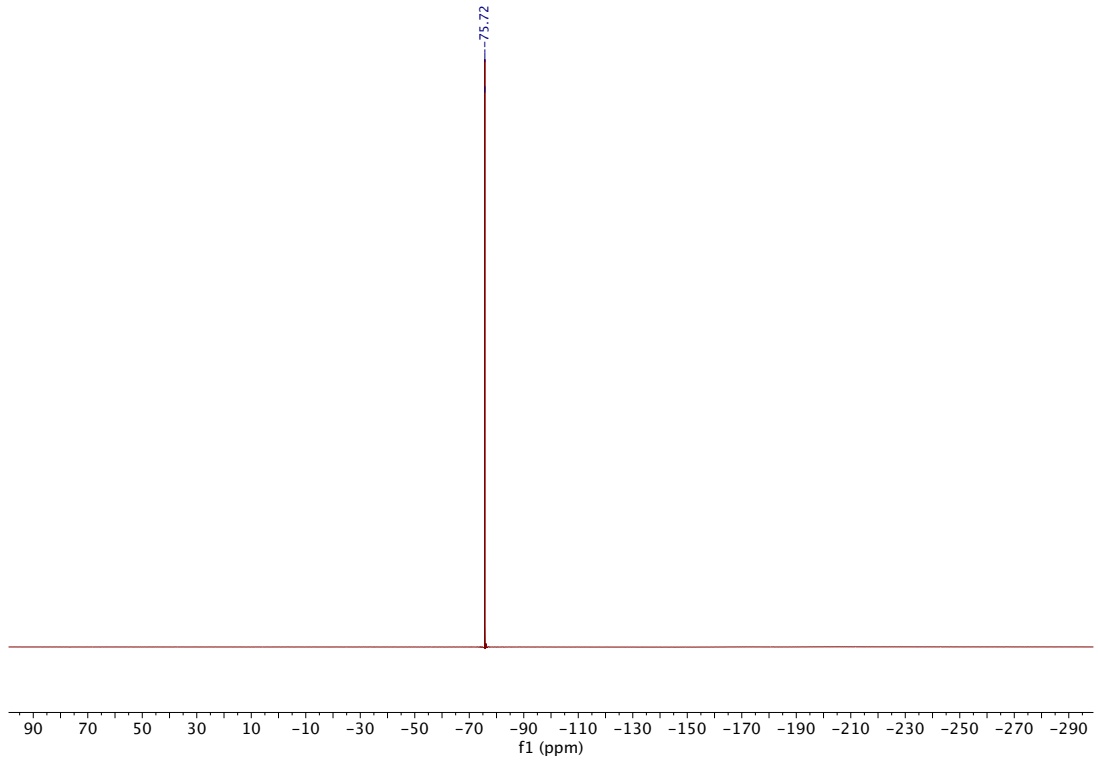


*Figure 8.56: $^{19}F\{^{1}H\}$ NMR spectrum (471 MHz, D_2O) of **PicAm-1**.*

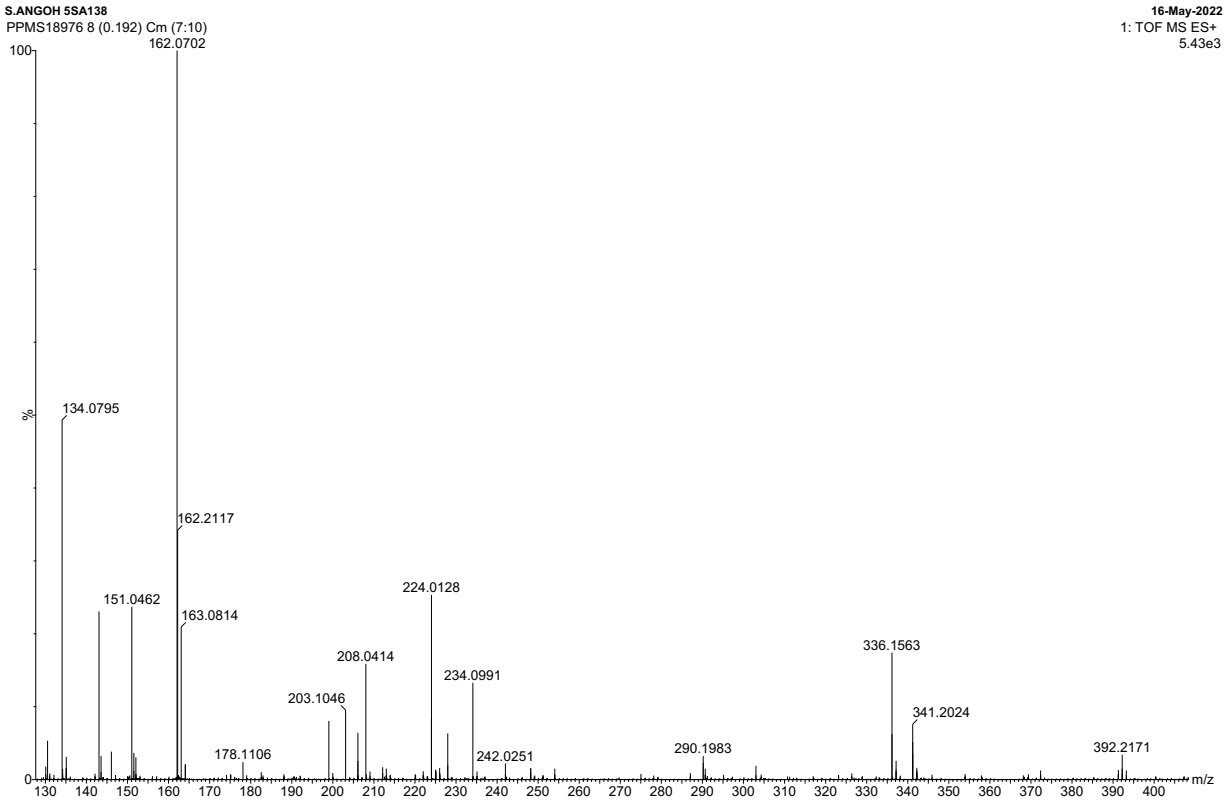


*Figure 8.57: HR-MS (ES^+ ToF) of **PicAm-1**.*

Di-*tert*-butyl 2,2'-((2-oxo-2-(((1*S*,2*S*)-2-(picolinamido)cyclohexyl)amino)ethyl)-azanediyl)di-acetate (PicAm-2-[t]Bu)

