Biophysical and Computational Studies of Biomacromolecular Systems

Dale Linsday Ang, B.Sc (Adv)(Hons),

MRACI CChem, AMRSC

ORCID: 0000-0001-5142-7555

School of Science and Health

Western Sydney University

A thesis submitted for the degree of

Doctor of Philosophy

July 2019

Acknowledgements

I am most grateful to my supervisor Professor Janice Aldrich-Wright for your never-ending enthusiasm, support and guidance. The many opportunities you have enabled for me throughout my candidature are truly appreciated. Thanks also to my cosupervisor Dr Christopher Gordon for the many helpful discussions.

A special thanks to Professor Alison Rodger for your guidance and inspiration throughout my time at Warwick University and since. Living in the UK for a year, so far from family, was a challenge but you made me feel welcome and included within your wonderful research group. It's an experience I will forever remember.

Thank you Harmony. It was only with your love and support that have I been able to complete this journey.

Statement of Originality

This thesis is submitted in fulfilment of the requirements for Doctor of Philosophy at Western Sydney University. The work presented in this thesis is, to the best of my knowledge and belief, original except where acknowledged in the text.

I hereby declare that I have not submitted this material, either in full or in part, for a degree at any other institution.

[redacted]

Dale Lindsay Ang

July 2019

Abstract

The binding of mono- and dinuclear platinum(II) complexes with both double-stranded and G-quadruplex DNA (QDNA) was explored using a combination of spectroscopic and computational techniques. The stabilising effects on G-quadruplex DNA were assessed and structure-activity relationships developed to guide the future development of QDNA selective complexes.

Additionally, two pieces of software were developed, each able to process spectroscopic information. The first application calculates binding constants from various spectroscopic data including circular dichroism and fluorescence. The second program, SOMSpec, uses a machine learning approach to elucidate biomacromolecular structure from circular dichroism and infrared spectra.

Contents

Introduction

These works represent an overarching study into the interactions of small molecules with biomacromolecules, such as DNA, for the purposes of understanding and directing the design of bioactive platinum(II) complexes with improved specificity, stabilisation properties and efficacy.[1] My particular interest has been in complexes which stabilise G-quadruplex DNA (QDNA). QDNA is formed from guanine-rich regions of DNA such as the telomeres of eukaryotic DNA, regions of linear non-coding $d(TTAGGG)_n$ tandem repeats.[2] G-rich regions self-associate through Hoogsteen base-pairing resulting in the formation of these non-canonical QDNA structures.[3] Importantly, the presence of these structures *in vivo*, particularly within telomeric[4] and oncogene promotor regions,[5,6] represent potential targets for the development of novel Platinum based anticancer agents.[7] Select topologies of these QDNA structures, which vary by sequence and length, are illustrated in Figure 1.

This thesis is arranged into six chapters, the first five are each based on a particular facet of my outputs during my candidature, followed by a sixth 'Conclusions and future work' chapter. Chapter 1 begins by exploring the binding affinity of a series of platinum(II)-phenanthroline complexes with double-stranded DNA (dsDNA) using synchrotron radiation circular dichroism (SRCD). A significant outcome of this work was the development of a new method to calculate equilibrium binding constants and a software implementation to simplify the process.[8] This laid the foundation for exploring DNA interactions with a class of platinum(II)-terpyridine complexes, as presented in Chapter 2.

In Chapter 2, I begin by evaluating the dsDNA and QDNA binding interactions

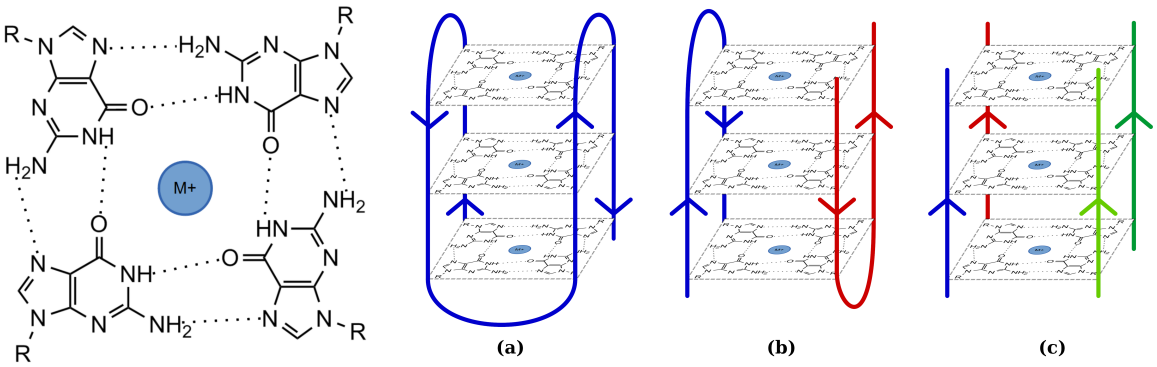


Figure 1: G-quadruplex guanine tetrad consisting of Hoogsteen-bound guanine bases stabilised by a central cation (left); schematics demonstrating the 5′–3′ arrangement of: (a) anti-parallel unimolecular; (b) anti-parallel bimolecular; and (c) tetramolecular parallel QDNA.

of a series of dinuclear platinum(II)-terpyridine complexes I had synthesised in my Honours year exhibiting QDNA affinity.[9] Ultimately, it was discovered that these complexes exhibited significant QDNA stabilisation, possessing impressive selectivity.[1] However, these complexes also were found to cause significant dsDNA precipitation, so my attention was then focussed toward the factors necessary for Pt(II) complexes to exhibit QDNA specificity. I also wanted to explore additional biophysical approaches to investigating these binding interactions. This resulted in the application of electrospray ionisation-mass spectrometry (ESI-MS) to a range of Pt(II)-phenanthroline complexes, culminating in Chapter 3.

Chapter 3 involves revisiting the platinum(II)-phenanthroline complexes of Chapter 1, where I investigated the interactions and selectivity of these with dsDNA and QDNA. Through the application of ESI-MS, CD, and computational docking simulations I was able to develop a greater understanding of the structure-activity relationship exhibited by the planar phenanthroline moiety on the strength and selectivity of binding with QDNA.

With the amount of CD and SRCD spectra obtained thus far I was curious of alternative means for exploring this data. Chapter 4 presents my efforts in applying more advanced computation methods to understanding biophysical data. In this research, I developed a machine learning method for predicting protein secondary

structure content from circular dichroism (CD) and also infrared (IR) and RAMAN spectra. This approach was intentionally made flexible enough to be applicable to other biophysical methods and with the acquisition of suitable training set data will be applied to a wider range of problems in the future.

Finally, Chapter 5 presents additional outputs that were tangential to my main research. Much of this was the result of collaborations where I contributed my expertise in either spectroscopy, synthesis, visualisation, or computer simulation. It includes additional published papers, conference proceedings, posters, and grants awarded during my candidature.

As a proponent of research transparency and reproducibility a substantial amount of my work presented here were created using only free software (specifically GPLv2 or later licensed). This is why the software developed during my candidature is GPLv3 licensed. It is important to define the phrase “free software” so I’ll do so now—think “free” as in liberty, not “no cost”. As stated by the Free Software Foundation (fsf.org) the four essential freedoms are: "The freedom to run the program as you wish, for any purpose (freedom 0); The freedom to study how the program works, and change it so it does your computing as you wish (freedom 1). Access to the source code is a precondition for this.; The freedom to redistribute copies so you can help others (freedom 2).; The freedom to distribute copies of your modified versions to others (freedom 3)."

To extend this same opportunity to others, I present here a subset of the software which has enabled this goal. My operating system of choice is Fedora Workstation. The computer I designed and built exclusively to handle my computational tasks, Arrakis, runs Ubuntu LTS. Molecular docking was performed by Autodock Vina.[10] Molecular dynamics simulations were all performed using GROMACS[11] compiled to enable CUDA acceleration, and prepared with a combination of Avogadro[12] and UCSF Chimera,[13] and analysed using VMD.[14] I utilised Inkscape, GIMP, Krita, and Blender for graphical work–particularly when generating the journal covers. Finally, this thesis and several of my journal manuscripts were typeset in LaTeX using TeXstudio, in combination with the Zotero[15] reference manager.

Chapter 1

Improved DNA equilibrium binding affinity determination of platinum(II) complexes using synchrotron radiation circular dichroism

Equilibrium binding constants are an important parameter used to assess many biomacromolecules interactions with small molecules. These are often investigated with a hope that they will correlate with biological activity and therefore provide a rationale to drug design. As my research focussed on the development of DNA binding small molecules it was imperative to be able to reliably and reproducibly measure these binding interactions.

This work begins by showing why the contemporary models for calculating binding affinity in biological systems are not representative of the physical phenomena. The literature methods of Scatchard,[16] Schmechel and Crothers,[17] and others determine the binding constant by extrapolating ligand concentration to zero. The approach presented here considers and calculates binding data simultaneously at all wavelengths of the experiment. In this work I chose to demonstrate the method using synchrotron circular dichroism data, but the method is valid for any type

of spectroscopic data containing binding information such as fluorescence and has been used successfully as such. Ultimately, I found that not only is this method far faster at producing results, allowing for near real time experimental feedback, it also revealed varying binding constant values across the various wavelengths. This finding is best stated by quoting the manuscript:

"As SRCD spectra is the results from all the combined couplings of transitions of all bases, we attribute the differences in K observed with changing wavelength to be a consequence of the interactions at specific bases. Although this interaction will change the overall spectra it may do more so in the region where the resulting couplings of the transitions from this interaction dominate. This suggests that... the binding may be more complex than this model is capable of describing as the definition is that K is a constant across all wavelengths."

As primary author, my contribution involved data acquisition, data analysis and processing, together with the development and implementation of the underlying code implementation—proof of concept code having been developed by Frank Stootman. I produced all artwork and wrote a majority of the text. The code is freely available, licensed under the GPLv2. This work was published in RSC Analyst and has been cited 2 times (as of November 2019). As this work is applicable to a broad range of researchers, the low citation count aludes to a challenge with the accessibility of the method. Likely, this may be attributed to the requirements of Mathematica and the need for a particular data format. This will be addressed in a future revision of the software. It has also become an important teaching tool in the third year undergraduate Pharmacological Chemistry unit at Western Sydney University where it is used to demonstrate calculation of binding constants from experimental fluorescence displacement data.

Paper reference

Ang, D. L.; Jones, N. C.; Stootman, F.; Ghadirian, B.; Aldrich-Wright, J. R.: Improved DNA Equilibrium Binding Affinity Determinations of Platinum(II) Complexes using Synchrotron Radiation Circular Dichroism. Analyst, 2016, 140, 4162–4169. DOI: 10.1039/C5AN00066A. Copyright The Royal Society of Chemistry. Reproduced with permission.[8]

Analyst

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2015, **140**, 4162

Improved DNA equilibrium binding affinity determinations of platinum(II) complexes using synchrotron radiation circular dichroism†

Dale L. Ang,[a] Nykola C. Jones,[b] Frank Stootman,[a] Bahman Ghadirian[a] and Janice R. Aldrich-Wright*[a]

Received 12th January 2015,
Accepted 14th April 2015

DOI: 10.1039/c5an00066a

www.rsc.org/analyst

The binding affinity of a series of square planar platinum(II) compounds of the type $[Pt(A_L)(I_L)]^{2+}$, where A_L is 1,2-diaminoethane and I_L are 1,10-phenanthroline (phen), 4-methyl-1,10-phenanthroline (4Mephen), 5-methyl-1,10-phenanthroline (5Mephen), 4,7-dimethyl-1,10-phenanthroline (47Me_2phen), 5,6-dimethyl-1,10-phenanthroline (56Me_2phen) or 3,4,7,8-tetramethyl-1,10-phenanthroline (3478Me_4phen) has been reinvestigated using Synchrotron Radiation Circular Dichroism (SRCD) spectroscopy. The additional peaks exhibited considerably greater intensity than those observed between 200 and 400 nm affording additional binding affinity determinations. In addition, the authors have reviewed the various mathematical approaches used to estimate equilibrium binding constants and thereby demonstrate that their mathematical approach, implemented with Wolfram *Mathematica*, has merit over other methods.

Introduction

Circular Dichroism (CD) spectroscopy is a well-established method used for biophysical examination of chemical and biological structures, however, in conventional CD instruments low flux of the light source in the far UV and vacuum UV wavelengths can limit the acquisition of information. These limitations may be overcome using Synchrotron Radiation Circular Dichroism (SRCD) spectroscopy.[1,2] The application of synchrotron radiation to CD was first reported over 20 years ago, the intense light source increasing effectiveness particularly as a spectroscopic technique for both structural and functional genomics.[3,4] An additional advantage of SRCD includes the ability to obtain data at shorter wavelengths encroaching into the vacuum UV region, and as a result, superior structural information.[3]

A comparison of the SRCD and conventional CD spectra of ct-DNA (Fig. 1), illustrates the extended wavelength range of SRCD. Note that the data is scaled to facilitate this comparison. These intensely absorbing bands in the vacuum and far UV region are seen for many compounds, including proteins and other biomolecules.[3,4] Chiral metal complexes (MCs), such as [(5,6-dimethyl-1,10-phenanthroline) (1*S*,2*S*-diamino cyclohexane) platinum(II)] dichloride (56MESS) and [(5,6-dimethyl-1,10-phenanthroline) (1*R*,2*R*-diaminocyclohexane) platinum(II)] dichloride (56MERR), also show additional intense absorption bands well into the far UV (Fig. 2).

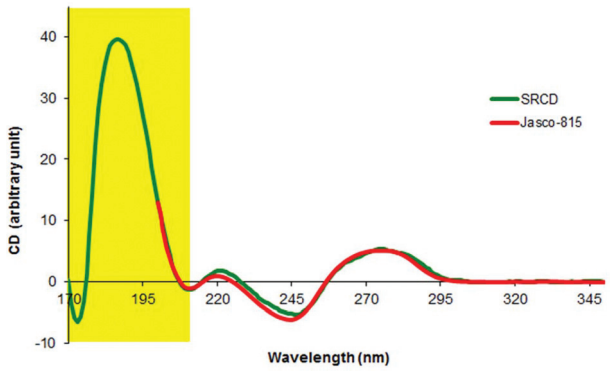


Fig. 1 The spectra of ct-DNA as measured on a conventional CD (red) and SRCD (green) spectrophotometer, where the spectra have been scaled to allow direct comparison. CD spectrum was recorded between 200–350 nm, 35 μM, 1.0 cm path length, ambient temperature and 10 accumulations. SRCD spectrum was recorded between 170–350 nm, 1.1 mM, 0.01 cm path length, 25 °C and 3 accumulations.

Equilibrium binding experiments are extensively used to investigate interactions between small molecules and biomacromolecules, such as proteins and DNA.[5–14] These interactions may involve intercalation, H-bonding, van der Waals

[a]*Nanoscale Organisation and Dynamics Group, School of Science and Health, University of Western Sydney, Penrith, NSW 2751, Australia.*
E-mail: J.Aldrich-Wright@uws.edu.au; Fax: (+61) 246203025; Tel: (+61) 246203218
[b]*ISA, Department of Physics and Astronomy, Aarhus University, DK 8000 Aarhus C, Denmark*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c5an00066a

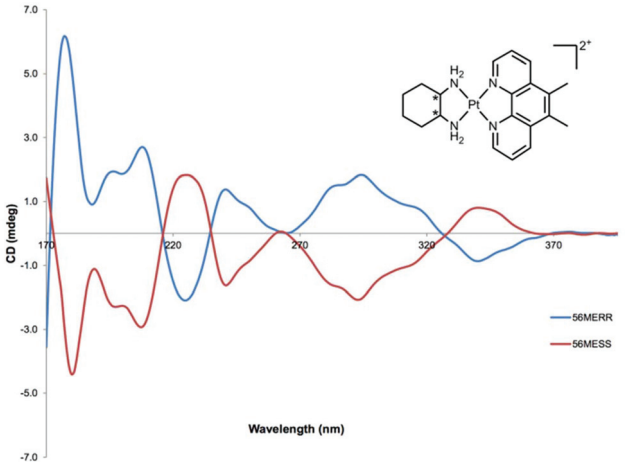


Fig. 2 The SRCD spectra of 56MESS and 56MERR. Wavelength = 170–400 nm, concentration = 4 mM, path length = 0.01 cm. Inset: the structure of 56MESS/56MERR, *denotes the chiral centres.

groove binding or π stacking. To evaluate the binding affinity of related small molecules for a particular biomacromolecule under specific experimental conditions the binding constant K_b is determined; often using methods that poorly estimate this value.

Determining an equilibrium binding constant

The equilibrium binding constant (K) and the number of bases per binding site (n) have heretofore been calculated using a variety of methods as developed by Scatchard,[15] Schmechel and Crothers,[16] McGhee and von Hippel,[17] Nordén,[18,19] Wolfe *et al.*,[20] Rodger,[7,21–23] Rodger and Nordén,[24] and Kumar and Asuncion,[25–27] among others. Some of these approaches share a common assumption that artificially determines the binding constant at very low bound-ligand concentrations (as demonstrated in the Supp data). Our method,[28] developed from first principles using the relation $L_F + S_F \underset{K}{\longleftrightarrow} L_B$ (where L_F and S_F are free ligand and free substrate concentrations respectively and L_B is the bound ligand concentration) and represented in eqn (1), obtains a value of K by accounting for the entire concentration range.

$$L_B = 0.5R\left(\frac{1}{K} + L_T + S_T - \sqrt{\left(\frac{1}{K} + L_T + S_T\right)^2 - 4S_T L_T}\right) \quad (1)$$

The fitting of eqn (1) to experimental data has been accomplished with a script (called a 'notebook') in *Mathematica* and is given in the ESI.† Of note is that eqn (1) has a mathematical point of inflexion which is indicated on the fitted curve as a red dot if a valid fit is achieved.

In this work we initially use this method in *Mathematica* for the fitting of eqn (1) to reprocess previously obtained CD binding data for [(1,10-phenanthroline)(1,2-diaminoethane)-platinum(II)] dichloride (PHENEN, 1) with ct-DNA to assure that the results are consistent. We also demonstrate that the method can also be used to determine the binding constant for UV data. Finally, we interrogate recently obtained SRCD binding data for ct-DNA and the platinum(II) complexes PHENEN, (**1**), [(4-methyl-1,10-phenanthroline)(1,2-diaminoethane)platinum(II)] dichloride (4MEEN, 2), [(5-methyl-1,10-phenanthroline)(1,2-diaminoethane) platinum(II)] dichloride (5MEEN, 3), [(4,7-dimethyl-1,10-phenanthroline)(1,2-diaminoethane)platinum(II)] dichloride (47MEEN, 4), [(5,6-dimethyl-1,10-phenanthroline)(1,2-diaminoethane)platinum(II)] dichloride (56MEEN, 5) and [(3,4,7,8-tetramethyl-1,10-phenanthroline)(1,2-diaminoethane)platinum(II)] dichloride (3478MEEN, 6) (Fig. 3) by using this improved method to observe what effect the increased beam intensity and shorter wavelengths contribute to the experimental results. Here we present a method, implemented with Wolfram *Mathematica*, for the quick and effective determination of the binding constant.

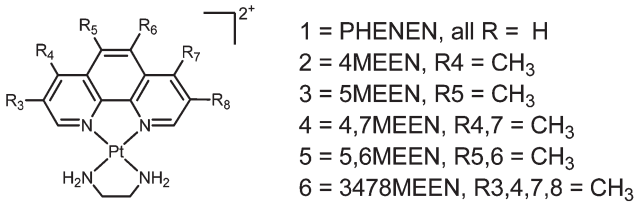


Fig. 3 The general structure of the platinum(II) complexes investigated.

Experimental

Materials

All platinum(II) complexes were synthesised and characterised as previously reported.[5] Ct-DNA, disodium hydrogen phosphate, sodium dihydrogen phosphate sodium fluoride and D-10-camphorsulfonic acid (CSA) were all purchased from Aldrich Chemical Company. Aqueous solutions were prepared with Milli-Q water (Millipore, MA). All other reagents and solvents were of analytical grade.

Instrumentation

Conventional CD were recorded as previously reported.[5] In summary: a JASCO J-810 CD spectropolarimeter, at the ambient temperature in the wavelength range of 200–350 nm, at ct-DNA concentrations of 35 μM (2600 μL) and MC solution of 10 mM, using a path length of 1 cm, a data pitch of 0.5 nm with 10 accumulations. Aliquots of MC (1 μL) were titrated into a cuvette containing the ct-DNA and the titration continued until no further change was observed in the CD spectrum, indicating that the saturation point of the DNA had been reached. Baseline subtraction was achieved by subtraction of the ct-DNA/buffer spectrum.

SRCD spectra were recorded at the ASTRID synchrotron at ISA, Aarhus University, Denmark (beamline CD1) operating at

Table 1 Prepared solutions for SRCD DNA binding study

Tube no.	Buffer/μL	DNA/μL	MC/μL	Buffer/μL	Total/μL	DNA/mM	MC/mM	Ratio
0	300	0	0	0	300	0	0.00	0
1	—	220	0	80	300	1.1	0.00	0
2	—	220	1	79	300	1.1	0.07	0.06
3	—	220	1.5	78.5	300	1.1	0.10	0.09
4	—	220	2	78	300	1.1	0.13	0.12
5	—	220	2.5	77.5	300	1.1	0.17	0.15
6	—	220	3	77	300	1.1	0.20	0.18
7	—	220	3.5	76.5	300	1.1	0.23	0.21
8	—	220	4	76	300	1.1	0.27	0.24
9	—	220	5	75	300	1.1	0.33	0.30
10	—	220	6	74	300	1.1	0.40	0.36
11	—	220	12	68	300	1.1	0.80	0.72

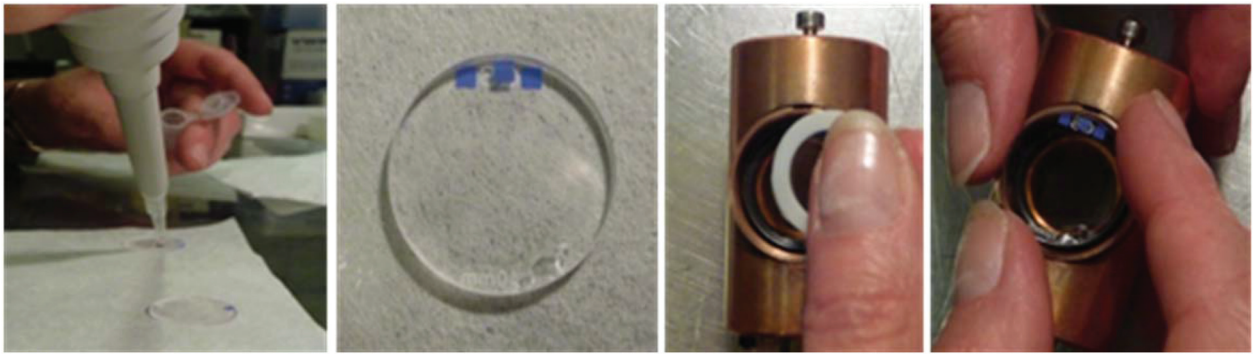


Fig. 4 Assembly of demountable 0.01 cm quartz cell of type QS124 (Hellma GmBH, Germany).

25 °C. The instrument was calibrated for magnitude and polarisation daily, using D-10-camphorsulfonic acid (CSA, 7.4 mg mL^{-1}). SRCD spectra were measured using a quartz cell of type QS124 (Hellma GmBH, Germany) with a path length of 0.01 cm. Spectra were measured over the wavelength range 170–350 nm, using a slit width of 0.25 nm, a data pitch of 1 nm, and an average dwell time of 2.0 s. Spectra were averaged from 3 accumulations and smoothed with the CD data processing software CDTool.[29]

A Thermo Scientific Evolution 300 UV-Vis spectrophotometer, with a 1 cm quartz cell at room temperature scanning over a 200–420 nm wavelength range, was used to determine the concentration of CSA (ε_{285} nm = 34.6 M^{-1} cm^{-1})[30] and ct-DNA (ε_{260} nm = 13 200 M^{-1} cm^{-1}).[31] For all experiments the ratio of the 192.5 : 290 nm peaks ranged between 2.04 and 2.06.

Binding experiments

Solution preparation. Some experimental changes to the published methods were made in order to reinvestigate the binding affinity of these platinum complexes with ct-DNA using SRCD.[5] Individual solutions (11 samples per experiment, 2 experiments per MC) were prepared for measurement to determine affinity constants for the MCs with ct-DNA. The DNA solution (1.1 mM) in phosphate buffer (3.6 mM) with NaF (36 mM) at pH 7.1 and MC solutions (2 mM) were prepared. Maintaining a constant volume of ct-DNA (220 μL), different volumes MC solutions (2 mM) were prepared off-site prior to the binding study (details in Table 1) and transported to the Synchrotron. An aliquot from the titration mixture (usually 25 μL) was loaded onto the base of the demountable cell (Fig. 4) by micropipette. The cover of the cell was then placed over the solution with care to avoid bubble formation. All samples were degassed before they were loaded into the cells. The spectra were recorded over the range 170–350 nm with 3 accumulations.

Processing of data. Processing of the experimental data was performed by fitting the data to eqn (1) across all wavelengths. This was implemented using Wolfram *Mathematica* version 9.‡

Results and discussion

Model verification

As a quality assurance test of the implementation, the previously published CD ct-DNA titration spectra[5,28] of **1** was reanalysed (Fig. 5). The *Mathematica* notebook produces an

‡The *Mathematica* binding notebook is provided in the ESI.† It can also be obtained by contacting the corresponding author.

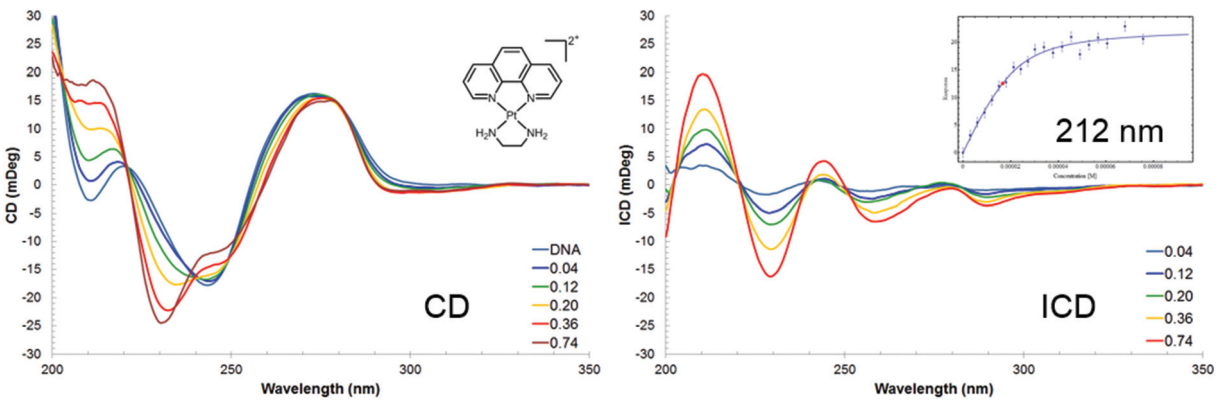


Fig. 5 CD (left) and ICD spectra (right) at varying ratios of [MC]/[DNA] (listed) [Pt(en)(phen)]$Cl_2 \cdot 2H_2O$ into ct-DNA, PS buffer. Binding curve at 212 nm (inset). The solid line is a fit of eqn (1), $K = 2.5 \pm 0.1 \times 10^5$ M. The red dot, at the point of inflexion, indicates a good fit.

Table 2 Summary of successful fits of previously reported CD spectra[28] of **1** titrated into ct-DNA, as determined by *Mathematica* notebook implementation of eqn (1), with associated error values and bases per binding site

Wavelength (nm)	K ($\times 10^5$)	n per complex
206.5	2.6 ± 0.1	3.8 ± 0.1
208.0	6.3 ± 0.4	3.2 ± 0.1
212.0	2.5 ± 0.1	3.7 ± 0.1
213.0	2.2 ± 0.1	3.8 ± 0.1
214.5	1.1 ± 0.1	5.0 ± 0.2
249.0	7.8 ± 0.1	1.1 ± 0.1
267.5	5.3 ± 0.2	2.6 ± 0.3
269.0	1.3 ± 0.1	2.8 ± 0.3
269.5	0.8 ± 0.1	2.8 ± 0.1

interactive dialogue allowing the entire wavelength range to be analysed and outputs this data in tabulated form. The results for **1** are listed in Table 2. Selection of an appropriate wavelength is evident when eqn (1) fits the curve, a red dot is drawn at the point of inflexion, as shown in Fig. 5, thus providing an unbiased evaluation of the data. Previously, the binding constant was determined to be $2.7 \pm 0.2 \times 10^5$ at 211 nm using the least squares method. Employing the method presented here the binding constant was calculated to be $2.5 \pm 0.1 \times 10^5$ at 212 nm which is in close agreement with the previously published value.[5]

It is noteworthy that this equation does not always achieve a valid fit of the experimental data. When the conditions are met the observation that the calculated equilibrium binding constant varies with wavelength is apparent and that not all wavelengths produce a mathematically sound fit, highlighting an important feature in our interpretation of the binding model with eqn (1). This demonstrates an important progression as it removes any influence of subjectivity in the selection, calculation and reporting of binding constants. As shown in the derivation of eqn (1) (ESI†), previous methods make assumptions to simplify the calculations and as such can only make approximations of the equilibrium binding constant at a given wavelength. As SRCD spectra is the results from all the combined couplings of transitions of all bases, we attribute the differences in K observed with changing wavelength to be a consequence of the interactions at specific bases. Although this interaction will change the overall spectra it may do more so in the region where the resulting couplings of the transitions from this interaction dominate. This suggests that the data may suffers from experimental noise or that the binding may be more complex than this model is capable of describing as the definition is that K is a constant across all wavelengths.

The data collected from the conventional CD also includes UV spectrum data which was likewise examined using the *Mathematica* notebook producing the UV spectra illustrated in Fig. 6. The inset depicts the response curves at peak maxima 227 and 274 nm which show that this data does not fit eqn (1). In fact for this binding interaction there was no wavelength for which a point of inflexion was obtained, *i.e.* the UV data does fit eqn (1).

SRCD experiments

The SRCD spectra of **1–6** each with ct-DNA were obtained in duplicate. The SRCD and induced SRCD (ISRCD) titration spectra for **1** are shown in Fig. 7. For each MC, a previously unseen and significantly more intense peak at ~185 nm was evident. The *Mathematica* notebook was used to interrogate the spectra of each MC; the red dot on the binding curve indicating a true fit of eqn (1). An example is illustrated in Fig. 7 (inset) with the data provided in Table 3 (Figures and Tables for the duplicate experiment for **1** and for **2–6** are provided in the ESI†). It is evident from this analysis that the resulting SRCD data did not cover a comprehensive concentration range; the samples were pre-prepared and transported to the synchrotron for measurement with no facility for further preparation. Despite this, the binding constant obtained at 211 nm was determined to be $1.6 \pm 0.2 \times 10^5$ M with the average $K = 2.2 \pm 0.3 \times 10^5$ M and $n = 3.0 \pm 0.1$ across repeats. Though slightly lower, this is comparable to the previously reported CD value for **1**.

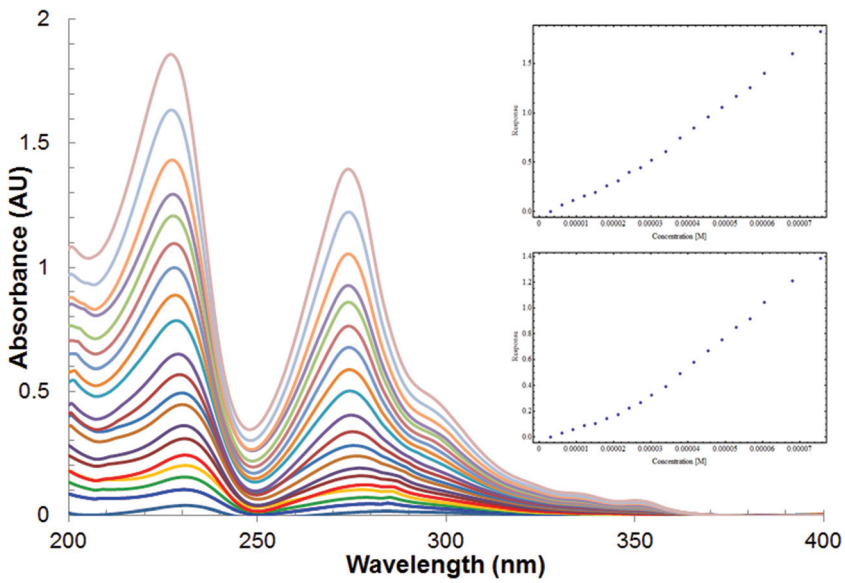


Fig. 6 UV spectra at varying ratios of [MC]/[DNA]. **1** into ct-DNA in PS buffer. Titration curves at 227 and 274 nm (inset). The lack of a curve indicates poor fit of eqn (1).

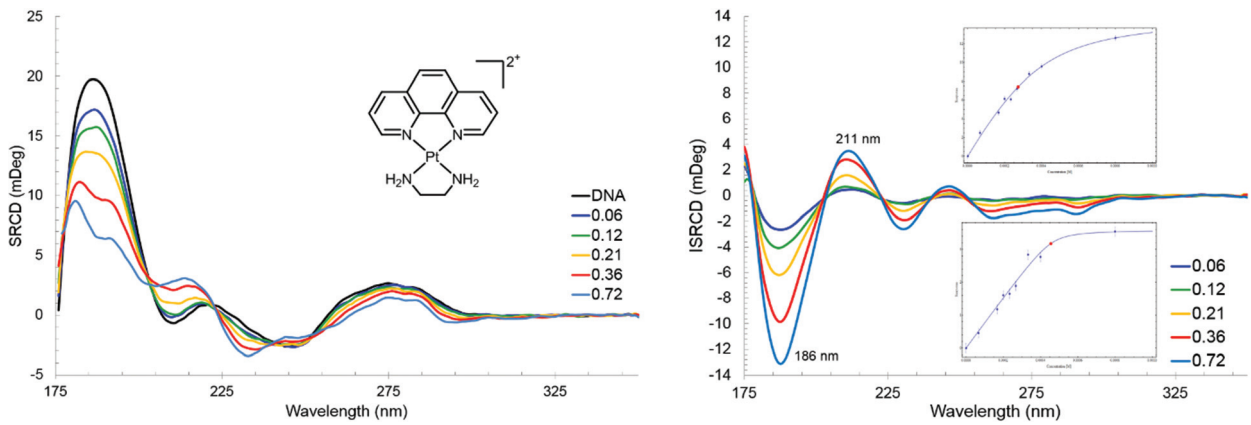


Fig. 7 Experiment A showing the SRCD and ISRCD spectra of varying concentrations of **1** with ct-DNA (1.1 mM) in 5 mM phosphate buffer, 50 mM NaF at 7.5 pH. Binding curve determined at 186 and 211 nm (inset). The solid line is a fit of eqn (2); the red dot indicates a good fit.

Table 3 Complex 1 binding data from SRCD experiment

Wavelength (nm)	$K \times 10^5$	n per complex
186	0.11 ± 0.0^a	3.0 ± 0.1
189	0.1 ± 0.0^a	2.8 ± 0.1
192	0.1 ± 0.0^a	3.4 ± 0.1
211	1.6 ± 0.4	2.4 ± 0.0^a
228	0.6 ± 0.1	2.0 ± 0.0^a
231	1.6 ± 0.3	2.2 ± 0.0^a
257	2.7 ± 0.5	2.8 ± 0.1
264	0.3 ± 0.0^a	2.1 ± 0.0^a

[a] Error less than 0.03.

The binding results from the SRCD (Table 4) are typically an order of magnitude lower compared with those determined from the CD binding data using the least squares method. Whereas previously determined CD binding constants were reported at a given wavelength (230 nm), a range of SRCD data values satisfied the mathematical fit of eqn (1), thus the average across all fitted wavelengths is reported. Despite the limitations of the titration concentrations the program was able to assess the binding affinity and number of binding sites for **1–6** is comparable to previously reported values.

The small shift in wavelength at peak maxima, which is apparent across all SRCD experiments, is attributable to DNA

Table 4 Summary of DNA binding affinities of the Pt(II) complexes determined by conventional CD (as previously reported)[5] and SRCD (using eqn (1))

	Compound	CD[a]			SRCD[b]	
		Nm	$K \times 10^6$	N	$K \times 10^5$	n
1	$[Pt(en)(phen)]Cl_2 \cdot 2H_2O$	230	2.2 ± 0.5	2.0 ± 0.3	0.9 ± 1.0	2.6 ± 0.5*
2	$[Pt(en)(4\text{-Mephen})]Cl_2$	230	1.2 ± 0.4	5.8 ± 1.1	1.8 ± 0.8	2.7 ± 0.6
3	$[Pt(en)(5\text{-Mephen})]Cl_2$	230	0.7 ± 0.2	3.8 ± 0.7	2.3 ± 1.0	3.6 ± 2.5
4	$[Pt(en)(4,7\text{-Me}_2\text{phen})]Cl_2$	230	1.3 ± 0.4	4.6 ± 0.1	0.3 ± 0.1	2.3 ± 0.3
5	$[Pt(en)(5,6\text{-Me}_2\text{phen})]Cl_2$	230	1.5 ± 1.1	4.9 ± 0.9	0.6 ± 0.4	3.0 ± 0.8
6	$[Pt(en)(3,4,7,8\text{-Me}_4\text{phen})]Cl_2$	230	0.7 ± 0.3	6.0 ± 0.4	1.0 ± 0.9	6.1 ± 2.2

[a] The intrinsic approach and Scatchard model[24] was used to determine K and n from the ct-DNA titration data. [b] Average of values across wavelengths determined using the *Mathematica* notebook implementation.

conformational changes induced by the MCs. The intercalating nature of the phenanthroline ligand causes elongation of the DNA and is especially evident with **1–5**. In previous methods these subtle shifts are disregarded and a wavelength value is chosen intuitively; this is avoided in our implementation as the fit may be at any wavelength provided that it is a mathematically sound. This is symptomatic of the difficulty with determining binding constants and is not always apparent by using other methods.

The shoulder appearing at higher concentrations, seen in Fig. 7 at ~180 nm, is evident in all but the spectra of **6**. This suggests an additional mode of binding other than intercalation. The binding of **6** may be by groove binding alone, the tetramethylated phenanthroline being too large to insert between the strands. This is also evidenced by the overall disparity in the titration spectra between **6** and the other MCs.

Obtaining ideal SRCD spectra

Collecting data on CD instruments below 190 nm is limited by a number of experimental conditions. The power of the synchrotron beam can be several orders of magnitude brighter than a Xenon lamp and does not experience a comparable flux decrease across the UV and vacuum UV wavelength ranges. Any reduction in light source intensity is amplified by absorptive losses within the quartz optics, atmospheric oxygen in the light path (despite nitrogen purging) as well as sample and buffer absorption.[32] These factors combine to limit the collection of meaningful data at wavelengths below 190 nm. Oxygen is excluded from the light path of SRCD as it is under vacuum, excepting the sample chamber which is constantly purged with dry nitrogen. Measurements in aqueous solutions can be extended down to 168 nm by short path length quartz sample cells (typically 0.01 cm) and the judicious choice of buffer.[32,33] The reduced sample volumes, accommodated by the improved signal-to-noise, allow samples to be measured even in absorbing buffers, lipids and detergents.

The shortest wavelength at which a spectrum can accurately be measured (cut-off wavelength) is influenced by salts, particularly chloride, and buffer composition (Table 5). If the components of the solution (*i.e.* buffers and salts) absorb

Table 5 Absorption properties of selected buffer and salts in the far UV[a]

Component/s	Light intensity cut-off/nm	Absorbance (10 mM solution, 0.1 cm path length cell)			
		180 nm	190 nm	200 nm	210 nm
NaF	170	>0.01	>0.01	>0.01	>0.01
KF	170	>0.01	>0.01	>0.01	>0.01
$NaClO_4$	170	>0.01	>0.01	>0.01	>0.01
Boric acid	180	>0.01	>0.01	>0.01	>0.01
NaH_2PO_4	195	0.15	0.01	>0.01	>0.01
Na borate (pH 9.1)	200	0.3	0.09	>0.01	>0.01
NaCl	205	>0.5	>0.5	0.02	>0.01
Na_2HPO_4	210	>0.5	0.3	0.05	>0.01
Cacodylate (pH 6.0)	210	>0.5	0.22	0.20	0.01
Na acetate	220	>0.5	>0.5	0.17	0.03
Tris/H_2SO_4 (pH 8.0)	220	>0.5	0.24	0.13	0.02
HEPES/Na^+ (pH 7.5)	230	>0.5	>0.5	0.5	0.37
MES/Na^+ (pH 6.0)	230	>0.5	0.29	0.29	0.07
NaOH	230	>2	>2	>2	0.5
MOPS (pH 7.0)	230	>0.5	0.28	0.34	0.10
Tricine (pH 8.5)	230	>0.5	>0.5	0.44	0.22
EDTA*	230	>0.5	>0.5	0.42	0.20

[a] Data are adapted from Kelly,[32] Schmid.[35] Additional data can be found in Rosenheck and Doty[36] and Buck *et al.*[37] *Data from this work.

strongly in the region of measurement then it should be substituted with a non-absorbing equivalent. The worst informal offender is chloride anion, so sodium chloride should be replaced where possible by either sodium fluoride or sodium sulfate. Buffers such as tris(hydroxymethyl)aminomethane (Tris) can be acidified with phosphoric or sulfuric acid in place of hydrochloric acid. Low pH buffers with carbonyl groups (citrate, acetate and glycine) and neutral buffers such as HEPES and tricine all absorb strongly, however low concentrations and short path length cells can minimise these effects on spectra.[32,34] Phosphate buffer (3.6 mM) replaced the Tris buffer that had been used in the previous experiment. Concentrations of up to 100 mM in path lengths of less than 50 μm can be used for a phosphate buffer.

While the concentrations of DNA and MC used in SRCD experiments is higher than comparable CD experiments, they are still relatively low (<∼1 mM) and the impact is minimal as the obtained spectra are equivalent. This is somewhat expected as the relative concentrations are equivalent and the viscosity of the solutions still close to that of the solvent.

Conclusions

The DNA binding affinity of a range of metal complexes was reinvestigated using SRCD and a new method was implemented for determining the binding constants. The implementation accommodates quick and effective determination of the binding constant across an entire spectrum, and importantly the ability to observe the potential for error in determining binding constants. The results compared favourably to previous methods but also highlight the potential shortcoming of reporting binding constants at particular wavelengths. Binding constants cannot, by definition, be wavelength dependent and should be constant. We have provided the means to investigate the entire spectrum and in doing so observed that the choice of wavelength makes a difference. Ct-DNA has many binding sites and when one is used up others will be occupied with different induced CD signals. It is only when the shape of the induced CD spectrum is the same at the mixing ratios that the data can used to determining binding constants. It is likely that the spectrum is complicated by signals from other binding interactions.

The *Mathematica* notebook implementation allows immediate processing and feedback of experimental data with the inclusion of an interactive spectrum navigator function and tabulated results. Implemented into an experimental technique this automated process has the advantage of avoiding insufficient or excessive data collection as well as identification of which wavelengths the binding constant can be determined with sufficient accuracy. Future refinements to the data processing procedure are planned including the ability to titrate DNA into a solution of MC and using the integral of the curve. The notebook is available upon request.

Acknowledgements

Our SRCD studies have been supported by grants from the Australian Synchrotron and RSC to JRAW. ISA, at Aarhus University for awarding beam time on the CD1 beam line, ASTRID. We thank members of the Aldrich-Wright lab group (UWS) for helpful discussions and Dr Søren Vrønning Hoffman (ISA) for their help and advice at the SRCD beamline. Dale Ang was supported by an Honours scholarship and research grant (no. 20752.64265).

Notes and references

1 P. A. Snyder and E. M. Rowe, *Nucl. Instrum. Methods*, 1980, **172**, 345–349.
2 J. C. Sutherland, E. J. Desmond and P. Z. Takacs, *Nucl. Instrum. Methods*, 1980, **172**, 195–199.
3 B. Wallace, *J. Synchrotron Radiat.*, 2000, **7**, 289–295.
4 B. A. Wallace and R. W. Janes, *Curr. Opin. Chem. Biol.*, 2001, **5**, 567–571.
5 C. R. Brodie, J. G. Collins and J. R. Aldrich-Wright, *Dalton Trans.*, 2004, 1145–1152, DOI: 10.1039/b316511f.
6 J. B. Chaires, in *Methods in Enzymology*, ed. M. J. W. Jonathan and B. Chaires, Academic Press, 2001, vol. 340, pp. 3–22.
7 M. A. Ismail, K. J. Sanders, G. C. Fennell, H. C. Latham, P. Wormell and A. Rodger, *Biopolymers*, 1998, **46**, 127–143.
8 X. Qu and J. B. Chaires, in *Methods in Enzymology*, ed. L. B. Michael and L. Johnson, Academic Press, 2000, vol. 321, pp. 353–369.
9 I. Haq, P. Lincoln, D. Suh, B. Norden, B. Z. Chowdhry and J. B. Chaires, *J. Am. Chem. Soc.*, 1995, **117**, 4788–4796.
10 T. D. Pollard, *Mol. Biol. Cell*, 2010, **21**, 4061–4067.
11 H. Ihmels and L. Thomas, *Org. Biomol. Chem.*, 2013, **11**, 480–487.
12 A. Granzhan, H. Ihmels and K. Jager, *Chem. Commun.*, 2009, 1249–1251, DOI: 10.1039/B812891J.
13 K. Benner, H. Ihmels, S. Kolsch and P. M. Pithan, *Org. Biomol. Chem.*, 2014, **12**, 1725–1734.
14 C. Ciatto, M. L. D'Amico, G. Natile, F. Secco and M. Venturini, *Biophys. J.*, 1999, **77**, 2717–2724.
15 G. Scatchard, *Ann. N. Y. Acad. Sci.*, 1949, **51**, 660–672.
16 D. E. Schmechel and D. M. Crothers, *Biopolymers*, 1971, **10**, 465–480.
17 J. D. McGhee and P. H. von Hippel, *J. Mol. Biol.*, 1974, **86**, 469–489.
18 B. Nordén and F. Tjerneld, *Biophys. Chem.*, 1976, **4**, 191–198.
19 P. E. Schipper, B. Nordén and F. Tjerneld, *Chem. Phys. Lett.*, 1980, **70**, 17–21.
20 A. Wolfe, G. H. Shimer and T. Meehan, *Biochemistry*, 1987, **26**, 6392–6396.
21 A. McCoubrey, H. C. Latham, P. R. Cook, A. Rodger and G. Lowe, *FEBS Lett.*, 1996, **380**, 73–78.
22 R. Rodger, *Methods Enzymol.*, 1992, **226**, 232.
23 A. Rodger, I. S. Blagbrough, G. Adlam and M. L. Carpenter, *Biopolymers*, 1994, **34**, 1583–1593.

24 A. Rodger and B. Nordén, *Circular dichroism and linear dichroism*, Oxford University Press, Oxford, 1997.
25 C. V. Kumar and E. H. Asuncion, *J. Am. Chem. Soc.*, 1993, **115**, 8547–8553.
26 C. V. Kumar, E. H. Asuncion and G. Rosenthall, *Microporous Mater.*, 1993, **1**, 299–308.
27 C. V. Kumar, R. S. Turner and E. H. Asuncion, *J. Photochem. Photobiol., A*, 1993, **74**, 231–238.
28 F. H. Stootman, D. M. Fisher, A. Rodger and J. R. Aldrich-Wright, *Analyst*, 2006, **131**, 1145–1151.
29 J. G. Lees, B. R. Smith, F. Wien, A. J. Miles and B. A. Wallace, *Anal. Biochem.*, 2004, **332**, 285–289.
30 A. J. Miles, F. Wien and B. A. Wallace, *Anal. Biochem.*, 2004, **335**, 338–339.
31 X. Qu, J. O. Trent, I. Fokt, W. Priebe and J. B. Chaires, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 12032–12037.
32 S. M. Kelly, T. J. Jess and N. C. Price, *Biochim. Biophys. Acta, Proteins Proteomics*, 2005, **1751**, 119–139.
33 A. J. Miles and B. A. Wallace, *Chem. Soc. Rev.*, 2006, **35**, 39–51.
34 F. Wien and B. A. Wallace, *Appl. Spectrosc.*, 2005, **59**, 1109–1113.
35 F. X. Schmid, in *Protein Structure: a Practical Approach*, Oxford Univiersity Press, Oxford, 1989, pp. 251–285.
36 K. Rosenheck and P. Doty, *Proc. Natl. Acad. Sci. U. S. A.*, 1961, **47**, 1775–1785.
37 R. P. Buck, S. Singhadeja and L. B. Rogers, *Anal. Chem.*, 1954, **26**, 1240–1242.

Chapter 2

Quadruplex DNA-stabilising dinuclear platinum(II) terpyridine complexes with flexible linkers

With an effective method to assess binding affinity developed and published, there was now a need to assess the double-stranded DNA (dsDNA) binding affinity of a series of platinum(II)-terpyridine complexes I had synthesised in my Honours year. These were expected to have significantly higher binding affinity and as a consequence improved biological activity over the platinum(II)-phenanthroline complexes studied in Chapter 1. However, they proved to bind almost quantitative to dsDNA through charge neutralisation with no apparent specificity, resulting in precipitation of the DNA. I therefore hypothesised that these planar molecules may preferentially stack with G-quadruplex DNA (QDNA) and stabilise the structure. Preliminary CD melt tests revealed this to indeed be the case, so I endeavoured to explore this phenomenon using various experimental and computational approaches.

In this work the stabilisation effects of four dinuclear platinum(II)-terpyridine complexes on QDNA and dsDNA was assessed. Techniques including synchrotron radiation circular dichroism (SRCD), fluorescent intercalator displacement (FID),

and fluorescence resonance energy transfer (FRET) melt studies. The dinuclear platinum(II)-terpyridine complexes, comprising two platinum(II) centres linked by a flexible central ligand, were found to be strongly stabilising, and highly selective for QDNA. The SRCD melting experiments showed up to 17°C stabilisation of QDNA melting temperature even in the presence of up to 600 fold excess of dsDNA.

As the primary author, my contribution involved synthesis of the complexes, data acquisition at the ISA synchrotron in Denmark, QDNA melting data analysis and processing, DFT and docking simulations, all artwork and graphics and writing a significant component of the text. This work was published in Chemistry–A European Journal and has been cited 22 times (as of November 2019). Importantly, this work was considered a "Hot Paper" for impact and a back cover to the journal, inspired by the work, was also published. The cover art can be found in Chapter 5.

Paper reference

Ang, D. L.; Harper, B. W., Cubo L.; Mendoza O.; Vilar R.; Aldrich-Wright J.: Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers. Chemistry – A European Journal, 2016, 22, 2317–2325. DOI: 10.1002/chem.201503663. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.[1]

Back cover reference

Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers (Chem. Eur. J. 7/2016, page 2540).