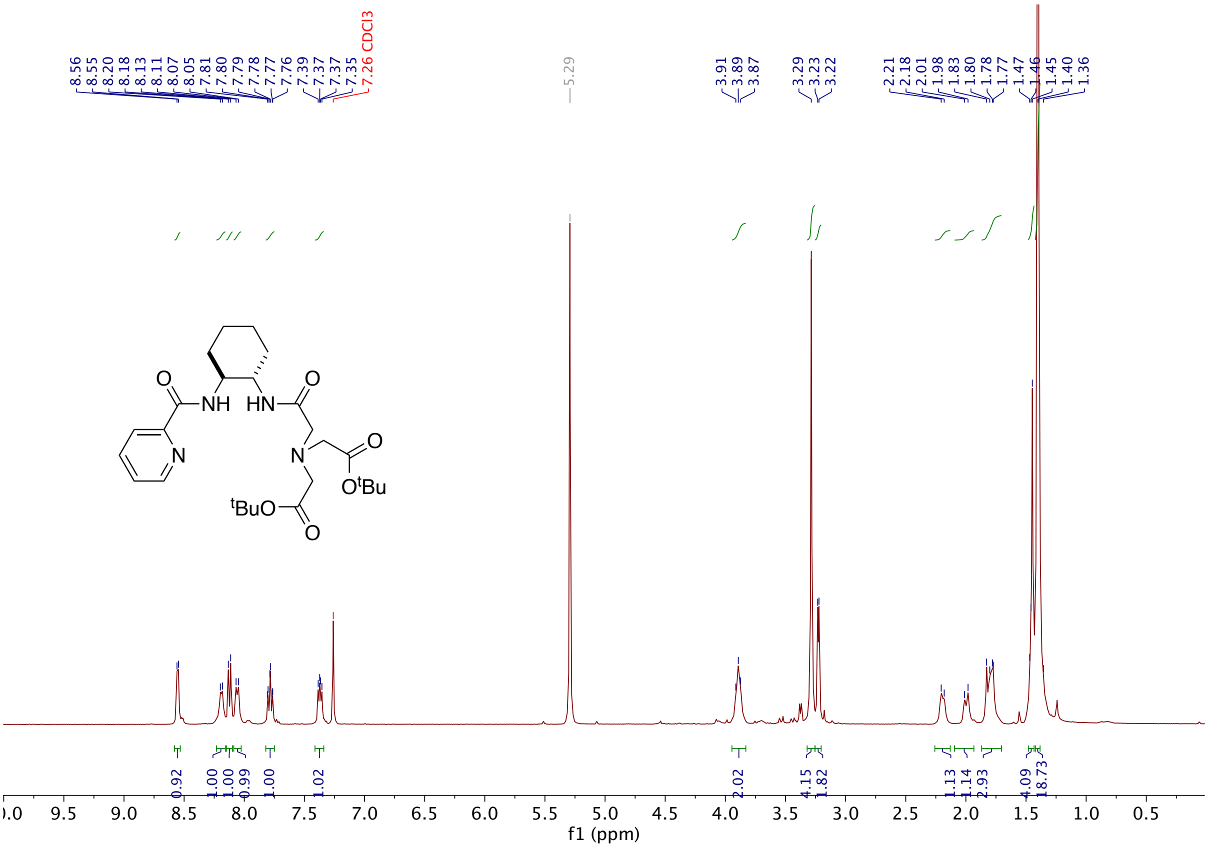


*Figure 8.58: [1]H NMR spectrum (400 MHz, CDCl₃) of **PicAm-2-[t]Bu**.*

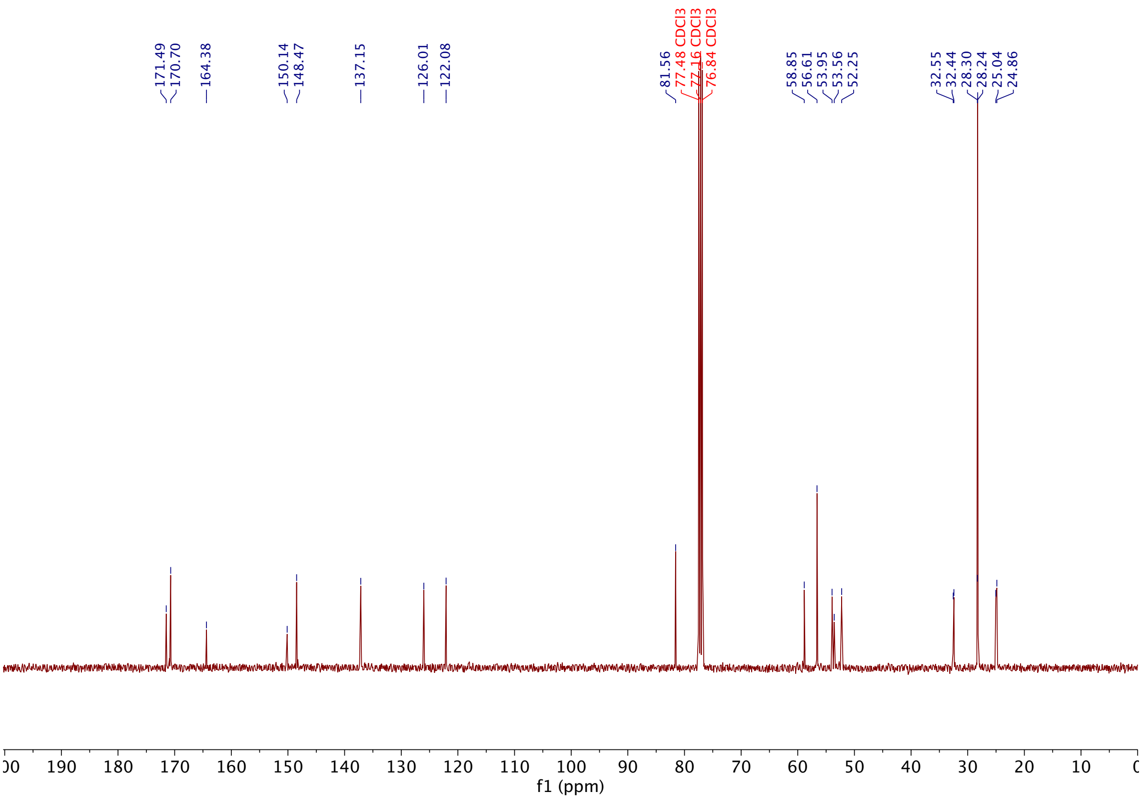


*Figure 8.59: [13]C{[1]H} NMR spectrum (101 MHz, CDCl₃) of **PicAm-2-[t]Bu**.*

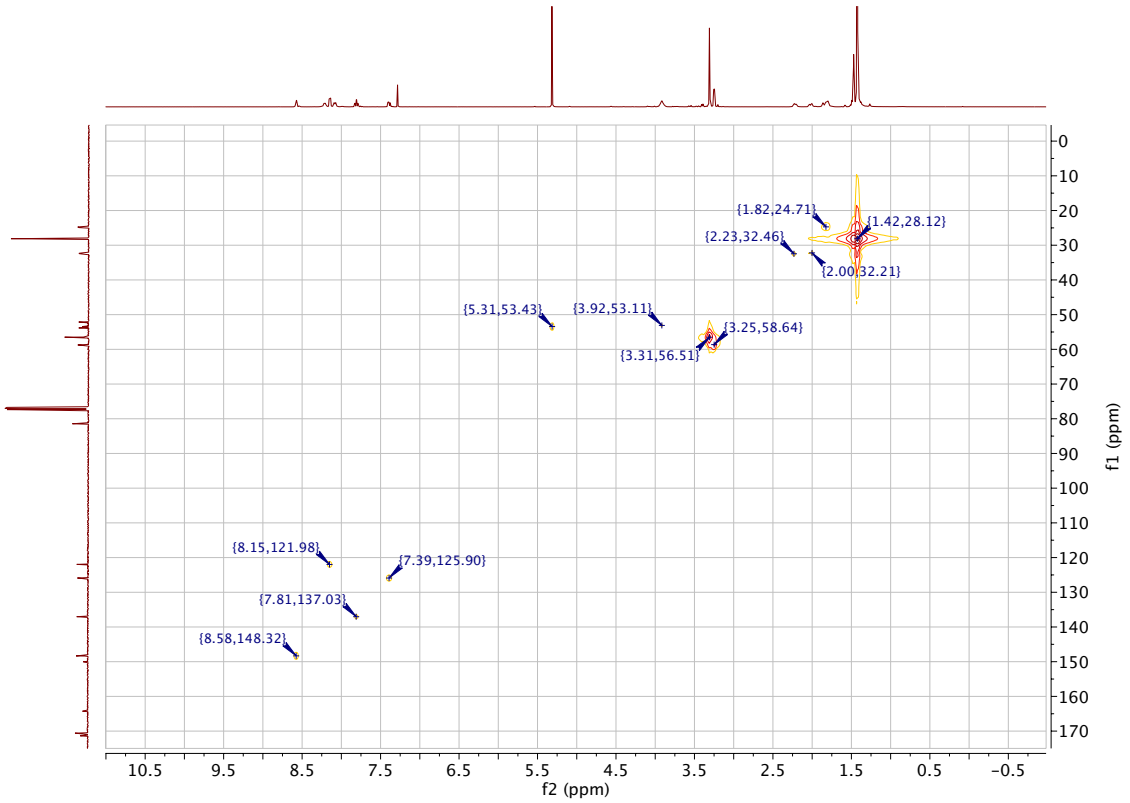


*Figure 8.60: HMQC NMR spectrum of **PicAm-2-^{t}Bu**.*

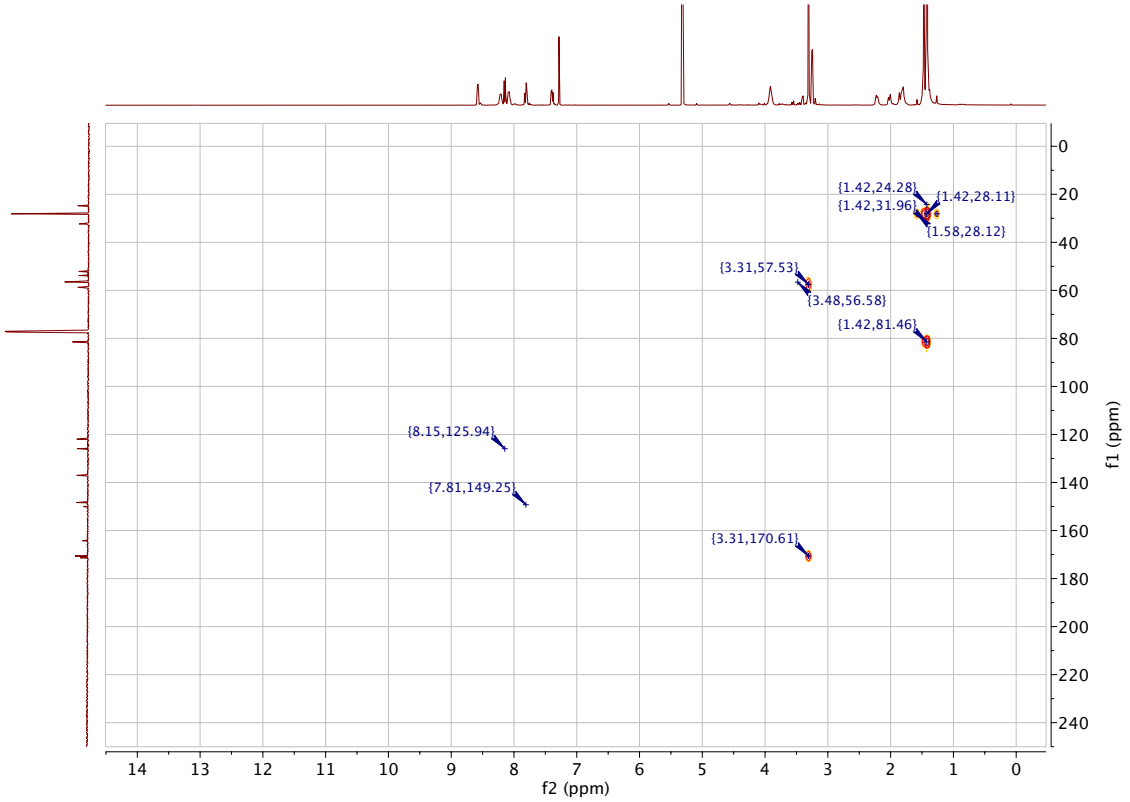


*Figure 8.61: HMBC NMR spectrum of **PicAm-2-^{t}Bu**.*

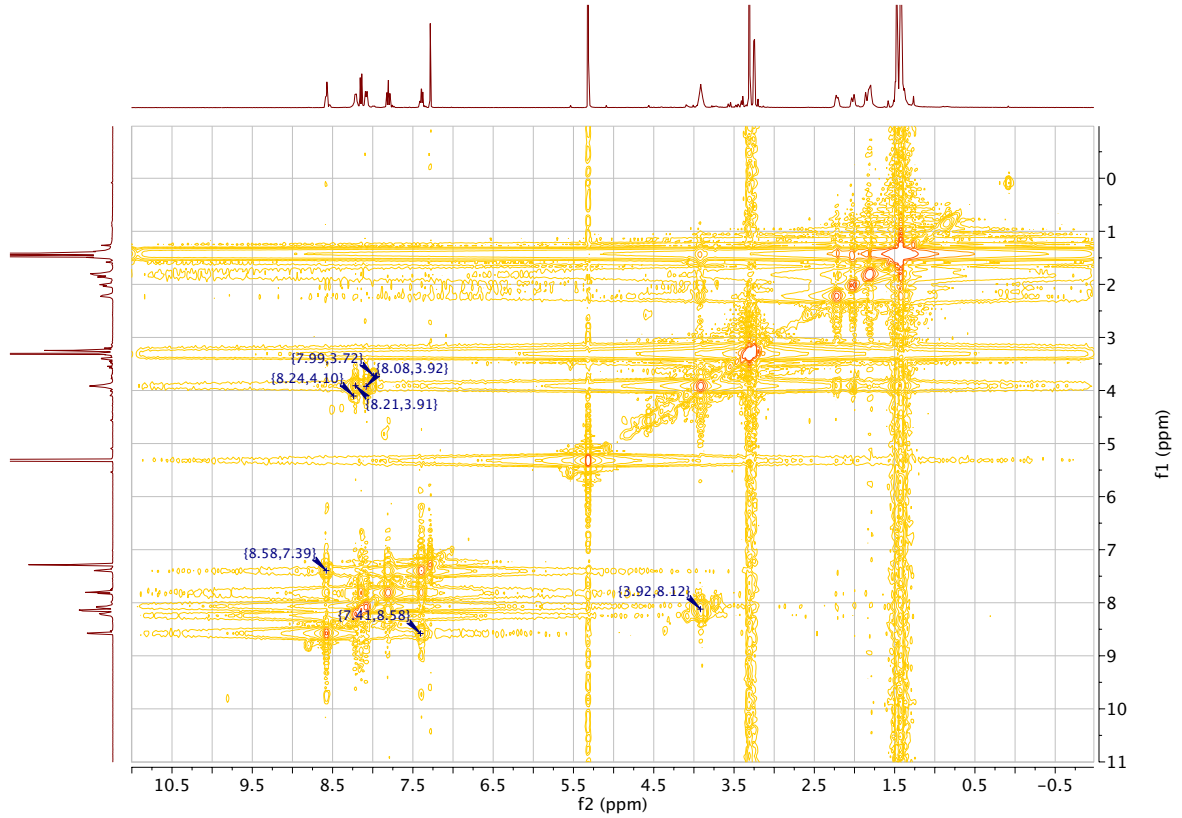


*Figure 8.62: COSY NMR spectrum (400 MHz, $CDCl_3$) of **PicAm-2-^{t}Bu**.*

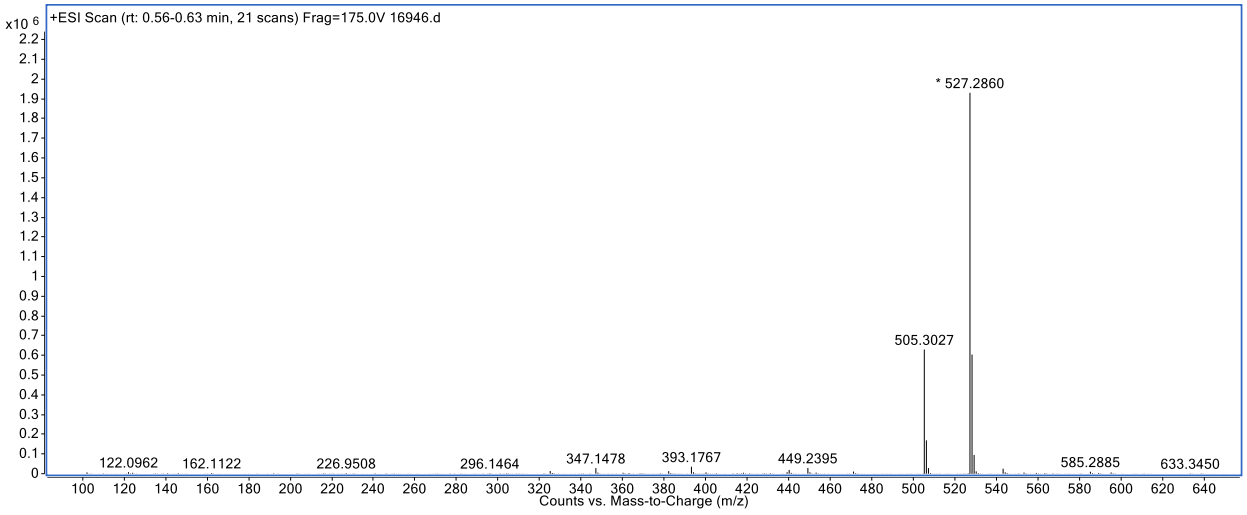


*Figure 8.63: HR-MS (ES^+ ToF) of **PicAm-2-^{t}Bu**.*

2,2'-((2-oxo-2-(((±)-trans-2-(picolinamido)cyclohexyl)amino)ethyl)azanediyl)diacetic acid trifluoroacetate (PicAm-2)