ChemPubSoc Europe
DOI: 10.1002/chem.201503663
CHEMISTRY A European Journal
Full Paper

Quadruplex DNA |*Hot Paper*|

Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers

Dale L. Ang,[a] Benjamin W. J. Harper,[a] Leticia Cubo,[b] Oscar Mendoza,[b] Ramon Vilar,*[b, c] and Janice Aldrich-Wright*[a]

Abstract: Four dinuclear terpyridineplatinum(II) (Pt–terpy) complexes were investigated for interactions with G-quadruplex DNA (QDNA) and duplex DNA (dsDNA) by synchrotron radiation circular dichroism (SRCD), fluorescent intercalator displacement (FID) assays and fluorescence resonance energy transfer (FRET) melting studies. Additionally, computational docking studies were undertaken to provide insight into potential binding modes for these complexes. The complexes demonstrated the ability to increase the melting temperature of various QDNA motifs by up to 17 °C and maintain this in up to a 600-fold excess of dsDNA. This study demonstrates that dinuclear Pt–terpy complexes stabilise QDNA and have a high degree of selectivity for QDNA over dsDNA.

Introduction

Guanine-rich sequences of DNA have the ability to self-assemble into non-canonical motifs such as G-quadruplexes (QDNA; Figure 1). Bioinformatic studies have shown that over 360 000 DNA sequences within the human genome have the potential to form QDNA[1] and in vitro experimental evidence supports their formation in mammalian cells.[2] These quadruplex-forming sequences are not randomly distributed, rather they are over-represented in gene promoter regions, suggesting a causative link with transcription.[3] G-quadruplexes have also been shown to assemble from the guanine-rich telomeric overhang sequences. Studies have demonstrated that formation of stable quadruplexes in telomeric regions prevents telomerase (over expressed in greater than 85% of cancer cells and plays an important role in cancer cell immortalisation) from extending the telomere, resulting in growth inhibitory effects.[4] Consequently there is growing interest in the development of small molecules that can bind and stabilise QDNA motifs, and in doing so, prevent unfolding of the secondary structure, thereby interfering with telomerase activity.

Structural and spectroscopic evidence indicating that large planar molecules interact with QDNA has been accumulated in the last decade.[6] As has been discussed elsewhere,[7] planar metal complexes are an attractive subclass of potential QDNA binders, which exhibit favourable electrostatic and π-stacking interactions with good solubility. In the mid-2000s, square planar nickel(II)–salphen,[8] platinum(II)–terpyridine[9] and manganese–porphyrin[10] complexes were reported to have high affinities towards QDNA. Since then, a significant number of metal complexes have also been reported as effective binders[7] with most reported examples being of monometallic complexes that form π–π stacking interactions with QDNA's guanine quartets. Although this has proven to be a successful strategy, there is a push to develop metal complexes with higher affinities and selectivity towards QDNA in comparison with other topologies, such as duplex DNA (dsDNA). One approach has been to develop polymetallic systems that combine the various modes of interaction into a single molecule by incorporation of a second metal centre. One such example is the combination of a known dsDNA intercalator, a platinum(II)–terpyridine moiety, with a dipicolyl- or cyclen-bound copper(II) or zinc(II) to produce non-symmetric dimetallic complexes.[11] These complexes were recently reported to possess greatly enhanced selectivity towards QDNA, with up to 100-fold preference over dsDNA.[11b] Several of these mixed dinuclear Cu^{II}, Pt^{II} and Zn^{II} complexes showed cytotoxicities similar to that of cisplatin and reinforced that the effort of combining properties of various metals can be advantageous. The aromatic and amine content of these compounds also substantiated and reinforced the importance of electrostatic and π-stacking interactions for selectively targeting QDNA.

Other polymetallic systems have been reported to demonstrate improved characteristics over their mononuclear counterparts. The Pt^{II} compound BBR3464, a cisplatin-like trinuclear compound,[12] is a prime example of this effort. In preclinical trials it proved more potent than cisplatin and effective against cisplatin-resistant cell lines. Recent phase II clinical trials of

[a] D. L. Ang, Dr. B. W. J. Harper, Prof. J. Aldrich-Wright
Nanoscale Organisation and Dynamics Group, School of Science and Health, Western Sydney University, Penrith, NSW 2751 (Australia)
E-mail: J.Aldrich-Wright@uws.edu.au

[b] Dr. L. Cubo, Dr. O. Mendoza, Prof. R. Vilar
Department of Chemistry, Imperial College London
London SW7 2AZ (UK)

[c] Prof. R. Vilar
Institute of Chemical Biology, Imperial College London
London SW7 2AZ (UK)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201503663.

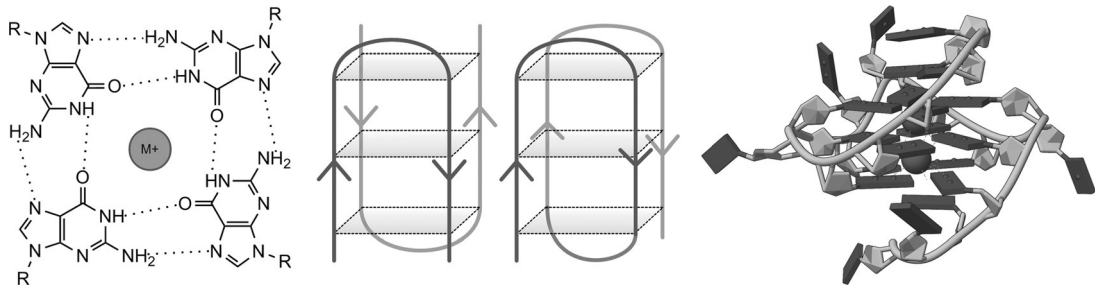


Figure 1. Left) QDNA tetrad of four Hoogsteen-bound guanine bases stabilised by a central cation; middle) examples of a two-strand and one-strand quadruplex structure; right) QDNA structure of human *c-Myc* promoter (PDB ID: 1XAV).[5]

BBR3464 have substantiated this cytotoxicity profile, the high toxicity due to the formation of DNA adducts promoted by the high flexibility of the molecule. A recent effort to modulate the cytotoxicity of BBR3464-like compounds focused on creating sterically hindered dinuclear complexes.[13] A significant reduction in reaction rate, and cytotoxicity in A2780 and A2780R cell lines, resulted from the sterically hindered equivalents, however resistance was not overcome.

Dinuclear Ru complexes have also exhibited significantly improved cytotoxicities in comparison to their mononuclear counterparts.[14] With similarities to BBR3464, Ru^{II}–arene compounds linked by bis(pyridinone) ligands of varying length have reportedly varied in their lipophilicities, influenced by the linker length with IC_{50} values similar to, and no cross-resistance with, oxoplatin. Dinuclear ruthenium complexes with pronounced QDNA binding affinities recently reported include the chiral ΔΔ- and ΛΛ-$[\{Ru(phen)_2\}_2(tpphz)]^{4+}$ (tpphz = tetrapyrido[3,2-*a*:2′,3′-*c*:3″,2″-*h*:2‴,3‴-*j*]phenazine).[15] These complexes have been shown to stack with the ends of QDNA with very high affinity and, in doing so, exhibit strong luminescence. The ΛΛ isomer is reported to bind approximately 40 times more tightly than the ΔΔ isomer, suggesting a strong role in the chirality and hence enantioselectivity with QDNA.

When studying the interactions of metal complexes with DNA, a range of spectroscopic techniques may be applied; simple UV/Vis titrations, fluorescent intercalator displacement (FID) assays, fluorescence resonance energy transfer (FRET) and circular dichroism (CD) melts are common. Although the results of these studies are highly valuable, recent developments in both techniques and instrumentation involving synchrotron radiation circular dichroism (SRCD) provides additional information for the analysis of macromolecular interactions, including secondary structural elucidation.[16] Although SRCD has been known for some time, it has only more recently been utilised for biological analysis, with instrumental refinement and calibration enabling improved reproducibility.[17] Advantages of SRCD over conventional CD for structural elucidation of biomolecules include the collection of additional data at shorter wavelengths. The radiance of synchrotron radiation is many orders of magnitude greater than that generated by xenon arc lamps[18] with the resulting spectra providing additional electronic transitions facilitating the differentiation of more secondary motifs and structures, which has proven particularly useful for protein structural analysis.[19] Other advantages include the acquisition of spectra with improved signal-to-noise ratios, and reduced sample size requirements.[20]

Herein we report the binding studies of four dinuclear Pt–terpy metal complexes **1–4** (Figure 2) with QDNA and dsDNA. Complexes **1–3** are dinuclear terpyridine-Pt^{II} joined by flexible thiol bridging ligands of increasing length, while complex **4** is joined by the flexible 4,4′-trimethylene-dipyridine ligand.[21] SRCD melting experiments of the various QDNA motifs were undertaken in order to determine relative binding affinity and secondary structure stabilisation effects induced by the complexes. The binding interactions of **2–4** were determined via FID assays and FRET melting studies. Docking simulations were also undertaken to provide insight into possible modes of action. These complexes exhibited high affinity and good selectivity for QDNA (human telomeric and *c-Myc*) over dsDNA.

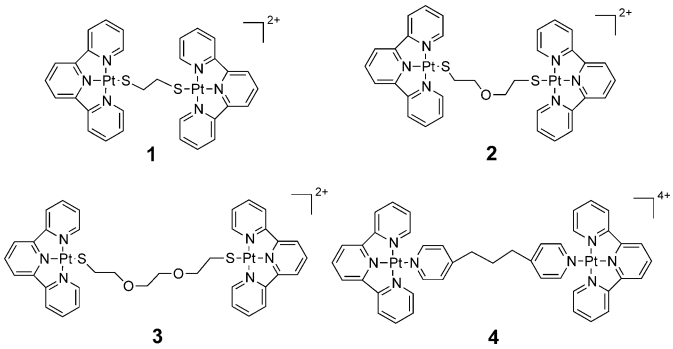


Figure 2. Structure of the Pt^{II}–terpy complexes used in this study. **1–3** vary by the dithiol linker lengths, whereas **4** includes a flexible propyl–dipyridyl linker.[22] Counter ions omitted for clarity.

Experimental Section

General procedures

All metal complexes were prepared as previously reported.[22] Buffers were prepared from standard lab reagents. The SRCD experiments were performed on three quadruplex-forming oligonucleotides. The lyophilised oligos were purchased from DNA Technology (Denmark): bimolecular **Q1**, 5′-$(G_4T_4G_3)_2$-3′, 22mer HTelo **Q2**, 5′-$(AG_3(T_2AG_3)_3)$-3′, and 22mer human oncogene promoter *c-myc* **Q3**, 5′-$(TGAG_3TG_3TAG_3TG_3TA_2)$-3′. These were annealed according to the following protocol: the oligonucleotide was dissolved in potassium phosphate buffer to approximately 1 mM, based on the manufacturer reported quantity, and heated at 95 °C for 15 min in a water

bath. The solution was removed from the bath and allowed to equilibrate to room temperature overnight. The stock solutions were stored at 4 °C before use. The potassium phosphate buffers used were KPB1 (10 mM, pH 7.0, 100 mM KF) for **Q1** and KPB2 (25 mM, pH 7.0, 70 mM KF) for **Q2** and **Q3**.

SRCD QDNA melts

Synchrotron experiments were performed at the Centre for Storage Ring Facilities, Aarhus, Denmark (ASTRID and ASTRID2). QDNA melting experiments were recorded from 5–85 °C in 5 °C intervals with an equilibration time of 5 min at each temperature. Data as averaged triplicate scans were collected from 170–350 nm in 1 nm steps. Samples were loaded into a circular sealed 0.1 mm path-length quartz cuvette. Determination of QDNA melting temperature (T_m) was achieved by plotting the peak intensity versus temperature to give a sigmoidal curve. A stock solution of metal complex (MC) in the same buffer as the oligo was added to the QDNA solution to achieve a 1:1 MC/QDNA ratio. The QDNA concentration was estimated from the HT of the SRCD scan at 25 °C using the ε_{260} values derived by applying the Tataurov, You and Owczarzy method.[23] The ε_{260} = 105 100, 228 500 and 228 700 $M^{-1}cm^{-1}$ for **Q1**, **Q2** and **Q3**, respectively. The final volume was adjusted to 120 μL with buffer, resulting in final concentrations of approximately 1×10^{-4} M.

Oligonucleotides used for FID and FRET experiments

Stock solutions were prepared and used in the FID and FRET DNA binding studies. The corresponding complex (**2–4**) was dissolved in DMSO to give a 1 mM stock solution. All solutions were stored at −20 °C, defrosted and diluted immediately before use. Non-labelled oligonucleotides and fluorophore labelled oligonucleotides of double HPLC-grade purity were purchased from Eurogentec (Belgium). For the QDNA, the 21mer double-labelled oligonucleotide **F21T-HTelo**, 5′-FAM-d($G_3(T_2AG_3)_3$)-TAMRA-3′ (donor fluorophore FAM, 6-carboxyfluorescein; acceptor fluorophore TAMRA, 6-carboxytetramethylrhodamine) was used for FRET studies and the 22mer **FID-HTelo** 5′-($AG_3(T_2AG_3)_3$)-3′ (identical sequence to **Q2**) for FID studies. For *c-Myc* DNA, 22mer double-labelled pu22 *myc* oligonucleotide **F-*myc*-T**, 5′-FAM-d($GAG_3TG_4AG_3TG_4A_2G$)-TAMRA-3′ was used for FRET studies and the 21mer **FID-*myc*** 5′-$G_4AG_3TG_4AG_3TG_3$-3′ for FID studies. For dsDNA, the **ds26** strand 5′-CAATCGGATCGAATTCGATCCGATTG-3′ was used for FID studies and the labelled strand **F-DS1-T**, 5′-FAM-d(TATAGCTATA-T_6-TATAGCTATA)-TAMRA-3′ for FRET studies. CT-DNA (sodium salt, Sigma-Aldrich) was used for the FRET competition assay.

FID measurements

Stock solutions (20 μM) were obtained by dissolving the corresponding oligonucleotide in MilliQ water, annealing to a 0.5 μM concentration in potassium cacodylate buffer (60 mM, pH 7.4) at 95 °C for 5 min and allowing to cool slowly to room temperature overnight. Thiazole Orange (TO; Sigma-Aldrich) was freshly prepared as a stock solution (10 mM) in DMSO. Metal complexes and TO were made up from stock solutions, to the final concentration in a potassium cacodylate buffer (60 mM, pH 7.4). Measurements were performed at a final 0.25 μM concentration of oligo, 0.5 μM concentration of TO, with varied metal complex concentrations. Samples were left to equilibrate for 3 min. Data was acquired at room temperature over a range of λ = 510–750 nm with excitation at λ = 501 nm.

FRET measurements

The labelled oligonucleotides **F21T-HTelo**, **F-*myc*-T** and **F-DS1-T** (see above) were chosen for FRET evaluation. Two different buffers were used for this assay: (a) a Caco.K90 buffer comprising lithium cacodylate (10 mM, pH 7.4), KCl (10 mM) and LiCl (90 mM) for **F21T-HTelo** and **F-DS1-T** oligos and (b) a Caco.K99 buffer comprising lithium cacodylate (10 mM, pH 7.4), KCl (1 mM) and LiCl (99 mM) for **F-*myc*-T** oligo. The labelled oligos were dissolved as a stock 20 μM solution in MilliQ water and then annealed at a 400 nM concentration in cacodylate buffer at 95 °C for 10 min and allowed to cool slowly to room temperature overnight. Metal complexes were dissolved from stock solutions to final concentrations in the corresponding cacodylate buffer. Measurements were performed on a RT-qPCR instrument (Agilent) with excitation at λ = 450–495 nm and detection at λ = 515–545 nm. Each well of a 96-well plate contained 200 nM oligo and metal complex to test. Fluorescence readings were taken at intervals of 0.7 °C between 25–95 °C. The melting temperatures of **F-DS1-T**, **F-*myc*-T** and **F21T-HTelo** were monitored in the presence of different concentration of the metal complexes from 0.2–4.0 μM. Final analysis of the data was carried out using OriginPro 8.5.

FRET competition assay

F21T-HTelo oligo was dissolved as a stock solution (20 μM) in MilliQ water, then annealed as a 400 nM concentration in Caco.K buffer [lithium cacodylate (10 mM, pH 7.4) and KCl/LiCl (10 mM/90 mM)] at 95 °C for 10 min, and allowed to cool slowly to room temperature overnight. A solution of CT-DNA (1.32 mg mL^{-1}) in water was stirred gently for 3 days and then centrifuged for 3 min (1200 rpm) to remove insoluble oligo. The concentration and purity of the supernatant was determined by absorbance reading on a CARY 300 BIO UV/Vis spectrophotometer and then diluted to obtain the final concentrations. The corresponding metal complexes were diluted from stock solutions to final concentrations in cacodylate buffer. Each well of a 96-well plate was prepared with a final 200 nM oligo concentration, 1 μM compound concentration, and the CT-DNA concentration to test. Measurements were performed under the same conditions as the FRET assay (see above). For **F-*myc*-T** a 20 μM stock solution in MilliQ water was diluted with Caco.K99 buffer.

DFT and docking Simulations

DFT Geometry optimisations of **1–4** were performed with Gaussian 09[24] in the gas phase at the B3LYP level of theory. The LANL2DZ basis set and effective core potentials were used for the Pt atoms; the 6-31G(d,p) basis set was used for all other atoms. The structures were created in Avogadro Version 1.1.1.[25] Preliminary geometry optimisation was performed by using the Universal Force Field (UFF).[26]

Docking simulations were performed by using SMINA docking software, a fork of Autodock Vina Version 1.1.2.[27] The dsDNA, d$(CGCGAATTCGCG)_2$ (PDB ID: 1BNA), and the following NMR solution structures of representative QDNA: **Q1** d$(G_4T_4G_4)_2$ (PDB ID: 156D), **HTelo** d$(A_3G_3(T_2AG_3)_3A_2)$ (PDB ID: 2HY9), ***c-Myc*** d$(TGAG_3TG_3TAG_3TG_3TA_2)$ (PDB ID: 1XAV) were used as the static substrates. The predicted docking conformations with the most favourable binding energies were averaged to obtain the final result; in all cases SD < 0.5 kcal mol^{-1}. Further details are provided in the Supporting Information.

Results and Discussion

SRCD QDNA melts

It is well established that the annealing regimen used to prepare QDNA is particularly critical to obtaining reproducible conformations. Polymorphism has been shown to be kinetically governed and the cooling ramp is critical in obtaining reproducible structures.[28] The key factor in preparation of **Q1** was allowing the sample to rest overnight; therefore progressing towards the more thermodynamically favoured topologies. This change was evidenced by changes in the relative intensities of the peaks at $\lambda \approx 290$ and 263 nm, which stabilised after storage to give the spectra shown in Figure 3. The appearance of a positive maxima at 290 and 263 nm is indicative of antiparallel and parallel topologies, respectively. The alternating *syn*- and *anti*-guanine conformations are suggested by the peak at $\lambda = 263$ nm, a positive maximum indicating a parallel and a negative maximum indicating an antiparallel tract.[29] Overall, these spectral signatures suggest that the final annealed **Q1** sample was a mixture of antiparallel, parallel and hybrid topologies. The final relative peak heights of the positive peaks at 290 and 263 nm correspond with the expected mixture of parallel, antiparallel and hybrid topologies for this sequence under these annealing conditions.[28]

The extended wavelength range of SRCD revealed several additional peaks compared with conventional benchtop CD machine; five unique peaks were observed for **Q1** (Figure 3, left). The SRCD spectra show a reduction in intensity across all peaks as the temperature is increased. The intensity changes of these CD peaks were plotted against temperature to give the melting curves shown in Figure 3 (right). The results revealed somewhat varying T_m values for each of the wavelength peaks.

Compounds **1**–**4** were each incubated with **Q1** and the melting process repeated. It was expected that as the metal complex bound to the quadruplex it would be stabilised and an associated increase in T_m would be observed. Additionally, changes in the SRCD spectra would result if the binding induced conformational changes in the quadruplex. The spectra with compounds **1** and **2** did not vary noticeably from that of the untreated **Q1** prior to the beginning of denaturation. This indicates that the metal complex interactions were likely surface interactions; the resultant interactions induced little conformation change. However, compounds **3** and **4** appeared to induce some spectral changes where the relative intensities of the 290 and 264 nm peaks varied. A reduction in the intensity of the 264 nm peak indicates a reduction in overall parallel topology, or conversely a shift towards antiparallel structure; a sign of inhomogeneity within the mixture. The isobestic points and melt curves remained consistent throughout. The presence of various topologies along with the variations observed in melting temperature for each peak highlights the importance of not simply averaging these results to obtain an overall T_m but also, perhaps more enlighteningly, provides some insight into the process of unfolding; the source of each CD peak representing a structural feature which may be associated with ligand stabilisation. The T_m and ΔT_m values are summarised in Table 1.

These results reveal there is little contribution towards stability of the positive peaks at $\lambda = 290$ or 206 nm with compounds **1** and **4** showing only a modest 2.9 ± 0.4 °C and 2.0 ± 0.3 °C increase in T_m, respectively. Interestingly, these are also the first to denature, suggesting that these spectral signatures originate from the same structural feature: the bimolecular antipar-

Table 1. T_m values for **Q1** with sign and wavelength of each peak, and ΔT_m values determined by SRCD upon addition of metal complex **1**–**4**.

QDNA/Complex	290 nm (+)	263 nm (+)	238 nm (−)	206 nm (+)	188 nm (+)
			T_m [°C]		
Q1	36.7 ± 0.1	42.2 ± 0.1	50.9 ± 1.1	38.3 ± 0.4	48.6 ± 0.6
			ΔT_m [°C]		
1	2.9 ± 0.4	1.0 ± 0.4	6.4 ± 4.8	0.2 ± 1.0	4.2 ± 1.1
2	0.3 ± 0.3	11.2 ± 1.6	2.5 ± 2.5	−0.1 ± 1.0	3.5 ± 1.1
3	−0.1 ± 0.4	2.9 ± 0.5	21.3 ± 12.0	−0.3 ± 1.0	7.1 ± 2.7
4	2.0 ± 0.3	0.6 ± 0.5	1.4 ± 2.3	0.6 ± 1.0	3.5 ± 1.1

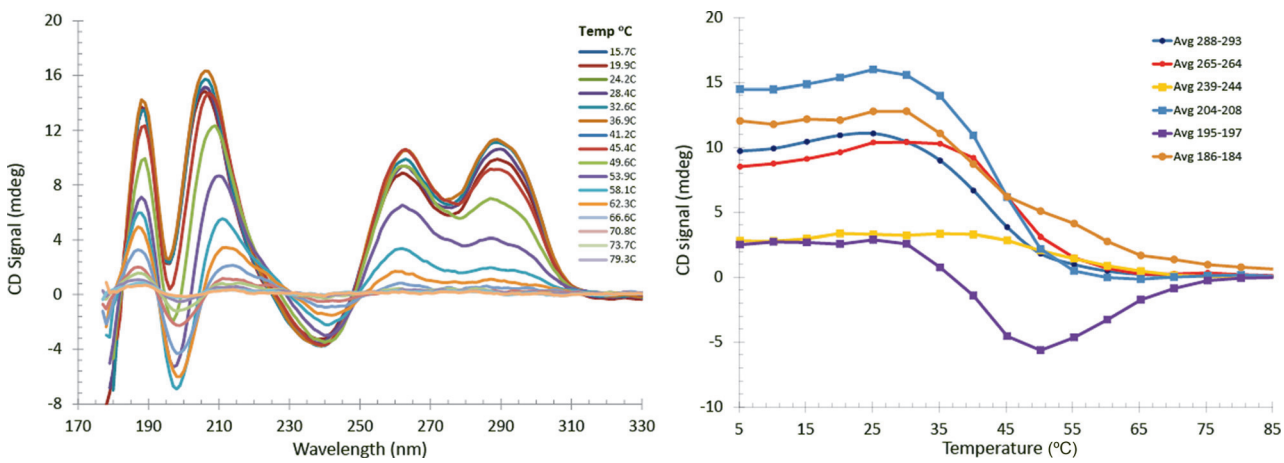


Figure 3. Left) The SRCD spectra of **Q1** at varying temperatures (average of three scans); right) the change in SRCD signal observed at each peak wavelength as a function of temperature.

allel quadruplex. This is understandable as the positive 206 nm peak may be indicative of an antiparallel quadruplex and the 290 nm peak has been linked to the alternating quartet polarity for this sequence type and also indicates antiparallel structure.[30] Complex **2** induced a change in T_m of 11.2±1.6 °C for the 263 nm peak. The large errors in the results for the negative peak at $\lambda = 238$ nm can be attributed to the low intensity of the spectrum and the evident wavelength shift upon melting; it was therefore excluded from subsequent analysis, the data being unreliable and of questionable utility. Lastly, single stranded thymine tracks are typically resolved around $\lambda \approx 190$ nm,[31] hence the large positive peak at $\lambda = 188$ nm for this sequence. The increased melt temperature across all complexes for the 188 nm peak suggests that it is a site of metal complex interactions, likely due to the exposed thymine surfaces. In summary, changes in T_m were most evident for the peaks at $\lambda = 263$ and 188 nm with **2** and **3**, respectively, inducing the greatest stabilisation. Due to limited instrument time, subsequent investigations were performed with complex **2** and **Q2** and **Q3** only. The aim was to further investigate whether similar stabilisation was evident with the more biologically relevant **Q2** and **Q3** quadruplex structures. The resulting melting spectra are shown in Figures 4 and 5 and the tabulated results are presented in Tables 2 and 3.

The 22mer **Q2** (**HTelo**) is representative of the human telomeric sequence, which consists of tandem repeating units of the single-stranded overhangs at the ends of chromosomes. This sequence has a propensity to fold into unimolecular quadruplex structures and in the presence of potassium cations has a relatively high T_m of about 50 °C. The SRCD spectrum of **Q2** shows a positive maximum peak at $\lambda = 292$ nm and a corresponding positive peak at $\lambda = 207$ nm. These both signify an antiparallel structure, as expected based on the reported NMR structure.[32] The small shoulder at 250 nm is likely representative of some parallel structure, although as indicated by the lower T_m could represent possible hybrid form resulting in an inhomogeneous mixture with varying stabilities, as reflected in the melting temperatures. Also evident are the characteristic

Table 2. T_m values for Q2 with sign and wavelength of each peak, and ΔT_m values determined by SRCD upon addition of metal complex 2.