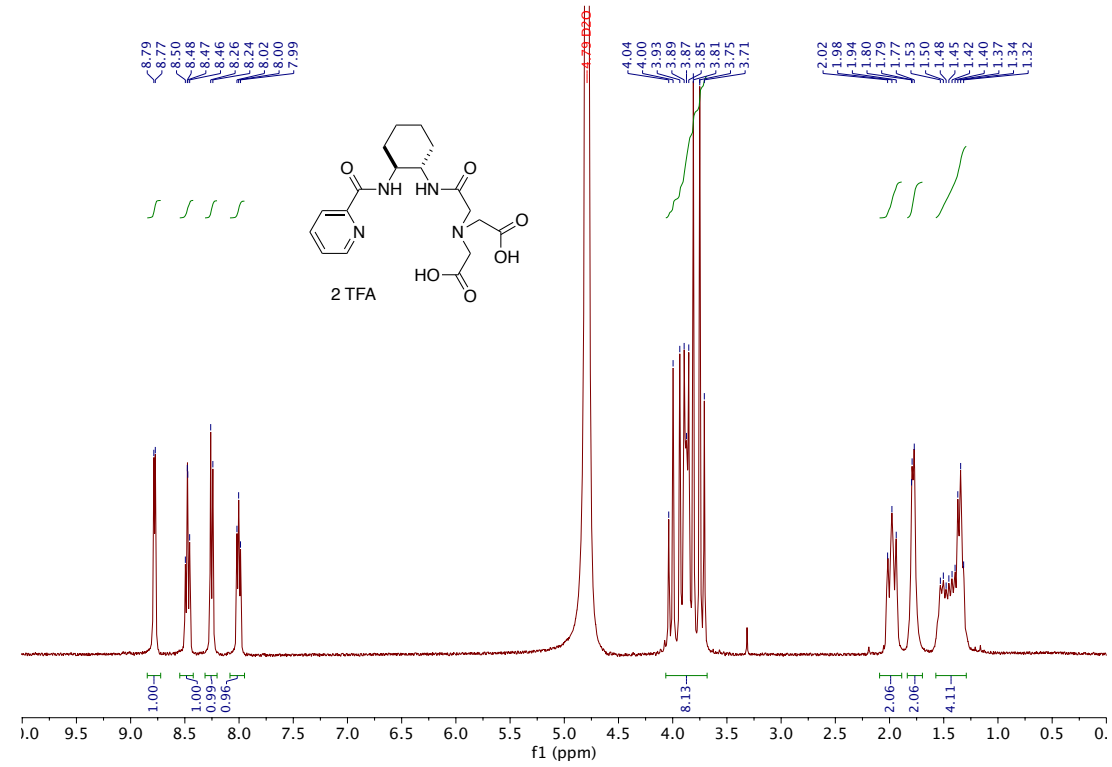


*Figure 8.64: ^{1}H NMR spectrum (400 MHz, D_2O) of **PicAm-2**.*

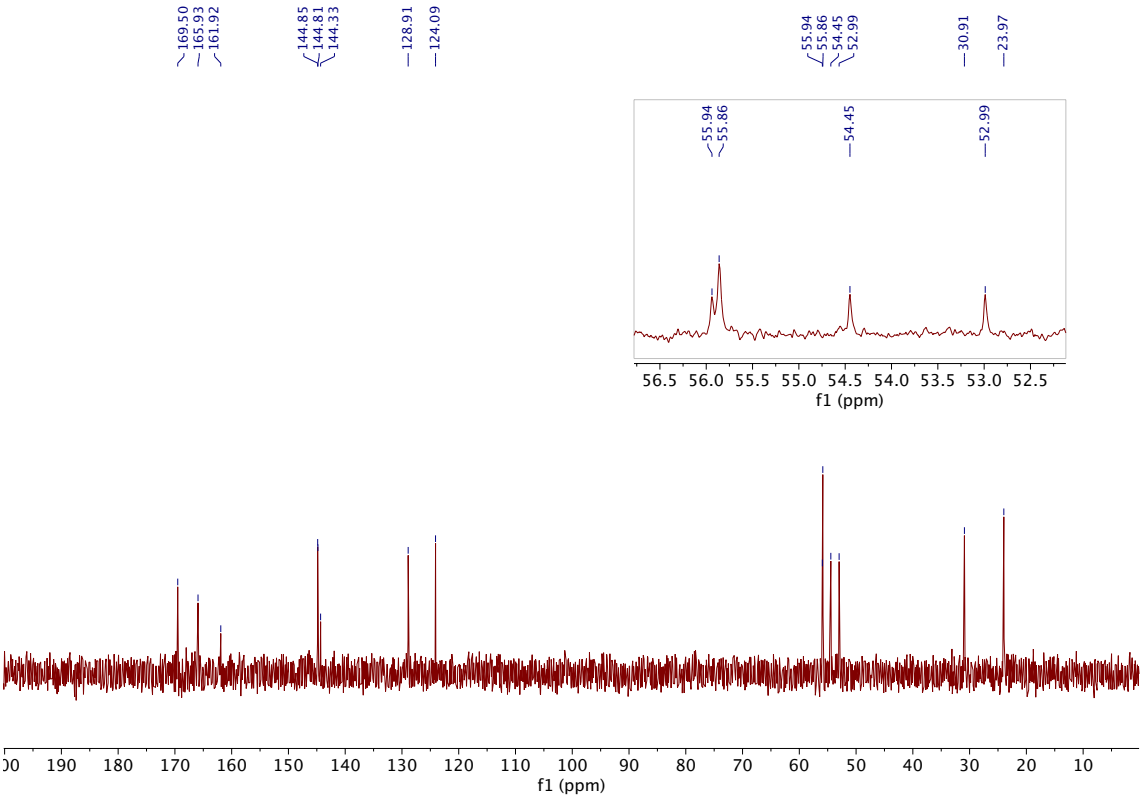


*Figure 8.65: $^{13}C\{^{1}H\}$ NMR spectrum (101 MHz, D_2O) of **PicAm-2**.*

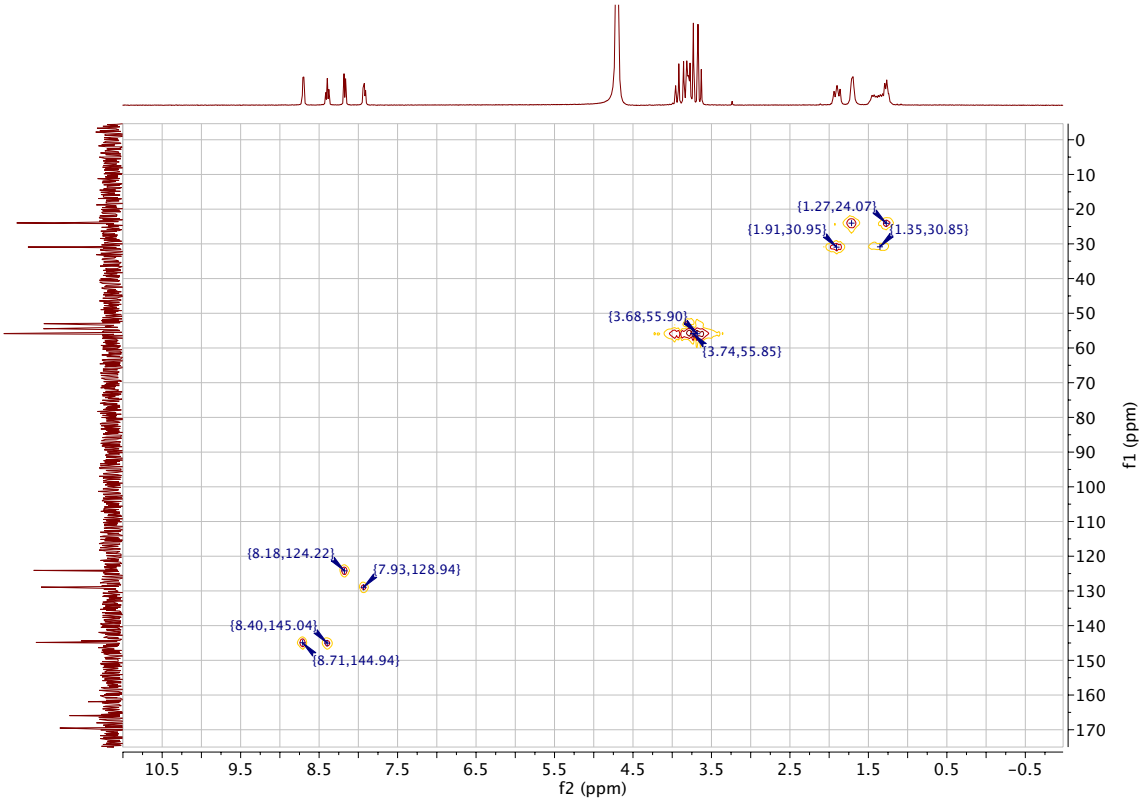


*Figure 8.66: HMQC NMR spectrum of **PicAm-2**.*

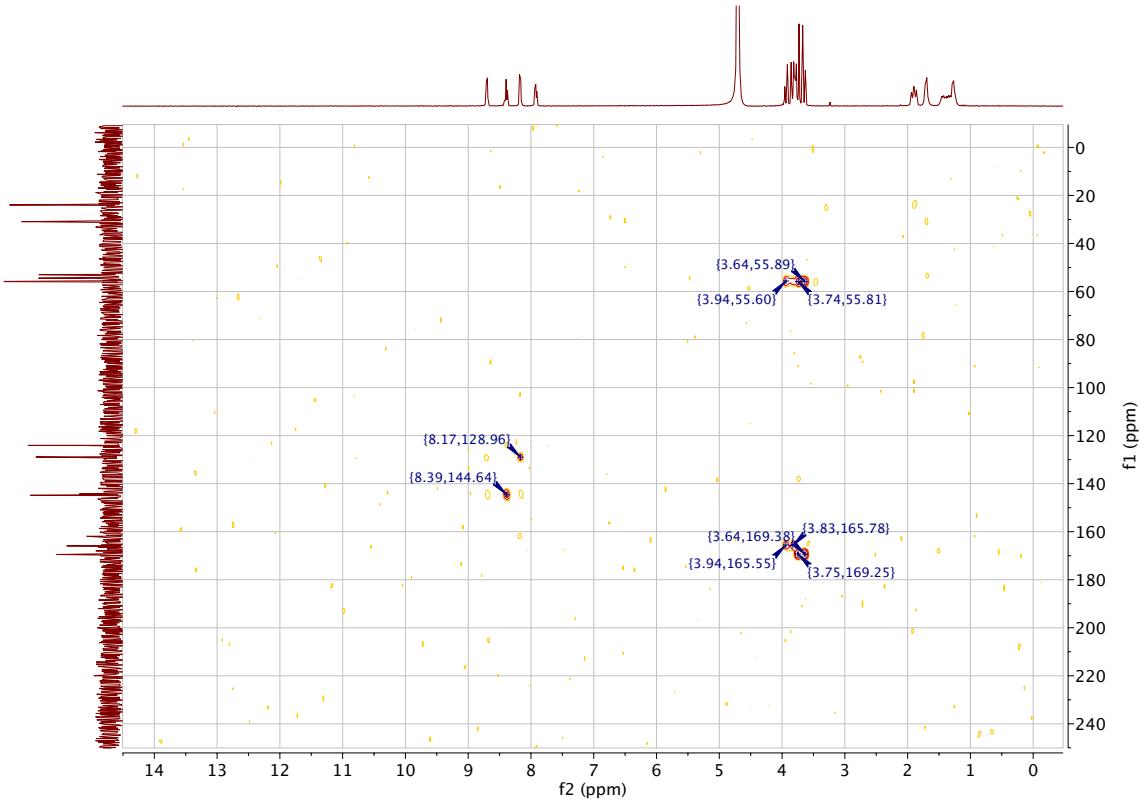


*Figure 8.67: HMBC NMR spectrum of **PicAm-2**.*

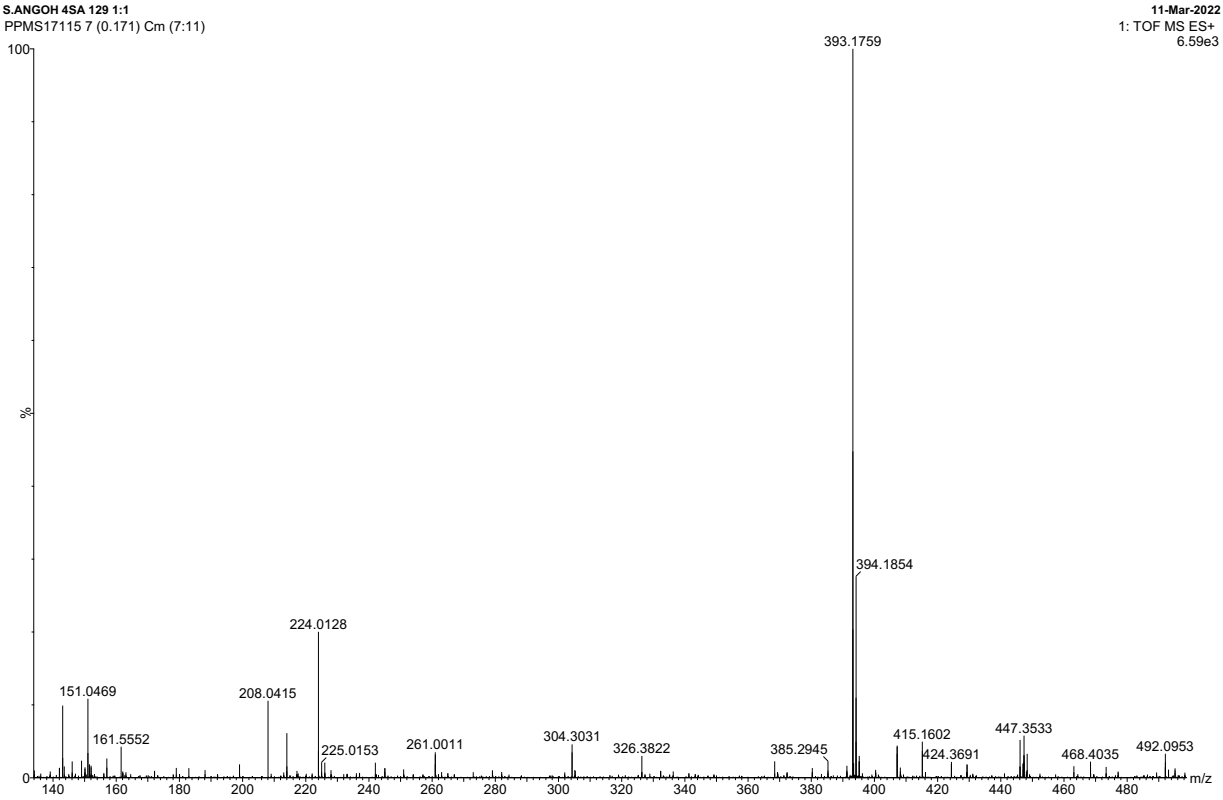


*Figure 8.68: HR-MS (ES$^+$ ToF) of **PicAm-2**.*

MnPicAm-2

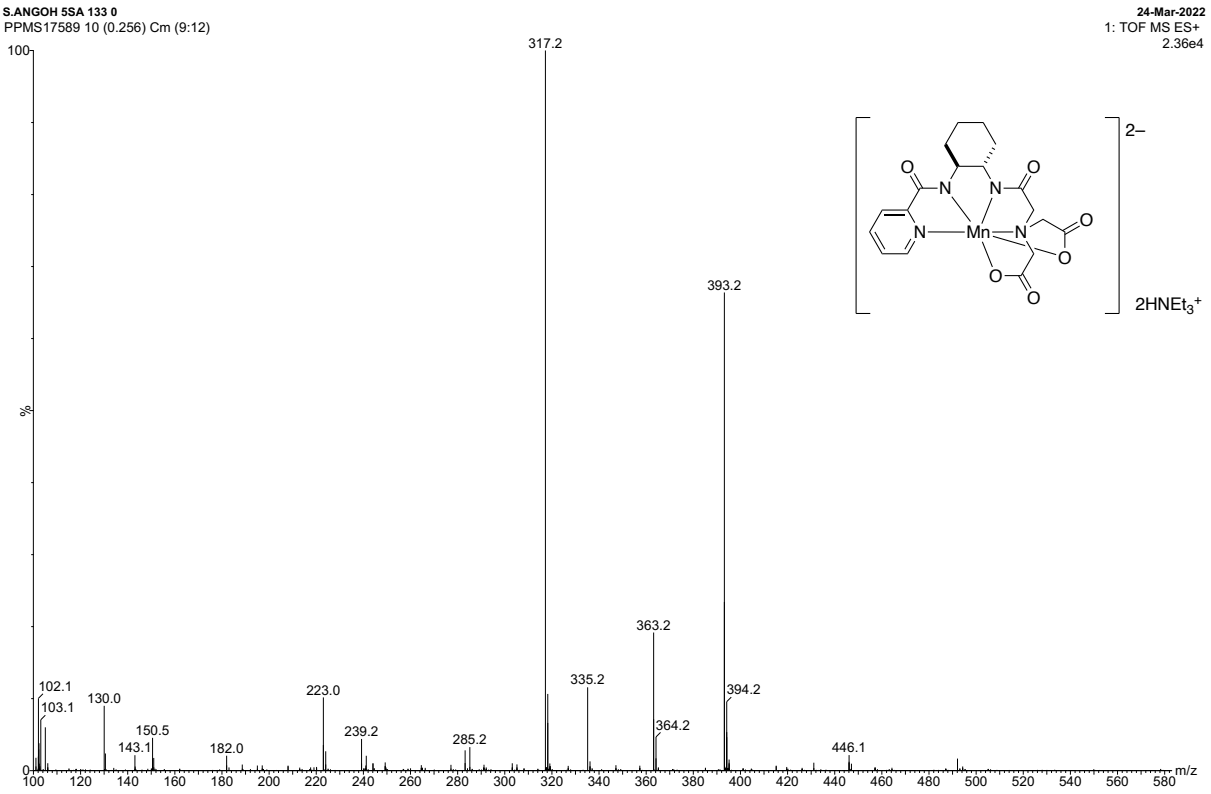


*Figure 8.69: HR-MS (ES$^+$ ToF) of **MnPicAm-2**.*

8.2 Experimental data for crystallography data

*Table 8.1: Summary of crystallographic refinement data for **PS1-salt**.*

Compound	**PS1-salt**
Chemical formula	$C_{37}H_{36}N_4P_2$, 0.15(CH_2Cl_2), 2(Br)
M (g mol^{-1})	771.19
T (K)	173
Diffractometer, wavelength (Å)	Agilent Xcalibur PX Ultra A, 1.54184
Crystal system	Triclinic
Space group	P-1
a (Å)	9.6790
b (Å)	12.3382
c (Å)	15.8588
a (deg)	79.118
b (deg)	80.214
g (deg)	75.829
V ($Å^3$)	1787.93
Z	2
ρ_{calc} (g mol^{-1})	1.432
μ (mm^{-1})	4.171
F(000)	785
Crystal colour/morphology	Colourless platy needles
Crystal size (mm^3)	0.201 x 0.170 x 0.086
θ range (deg)	3.740 to 73.323
Index ranges	-11<=h<=7, -15<=k<=13, -18<=l<=19
Reflns collected/unique	10356 / 6822 [R(int) = 0.0460]
Reflns observed [F>4σ(F)]	4912
Absorption correction	Analytical
Max. and min. transmission	0.786 and 0.610
Refinement method	Full-matrix least-squares on F^2
Data / restrains / parameters	6822 / 27 / 439
Goodness-of-fit on F^2	1.017
Final R indices [F>4σ(F)]	R1 = 0.0464, wR2 = 0.1107
R indices (all data)	R1 = 0.0718, wR2 = 0.1303
Largest diff. peak, hole	0.907, -0.521 $eÅ^{-3}$
Mean and maximum shift / error	0.000 and 0.002

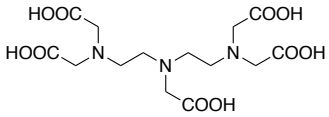
Table 8.2: Full list of bond lengths (Å) and angles (°) for ***PS1-salt.***

N(1)-C(2)	1.472(5)	C(32')-C(37')	1.28(6)
N(1)-P(6)	1.622(3)	C(32')-C(33')	1.33(6)
C(2)-C(3)	1.523(5)	C(33')-C(34')	1.35(5)
C(3)-C(4)	1.530(6)	C(34')-C(35')	1.42(5)
C(4)-N(5)	1.459(6)	C(35')-C(36')	1.29(5)
N(5)-P(25)	1.623(3)	C(36')-C(37')	1.40(4)
P(6)-C(13)	1.783(5)	C(38)-C(39)	1.366(7)
P(6)-C(19)	1.788(4)	C(38)-C(43)	1.406(7)
P(6)-C(7)	1.808(4)	C(39)-C(40)	1.388(7)
C(7)-N(8)	1.343(6)	C(40)-C(41)	1.361(9)
C(7)-C(12)	1.374(6)	C(41)-C(42)	1.378(9)
N(8)-C(9)	1.351(6)	C(42)-C(43)	1.389(8)
C(9)-C(10)	1.373(7)	C(50)-Cl(52)	1.77(5)
C(10)-C(11)	1.383(7)	C(50)-Cl(51)	1.82(5)
C(11)-C(12)	1.389(6)		
C(13)-C(18)	1.400(6)	C(2)-N(1)-P(6)	124.8(2)
C(13)-C(14)	1.403(7)	N(1)-C(2)-C(3)	111.0(3)
C(14)-C(15)	1.382(8)	C(2)-C(3)-C(4)	109.6(4)
C(15)-C(16)	1.389(10)	N(5)-C(4)-C(3)	109.5(4)
C(16)-C(17)	1.400(11)	C(4)-N(5)-P(25)	121.6(3)
C(17)-C(18)	1.372(8)	N(1)-P(6)-C(13)	108.22(18)
C(19)-C(20)	1.389(6)	N(1)-P(6)-C(19)	114.87(18)
C(19)-C(24)	1.404(6)	C(13)-P(6)-C(19)	110.8(2)
C(20)-C(21)	1.389(6)	N(1)-P(6)-C(7)	106.05(18)
C(21)-C(22)	1.380(8)	C(13)-P(6)-C(7)	109.73(19)
C(22)-C(23)	1.381(9)	C(19)-P(6)-C(7)	107.04(19)
C(23)-C(24)	1.380(7)	N(8)-C(7)-C(12)	123.4(4)
P(25)-C(38)	1.795(4)	N(8)-C(7)-P(6)	114.3(3)
P(25)-C(26)	1.799(4)	C(12)-C(7)-P(6)	122.1(3)
P(25)-C(32)	1.808(3)	C(7)-N(8)-C(9)	117.0(4)
P(25)-C(32')	1.813(18)	N(8)-C(9)-C(10)	123.0(4)
C(26)-N(27)	1.359(6)	C(9)-C(10)-C(11)	119.1(4)
C(26)-C(31)	1.366(7)	C(10)-C(11)-C(12)	118.5(4)
N(27)-C(28)	1.348(8)	C(7)-C(12)-C(11)	118.8(4)
C(28)-C(29)	1.366(9)	C(18)-C(13)-C(14)	119.5(4)
C(29)-C(30)	1.377(8)	C(18)-C(13)-P(6)	121.2(3)
C(30)-C(31)	1.398(8)	C(14)-C(13)-P(6)	119.2(3)
C(32)-C(33)	1.390	C(15)-C(14)-C(13)	120.2(5)
C(32)-C(37)	1.390	C(14)-C(15)-C(16)	119.8(6)
C(33)-C(34)	1.390	C(15)-C(16)-C(17)	120.2(6)
C(34)-C(35)	1.390	C(18)-C(17)-C(16)	120.2(6)
C(35)-C(36)	1.390	C(17)-C(18)-C(13)	120.0(5)