QDNA/Complex	292 nm (+)	250 nm (+)	207 nm (+)	198 nm (−)	189 nm (+)
			T_m [°C]		
Q2	51.3±1.0	42.8±1.1	50.6±1.2	48.3±3.0	47.2±0.8
			ΔT_m [°C]		
2	1.5±1.6	12.7±3.5	5.2±1.8	10.1±3.9	3.8±1.1

Table 3. T_m values for Q3 with sign and wavelength of each peak, and ΔT_m values determined by SRCD upon addition of metal complex 2.

QDNA/Complex	265 nm (+)	244 nm (−)	207 nm (+)	185 nm (−)
		T_m [°C]		
Q3	50.0±1.2	50.1±1.3	49.2±1.0	50.6±1.4
		ΔT_m [°C]		
2	5.8±1.9	5.0±2.8	7.1±2.0	6.7±1.7

thymine single stranded loops with a positive maximum at $\lambda \approx 189$ nm. When treated with **2**, increases are seen at $\lambda = 250$ and 198 nm with ΔT_m values of 12.7±3.5 and 10.1±3.9, respectively. The 189 nm peak also undergoes a small increase in T_m, as was seen with **Q1**.

Q3 is representative of the human *c-Myc* promoter, and plays a role in promoting the transcription of the *c-Myc* oncogene.[33] Therefore is an important target for research as stabilisation of the quadruplex structure and subsequent suppression of the oncogene could lead to the development of new anticancer agents.[34] The SRCD spectra of **Q3** and of **Q3** in the presence of a 1:1 molar ratio of **2** are shown in Figure 5. The NMR structure of **Q3** is shown as an inset, comprising a unimolecular primarily parallel-stranded topology in potassium solution.[5] A strong positive peak at $\lambda = 265$ nm and negative maximum at $\lambda \approx 244$ nm are strong indicators of parallel-type quadruplexes.[29] The positive maximum at $\lambda = 207$ nm and negative

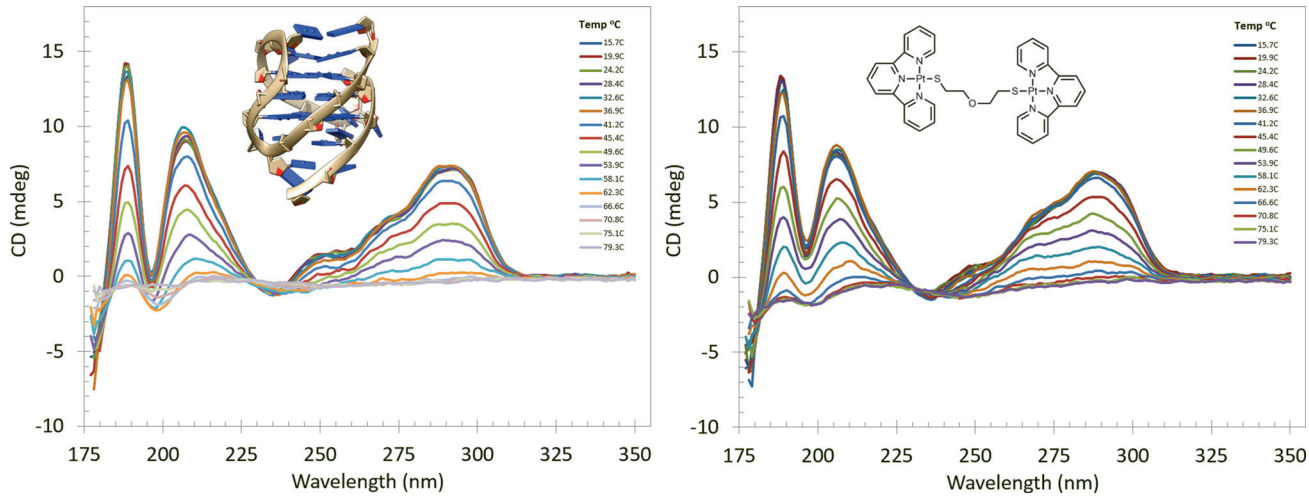


Figure 4. Left) SRCD melt spectra of **Q2** (inset: PDB ID: 2HY9); right) melt spectra of **Q2** with compound **2**.

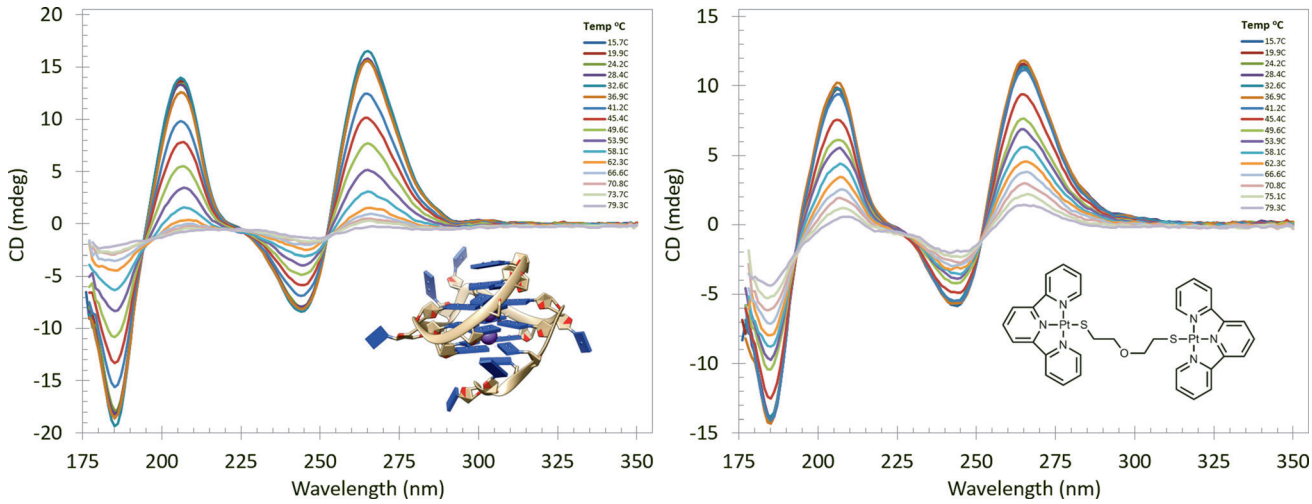


Figure 5. Left) SRCD melt spectra of **Q3** (inset: PDB ID: 1XAV); right) melt spectra of **Q3** with compound **2**.

maximum at $\lambda = 185$ nm may be due to the lack of thymine loops. Interestingly all four peaks possess melt temperatures within 1 °C suggesting a strongly topologically homogeneous solution. The lack of peak shift and clear isobestic points also confirm the homogeneity of the solution.

Upon addition of **2**, the melt temperatures all increase by between roughly 5 and 7 °C, which is a lesser change than that seen with **Q1** and **Q2** but consistent with stabilising interactions. It is important to note the heating block within the instrument was unable to heat above 80 °C. This lack of additional data resulted in a poorer fit of the data points and hence larger errors, but it can be clearly seen in the SRCD spectrum where some CD signal is present even at the highest attainable temperature.

The stabilising effects of **1**, the complex with the shortest linker, on **Q1** were modest, with a maximum observed ΔT_m of 4.2 °C. Complexes **2** and **3**, containing linkers of intermediate length and with the potential for H-bonding interactions, displayed the largest quadruplex stabilisations, the greatest value of ΔT_m being 11.2 °C for **2**. The potential for additional stacking interactions of the dipyridyl linker of **4** was not reflected in the T_m values with a maximum of ΔT_m of 3.5 °C. For complexes **1**, **3** and **4**, the largest stabilising effect was on the positive 188 nm peak, indicating interactions with the thymine loops, whereas that for **2** was with the positive 263 nm peak, corresponding to the parallel structure of the QDNA motif. These results indicate that there may be an ideal length for optimal binding with **Q1**, the increased linker lengths of **2** and **3** to both stack and groove bind to the quadruplex motif. **1** lacks the length and flexibility and thus may only stack at one end.[6c,35] These results are supported by literature reports identifying the importance of planar metal geometry for binding and nucleic acid self-assembly of Pt–terpy complexes with QDNA.[9,36]

FID Competition

Following on from the SRCD results, the binding affinities of **2**–**4** were screened by using fluorescent indicator displacement (FID) assays to determine the affinity of the MCs towards QDNA (HTelo and *c-Myc*; see Experimental Section for details of the sequences) and dsDNA. Compound **1** was not tested due to its lack of activity (as established by SRCD melting experiments), the poor solubility of the complex and its propensity for precipitating dsDNA from solution. Upon addition of **2**, **3** and **4** to the corresponding DNA structures previously incubated with thiazole orange (TO), a concomitant decrease in TO's fluorescence intensity was observed (Table 4 and Figure S2 in the Supporting Information). The degree of displacement of TO by these complexes, which in turn is a good indication of the ability of a compound to interact with a given DNA structure, was evaluated from the concentration of the compounds required to give a 50% decrease in dye fluorescence (DC_{50} values, shown in Table 4). The DC_{50} values suggest that the three complexes have high affinity (especially **2**) towards quadruplex HTelo and *c-Myc* DNA. These results reflect the SRCD outcomes showing **2** inducing the greatest ΔT_m. It is generally accepted that molecules displaying DC_{50} values equal to or less than 0.5 μM are excellent QDNA binders.

Table 4. DC_{50} values [μM] for *c-Myc*, HTelo and ds26 DNA determined by FID upon addition of complexes 2, 3, and 4.

	TO displacement (DC_{50})			Selectivity (DC_{50})	
Complex	*c-Myc*	HTelo	ds26	ds26/*c-Myc*	ds26/HTelo
2	0.16 ± 0.01	0.29 ± 0.03	0.88 ± 0.01	5.5	3.0
3	0.46 ± 0.03	0.94 ± 0.03	1.19 ± 0.10	2.6	1.3
4	0.42 ± 0.03	0.64 ± 0.10	0.95 ± 0.10	2.3	1.5

FRET melting assay

The ability of **2**, **3** and **4** to stabilise QDNA was also investigated by using FRET assays (see Figures 6 and Figures S3 and S4 in the Supporting Information, respectively). Table 5 summarises the ΔT_m values ranging between 13.4 and 17.2 °C, confirming the good affinity of **2**–**4** towards these two QDNA motifs.

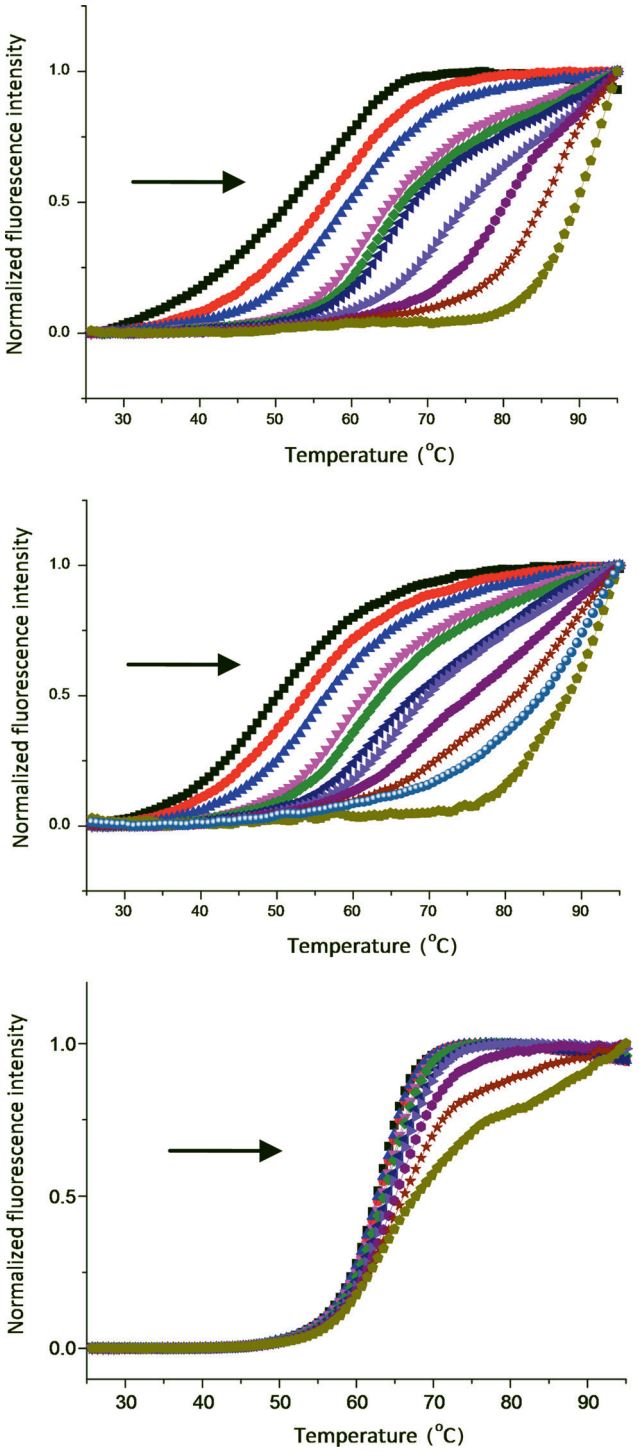


Figure 6. FRET melting results of experiments carried out with **F-*myc*-T** (top), **F21T-HTelo** (middle) and **F-DS1-T** (bottom) with increasing concentration of **4** (0–4 μM).

FRET competition assays were also carried out to study further the selectivity displayed by these metal complexes. More specifically, QDNA ΔT_m values were measured for a fixed concentration (1 μM) of added metal complex in the presence of an increasing excess of CT-DNA (0.6 μM to 120 μM). Figure 7 summarises the result of the competitive FRET assay. The data demonstrates the selectivity of these complexes as the ΔT_m values of **F-*myc*-T** and **F21T-HTelo** did not decrease upon addition of up to 300-fold excess of CT-DNA (competitor) to labelled QDNA (60 μM base pair concentration). Particularly noticeable is the ability of **4** to selectively stabilise **F-*myc*-T** even at higher concentrations of CT-DNA (120 μM, equivalent to a 600-fold excess), since the **F-*myc*-T** QDNA melting temperature does not decrease. In the case of **2** and **3**, no changes in QDNA melting temperature (for either sequence studied) were observed for up to 6 μM CT-DNA. However, addition of 120 μM CT-DNA (equivalent to 600-fold excess) decreased the melting temperatures by between 5 and 8 °C. Once again, QDNA stacking interactions dominate and are especially evident for **4** with the aromatic linker.

Table 5. ΔT_m values for F-*myc*, HTelo and ds26 DNA determined by FRET upon addition of compounds 2, 3, and 4.

Complex	ΔT_m [°C] F-*myc*-T	F21T-HTelo	F-DS1-T
2	13.5 ± 1.4	17.2 ± 1.6	1.2 ± 1.0
3	15.0 ± 1.3	17.1 ± 0.8	0.3 ± 0.1
4	14.8 ± 0.3	13.4 ± 0.2	0.8 ± 0.1

Computational studies

Analysis of the docking results revealed that **1** exhibited poor affinity for both types of DNA, self-stacking in preference to DNA interaction in all docking runs. With the dsDNA substrate (Figure 8), **2** and **3** displayed an affinity for the minor groove; the linkers of both **2** and **3** sitting deeply along the minor groove of dsDNA. This effect has been seen in previous docking studies.[37] With **4**, the major groove was preferred. In each case, the terpyridine moieties participated in an edge-to-face stacking interaction with the aromatic base residues. The linkers lay along the grooves of the DNA, whereas the more positively charged Pt centres interacted preferentially with regions of the negatively charged DNA backbone or aromatic residues.

With respect to the QDNA interactions of **2**–**4** (Figure 9), there is an apparent affinity between Pt and the central cavity, effectively taking the place of the K^+ ions; the Pt–terpy moiety is clearly end-stacking with the QDNA structure. This interaction has been suggested to have a vital effect on stabilising these QDNA.[38] The second metal centre typically lies in the groove, where there are likely electrostatic interactions between the metal centre and the widely exposed negatively charged backbone. The enhanced affinity of **4** for QDNA may also be attributed to the stacking interactions of the linker. Overall, these results suggest that the aromatic linker of **4** improved the affinity for QDNA through increased potential to π-stack with exposed bases of the QDNA structure, whereas the chain lengths of **2** and **3** are optimal for face and groove interactions. Although not reflected in the absolute predicted binding values, it is clear that the increased planar surface area and positive charge contribute to the overall affinity with QDNA. The binding energies with dsDNA and QDNA are summarised in Table 6.

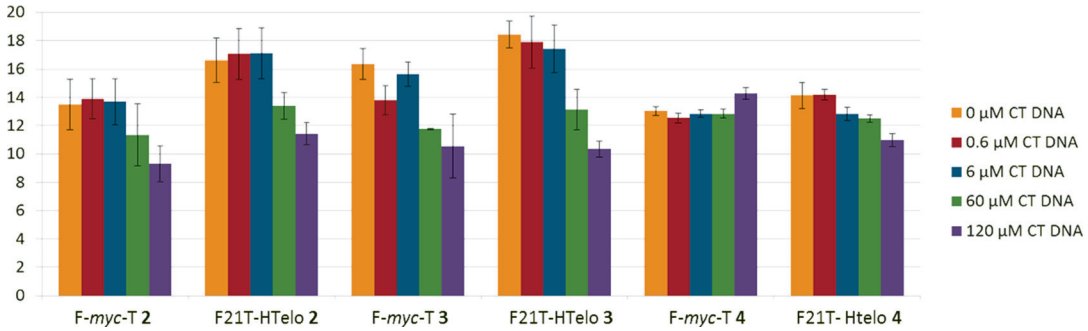


Figure 7. FRET competition experiments of selected complexes at 1 μM concentration showing the reduction of **F-*myc*** and **F21T-HTelo** DNA melting temperature with increased concentration of CT-DNA.

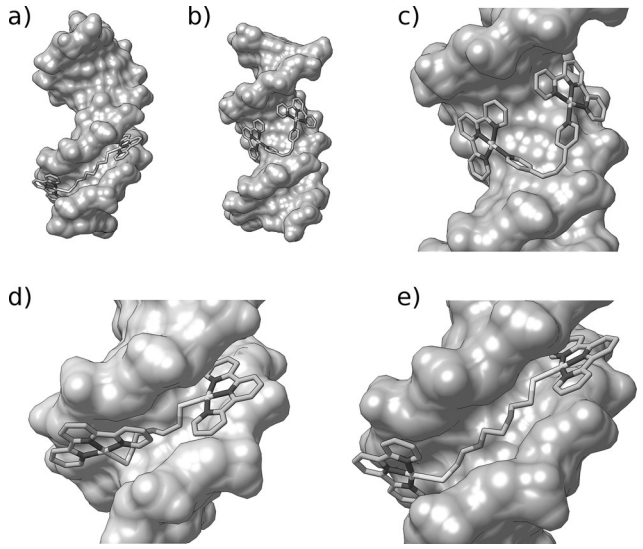


Figure 8. Predicted docking positions for MC–dsDNA interactions. Overview of minor and major groove binding: a) Complex **3**; b) complex **4**. Close up views of the interactions: c) complex **4**; d) complex **2**; e) complex **3**.

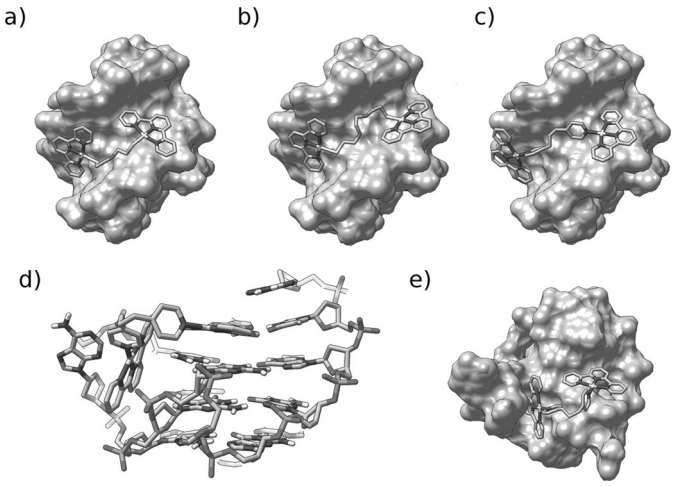


Figure 9. Predicted docking positions for MC–QDNA interactions. Overviews of the stacking interactions with **Q1**: a) Complex **2**; b) complex **3**; c) complex **4** (PDB ID: 156D). d) Side view and e) overview of interactions of **4** with *c-Myc* (PDB ID: 1XAV).

Table 6. Summary of results for docking simulations of **1–4** with substrates dsDNA and QDNA.

	Predicted Binding Energy $\Delta G \pm$ SD [kcal mol^{-1}]			
	dsDNA		QDNA	
Complex	12mer	**Q1**	**Q2** (HTelo)	**Q3** (*c-Myc*)
1	–	–	-6.4 ± 0.1	-6.9 ± 0.0
2	-5.7 ± 0.0	-8.3 ± 0.1	-6.7 ± 0.1	-6.2 ± 0.1
3	-8.3 ± 0.1	-7.9 ± 0.1	-6.3 ± 0.1	-7.3 ± 0.4
4	-9.3 ± 0.0	-10.0 ± 0.0	-7.6 ± 0.2	-9.1 ± 0.0

Conclusion

The process of identifying small molecules that bind to and stabilise biologically relevant quadruplex structures is an area of intense interest; with this study we have demonstrated that dinuclear Pt–terpy complexes exhibit such properties. They possess the ability to bind and stabilise QDNA structures, increasing the melting temperatures by up to about 13 °C, while not significantly affecting the conformation, as evidenced through SRCD experiments. The SOS-linked **2** was shown to favour the parallel topologies of **Q1** and **Q2** as evidenced by the large increase in T_m: 11.2 °C for the positive 263 nm peak of **Q1** and 12.7 °C for the positive 250 nm peak of **Q2**. Similar results were obtained for the FRET studies with compounds **2**, **3** and **4** all increasing the T_m by up to 17 °C for QDNA but only a marginal increase of between 0.3 and 1.2 °C for dsDNA. The FID experiments confirmed that the DC_{50} of **1** was 0.16 μM towards *c-myc* with all compounds below 0.5 μM. Competitive FRET experiments confirmed these findings, with compound **4** exhibiting especially high selectivity towards QDNA in the presence of up to a 600-fold concentration of dsDNA, while no decrease in the T_m of the QDNA was observed.

Additionally, SRCD spectra of the melting profiles for both **HTelo** (**Q2**) and ***c-Myc*** (**Q3**) motifs were recorded and analysed at biologically relevant buffer concentrations. Finally, through molecular modelling experiments, the potential stabilisation effects through stacking and electrostatic interactions were visualised. The predicted binding energies mirrored the findings of the experimental methods applied herein.

Further work towards extending this study to the other linked compounds **1**, **3** and **4** for further SRCD studies (when access allows) and further molecular dynamics simulations in-

cluding simulated annealing would assist in further elucidating the modes of interaction. Additionally, substituted terpyridine moieties could improve selectivity towards QDNA motifs. Finally, the solubility of these compounds must be addressed; likely areas for this are different counter ions or the addition of solubilising groups. Work is already underway, with some progress made towards more selective complexes of this type.

Acknowledgements

We would like to thank ISA, Department of Physics and Astronomy, for beam time on ASTRID and the Australian Synchrotron for providing International Synchrotron Access project travel funds. D.L.A and B.W.J.H were supported by an Australian Postgraduate Award from the Western Sydney University. We wish to acknowledge the Western Sydney University for funding through internal research grants and the School of Science and Health for support through a New Research Project Development Grant. We are grateful to the Engineering and Physical Sciences Research Council (EPSRC) for funding (grant number EP/H005285/1) and the Spanish Government and European Commission for fellowships for L.C.M. and O.M. respectively.