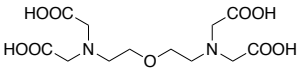
C(36)-C(37)	1.390	C(20)-C(19)-C(24)	120.0(4)
C(20)-C(19)-P(6)	118.6(3)	C(33)-C(32)-P(25)	120.4(3)
C(24)-C(19)-P(6)	121.0(3)	C(37)-C(32)-P(25)	119.5(3)
C(21)-C(20)-C(19)	119.6(4)	C(34)-C(33)-C(32)	120.0
C(22)-C(21)-C(20)	120.0(5)	C(33)-C(34)-C(35)	120.0
C(21)-C(22)-C(23)	120.7(4)	C(34)-C(35)-C(36)	120.0
C(24)-C(23)-C(22)	120.2(5)	C(37)-C(36)-C(35)	120.0
C(23)-C(24)-C(19)	119.5(5)	C(36)-C(37)-C(32)	120.0
N(5)-P(25)-C(38)	110.01(17)	C(37')-C(32')-C(33')	124(3)
N(5)-P(25)-C(26)	108.6(2)	C(37')-C(32')-P(25)	114(3)
C(38)-P(25)-C(26)	108.1(2)	C(33')-C(32')-P(25)	116(3)
N(5)-P(25)-C(32)	111.2(2)	C(32')-C(33')-C(34')	114(3)
C(38)-P(25)-C(32)	108.7(2)	C(33')-C(34')-C(35')	121(3)
C(26)-P(25)-C(32)	110.2(3)	C(36')-C(35')-C(34')	120(3)
N(5)-P(25)-C(32')	120.4(18)	C(35')-C(36')-C(37')	117(3)
C(38)-P(25)-C(32')	106.1(18)	C(32')-C(37')-C(36')	120(3)
C(26)-P(25)-C(32')	103.0(19)	C(39)-C(38)-C(43)	119.5(4)
N(27)-C(26)-C(31)	124.1(4)	C(39)-C(38)-P(25)	122.5(4)
N(27)-C(26)-P(25)	112.7(4)	C(43)-C(38)-P(25)	117.9(4)
C(31)-C(26)-P(25)	123.3(3)	C(38)-C(39)-C(40)	120.6(5)
C(28)-N(27)-C(26)	115.9(5)	C(41)-C(40)-C(39)	120.0(5)
N(27)-C(28)-C(29)	124.1(5)	C(40)-C(41)-C(42)	120.7(5)
C(28)-C(29)-C(30)	118.8(5)	C(41)-C(42)-C(43)	119.9(6)
C(29)-C(30)-C(31)	119.1(5)	C(42)-C(43)-C(38)	119.3(5)
C(26)-C(31)-C(30)	118.0(4)	Cl(52)-C(50)-Cl(51)	110(3)
C(33)-C(32)-C(37)	120.0		

8.3 List of MnBCA ligands

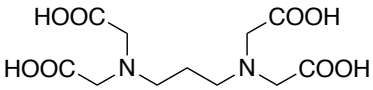
List of MnBCAs compiled by Uzal-Varela and co-workers in 2022.[1]