Keywords: circular dichroism · DNA · FRET · G-quadruplexes · platinum

[1] a) H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414–422; b) Q. Yang, J. Xiang, S. Yang, Q. Zhou, Q. Li, Y. Tang, G. Xu, *Chem. Commun.* **2009**, 1103–1105; c) J. Bidzinska, G. Cimino-Reale, N. Zaffaroni, M. Folini, *Molecules* **2013**, *18*, 12368–12395.
[2] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–186.
[3] S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275.
[4] S. M. Gowan, R. Heald, M. F. G. Stevens, L. R. Kelland, *Mol. Pharmacol.* **2001**, *60*, 981–988.
[5] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
[6] a) F. X. Han, R. T. Wheelhouse, L. H. Hurley, *J. Am. Chem. Soc.* **1999**, *121*, 3561–3570; b) D. Monchaud, M.-P. Teulade-Fichou, *Org. Biomol. Chem.* **2008**, *6*, 627–636; c) G. W. Collie, R. Promontorio, S. M. Hampel, M. Micco, S. Neidle, G. N. Parkinson, *J. Am. Chem. Soc.* **2012**, *134*, 2723–2731.
[7] S. N. Georgiades, N. H. Abd Karim, K. Suntharalingam, R. Vilar, *Angew. Chem. Int. Ed.* **2010**, *49*, 4020–4034; *Angew. Chem.* **2010**, *122*, 4114–4128.
[8] J. E. Reed, A. A. Arnal, S. Neidle, R. Vilar, *J. Am. Chem. Soc.* **2006**, *128*, 5992–5993.
[9] H. Bertrand, D. Monchaud, A. De Cian, R. Guillot, J.-L. Mergny, M.-P. Teulade-Fichou, *Org. Biomol. Chem.* **2007**, *5*, 2555–2559.
[10] I. M. Dixon, F. Lopez, A. M. Tejera, J.-P. Estève, M. A. Blasco, G. Pratviel, B. Meunier, *J. Am. Chem. Soc.* **2007**, *129*, 1502–1503.
[11] a) K. Suntharalingam, A. J. White, R. Vilar, *Inorg. Chem.* **2010**, *49*, 8371–8380; b) V. S. Stafford, K. Suntharalingam, A. Shivalingam, A. J. P. White, D. J. Mann, R. Vilar, *Dalton Trans.* **2015**, *44*, 3686–3700.
[12] T. Servidei, C. Ferlini, A. Riccardi, D. Meco, G. Scambia, G. Segni, C. Manzotti, R. Riccardi, *Eur. J. Cancer* **2001**, *37*, 930–938.
[13] S. D. Brown, K. D. Trotter, O. B. Sutcliffe, J. A. Plumb, B. Waddell, N. E. B. Briggs, N. J. Wheate, *Dalton Trans.* **2012**, *41*, 11330–11339.
[14] M. G. Mendoza-Ferri, C. G. Hartinger, M. A. Mendoza, M. Groessl, A. E. Egger, R. E. Eichinger, J. B. Mangrum, N. P. Farrell, M. Maruszak, P. J. Bednarski, F. Klein, M. A. Jakupec, A. A. Nazarov, K. Severin, B. K. Keppler, *J. Med. Chem.* **2009**, *52*, 916–925.
[15] T. Wilson, P. J. Costa, V. Félix, M. P. Williamson, J. A. Thomas, *J. Med. Chem.* **2013**, *56*, 8674–8683.
[16] a) P. A. Snyder, E. M. Rowe, *Nucl. Instrum. Methods* **1980**, *172*, 345–349; b) J. Clark Sutherland, E. J. Desmond, P. Z. Takacs, *Nucl. Instrum. Methods* **1980**, *172*, 195–199.
[17] a) A. J. Miles, F. Wien, J. G. Lees, A. Rodger, R. W. Janes, B. Wallace, *Spectroscopy* **2003**, *17*, 653–661; b) A. J. Miles, F. Wien, J. G. Lees, B. Wallace, *Spectroscopy* **2005**, *19*, 43–51.
[18] B. A. Wallace, R. W. Janes, *Modern Techniques for Circular Dichroism and Synchrotron Radiation Circular Dichroism Spectroscopy, Vol. 1*, IOS Press Inc, Amsterdam, **2009**.
[19] a) A. J. Miles, B. Wallace, *Chem. Soc. Rev.* **2006**, *35*, 39–51; b) C. Dicko, M. R. Hicks, T. R. Dafforn, F. Vollrath, A. Rodger, S. V. Hoffmann, *Biophys. J.* **2008**, *95*, 5974–5977.
[20] a) B. Wallace, R. W. Janes, *Curr. Opin. Chem. Biol.* **2001**, *5*, 567–571; b) A. J. Miles, S. V. Hoffmann, Y. Tao, R. W. Janes, B. A. Wallace, *Spectroscopy* **2007**, *21*, 245–255.
[21] L. Hu, C. Slebodnick, R. D. Gandour, B. E. Hanson, *Inorg. Chim. Acta* **2008**, *361*, 2439–2446.
[22] B. W. J. Harper, J. R. Aldrich-Wright, *Dalton Trans.* **2015**, *44*, 87–96.
[23] A. V. Tataurov, Y. You, R. Owczarzy, *Biophys. Chem.* **2008**, *133*, 66–70.
[24] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision D.01, Gaussian Inc., Wallingford, CT, USA, **2009**.
[25] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, *J. Cheminf.* **2012**, *4*, 17–17.
[26] A. K. Rappé, M. A. Pietsch, D. C. Wiser, J. R. Hart, L. M. Bormann-Rochotte, W. M. Skiff, *Mol. Eng. Mol. Engin.* **1997**, *7*, 385–400.
[27] O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455–461.
[28] I. Prislan, J. Lah, M. Milanic, G. Vesnaver, *Nucleic Acids Res.* **2011**, *39*, 1933–1942.
[29] J. Kypr, I. Kejnovská, D. Renčiuk, M. Vorlíčková, *Nucleic Acids Res.* **2009**, *37*, 1713–1725.
[30] P. Schultze, F. W. Smith, J. Feigon, *Structure* **1994**, *2*, 221–233.
[31] A. I. S. Holm, L. M. Nielsen, B. Kohler, S. V. Hoffmann, S. B. Nielsen, *Phys. Chem. Chem. Phys.* **2010**, *12*, 3426–3430.
[32] J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 2440–2450.
[33] L. M. Facchini, L. Z. Penn, *FASEB J.* **1998**, *12*, 633–651.
[34] C. Böckelman, S. Koskensalo, J. Hagström, M. Lundin, A. Ristimäki, C. Haglund, *Cancer Biol. Ther.* **2012**, *13*, 289–295.
[35] a) S. M. Hampel, A. Sidibe, M. Gunaratnam, J.-F. Riou, S. Neidle, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 6459–6463; b) S. M. Hampel, A. Pepe, K. M. Greulich-Bode, S. V. Malhotra, A. P. Reszka, S. Veith, P. Boukamp, S. Neidle, *Mol. Pharmacol.* **2013**, *83*, 470–480.
[36] a) K. Suntharalingam, A. J. White, R. Vilar, *Inorg. Chem.* **2009**, *48*, 9427–9435; c) C. Yu, K. H.-Y. Chan, K. M.-C. Wong, V. W.-W. Yam, *Chem. Commun.* **2009**, 3756–3758.
[37] K. Suntharalingam, O. Mendoza, A. A. Duarte, D. J. Mann, R. Vilar, *Metallomics* **2013**, *5*, 514–523.
[38] N. H. Campbell, N. H. A. Karim, G. N. Parkinson, M. Gunaratnam, V. Petrucci, A. K. Todd, R. Vilar, S. Neidle, *J. Med. Chem.* **2012**, *55*, 209–222.

Received: September 12, 2015
Published online on December 16, 2015

Chapter 3

A study of Pt(II)-phenanthroline complex interactions with double-stranded and G-quadruplex DNA by ESI-MS, circular dichroism, and computational docking

Expanding on my previous work applying SRCD and other spectroscopic techniques, I was intrigued by the idea that mass spectrometry could be used to study the binding of small molecules with biomacromolecules such as DNA and proteins. Due to the harsh ionisation processes typically required in mass spectroscopy, resulting in fragmentation of the analyte of interest, a milder ionisation method is required. Electrospray ionisation (ESI) is a relatively soft ionisation process where a strong electric field is applied to a nebulised flow of liquid, creating a fine aerosol of charged ions in the gas phase. These charged particles are attracted to the oppositely charged spectrometer inlet where they are analysed.

ESI is the predominant method of ionisation when the analyte of interest is in

liquid, as is the case when coupled to a liquid chromatography system. In this way biomacromolecules such as DNA and proteins may survive intact and be observed without fragmentation. The application of ESI-MS is not a quantifiable method but a hierarchy of binding may be developed for a series of small molecules. My initial training in the technique was performed by our colleagues at Wollongong University. From there it was developed in house with the assistance of the mass spectrometry facility in the School of Medicine.

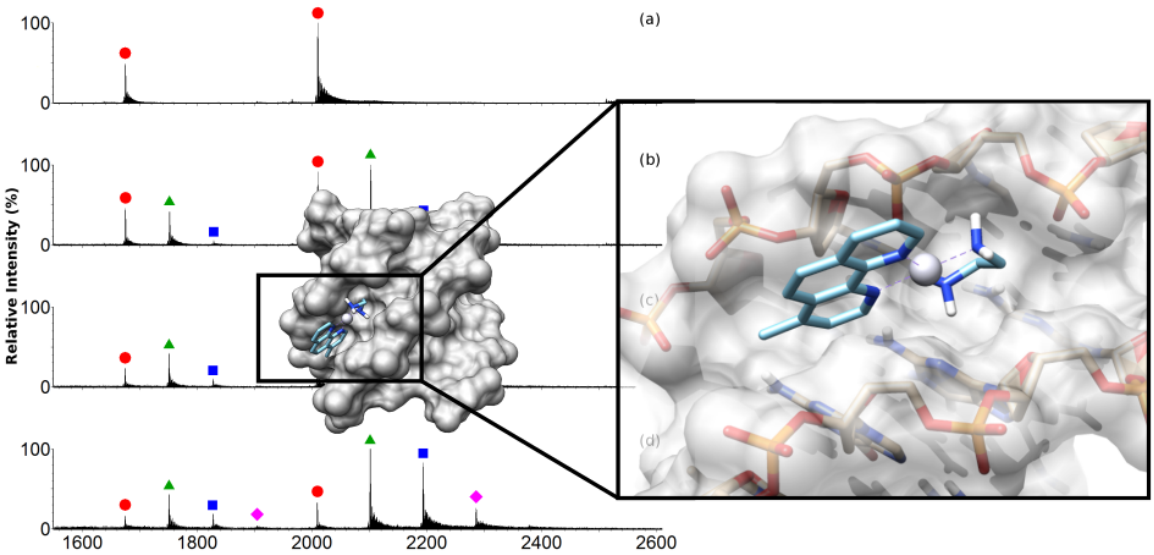
An alternative ionisation method to ESI is matrix-assisted laser desorption/ionization (MALDI). Both are considered soft ionisation methods, with resultant low fragmentation of analytes, but MALDI is less likely to produce multivalent ions due to the ionisation mechanism. This may limit the detection of higher molecular weight analytes, as is the case for G-quadruplex complexes with molecular weights often >8000 Da. ESI has also been found to produce higher resolution spectra with lower degree of fragmentation when applied to the investigation of platinum complexes with biomacromolecules.[18] The identity of ions resulting from the ESI-MS experiments are readily deconvoluted with the number of bound complexes and included ammonium ions observable.

In this work, a range of platinum(II)-phenanthroline complexes matching those in Chapter 1 were synthesised and characterised according to literature method. A new method of preparing DNA for ESI-MS analysis was developed and applied successfully, along with several computational techniques and CD incubation studies. One of the outcomes of this study showed that the ESI-MS experiment binding affinity correlated with the published known binding strengths, therefore showing that ESI-MS can be applied to qualitatively determine the binding characteristics of these platinum(II)-phenanthroline complexes. I performed all experimental work, DFT and docking simulations, generated all artwork and graphics and wrote the entirety of the text.

Paper reference

Ang, D. L.; Kelso, C.; Beck, J.; Ralph, S.; Harman, D. G.; Aldrich-Wright J.: A study of Pt(II)-phenanthroline complex interactions with double-stranded and G-quadruplex DNA by ESI-MS, circular dichroism, and computational docking. JBIC Journal of Biological Inorganic Chemistry, 25, 429–440(2020). DOI: 10.1007/s00775-020-01773-4.[19]

Graphic abstract



A study of Pt(II)-phenanthroline complex interactions with double-stranded and G-quadruplex DNA by ESI-MS, circular dichroism, and computational docking.†

Dale L. Ang,[a] Celine Kelso,[c] Jennifer L. Beck,[c] Stephen F. Ralph,[c] David G. Harman,[b] Janice R. Aldrich-Wright[ab*]

Abstract

The binding interactions of a series of square planar platinum(II)-phenanthroline complexes of the type $[Pt(P_L)(A_L)]^{2+}$ (where P_L = variously methyl-substituted 1,10-phenanthroline (phen) and A_L = ethane-1,2-diamine (en)) were assessed with a G-quadruplex DNA (5′-TTG GGG GT-3′, G4DNA) and a double-stranded DNA (5′-CGC GAA TTC GCG-3′, dsDNA) sequence by ESI-MS. The results indicate a strong correlation between G4DNA affinity and increasing phenanthroline methyl substitution. Circular dichroism (CD) spectroscopy and molecular docking studies also support the finding that increased substitution of the phenanthroline ligand increased selectivity for G4DNA.

1 Introduction

In the presence of appropriately sized cations, guanosine-rich DNA regions self-assemble to form planar guanine tetrad structures, commonly referred to as G-quadruplex DNA (G4DNA).[12,20] The nature and importance of these non-canonical DNA structures is becoming increasingly apparent with early estimates on the number of G-quadruplex-forming sequences in the human genome of ∼20,000.[25] More recent efforts to estimate the number of possible G-quadruplex forming sequences have significantly revised this number to well over 300,000.[7] Combined with the discovery that G4 motifs are also present in heterochromatin,[18] compounds that selectively target and bind these unique DNA structures may lead to new and innovative chemotherapeutic strategies.[38]

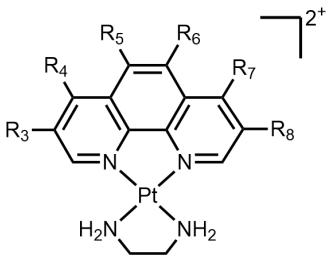
Approximately half of all chemotherapeutic regimens involve platinum-based compounds and cis-

[a] *Nanoscale Organisation and Dynamics Group, School of Science, Western Sydney University, Locked Bag 1797 Penrith South DC, NSW 2751, Australia. Fax: +61 2 4620 3025; Tel: +61 2 4620 3218; Email: j.aldrich-wright@westernsydney.edu.au.*
[b] *School of Medicine, Western Sydney University, Australia.*
[c] *School of Chemistry and Molecular Bioscience, University of Wollongong, Northfields Avenue, Wollongong, 2522, Australia.*
† Electronic Supplementary Information (ESI) available: [details of any supplementary information included here]. See DOI: 10.xxxx/x000000x/

platin is the leading treatment.[29] The typical mode of action of platinum complexes is to form DNA adducts with the N_7 of guanine, interfering with replication, causing cell death. The non-selective nature of these chelates often results in debilitating side effects and the development of resistance. As such, there is a strong focus to develop small molecules that selectively target non-canonical DNA motifs such as G4DNA to achieve higher efficacy while reducing negative side-effects.[8]

The various structural topologies afforded by G4DNA open up numerous targeting opportunities. Conceptually, by stabilising the G4DNA structures selectively, the transcription of genes such as the *c-myc* oncogene[9,33] and even telomerase activity, may be modulated.[13] Recently, regulation of gene expression, by manipulating the stability of a modified G4DNA, was demonstrated in zebrafish embryos.[27] The technique involved reversible *cis–trans* photoisomerisation of a photochromic nucleobase to induce a hyperstable G4DNA state that inhibited transcription. Dinuclear Pt(II)-terpyridine complexes have been shown to stablilise G4DNA melting temperatures by up to 17 °C, suggesting similar inhibition effects may be possible with Pt(II) complexes.[1]

In vitro experimental evidence, as obtained by linear dichroism, suggests that the chief mode of interaction between planar aromatic cationic metal complexes and double-stranded DNA (dsDNA) is via intercalation.[30] In contrast, G4DNA, with strongly interacting strands and an internal cationic core, would be expected to be relatively immune to intercalation: stacking and groove binding interactions would dominate. Previous investigations with a subset of these platinum complexes (PCs) examined interactions with G4DNA at extremely high PC:DNA ratios and found the square-planar geometry to be critical to stacking interactions and thus favourable with the planar tetrad G4DNA structures.[35]



1 = PHENEN, all R = -H
2 = 4MEEN, R_4 = -CH_3
3 = 5MEEN, R_5 = -CH_3
4 = 47MEEN, $R_{4,7}$ = -CH_3
5 = 56MEEN, $R_{5,6}$ = -CH_3
6 = 3478MEEN, $R_{3,4,7,8}$ = -CH_3

Figure 1 The six square planar Pt(II)-phenanthroline complexes studied, along with the abbreviated naming conventions used. All were prepared as dichloride salts according to literature methods.[10]

Here we report the interactions of a range of biologically active square planar PCs comprising a substituted 1,10-phenanthroline (phen) pharmacophore, I_L, together with an achiral bidentate ethane-1,2-diamine (*en*) ancillary ligand, A_L. Table 1 lists the IC_{50} (inhibitory concentration) values (L1210, murine leukemia cell line) and dsDNA binding constants (K_b) for these complexes.[10] Through the

application of ESI-MS binding experiments, circular dichroism spectroscopy, and molecular docking simulations we present evidence that planar molecular surface area is a critical component of effective G4DNA interactions. Additionally, the contribution of lipophilicity to G4DNA binding is considered. Finally, a streamlined purification method is presented, simplifying the method and reducing the time and cost of preparing DNA oligomers for ESI-MS experimentation.

Table 1 List of IC_{50} values,[10] dsDNA binding constants, K_b,[10] HPLC retention time, and predicted binding affinity. *Errors were < 0.1 kcal/mol

ID	Complex	IC_{50} (μM, L1210)	$K_b \times 10^6$ (230 nm)	Retention time (k′, min)	ΔG (kcal/mol)*
–	Cisplatin	0.5	–	–	–
1	[Pt(en)(phen)]Cl_2	9.7 ± 0.4	2.2 ± 0.5	4.736	–6.4
2	[Pt(en)(4-mephen)]Cl_2	> 50	1.2 ± 0.4	5.788	–6.8
3	[Pt(en)(5-mephen)]Cl_2	2.8 ± 0.8	0.7 ± 0.2	5.938	–6.5
4	[Pt(en)(4,7-me_2phen)]Cl_2	> 50	1.3 ± 0.4	6.828	–6.5
5	[Pt(en)(5,6-me_2phen)]Cl_2	1.5 ± 0.3	1.5 ± 1.1	6.987	–6.8
6	[Pt(en)(3,4,7,8-me_4phen)]Cl_2	> 50	0.7 ± 0.3	8.434	–6.9

2 Results and Discussion

2.1 Preparation of oligomers for ESI-MS

G4DNA is stabilised by the presence of K^+ ions within the central cavity of the tetrads. Substituting K^+ with Na^+ adversely affects the stability of the G4 structure due to the smaller ionic radius of Na^+ (K^+ = 138 pm, Na^+ = 102 pm). Conveniently, the ammonium ion (NH_4^+) is an excellent substitute for K^+ as the ionic radius for NH_4^+ is ∼140 pm.[32,34] In our experience, the G4DNA stability of the K^+ and NH_4^+ adducts are similar, as measured by melting temperature. It is important to minimise the presence of Na^+ and K^+ salt adducts when preparing highly anionic nucleic oligomers for sensitive techniques such as ESI-MS; these adducts interfere to such an extent that it is difficult to quantify the results effectively. Therefore, it is necessary to replace these with a volatile salt before analysis—ammonium in the form of ammonium acetate buffer. To this end, contemporary DNA oligomer purification methods for ESI-MS experiments are time consuming. These processes usually consist of combining multiple HPLC injections (to limit column overloading) in ammonium acetate buffer followed by concentrating the collected fractions down to a single sample; this process typically requires several days and leads to inevitable sample loss.

We introduce a simple and robust purification/cation exchange step that retains purity and minimises sample loss, using a commercially available cation exchange column coupled to a peristaltic pump and simple UV detector. The column was first saturated with ammonium acetate buffer, the G4DNA loaded, subsequently eluted, and then lyophilised as one fraction. The total process, exclud-

ing the lyophilisation step, can be completed within 30 min. Typically, multiple Na^+ and K^+ salt adduct peaks are observed in ESI mass spectra. The absence of these factors in the negative ion mass spectra (Figure 2) is evidence of this refined method. D2 was subjected to the same chromatographic purification and cation exchange process. In the spectrum for D2 (Figure 2 upper) m/z 1214.2, 1821.9, and 2429.8 were assigned to $[ssD2-3H]^{3-}$, $[D2-4H]^{4-}$, and $[D2-3H]^{3-}$, respectively. Some unassignable peaks are observed, however the intensities are <20% of the total ion current (TIC). Peak tailing, signifying adducts of mostly K^+ and Na^+, is minimal. A peak assignable to the 3^- charge state at m/z 2429.8 is also visible but at a very low ion count. The G4 spectrum is similarly assignable. Ions of high abundance at m/z 1674.7 and 2009.8 are assigned to $[G4+4NH_4^+-10H]^{6-}$ and $[G4+4NH_4^+-9H]^{5-}$ respectively with the predominant ions in the 6^- and 5^- charge states. There are no non-assignable peaks is this spectra and minimal adducts. Individual spectra are provided in the SI.

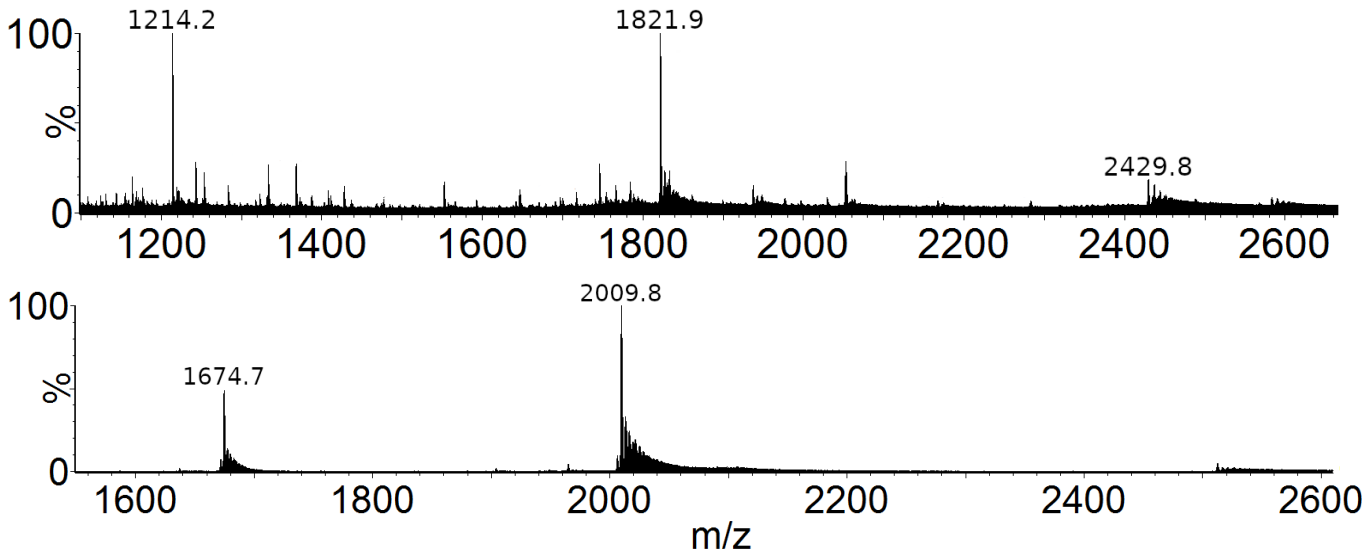


Figure 2 Negative ion ESI-MS spectrum of D2 (upper) and G4 (lower). D2: Peaks at m/z 1214.2, 1821.9, and 2429.8 were assigned to $[ssD2-3H]^{3-}$, $[D2-4H]^{4-}$, and $[D2-3H]^{3-}$ ions, respectively. G4: Peaks at m/z 1674.7, and 2009.8 correspond to the $[G4+4NH_4^+-10H]^{6-}$ and $[G4+4NH_4^+-9H]^{5-}$ ions, respectively.

2.2 Interactions of platinum(II) complexes with double-stranded and G-quadruplex DNA

The PCs analysed were chosen as they demonstrate *in vitro* anticancer activity and the dsDNA binding constants have been quantified (Table 1).[10] The series exhibit structurally distinct modifications to the planar aromatic ligand, specifically the number and position of methyl moieties. This provides the ability to compare the effects of these modifications from a structure activity relationship (SAR) analysis perspective, directly. Furthermore, these PCs are well-characterised, very stable in solution,

and contain no chiral centres that may interfere with CD analysis. The G4 and D2 oligomer sequences were chosen in order to provide thermodynamically stable non-covalent structures unlikely to dissociate during drug binding. High G/C content D2, and a previously reported G4 were chosen specifically for this purpose.[6]

To begin with, the interactions of **5** with G4 were examined using negative ion ESI-MS (Figure 3). An important point to make is that with increasing concentrations of **5** the intensity of the free G4 5^- ion (red circle, *m/z* 2010.0) decreases visibly but does not disappear entirely. The abundance of ions where one PC is bound (green triangle, *m/z* 2102.2) stays relatively stable, and is the dominant ion across all concentration ratios. This is followed closely by two PCs bound to the G4 (blue square, *m/z* 2194.5) and lastly at the very high 10:1 PC:DNA ratio, the three PCs bound G4 ion is clearly evident at (pink diamond, *m/z* 2286.0). Across the board, the binding affinity with G4 shows that as the phenanthroline methylation increases, so too does G4DNA binding (Figure 4).When exposed to the PCs with dimethylated phenanthrolines, **4** and **5**, there is less than 20% free G4 remaining at the highest concentration, bested only by tetramethylated phenanthroline PC (**6**) which shows strong interaction at all ratios. Nonetheless, there remains a detectable portion of free G4 in solution even at the highest concentrations of PC.

When the ESI mass spectra of D2 with the PCs are examined, it becomes readily apparent that the 5:1 ratio is more than adequate to fully bind all available DNA, resulting in no discernible free D2 ion intensity. At the 1:1 PC:D2 ratio PCs **2**, **4**, and **5** demonstrate the highest affinity with ∼30% unbound. This is followed by **3** with ∼45% unbound, and lastly **1** and **6** with ∼50% unbound D2 at this ratio. The experimental K_b values for dsDNA also reflect this trend (Table 1). It is hypothesised that with limited planar surface area for intercalation, in this scenario groove binding and electrostatic interactions dominate, resulting in a reduced chance of surviving the ionisation process required for ESI-MS analysis. PC **6** also interacts poorly; the width of the tetra-methylated phenanthroline ligand sterically limits access to the intercalation site. At concentrations of 5:1 PC:DNA and above, all available dsDNA was bound to a PC.