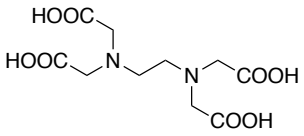
L01



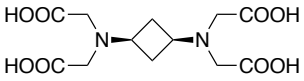
L02



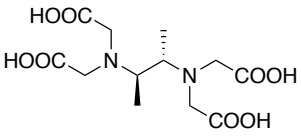
L03



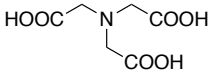
L04



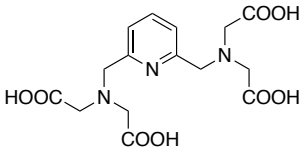
L05



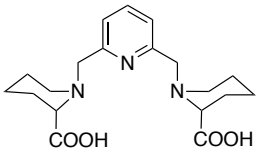
L06



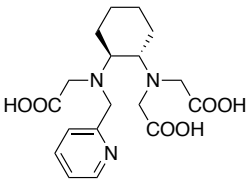
L07



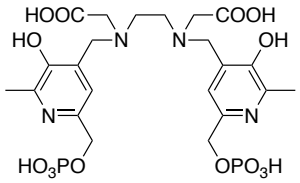
L08



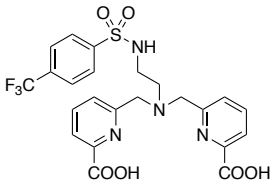
L09



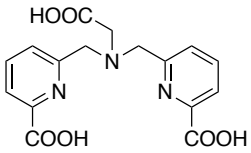
L10



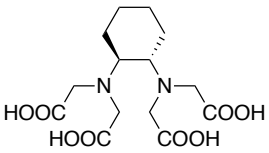
L11



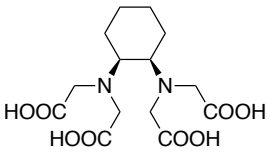
L12



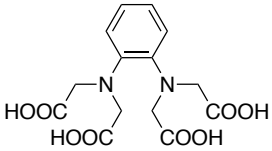
L13



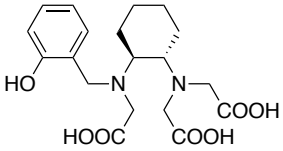
L14



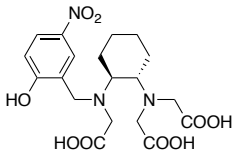
L15



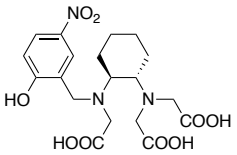
L16



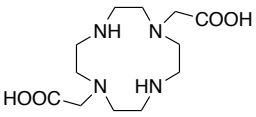
L17



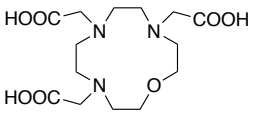
L18



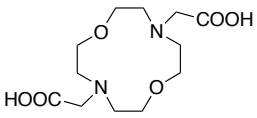
L19



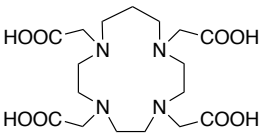
L20



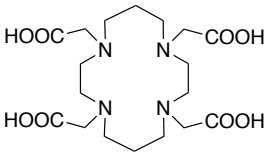
L21



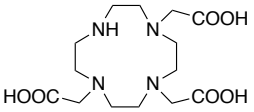
L22



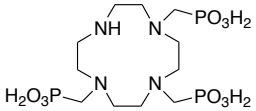
L23



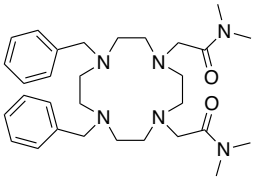
L24



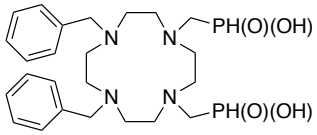
L25



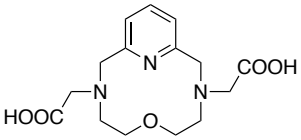
L26



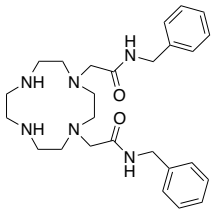
L27



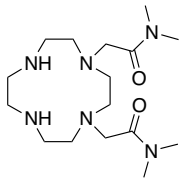
L28



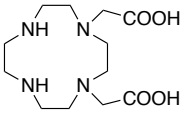
L29



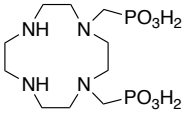
L30



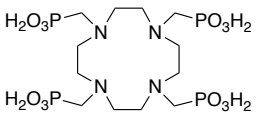
L31



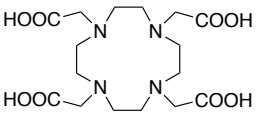
L32



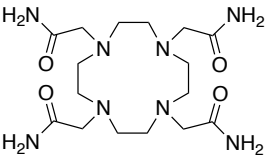
L33



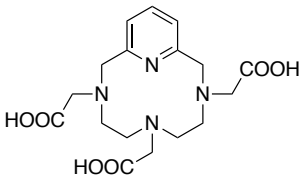
L34



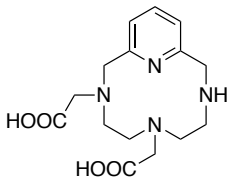
L35



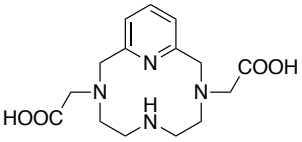
L36



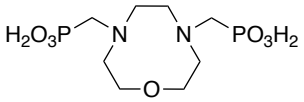
L37



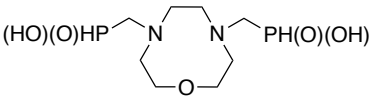
L38



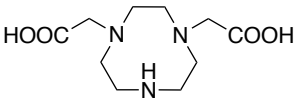
L39



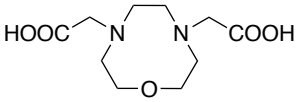
L40



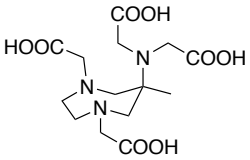
L41



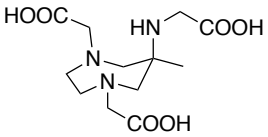
L42



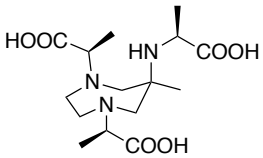
L43



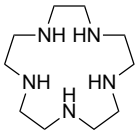
L44



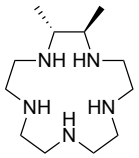
L45



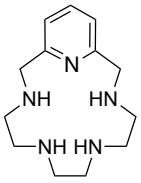
L46



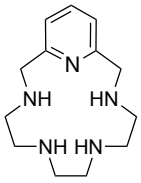
L47



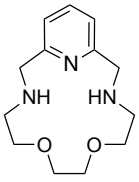
L48



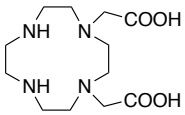
L49



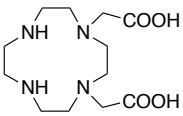
L50



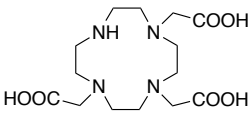
L51



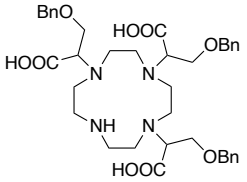
L52 (*L32*)



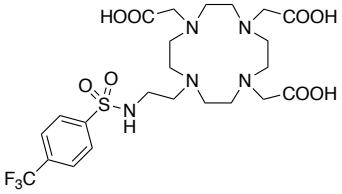
L53 (*L32*)



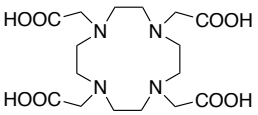
L54 (*L25*)



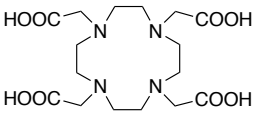
L55



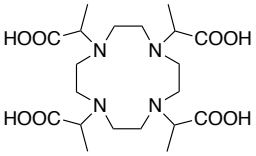
L56



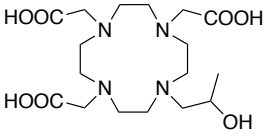
L57 (*L35*)



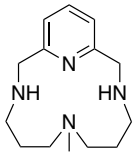
L58 (*L35*)



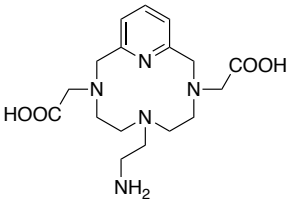
L59



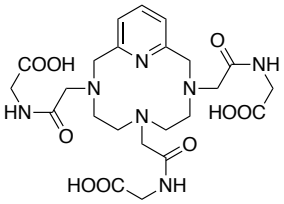
L60



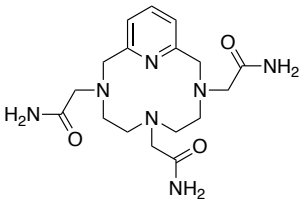
L61



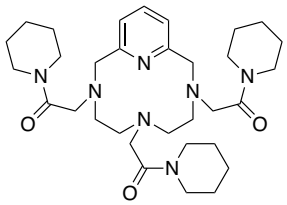
L62



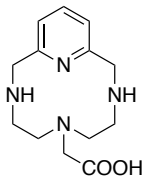
L63



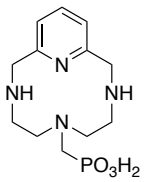
L64



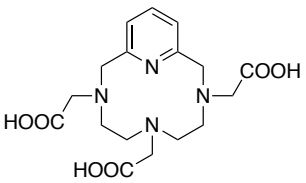
L65



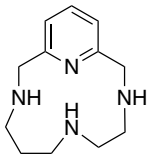
L66



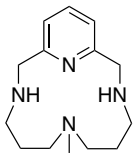
L67



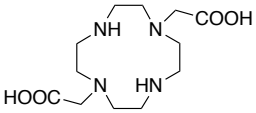
L68 (*L37*)



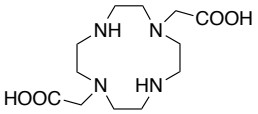
L69



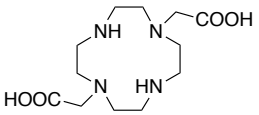
L70 (*L61*)



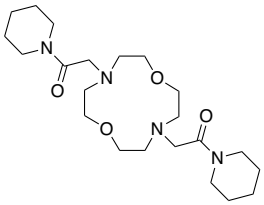
L71 (*L20*)



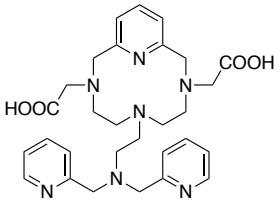
L72 (*L20*)



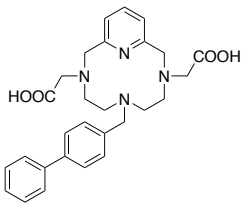
L73 (*L20*)



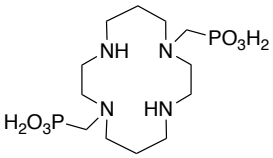
L74



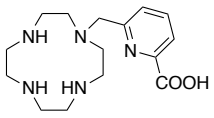
L75



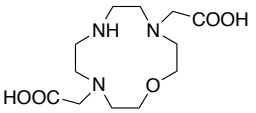
L76



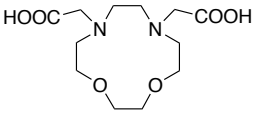
L77



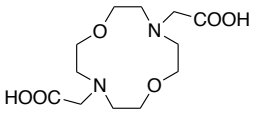
L78



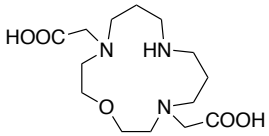
L79



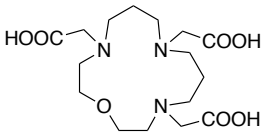
L80



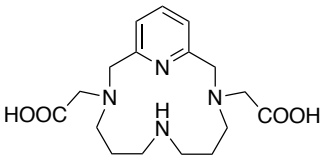
L81 (*L22*)



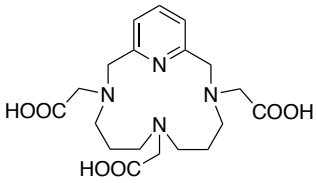
L82



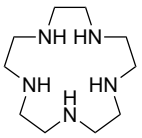
L83



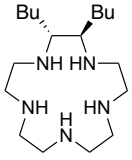
L84



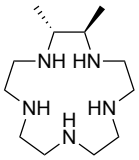
L85



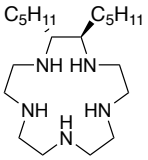
L86 (*L47*)



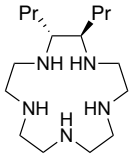
L87



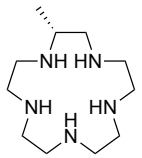
L88 (*L48*)



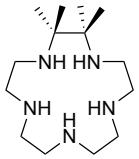
L89



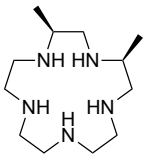
L90



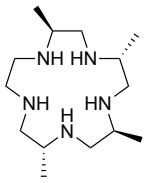
L91



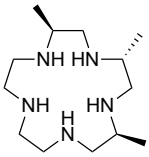
L92



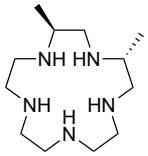
L93



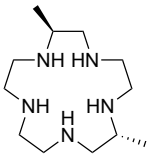
L94



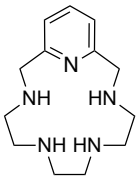
L95



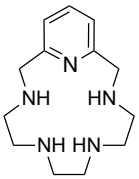
L96



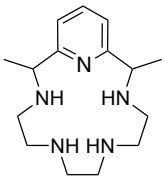
L97



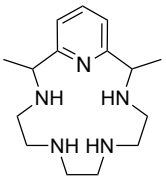
L98 (*L49*)



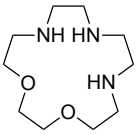
L99 (*L49*)



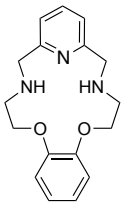
L100



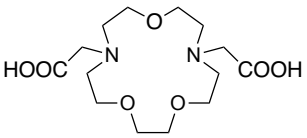
L101 (*L100*)



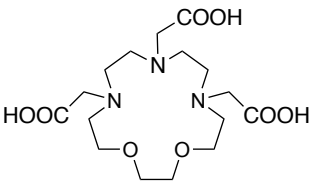
L102



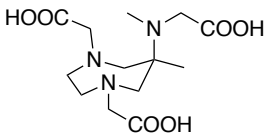
L103



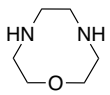
L104



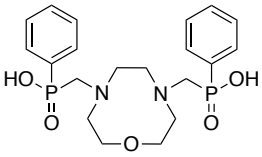
L105



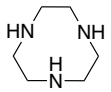
L106



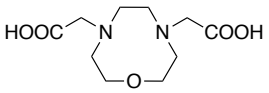
L107



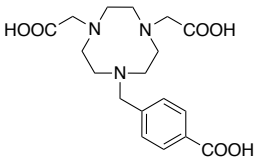
L108



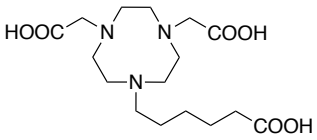
L109



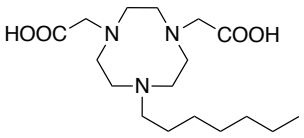
L110 (*L43*)