G4 binding follows a different trend: at all PC:G4 ratios, the trend for binding follows **1** < **2** < **3** < **4** ≃ **5** < **6**. When we consider the surface area of the aromatic ligand it is clear there is a direct association with G4 binding affinity. It is of note that even at 10:1 PC:G4 ratios there is still >10% free unbound DNA. The possible cause of this is that the ESI process imparts too much energy for the low binding affinity of the PC:G4 association to survive intact. A summary of the ESI-MS results of PCs **1**–**6** with G4 (upper) and D2 (lower) presented as a percentage of free unbound DNA (calculated

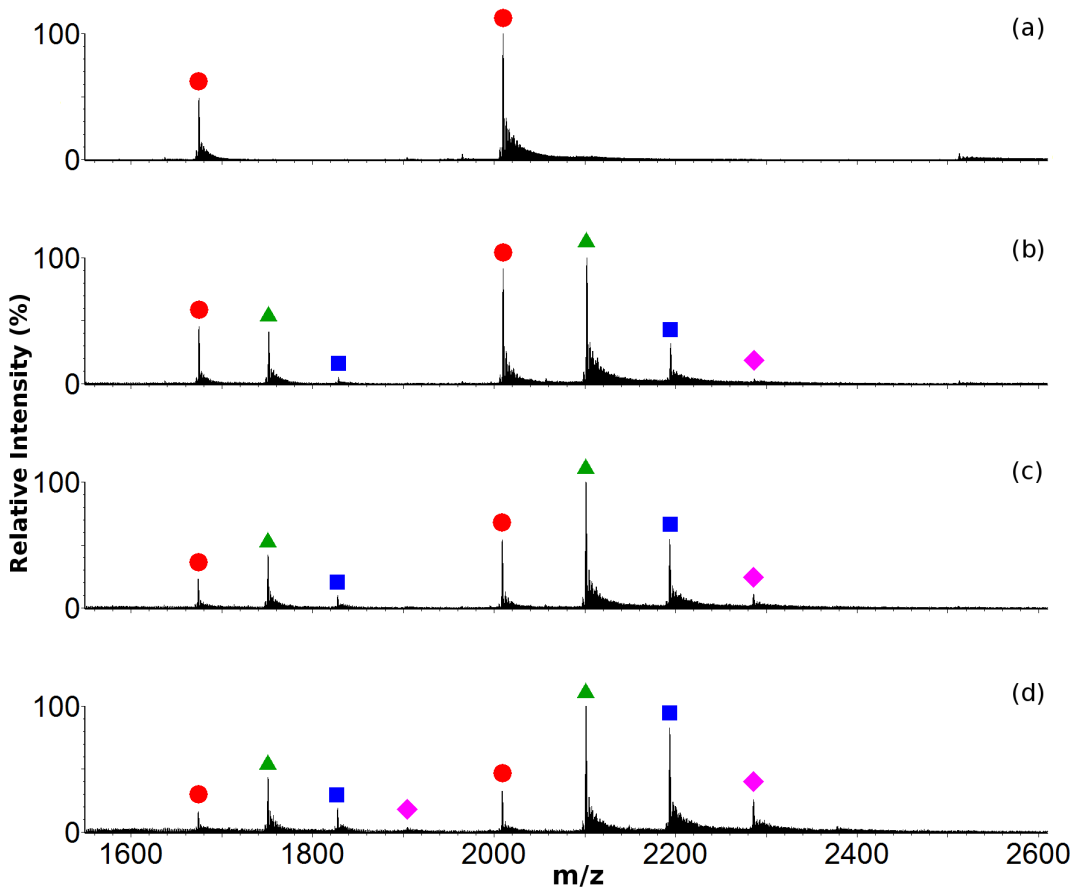


Figure 3 Negative ion ESI mass spectra of free G4 and solutions containing G4 with increasing PC:G4DNA ratios with **5**. (a) G4 only; (b) 1:1; (c) 5:1; (d) 10:1. Red circle: G4 only; green triangle: G4 + 1PC; blue square: G4 + 2PC; pink diamond: G4 + 3PC.

from the sum of all relative ion intensities) are presented in Figure 4.

From these results it can be summarised that the factors which favour binding with D2 are not necessarily the same as those that contribute to G4 affinity. As the most favourable G4 binder, **6**, is equal poorest with **1** when interacting with D2. The appearance of free G4 ions even at the highest ratios of PCs also suggested that the types of interactions with G4 are far weaker than interactions with D2. A likely explanation for this trend is that intercalation is not feasible within the G4 structure, leaving only groove binding and stacking interactions as possible alternatives. Groove binding and stacking require lower binding energy and therefore are inherently less stable than intercalation and so less likely to survive the ionisation process.

It should be noted that at high PC concentrations, as in the case of 10:1 PC:DNA experiments here, ion suppression may be a factor affecting the validity of the mass spectra observed. Ion suppression occurs when the metal complex is more readily ionised than the DNA phosphate groups, resulting

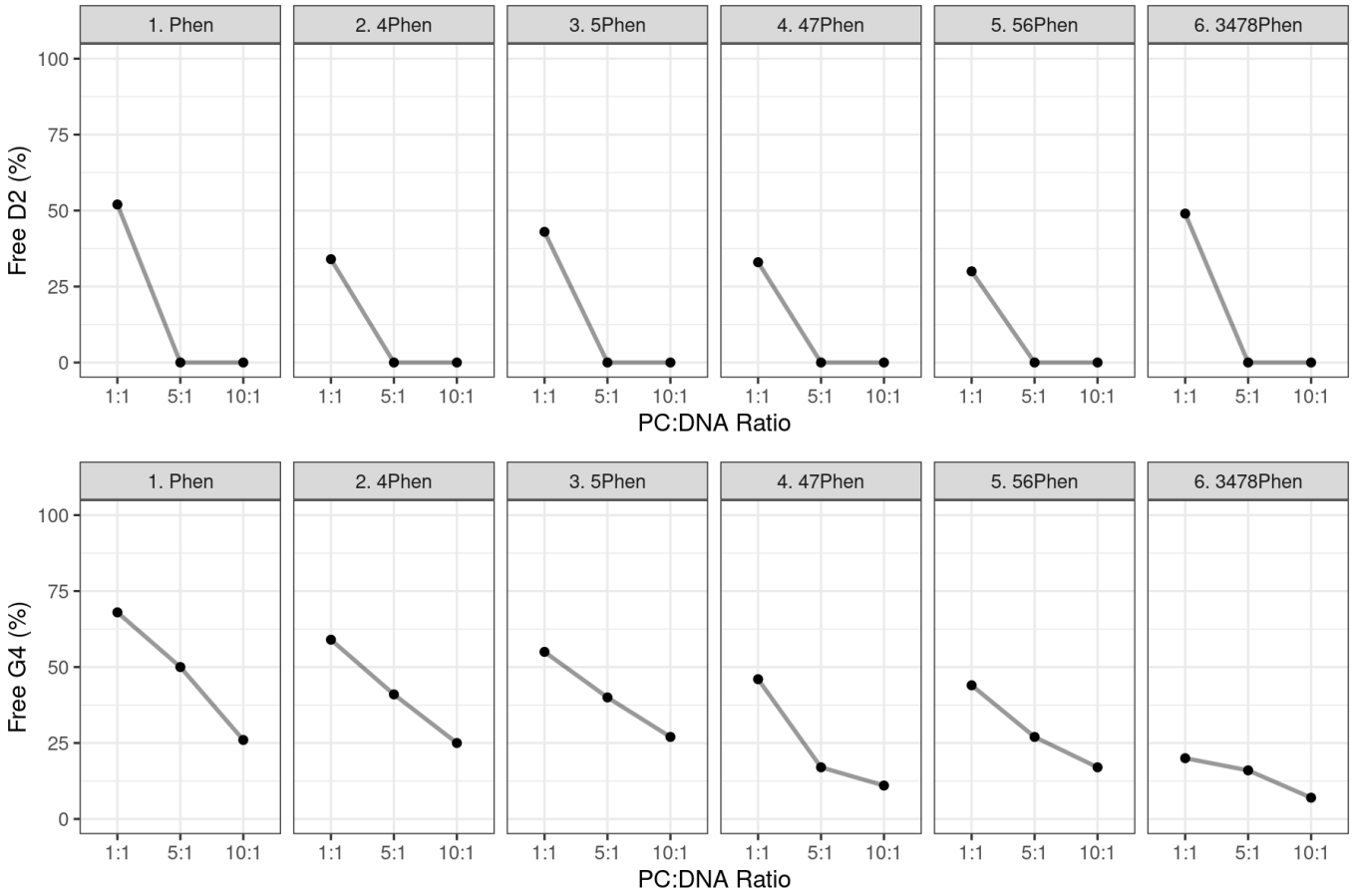


Figure 4 Percentage of free D2 and G4 at various PC:DNA ratios for each of the PCs studied, as measured by ESI-MS.

in preferential ionisation, effectively stealing the charge.[2,3,11] This was observed as a lowering of the overall S/N ratio of the spectrum. Additionally, non-volatile components make droplet formation or droplet evaporation less efficient which, in turn, affects the amount of charged G4 reaching the detector. This phenomenon is exacerbated as PC:DNA ratios increase.

2.3 Induced Circular Dichroism by metal complex

In order to observe if any change in structure occurred, G4 was treated with each PC at what may be considered saturation (10:1 PC:G4) and a CD spectrum measured (Figure 5 upper). A positive peak at 213 nm, negative peak at 244 nm and strong positive peak apparent at ∼265 nm are all indicative of parallel quadruplex formation[19] and is observed clearly in these CD spectra. Upon the addition of PCs, conformational changes are apparent; loss of CD signal magnitude follows the sequence **1** (phen) < **2** (4phen) < **3** (5phen) = **5** (56phen) < **4** (47phen) < **6** (3478phen). When the stacking

interaction sites are filled, additional interaction through groove binding occurs, thereby elongating the G4 structure, with a resulting loss of optical activity and reduction in magnitude.

The appearance of a shoulder at ∼300 nm is noteworthy as it is absent in the free G4, and increases in intensity with the increasing binding affinity of the PCs. This is in agreement with the behaviour observed when a ruthenium dipyridophenazine (dppz) complex was incubating with a similar parallel G-quadruplex sequence at a 4:1 ratio of metal complex resulting in the DNA adopting an antiparallel topology.[24] The conformational change induced by the PC supports the conclusions in the ESI-MS studies where **4** and **6** bind most strongly at these high ratios, so much so that they induce a change in the G4 structure.

The induced CD spectra (ICD, Figure 5 lower) (the G4+PC CD spectrum minus the G4 CD spectrum) increases proportional to molecular surface area. PCs **3** and **5** exhibit near identical ICD, which is as expected as each is methylated in the 5 position (5-methylphen and 5,6-dimethylphen ligands, respectively). Across all PCs, all peaks decrease in intensity proportionally with the strength of binding, as supported by the ESI-MS results. From the ICD it is evident that the reduction in magnitude of all peaks, and in particular the ∼265 nm peak, equates to an overall reduction of parallel quadruplex structure.

2.4 Hydrophopbicity determination by Reverse-Phase HPLC

Reverse-phase HPLC (RP-HPLC) is a separation technique which uses a non-polar C18 stationary and polar mobile phase. This results in the preferential partitioning between the two phases based on molecular hydrophobicity. The more hydrophilic the analyte, the more it would partition into the polar mobile phase, reducing retention times. Using this principle, the measured retention time of a series of related analytes provides a first order estimate of the relative hydrophobicity of a compound.[4]

Table 1 lists the corrected retention time, k′, of each PC. PC **1** elutes first (4.74 min) followed by the monomethyl substituted **2** and **3** (5.79 and 5.94 min, respectively), then the dimethyl substituted **4** and **5** (6.83 and 6.99 min, respectively) and finally tetramethyl-substituted **6** (8.43 min). This increase in elution time can be attributed to the increasing surface area of the phen moiety, and the electron density and charge distribution of the complex.

Correlating retention time with percent bound DNA (at a 1:1 PC:DNA ratio) provides an insight into binding specificity. For G4 there is very strong correlation (Pearson correlation coefficient (PCC) = 0.98, R^2 = 0.9778), contrasting with very low correlation for D2 (PCC = 0.14, R^2 = 0.0205)

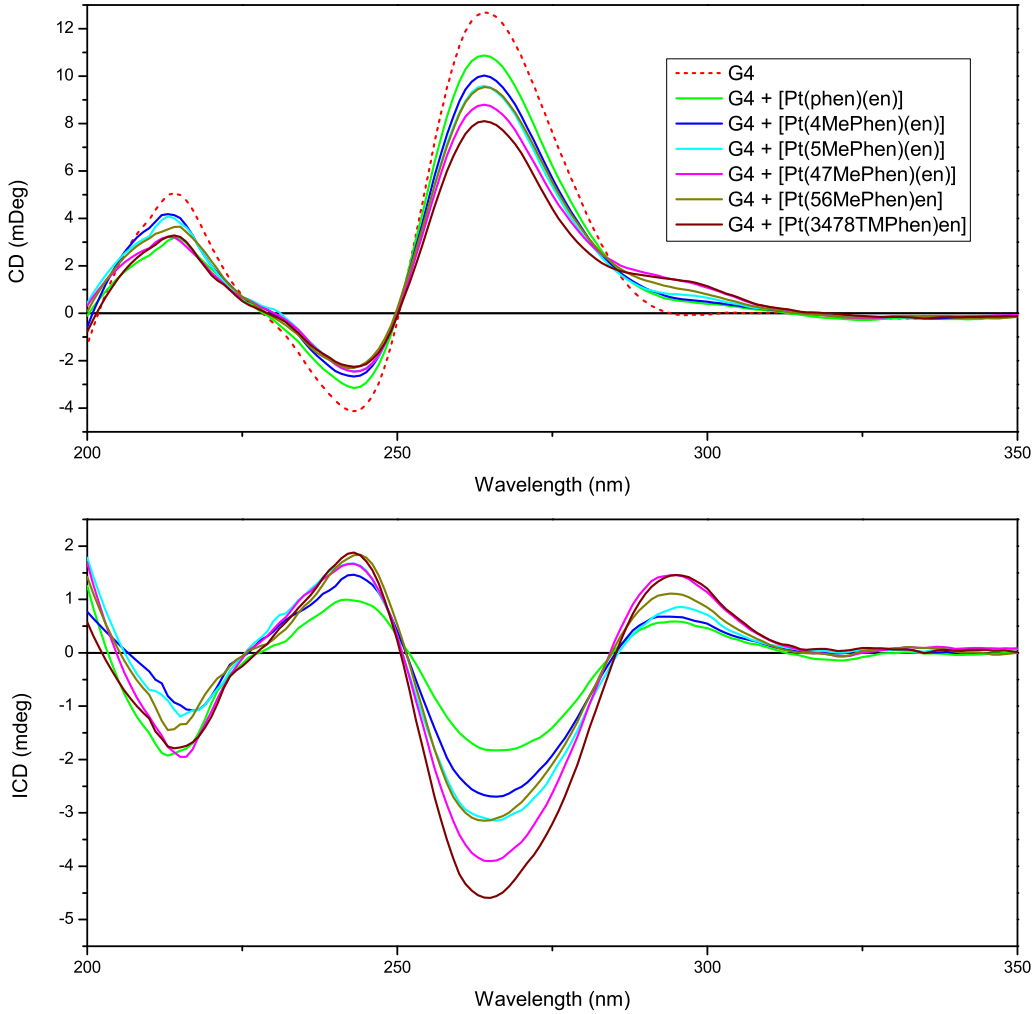


Figure 5 CD (upper) and induced circular dichroism (ICD, lower) spectra of G4 incubated with PCs at 10:1 MC:DNA ratio in buffer (NH_4OAc, pH 7.0) at an oligomer concentration of 5 μM. Samples were incubated at room temperature for 1 h prior to measurement. Cell pathlength 0.1 cm, 1 nm data pitch, 100 nm min^{-1} scan speed, 20 accumulations. Data were smoothed once with a 7 point Savitzky-Golay algorithm.

(Figure S19). However, if PC **6** is excluded from the correlation calculation, the D2 correlation improves substantially to 0.89. This highlights the specificity of **6** for G4, qualifying the tetra-methylated phenanthroline moiety for investigation in future G4 selective binders.

2.5 Computational docking and DFT

Docking simulations were performed using the representative PDB structures 1BNA for D2 and 2JT7 for G4 (Figure 6). The rigid structure of the square planar complexes equates only a small conformational sampling space for the docking simulations. It was expected that the conserved ethane-1,2-

diamine moiety would contribute in a generally equal way to the overall binding affinity across the complexes and resulted in extremely reproducible docking binding energy and conformation predictions across replicates (<1% variation). This thereby restricted the study wholly to the contribution of the substitution effects of the phenanthroline moiety. Predicted binding energy results of the computational docking are summarised in Table 1. The predicted binding generally follows **1** < **3** ≃ **4** < **2** ≃ **5** < **6**, a slight deviation from the retention time and ESI-MS results. These theoretical predictions generally follow the substitution of the phenanthroline.

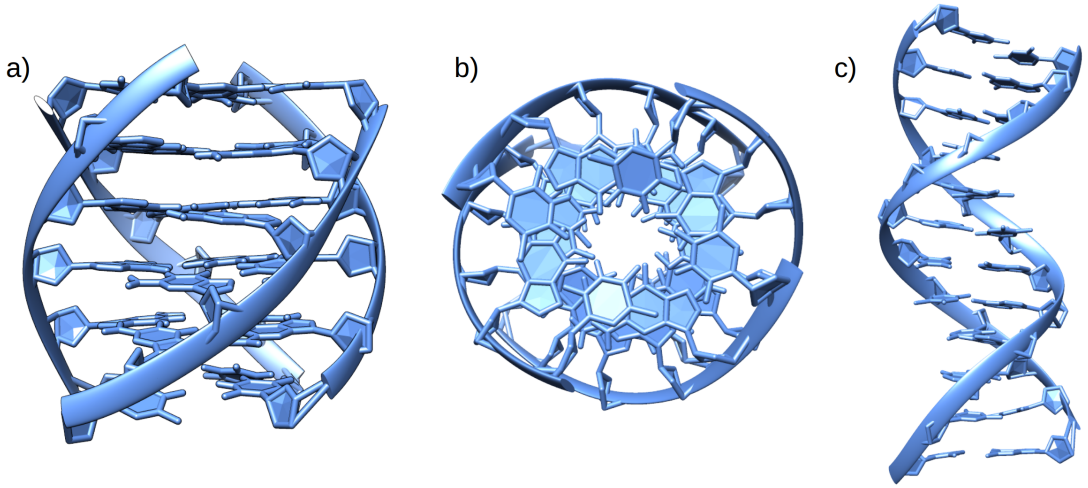


Figure 6 The models used in the docking simulations. Side view (a) and top view (b) of G4 tetramolecular parallel quadruplex (PDB ID: 2JT7); (c) 12mer D2 (PDB ID: 1BNA).

Analysis of the predicted binding conformations (Figure 7) reveals a subtle interplay between electrostatic and hydrophobic/π-stacking interactions. The non- and mono-methylated PCs (**1**, **2**, and **3**) interact with the G4 via groove binding with the methyl moiety oriented inwards, away from the bulk water, while the positively charged *en* moiety is observed to interact with the negatively charged backbone phosphates. This is in contrast to what is seen with the di- or tetra-methylated PCs (**4**, **5** and **6**) where the effect of an increased neutral hydrophobic region dominates in the interaction, resulting in a preference for stacking with the top and/or bottom tetrads formed by the guanine bases. Thus, the increased hydrophobic surface area of the methylation drives the interaction towards a preference for stacking over groove binding. The strong stability of these parallel tetramolecular quadruplexes precludes the capacity for intercalation, which is the predominant form of interaction of these PCs with double-stranded DNA.[10]

The distribution of partial atomic charges is summarised as a boxplot (Figure S20) and the calculated electrostatic potential visualised as an isosurface (Figure S21). Figures S20 & S21 serve to

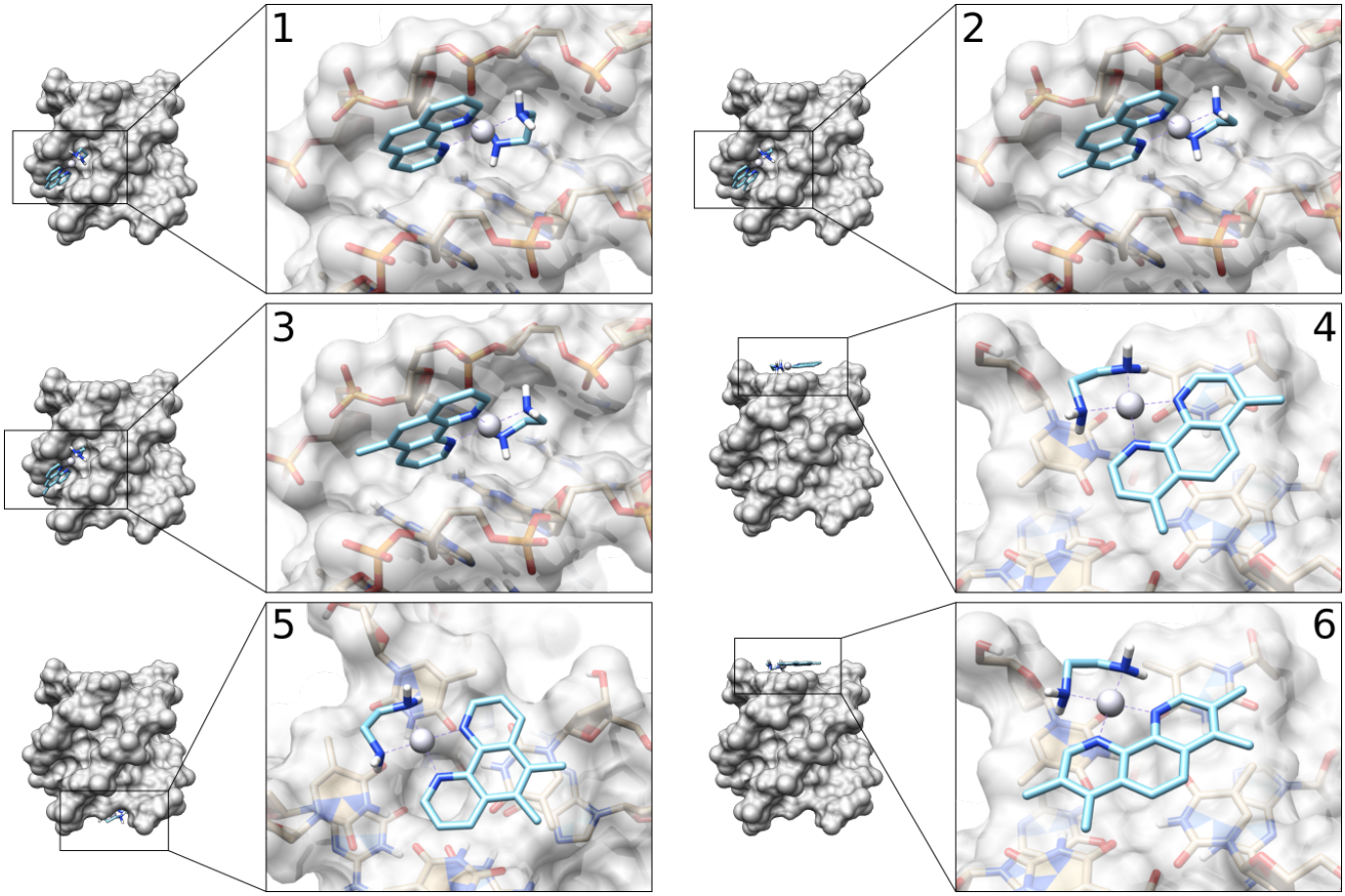


Figure 7 Selected poses of predicted docking conformations of the PCs (**1**–**6**) with G4DNA PDB 2JT7.

illustrate that the subtle change in charge distribution is attributable to the degree of methylation, and in the interaction with G4 this manifests as a transition from groove binding to π-stacking. Polarisation and π-stacking interactions have been shown to be influenced directly by substituents,[28] which is demonstrated here with the increasing methylation reducing the charge density on the central platinum atom. The dominant charges are those for the platinum ($\sim$+0.72) and the two nitrogens of the phenanthroline ($\sim$–0.67 each). The majority of the remaining positive charge is carried by the central phenanthroline carbons (atoms 3, 7, 8, 14) and the ethane-1,2-diamine carbons (atoms 17, 18). As expected, the largest fluctuations in charge are from the carbons where the methyl substituents attach (atoms 4, 5, 10, 11, 12, and 13). Therefore, it is readily apparent that the methyl substituents play a role in the overall ESP. The calculated Mulliken charges are listed in Table S3. Overall, the molecular charge is heavily biased towards the *en* moiety with methylation affecting the charge distribution. Increasing methylation shifts the charge density towards the *en* moiety, exposing more neutral phenanthroline surface and hence biasing stacking interactions over the more electrostatically driven

groove binding.

3 Conclusions

Here we analysed, with the application of ESI-MS, molecular docking simulations and associated techniques, how the variously methyl-substituted planar aromatic phenanthroline moieties in a series of platinum(II) complexes interact with dsDNA and G4DNA. Also presented is a fast method for the preparation of DNA for use in ESI-MS experiments. The results highlight some important structural characteristics required in designing a new generation of chemotherapeutic agents to target G4DNA. Specifically, that lipophilic aromatic pharmacophores play an important role in binding with G4DNA, likely primarily through a π-stacking mechanism with the tetrad. This may mean modulating the π-stacking effect and charge distribution through variously charged substituents.

Interestingly, there is little correlation between the published IC_{50} values of the PCs tested and specificity towards G4DNA. Conversely, a strong correlation between the lipophilicity of the complexes and binding was observed for G4DNA but not dsDNA. The main *in vitro* mode of action of these PCs is by interaction with the cytoskeletal proteins, resulting in stalled cell replication leading to cell death.[21] This is fortuitous as it highlights that features optimal for this type of interaction are not necessarily the same as required for G4DNA binding. Furthermore, these PCs may shed some light on possible leads for further functionalisation with the ultimate goal of highly specific targeting of G4DNA.

Despite the docking analysis of PCs 1–3 showing groove binding, the experimental CD evidence suggests end-capping as the main mode of interaction for these PCs. In general, platinum(II) complexes are classified as G4DNA end-capping with tetrad π-stacking type binding.[22] The discrepancy can be attributed to limitations of the docking simulations of this type with rigid substrates. However, it still captures a very important trend in that methylation of the phenanthroline plays a role in the end-capping interactions, as complexes 4–6 display.

While this family of PCs are a starting point for understanding G4DNA binding from a structural perspective, derivatives of similar square planar metal complexes have demonstrated greater affinity for the planar tetrad structure of G4DNA.[14] Platinum(II) is a strong candidate due to being more stable than equivalent Ni(II) and Pd(II) complexes,[5,26,31] and ability to be oxidised to a Pt(IV) prodrug which then reduces to the active form in vivo.[15] However, we have not considered the possibility of further interactions with the G4DNA structure at this point. Hydrogen bonding plays a critical role in biomolecular interactions and selectivity may be improved with strategically placed H-bond donors and acceptors. Future work will expand the range of G4DNA substrates studied through further

synthesis and molecular dynamics simulations on these systems.

4 Experimental

4.1 Pt(II) complexes

The PCs, presented in Figure 1, were prepared by literature methods as dichloride salts.[37] The PoraPak Rxn CX columns were purchased from Waters. The oligomer 5′-TTG GGG GT-3′ (G4, 2496.7 Da) was purchased from GeneWorks, Thebarton, South Australia. The self-complementary oligomer 5′-CGC GAA TTC GCG-3′ (D2, 3646.4 Da) was purchased from DNA Technology, Denmark. Ammonium acetate was purchased from APS Chemicals and HPLC grade acetonitrile, methanol and DMSO from Merck chemicals.

4.2 DNA purification and cation exchange

The dsDNA (D2) was composed of the self-complementary oligomer 5′-CGC GAA TTC GCG-3′, and the G4DNA (G4) oligomer sequence was 5′-TTG GGG GT-3′. Oligomer purification and cation exchange were performed using a Waters Porapak Rxn CX 20 cc column. A constant flow rate of 1 mL min^{-1} was controlled using a Bio-Rad low pressure peristaltic pump and detection was achieved with an inline Bio-Rad UV detector at 260 nm. The column was prepared by eluting methanol (10 mL) then water (20 mL), followed by equilibration with ammonium acetate solution (100 mL, 150 mM, pH 7.0). DNA, as supplied, was dissolved in ammonium acetate buffer (NH_4OAc, 300 μL, 150 mM, pH 7.0). The DNA solution was loaded onto the equilibrated column and eluted with buffer so that the peak could be collected. The resulting ∼5 mL purified DNA solution was lyophilised. Final concentration of the DNA was spectrophotometrically determined by UV absorbance using the following molar absorptivities: G4 ε_{260} = 76,000 M^{-1} cm^{-1}, D2 ε_{260} = 191,500 M^{-1} cm^{-1}. Concentrations were then adjusted to achieve a stock concentration of 1 mM. The final tetramolecular G4 and double-stranded D2 DNA were obtained by annealing in a water bath at 95 °C for 5 min and cooling naturally to room temperature over several hours. Final concentrations of the annealed DNA stocks were confirmed spectrophotometrically and stored at 4 °C prior to use.

4.3 ESI-MS

Mass spectra were obtained in ESI negative mode using a Waters XEVO QToF ESI mass spectrometer with a Z-spray ionisation source. Spectra were obtained with an MCP potential of 2.4 kV, a cone

voltage of 25 V, capillary tip potential of 2.3 kV and the extraction cone at 4.0 V. The desolvation gas (nitrogen) flow was 150 L hr^{-1} at 150 °C, cone gas flow 20 L hr^{-1}, with a source temperature of 70 °C. Sample solutions were infused using a KD Scientific Model 100 syringe pump, at a flow rate of 5 μL min^{-1}. A lockspray solution of 200 pg μL^{-1} leucine enkephalin in 50% aqueous acetonitrile plus 0.1% formic acid was infused at 5 μL min^{-1} simultaneously, in order to enable mass accuracy correction. All sample spectra were obtained in ESI negative ion mode over a 50–4000 m/z range. MassLynx v4.1 was used for the processing of spectra. A 5 min window of the total ion chromatogram was combined, and the region from 1200–3000 m/z chosen so as to encompass all visible charge states. All samples were prepared at 1:1, 5:1 and 10:1 PC:DNA ratios, vortexed and adjusted to a final volume of 25 μL. The solutions were vortexed and equilibrated at room temperature for 5 min before infusion. DNA concentration in each sample was 10 μM. The solution was vortexed and equilibrated at room temperature for 5 min before infusion.

4.4 Circular dichroism

Circular dichroism (CD) spectra (200–400 nm) were obtained using a Jasco J-810 spectropolarimeter and 0.1 cm pathlength quartz cell, a data pitch of 1 nm, scan speed of 100 nm min^{-1} and 20 accumulations. Data were smoothed once with a 7 point Savitzky-Golay algorithm. Solutions were prepared at a 10:1 PC:G4DNA ratio in buffer (NHOAc, pH 7.0) at an oligomer concentration of 5 μM and incubated at room temperature for 1 h prior to measurement.

4.5 HPLC

Analytical reverse-phase HPLC (RP-HPLC) of the PCs was performed on an Agilent Technologies 1260 Infinity, equipped with a Phenomenex C18 reverse phase column (4.6 $\times$ 150 mm, 5 μm). The mobile phase consisted of solvent A (0.06 % TFA in water) and solvent B (0.06 % TFA in ACN:H_2O (90:10)). In each case 10 μL samples were injected and eluted with a gradient of 0–40 (% B) over 15 min at a flow rate of 1 mL min^{-1}. Detection was by photodiode array at 254 and 214 nm. Reported retention times are corrected by subtracting the elution time of a non-retained compound.

4.6 Computational methods

Computational binding affinity and docking analysis was performed using a custom version of AutoDock Vina.[36] The substrate DNA models were based on the protein data bank structures G4 (PDB ID:

2JT7)[23] and D2 (PDB ID: 1BNA).[16] The search space was specified so as to encompass the entire substrate. Five runs of each docking were performed, the results averaged and the PCs ranked by reported binding energy. The predicted docking conformations with the most favourable binding energies were averaged to obtain the final result; in all cases SD < 0.1 kcal mol^{-1}. DFT geometry optimisations and Mulliken population analysis were performed with Gaussian 09[17] in the gas phase at the B3LYP level of theory. The LANL2DZ basis set and effective core potentials were used for the Pt atoms; the 6-31G(d,p) basis set was used for all other atoms. Mulliken charges are the reduced form—hydrogens are summed into the heavy atoms. Input files are listed in the supplementary information.

5 Acknowledgments

Thanks to WSU Mass Spectrometry Facility for access to the ToF mass spectrometer. Thanks also to WSU Summer Scholarship students Yuliani Yuliani and Sandy Craze for contributions to the experimental component and Dr Elise Wright for helpful discussions. D.L.A. was supported by an Australian Postgraduate Award with WSU topup.

6 Conflicts of interest

The authors declare that they have no conflict of interest.

References

1 Ang DL, Harper BWJ, Cubo L, Mendoza O, Vilar R, Aldrich-Wright J (2016) Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers. Chemistry – A European Journal 22(7):2317–2325, DOI 10.1002/chem.201503663

2 Annesley TM (2003) Ion Suppression in Mass Spectrometry. Clinical Chemistry 49(7):1041–1044, DOI 10.1373/49.7.1041, 00933

3 Antignac JP, de Wasch K, Monteau F, De Brabander H, Andre F, Le Bizec B (2005) The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis. Analytica Chimica Acta 529(1–2):129–136, DOI 10.1016/j.aca.2004.08.055, 00224

4 Ayouni L, Cazorla G, Chaillou D, Herbreteau B, Rudaz S, Lantéri P, Carrupt PA (2005) Fast Determination of Lipophilicity by HPLC. Chromatographia 62(5):251–255, DOI 10.1365/s10337-005-0608-6

5 Bandeira S, Gonzalez-Garcia J, Pensa E, Albrecht T, Vilar R (2018) A Redox-Activated G-

Quadruplex DNA Binder Based on a Platinum(IV)–Salphen Complex. Angewandte Chemie International Edition pp 310–313, DOI 10.1002/anie.201709968@10.1002/(ISSN)1521-3773.EuroBIC14

6 Beck JL (2011) Developments in Electrospray Ionization Mass Spectrometry of Non-Covalent DNA–Ligand Complexes. Australian Journal of Chemistry 64(6):705–717, 00002

7 Bedrat A, Lacroix L, Mergny JL (2016) Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Research 44(4):1746–1759, DOI 10.1093/nar/gkw006

8 Betzer JF, Nuter F, Chtchigrovsky M, Hamon F, Kellermann G, Ali S, Calméjane MA, Roque S, Poupon J, Cresteil T, Teulade-Fichou MP, Marinetti A, Bombard S (2016) Linking of Antitumor trans NHC-Pt(II) Complexes to G-Quadruplex DNA Ligand for Telomeric Targeting. Bioconjugate Chemistry 27(6):1456–1470, DOI 10.1021/acs.bioconjchem.6b00079

9 Biffi G, Tannahill D, McCafferty J, Balasubramanian S (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. Nature Chemistry 5(3):182–186, DOI 10.1038/nchem.1548

10 Brodie CR, Collins JG, Aldrich-Wright JR (2004) DNA binding and biological activity of some platinum(II) intercalating compounds containing methyl-substituted 1,10-phenanthrolines. Dalton Transactions (8):1145–1152, DOI 10.1039/B316511F, 00052

11 Buhrman DL, Price PI, Rudewiczcor PJ (1996) Quantitation of SR 27417 in human plasma using electrospray liquid chromatography-tandem mass spectrometry: A study of ion suppression. Journal of the American Society for Mass Spectrometry 7(11):1099–1105, DOI 10.1016/S1044-0305(96)00072-4, 00307

12 Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: Sequence, topology and structure. Nucleic Acids Research 34(19):5402–5415, DOI 10.1093/nar/gkl655, cited by 0637

13 Charif R, Granotier-Beckers C, Bertrand HC, Poupon J, Ségal-Bendirdjian E, Teulade-Fichou MP, Boussin FD, Bombard S (2017) Association of a Platinum Complex to a G-Quadruplex Ligand Enhances Telomere Disruption. Chemical Research in Toxicology 30(8):1629–1640, DOI 10.1021/acs.chemrestox.7b00131

14 Davis KJ, Richardson C, Beck JL, Knowles BM, Guédin A, Mergny JL, Willis AC, Ralph SF (2015) Synthesis and characterisation of nickel Schiff base complexes containing the meso-1,2-diphenylethylenediamine moiety: Selective interactions with a tetramolecular DNA quadruplex. Dalton Transactions 44(7):3136–3150, DOI 10.1039/C4DT02926G

15 Deo KM, Ang DL, McGhie B, Rajamanickam A, Dhiman A, Khoury A, Holland J, Bjelosevic A, Pages

B, Gordon C, Aldrich-Wright JR (2018) Platinum coordination compounds with potent anticancer activity. Coordination Chemistry Reviews 375:148–163, DOI 10.1016/j.ccr.2017.11.014

16 Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE (1981) Structure of a B-DNA dodecamer: Conformation and dynamics. Proceedings of the National Academy of Sciences 78(4):2179–2183, 00775 PMID: 6941276

17 Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery Jr JA, Peralta JE, Ogliaro F, Bearpark MJ, Heyd J, Brothers EN, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09. Gaussian, Inc.

18 Hoffmann RF, Moshkin YM, Mouton S, Grzeschik NA, Kalicharan RD, Kuipers J, Wolters AHG, Nishida K, Romashchenko AV, Postberg J, Lipps H, Berezikov E, Sibon OCM, Giepmans BNG, Lansdorp PM (2016) Guanine quadruplex structures localize to heterochromatin. Nucleic Acids Research 44(1):152–163, DOI 10.1093/nar/gkv900

19 Holm AIS, Kohler B, Hoffmann SV, Brøndsted Nielsen S (2010) Synchrotron radiation circular dichroism of various G-quadruplex structures. Biopolymers 93(5):429–433, DOI 10.1002/bip.21354, 00000

20 Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Research 33(9):2908–2916, DOI 10.1093/nar/gki609

21 Kostrhunova H, Zajac J, Novohradsky V, Kasparkova J, Malina J, Aldrich-Wright JR, Petruzzella E, Sirota R, Gibson D, Brabec V (2019) A Subset of New Platinum Antitumor Agents Kills Cells by a Multimodal Mechanism of Action Also Involving Changes in the Organization of the Microtubule Cytoskeleton. Journal of Medicinal Chemistry 62(10):5176–5190, DOI 10.1021/acs.jmedchem.9b00489

22 Liu W, Zhong YF, Liu LY, Shen CT, Zeng W, Wang F, Yang D, Mao ZW (2018) Solution structures of multiple G-quadruplex complexes induced by a platinum(II)-based tripod reveal dynamic binding.

Nature Communications 9(1):1–11, DOI 10.1038/s41467-018-05810-4

23 Martino L, Virno A, Pagano B, Virgilio A, Di Micco S, Galeone A, Giancola C, Bifulco G, Mayol L, Randazzo A (2007) Structural and Thermodynamic Studies of the Interaction of Distamycin A with the Parallel Quadruplex Structure [d(TGGGGT)]4. Journal of the American Chemical Society 129(51):16048–16056, DOI 10.1021/ja075710k, 00064

24 McQuaid K, Abell H, Gurung SP, Allan DR, Winter G, Sorensen T, Cardin DJ, Brazier JA, Cardin CJ, Hall JP (2019) Structural Studies Reveal Enantiospecific Recognition of a DNA G-Quadruplex by a Ruthenium Polypyridyl Complex. Angewandte Chemie International Edition 58(29):9881–9885, DOI 10.1002/anie.201814502

25 Morel E, Poyer F, Vaslin L, Bombard S, Teulade-Fichou MP (2016) Photoactivatable platinum(II) terpyridine derivatives for G-quadruplex DNA double anchoring. Inorganica Chimica Acta 452:152–158, DOI 10.1016/j.ica.2016.02.033

26 Morel E, Beauvineau C, Naud-Martin D, Landras-Guetta C, Verga D, Ghosh D, Achelle S, Mahuteau-Betzer F, Bombard S, Teulade-Fichou MP (2019) Selectivity of Terpyridine Platinum Anticancer Drugs for G-quadruplex DNA. Molecules 24(3):404, DOI 10.3390/molecules24030404

27 Ogasawara S (2018) Transcription Driven by Reversible Photocontrol of Hyperstable G-Quadruplexes. ACS Synthetic Biology 7(11):2507–2513, DOI 10.1021/acssynbio.8b00216

28 Parrish RM, Sherrill CD (2014) Quantum-Mechanical Evaluation of π–π versus Substituent-π Interactions in π Stacking: Direct Evidence for the Wheeler–Houk Picture. Journal of the American Chemical Society 136(50):17386–17389, DOI 10.1021/ja5101245, 00000

29 Reedijk J, Lohman P (1985) Cisplatin: Synthesis, antitumour activity and mechanism of action. Pharmacy World & Science 7(5):173–180, DOI 10.1007/BF02307573, cited by 0082

30 Rodger A, Dorrington G, L Ang D (2016) Linear dichroism as a probe of molecular structure and interactions. Analyst 141(24):6490–6498, DOI 10.1039/C6AN01771A

31 Sabharwal NC, Mendoza O, Nicoludis JM, Ruan T, Mergny JL, Yatsunyk LA (2016) Investigation of the interactions between Pt(II) and Pd(II) derivatives of 5,10,15,20-tetrakis (N-methyl-4-pyridyl) porphyrin and G-quadruplex DNA. JBIC Journal of Biological Inorganic Chemistry 21(2):227–239, DOI 10.1007/s00775-015-1325-8

32 Shannon RD, Prewitt CT (1970) Revised values of effective ionic radii. Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry 26(7):1046–1048, DOI 10.1107/S0567740870003576

33 Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proceedings of the National Academy of Sciences 99(18):11593–11598, DOI 10.1073/pnas.182256799, 00952

34 Sidey V (2016) On the effective ionic radii for ammonium. Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials 72(4):626–633, DOI 10.1107/S2052520616008064

35 Talib J, Green C, Davis KJ, Urathamakul T, Beck JL, Aldrich-Wright JR, Ralph SF (2008) A comparison of the binding of metal complexes to duplex and quadruplex DNA. Dalton Transactions (8):1018–1026, DOI 10.1039/B715458E, cited by 0019

36 Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31(2):455–461, DOI 10.1002/jcc.21334, cited by 0949

37 Wheate NJ, Taleb RI, Krause-Heuer AM, Cook RL, Wang S, Higgins VJ, Aldrich-Wright JR (2007) Novel platinum(II)-based anticancer complexes and molecular hosts as their drug delivery vehicles. Dalton Transactions (43):5055–5064, DOI 10.1039/B704973K, 00055

38 Zhai Q, Gao C, Ding J, Zhang Y, Islam B, Lan W, Hou H, Deng H, Li J, Hu Z, Mohamed HI, Xu S, Cao C, Haider SM, Wei D (2019) Selective recognition of c-MYC Pu22 G-quadruplex by a fluorescent probe. Nucleic Acids Research 47(5):2190–2204, DOI 10.1093/nar/gkz059

Chapter 4

SOMSpec: a machine learning approach to protein secondary structure prediction from Infrared and Circular Dichroism spectra

Figure 4.1: SOMSpec logo.

During my time at Warwick University, in addition to designing and performing molecular dynamics simulations, there was a request to redesign some existing inhouse software for protein secondary structure prediction from circular dichroism (CD) spectra. The existing program, named SSNN,[20] was written a couple of years before by a student of the group. The program was specific to CD data, was slow

(requiring multiple hours to train), required data to be in a very strict format and lacked any way to add additional features, such as support for variable step sizes, or plotting of melt curves. As I had amassed a large number of CD and SRCD spectra comprising various DNA motifs, I also saw the potential in predicting secondary structure of nucleic acids. There was the additional desire for improved flexible and potentially enable prediction from IR and RAMAN spectra. As a result, I wrote SOMSpec. Over time SOMSpec has evolved into a fully fledged software suite able to be trained on any type of spectral data, and predict any number of properties provided an appropriate validated training set exists. It can be operated either by command line or graphical interfaces. SOMSpec features various modules for pre- and post- data formatting. The implementation details are included in the manuscript and an installer for Linux or Windows can be downloaded from somspec.org. It is free software licensed under GPLv3. The architecture of SOMSpec is outlined in the flowchart in Figure 4.2.

Following is a brief description of how SOMSpec operates. The implementation specifics are described in the manuscript. Consider a single CD spectrum of a protein. By stripping the wavelength data from this it becomes a simple one dimensional vector of values, with each value representing a given magnitude for each step along the wavelength range. If enough of these vectors are placed in a grid, oriented along the z axis, and a machine learning algorithm allowed to recursively operate on and change these vectors, or nodes, an interesting emergent property is exhibited. The nodes, each representing a single spectrum, are topologically grouped by features. This is the basis for what are known as Kohonen self organising maps (SOM).[21] A pictorial representation of a self organising map is presented as Fig. 1 in the manuscript. What is important to note is that each node represents an arbitrarily large vector and is equally spaced across a cartesian addressable grid. If each node was summed and plotted as a 3D surface or heat map it would look like Figures 4.3 and 4.4, respectively.

In Figure 4.4 the topological grouping of secondary structure properties is readily apparent. The midpoints and vector sum plots are simplistic representations

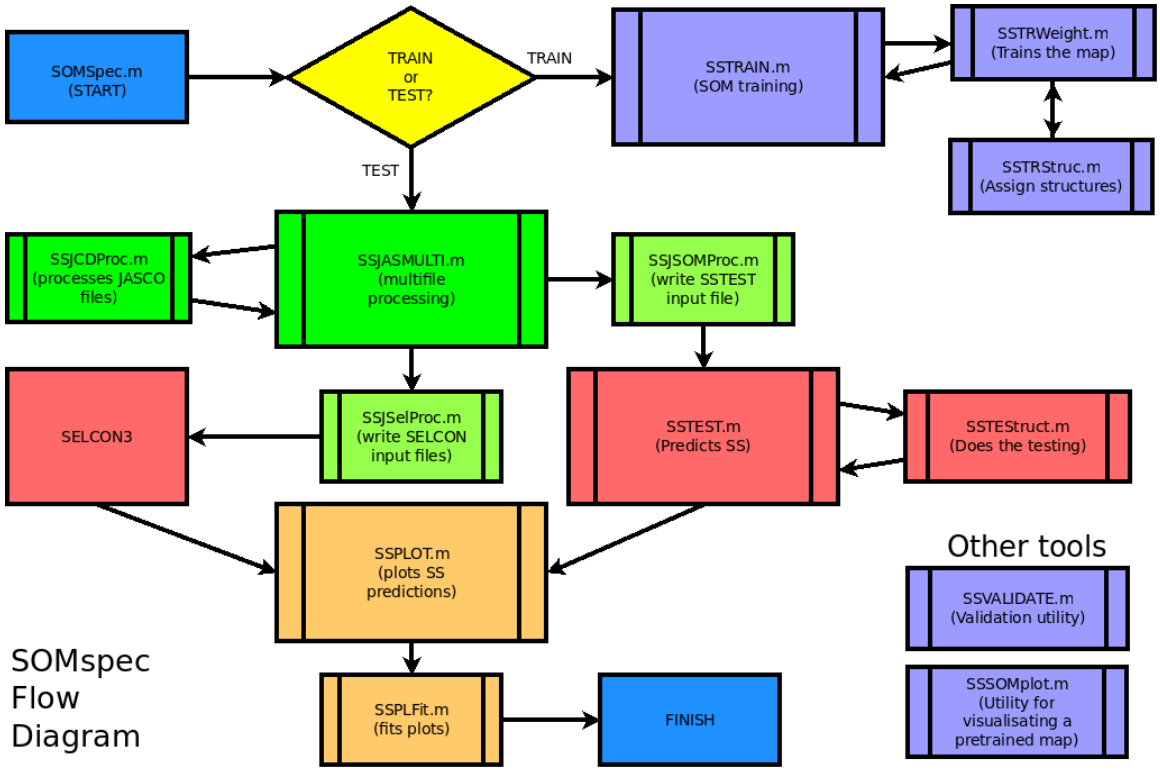


Figure 4.2: Flowchart of how the main SOMSpec program operates. The two main decision trees are TRAIN, where a map is created from the training set, and TEST, where a trained map is used to make predictions from a user provided spectrum.

of the midpoint value and sum of each vector, respectively. The Property 1–5 plots represent the distribution of α-helix, β-sheet, bends, turns and other secondary structure content of the trained map. It is evident how nodes containing significant Property 1 (α-helix) are arranged topologically away from those containing high Property 2 (β-sheet), as is expected. Property 6 indicates the locations of each training set spectrum within the map (yellow squares).

A useful feature of the representation in Figure 4.4 is the ability to quickly analyse a training set for deficiencies or odd spectra. An example in this case is node (14,22) of the Property 1 plot (dark blue square upper centre). This node is part of the training set, as can be confirmed by checking the Property 6 plot, and stands out for not containing the same percentage α-helix secondary structure as the surrounding nodes. This indicates the spectrum is either of poor quality or possesses incorrectly annotated secondary structure information.

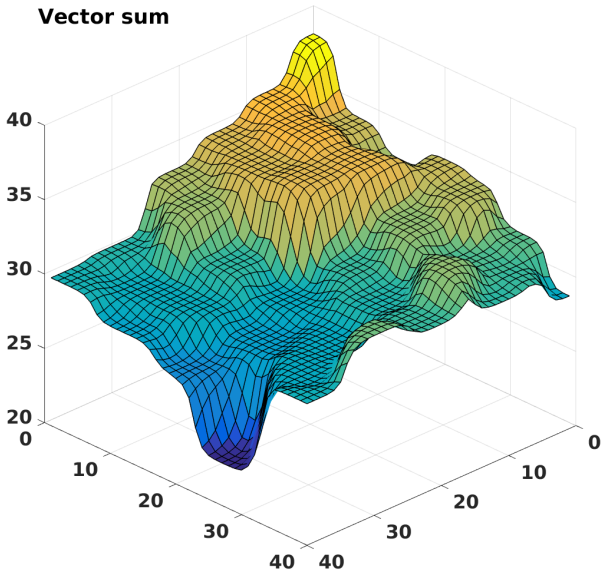


Figure 4.3: 3D surface plot of a 40 × 40 node self organising map trained on 47 protein training set of infrared spectra. The values for each node are summed and presented as a single value.