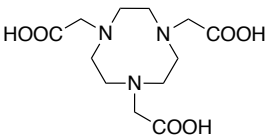
L111



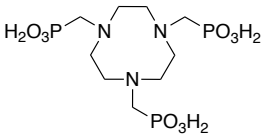
L112



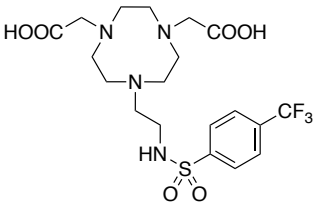
L113



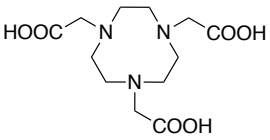
L114



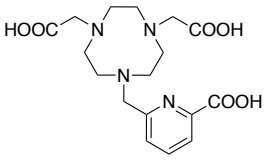
L115



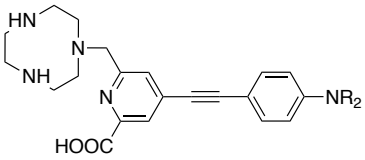
L116



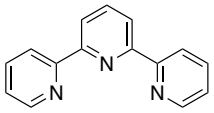
L117 (*L114*)



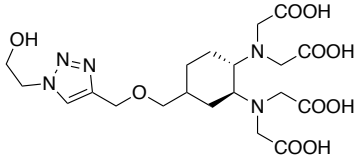
L118



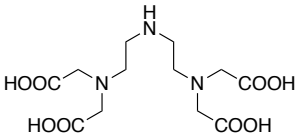
L119



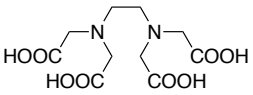
L120



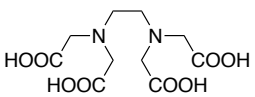
L121



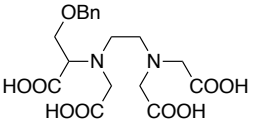
L122



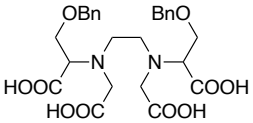
L123 (*L04*)



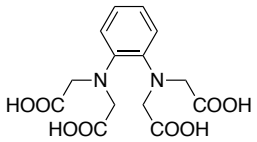
L124 (*L04*)



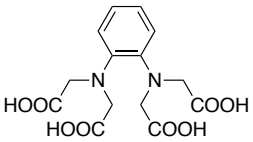
L125



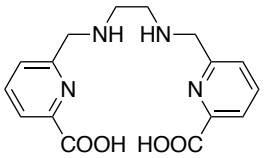
L126



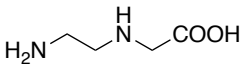
L127 (*L16*)



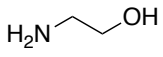
L128



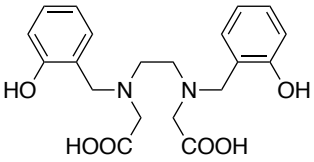
L129



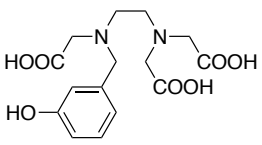
L130



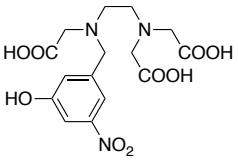
L131



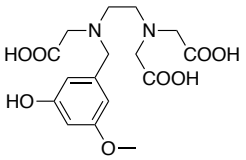
L132



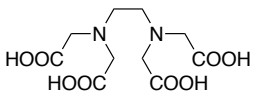
L133



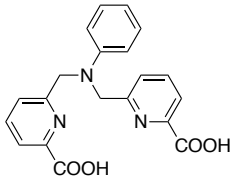
L134



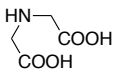
L135



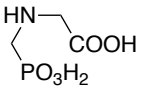
L136 (*L04*)



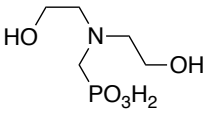
L137



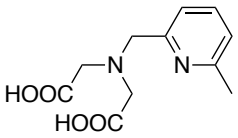
L138



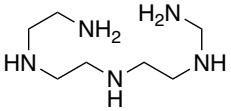
L139



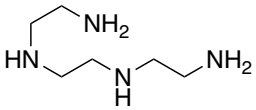
L140



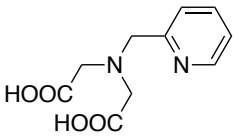
L141



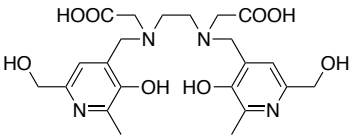
L142



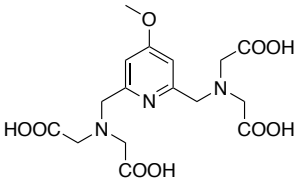
L143



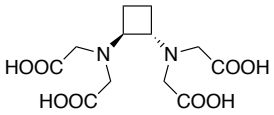
L144



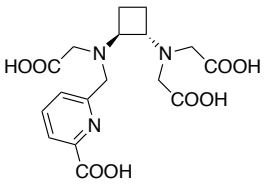
L145



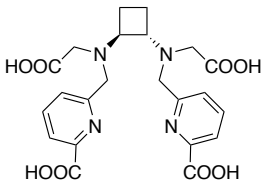
L146



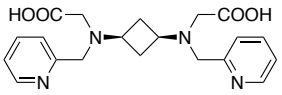
L147



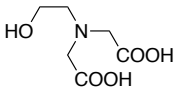
L148



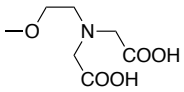
L149



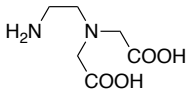
L150



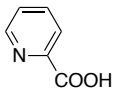
L151



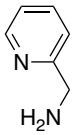
L152



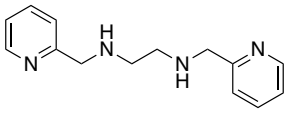
L153



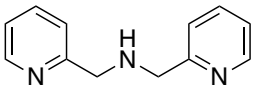
L154



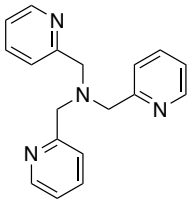
L155



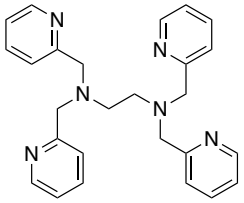
L156



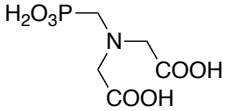
L157



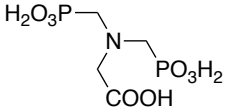
L158



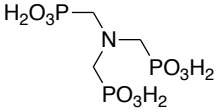
L159



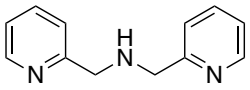
L160



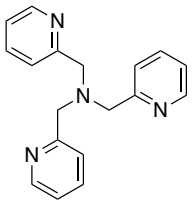
L161



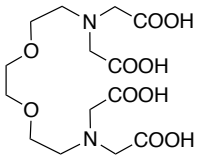
L162



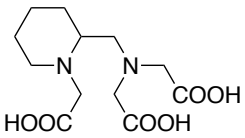
L163 (*L157*)



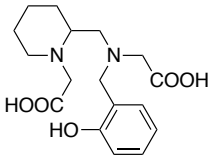
L164 (*L158*)



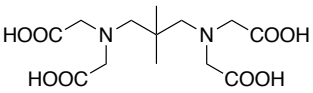
L165



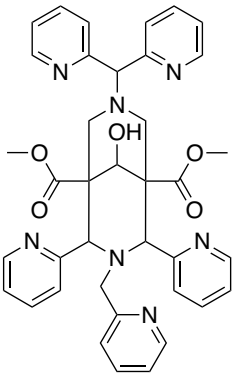
L166



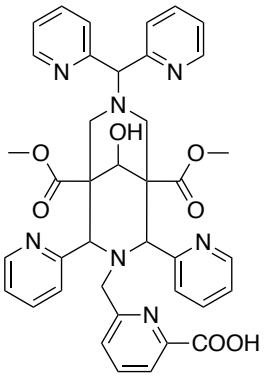
L167



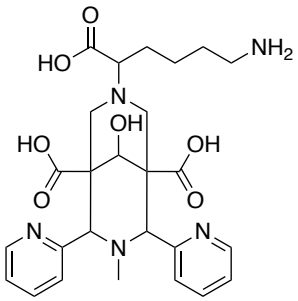
L168



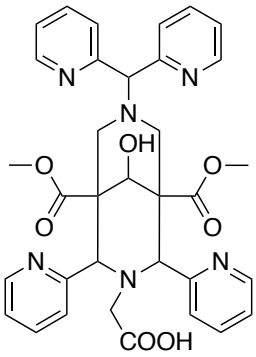
L169



L170



L171



L172

8.3.1 References

1 R. Uzal-Varela, F. Pérez-Fernández, L. Valencia, A. Rodríguez-Rodríguez, C. Platas-Iglesias, P. Caravan and D. Esteban-Gómez, *Inorg. Chem.*, 2022, **61**, 14173–14186, DOI: 10.1021/acs.inorgchem.2c02364.

8.4 Optimised geometry coordinates of DFT complexes

.log files containing the optimised geometry coordinates and frequency calculations are provided in a supplementary zipped folder.