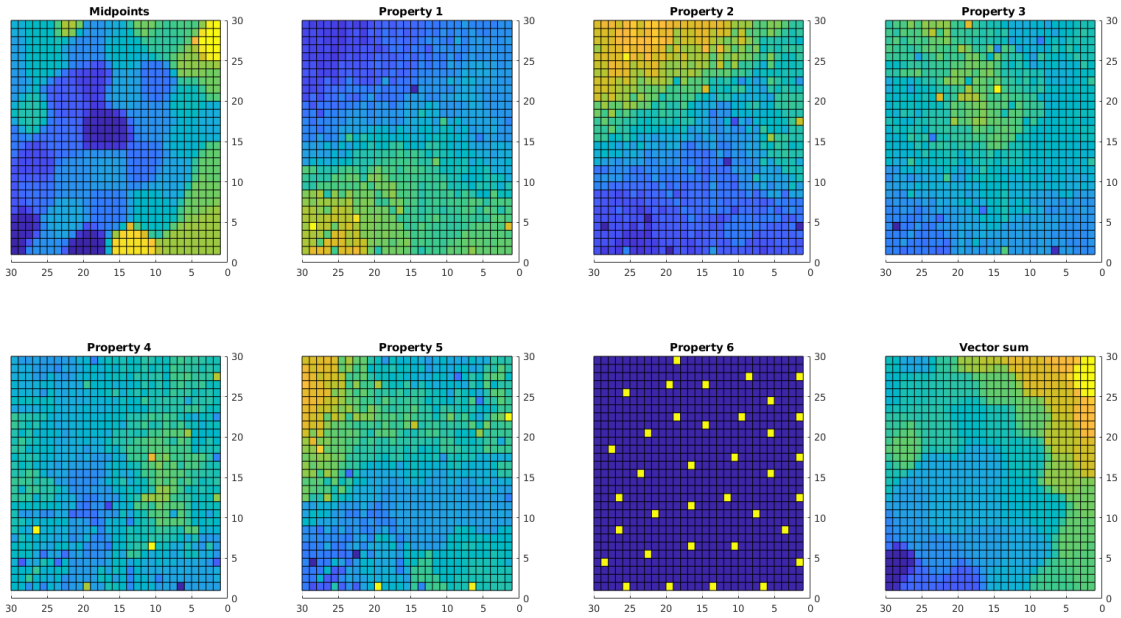


Figure 4.4: Various plots of a 30 × 30 node self organising map displaying the emergent topological grouping of information. The plots labelled Property 1–5 represent the total α-helix, β-sheet, bends, turns and other secondary structure content of each node on the trained map. The SOM was trained on 47 transmission infrared protein spectra annotated with five secondary structure properties.

SOMSpec utilises this emergent property of self organising maps. It reads in a set of spectra with known structure properties—the training set—and generates a map which can be used to match with experimental spectra and therefore make predictions on the properties of that spectra. Clearly, the choice of algorithms and data format is key to the success of this method. Significant code review and validation testing has so far been performed, and although there are a number of variables which may be chosen when training, many are interdependent and are automatically calculated in this implementation. The user only needs to decide on the map size, how many properties to predict (which is ultimately predicated on the quality of the training set) and number of iterations. Extensive validation testing at map sizes between 10–100 and iterations between 10,000–100,000 has so far revealed the root mean square deviation (RMSD) in predictions varies less than 10%. This is in part due to the robust design and scaling of the SOMSpec code, so that in practice there is little change in the accuracy of the results when varying the map size or number of iterations in the training process.

A suitable training set consists of multiple spectra with annotated structure properties. Initial development used an existing CD training set of 47 proteins annotated with five secondary structure properties: α-helix, β-sheet, turns, bends, and other.[20] Subsequent development utilised the globular protein SP175 dataset, comprising 71 globular protein CD spectra from 175–240 nm. This reference set is considered the gold standard, and is published on the Protein Circular Dichroism Data Bank (PCDDB).[22] The standard PCDDB annotations are from DSSP.[23] These properties were combined into the five standard classes as follows: helix (H+I+G), sheet (B+E), bonded turn (T), bend (S), loop or irregular (C). A curated training set of various double-stranded and G-quadruplex DNA sequences is currently in progress but should provide the ability to somewhat predict DNA structure, or at the very least discriminate between parallel and anti-parallel G-quadruplex topologies.

When choosing a training set it is important to validate suitability, specifically the informational content of the set. SOMSpec incorporates an automated leave-one-out-validation process (LOOV) for this purpose. The process involves removing

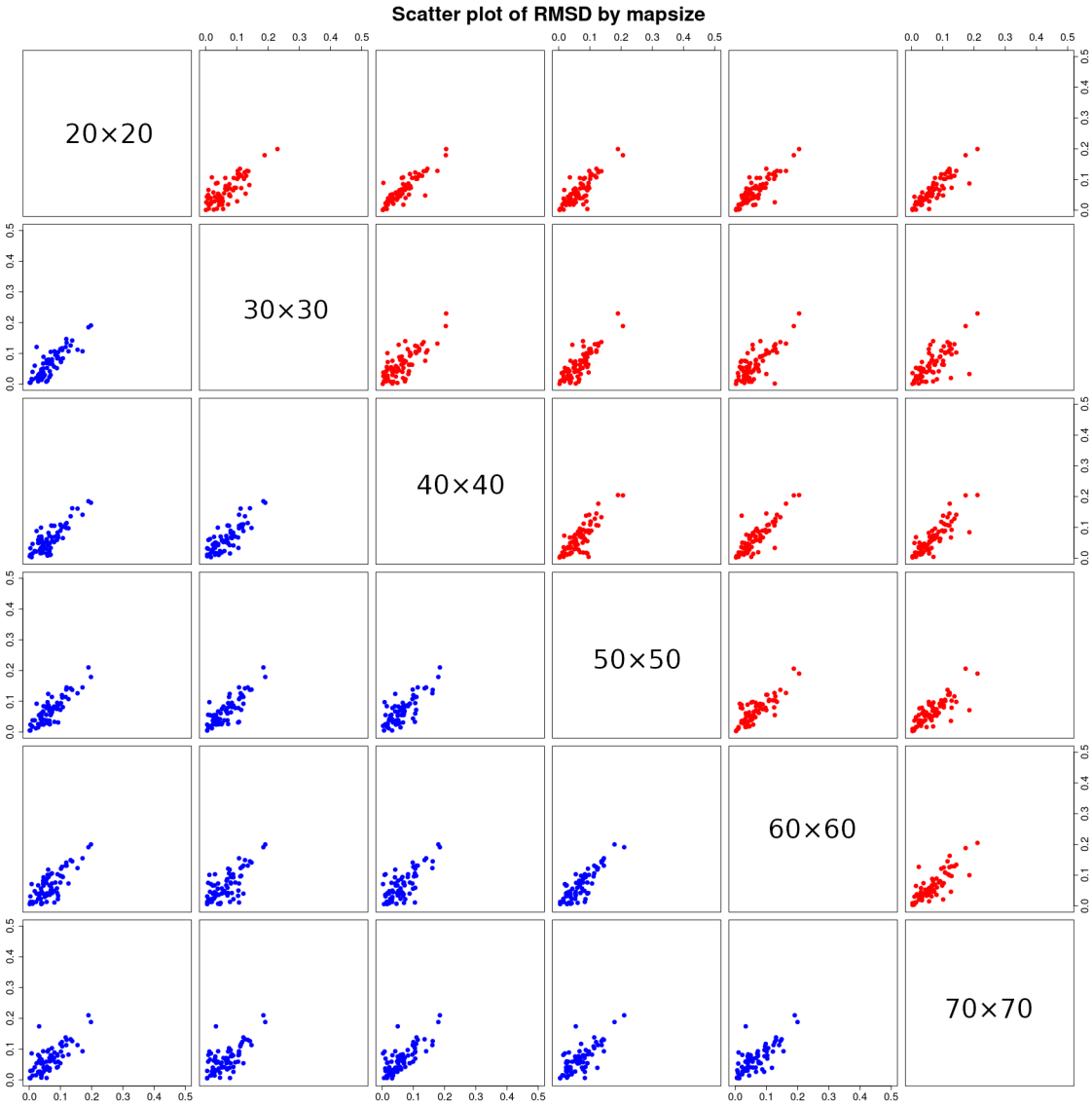


Figure 4.5: Scatter plots depicting the root mean square deviation of predicted secondary structure properties of the SP175 training set at map sizes of 20×20–70×70 and 20,000 (red) and 60,000 (blue) iterations.

one spectrum from the training set, training on the remaining set, then predicting the properties of the removed spectrum, and finally, comparing the predicted results with the sets own annotation values. The validation results of the SP175 data set at two epoch levels and a range of map sizes and are presented in Figure 4.5. Each data point represents one protein spectrum from the set, with RMSD of the predicted vs real properties plotted. The majority of points are clustered between

0.0–0.1 RMSD. It is difficult to graphically represent such multidimensional data but this representation is useful to compare the effects of varying map sizes and iterations. It also allows the identification of proteins with potentially poor annotations.

Choice of the IR training set was guided by best practice. Specifically, the 1600–1700 cm^{-1} region was chosen, corresponding to the protein Amide I band. It is well understood that the exact band maximum is determined by the hydrogen bonding and backbone conformation of the protein, hence it contains much of the secondary structure information of a protein. This region also exhibits the most intense absorption in a protein IR spectrum. All spectra were normalised to create the training set.

A draft of the manuscript follows. This document outlines the underlying machine learning theory, based on Kohonen maps.[21] As the primary author, I designed and developed all software, artwork and graphics and wrote all of the text, excluding the transmission IR and RAMAN training set acquisition which was performed by Marco Pinto-Corujo (University of Warwick). I also designed, developed and host the associated website somspec.org where the software will be distribute in binary and source code formats along with training sets where available with source code hosted on Gitlab. An important planned feature includes the prediction of CD spectra given a PDB structure, or simply secondary structure content, which would be implemented as a reversal lookup of a trained map. This work will be submitted to the RSC journal PCCP in the near future.

Cite this: DOI: 10.1039/xxxxxxxxxx

SOMSpec: a machine learning approach to protein secondary structure prediction from Infrared and Circular Dichroism spectra

Dale L. Ang,*[a] Marco Pinto-Corujo,[b‡] Nikola Chmel,[b‡] Alison Rodger[c] and Janice Aldrich-Wright[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Protein secondary structure prediction is achievable from infra red (IR) spectra (transmission and Raman) using a machine learning algorithm based on self organising maps. We show that this method can predict protein secondary structures information from IR data within <5%. With appropriate training sets, SOMSpec can be directly used for a wide range of other applications. We show that our approach and our implementation of it is flexible enough to accept non-spectral data such as the Iris flower data set as well as undertake secondary structure fitting from circular dichroism spectra. We anticipate direct application to structure prediction using other spectroscopic techniques such as Raman and Raman Optical activity and to other macromolecules such as DNA when high quality training sets that cover the structural space become available. Two two high quality reference sets of protein spectra - set of 47 transmission IR spectra and 46 RAMAN spectra have also been acquired and are provided with this software as robust training sets. Applications range from fast protein structure prediction to antibody testing, forensics and other fields. The software is licensed under GPL3 and we encourage contribution from other researchers.

Introduction

Interpretation of data obtained via spectroscopic techniques such as infrared (IR) and circular dichroism (CD) spectroscopy, is of great importance to industrial and pharmacological research, particularly in quantifying the secondary structure (SS) content of biomolecules such as proteins and nucleic acids. Specifically, applications include drug design, and biopharmaceutical quality control due to the inherent structure-function relationships of biomolecules. The human genome is estimated to contain around 21,000 protein-encoding genes while the total number of proteins in human cells is estimated to be upwards of 250,000. Currently, the protein data bank contains structures for approximately 10% (∼28,000)[1] of these proteins with most of the structures being crystallographic so not necessarily representative of solution behaviour. Given this, there is a significant deficiency in the structural knowledge for a vast array of proteins and environments.

The two most common methods of obtaining protein structure, XRD and NMR, have a number of drawbacks making them unsuitable for many proteins. X-ray diffraction requires crystals of sufficient quality which often very difficult to obtain, and NMR requires high quantities of stable protein for the extended time scales required of the experiments. These techniques are also limited in terms of concentration and environment. The circular dichroism spectrum of a protein, while not giving an atomic resolution structure, can often be analysed to provide reasonable estimates of secondary structure content of that protein as long as the concentration is between 0.01 mg/mL and 10 mg/mL and the buffer components do not absorb much light down to 190 nm. A variety of computational methods exist to tease out reasonable estimates of the SS content from CD spectra including SELCON3, CONTIN, CDsstr and our own SSNN (secondary structure neural network).[2] Other neural network approaches such as K2d[3] and SOMCD[4] might be as successful but are only available as pretrained programs with dated and limited reference sets.

We recently turned to infrared spectroscopy to cope with samples with significantly absorbing buffer components or concentrations above 10 mg/mL and were surprised to find a lack of robust

[a] Nanoscale Organisation and Dynamics Group, Western Sydney University, Campbelltown, NSW, 2560, Australia.
[b] Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom.
[c] Molecular Sciences, Macquarie University, NSW, 2109, Australia.

† Electronic Supplementary Information (ESI) available: training data sets and validation results. Source-code and precompiled binaries for Linux and Windows are available for download from https://www.somspec.org. See DOI: 10.1039/b000000x/

validated methods to extract SS despite many people using it for the purpose. Most approaches that are used involve curve fitting to the spectrum or its second derivative and assigning each resulting Gaussian or Lorenzian to a secondary structure motif. It logically seemed to us that an approach analogous to those used with CD data ought to be preferable. However, this requires both a good reference set of spectra covering the structural space and a methodology to extract SS.

In this paper we present SOMSpec, the results of our effort to improve our previous algorithm SSNN[2] to encompass infrared protein spectra and indeed to almost anything else where the reference training set and test data (such as an IR or CD spectrum) is associated with structural (or other descriptor) data. SOMSpec incorporates a suite of programs for automating the entire process of training and prediction and validation tools for training set development. SOMSpec is currently implemented as Matlab compatible code making a reproducible, robust and reliable approach to protein secondary structure analysis from infrared absorbance data. Our approach was to provide a classical SOM-based tool for the classification and property prediction of any range of inputs, with a focus on protein secondary structure from spectral data. Our previous implementation focused solely on predicting the SS of a protein from its CD spectrum with predetermined un-flexible parameterisation. Our goal here was to provide a tool that was directly applicable to other spectral, with immediate emphasis on IR absorbance but allowing for more generalised inputs. We have therefore generalised the approach and improved the algorithms used.

The program suite.

SOMSpec can be run either through a graphical user interface (GUI) or from a terminal. The terminal version is currently the more flexible and the only way to access some of the more advanced features. Several of the main programs are briefly described.

Features of the main SOMSpec application include:

- All input file formats for training and predicting are column based plaintext comma, tab or space separated. Number of columns and rows is arbitrary. Each column represents a single spectrum and the software automatically determines the number of columns and field separator at runtime.
- Input formats are internally interpolated so there is no need to match steps as long as wavelength domain is the same
- Number of properties is flexible, from 1 to many but in practice <6.
- Visualisation options: realtime graphical representation of map training and completed maps, interpolated spectra, all switchable at runtime. The realtime visualising of the map training is great for teaching/understanding SOM concepts and how training/learning works. Also useful for debugging.
- Automated leave-one-out-validation (LOOV). This means it automatically processes a training set and trains, then tests.
- Scaling of input spectra. this is especially useful for CD as alpha content affects magnitude

Additional tools included with the SOMSpec suite:

- automatically build training sets from PCDDB data, evaluate melts with plotting of SS vs Temp, generation of SELCON input files for validation against existing toolkits/software
- GUI access to most tools (common) from Windows/Linux executable

Program descriptions.

The following is a list of programs comprising the SOMSpec suite.

`SOMSpec.m`: Command-line version of the main program and the most fully featured version. Ability to controlling all training/predicting and validation operations. Additional features include the ability to observe the training in real time, automatic concentration/normalisation of spectra, support for any number of discrete property features, and several other more advanced features. Useful for running on headless servers and remote computers.

`SOMSpecGUI.m`: Graphical interface version of the main program. Limited in features compared to the command-line version but provides the most commonly used features such as training, prediction, plotting of trained self-organising maps, and validation.

`SSJAS.m`: Produces correctly formatted SELCON input files from Jasco ascii files. Useful for testing and comparing results between SOMSpec and SELCON. Input is 1 CD spectrum.

`SSJASMULTI.m`: Produces correctly formatted SELCON input files from Jasco ascii files. Useful for testing and comparing results between SOMSpec and SELCON. Ability to process multiple spectra at once. Input is a directory of CD spectra. Data must be formatted from long to short wavelength, ie. 240 - 190 nm and span the required training sets.

`SSJCDProc.m`: Reads and processes standard Jasco Spectra Manager output files. Flexible enough to accept any version. Provides automatic translation from Jasco format to SELCON or SOMSpec input.

`SSPCDDB.m`: Produces correctly formatted training sets from PCDDB .pcd files. Input is a directory of CD spectra. Data MUST be formatted from long to short wavelength, ie. 240 - 190 nm and span the required training sets.

`SSPLFIT.m`: Plot the results of multiple spectral predictions, fit sigmoidal curves and output values. Extremely useful for processing CD melts.

`SSPLOT.m`: Plots the results from a multi-file processing. See SSPLFIT.m.

`SSSOMPLOT.m`: Reads in a trained SOM and plot the results. Similar output to the real-time plots option but accessible after a run. Also accessible from the GUI.

`SSVALIDATE.m`: Automated Leave-One-Out-Validation (LOOV) program for SOMSpec. Creates properly formatted training set and input files that allow iteration through a training set.

Implementation

Self-organising map and protein secondary structure prediction.

A self-organising map (SOM),[5] is a type of artificial neural network that can be trained in an unsupervised manner. Unsupervised learning is where unclassified data (in our case the spectrum of a protein, nucleic acid, and so on) is sorted by identifying commonalities.

To predict the secondary structure content of a protein given it's CD spectrum our implementation takes several steps. 1) training a SOM from a reference set, 2) using that SOM as the basis of predicting the SS content from a "test" spectrum. Pictorially, our SOM is represented in figure 1. In practice each step is complicated by many factors.

A SOM achieves this first by organising a reference set into a 2 dimensional map (training)–the distance between nodes reflecting how different the data sets are. The missing nodes are filled by creating weighted averages of the reference set nodes. All nodes are then assigned SS properties matching as closely as possible. The result is a topologically ordered map which mimics the input training set with regions corresponding to particular features of the input space. The prediction step involves finding the most closely representative nodes to the given test spectrum and calculating the secondary structure as a weighted sum of those nodes as output.

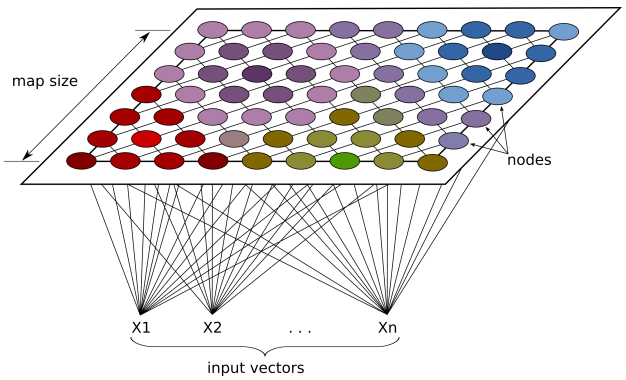


Fig. 1 Pictorial representation of a self-organising map.

The three phases.

Training occurs in two phases. Phase 1: map construction, and Phase 2: structure assignment. Phase 3 can be considered the prediction of properties given a test spectrum.

Phase 1: map construction. Phase 1 begins with the initialisation of an untrained randomised map. The reference set (input vector space), consisting of columns representing each spectrum, is randomly sampled a user-definable number of iterations. For each iteration a single node of the map is identified as the best matching unit (BMU) of the input vector under consideration calculated as the Euclidean distance between the node vector and the input vector (equation 1). The closest node becomes the BMU.

$$Dist = \sqrt{\sum_{i=0}^{n}(I_i - W_i)^2} \tag{1}$$

where I_i and W_i are the current input (reference set member) and node vectors respectively, and n is the number of values in the vector. All other nodes within the current neighbourhood of the BMU are influenced according to equation 2

$$W(t+1) = W(t) + \theta(t)L(t)(I(t) - W(t)) \tag{2}$$

Here, t is the current time step/iteration, $\theta(t)$ is the radial bias function (responsible for the spatial convolution of the xxx) (equation 3) which is the region of influence of the current learning step, also known as the neighbourhood, L is the learning rate. Setting the BMU as the active node, all other map nodes within the neighbourhood are biased in a distance dependent manner using the radial biasing function

$$\theta(t) = \exp\left(\frac{-(distance\ from\ BMU)^2}{2 \times radius^2}\right) \tag{3}$$

The radius value initialises at half the map size. Values for both the radius and learning rates decay exponentially following the general equation:

$$Z(t+1) = Z(t)\exp\left(\frac{-t}{c}\right) \tag{4}$$

where Z represents the radius or learning rate. An important aspect of this decay function is its reliance on the time constant, c. The time constant is calculated automatically according to equation 5.

$$c = \frac{Number\ of\ iterations}{ln(radius)} \tag{5}$$

Once the total number of iterations is reached Phase 2 commences.

Phase 2: assigning SS content to nodes. In Phase 2 begins the input space (training set) is sequentially sampled and a best matching node or unit (BMU) is identified for each input vector according to a NRMSD calculation (equation 6). This node is then assigned structural parameters accordingly. Surrounding nodes are assigned structural parameters accordingly.

Once the input space is exhausted, training is complete. Various plots of the trained map and each property are displayed for convenience and as a visual check that the training completed successfully (see ESI). The map weighting and property data along with a parameters file containing the map details such as learning rate, map size, number of iterations, and time constant is output as plain comma separated text files. Realtime visualisation is a useful feature which, when enabled, assists the user to understand the processes involved in SOM training.

Phase 3: predicting SS content from spectra. A user definable number of BMUs may be specified (default is 5) and are ranked according to similarity to the input vector. Once the desired number of BMUs is found, the property contribution of each BMU to the final node value is calculated using a distance dependent

weighting.

Prediction of unknown properties for a test data set (in our case a measured spectrum from which we wish to determine protein SS), follows in a similar way to the property assignment training. The only current caveat, apart from issues of data quality, is that the row (spectral) dimension of the test set must match the training set used. Any number of columns can be present and is automatically calculated at run time. Tests are run iteratively against each SOM node to identify a user-definable number of BMUs for each input test item. These BMUs are ranked by similarity to the input and the contribution of each is calculated in a distance dependent manner following equation (1.3). In the final step the results are normalised and assigned as the property prediction. Estimation of spectra is performed by interpolating across the BMUs and is visualised in the output along with a two dimensional representation of the SOM with BMUs present (as illustrated in Figure 2). The predicted spectral NRMSD value is calculated as equation 6

$$NRMSD = \frac{\sqrt{\frac{\sum(X-Y)^2}{N}}}{M-m} \tag{6}$$

Figure 2 BMU positions and predicted spectrum output after a prediction The top BMUs contributing to the property prediction are labelled displayed. The predicted spectrum (blue) overlays the input spectrum (black) along with residuals (red).

It should be noted that in this implementation the only necessary user defined options are the map size and number of iterations. All other parameters are calculated from these at runtime. For research and testing purposes the code is readily reconfigurable with respect to other parameters such as the number of BMUs and the size of the map; additional biasing functions are present in the code to be implemented if desired. Recommendations on approximate values for various data sizes can be found in the ESI.

Results and discussion.

Leave one out validation.

The SOMSpec code was validated using a leave-one-out protocol. (note here that the results are not directly comparable to previous code as the NRMSD is calculated differently). Comparisons with respect to our previous code from CD fitting data shows that SOMSpec more consistently predicts more accurately that SSNN.

IR data set.

A primary driving force for the development of SOMSpec was the ability to predict protein SS content from IR spectra. Such an approach has only recently become possible due to the availability of a high quality (47 protein) reference set collected on 10 mg mL^{-}1 protein solutions in a CaF_2 cell collected at 2 cm^{-1} resolution.

Following literature precedent we normalised the spectra to have maximum magnitude of 1 between 1600 and 1700 cm^{-1} and considered only the data in this range. A leave-one-out-validation validation (LOOV) was performed on this data set for a variety of map sizes and iterations. Complete results are provided in Table 1 of the ESI. It was found that for this size dataset (46 proteins) with this level of information content (6 structure motifs) a map size of 40×40 with 40,000 iterations gave the best trade-off between training time and accuracy of results in terms of final NRMSD values.

Current limitations

The application is single threaded. Training a large map only takes a few minutes on a recent processor (eg. Intel Core i5-8250U). Predictions only take less than two seconds. The code has been validated to very large maps (80×80) and a high number of iterations (100,000). There is no technical reason (other than memory usage) why it couldn't be extended to further as the code supports arbitrary sizes.

Conclusions

We have developed a software package for the prediction of protein SS from IR spectra. The method is flexible enough that various other spectrophotometric or other classification data may be fed in and property predictions obtained as, internally, the code does not differentiate between protein structure or any other property. SOMSpec aims to be an easy to use program for training neutral networks tailored towards the spectroscopist. There are several useful modules with more in the planning and implementation stage as the needs arise. SOMSpec was developed to scratch an itch in our own research and has quickly been adopted by as a rapid and easy to use tool for secondary structure prediction and it is hoped will be similarly adopted to the wider scientific community. Work has already begun on further reference training sets and in the future it is envisaged that Raman and any other spectroscopic techniques will be exploited for their property prediction using this tool. The SOMSpec code is currently compatible with any platform supporting recent versions of Matlab (tested on Matlab 2018).

Methods.

Protein samples

A set of 33 Lyophilised protein powders from Sigma Aldrich for solid state infrared and a different set of 32 for Raman prepared as described next.

Transmission infrared sample preparation and instrumentation

Between 1 and 5 mg of the protein powder was first grounded in a mortar and after mixed with also ground KBr (s) to obtain a 1 to 10% dilution. Next, the KBr and protein mixture was compressed by means of a hydraulic press between 5 and 10 kpsi for 1 min and a pellet was obtained. NaCl windows in a PIKE Technologies cell were used to hold the pellet in the way of the beam.

The measurements were carried out with a Jasco FT/IR-4200 spectrometer with a 4 cm^{-1} resolution, cosine apodization, TGS detector and accumulated over 100 scans in the range between 400–4000 cm^{-1}. A 10 L min^{-1} flow rate N_2(g) was used to purge the instrument from water vapour: the interferometer first until the level of water vapour reached a minimum then only the sample chamber during the measurements.

A water spectrum was collected separately with CaF_2 windows was subtracted using the small water band in the region between 1800–2400 cm^{-1} as a reference.

Raman sample preparation and instrumentation

Between 5 and 10 mg of the protein powder was first ground in a mortar and then introduce in a 4×4 mm Starna quartz cuvette for the measurement. The instrument used was a Thermo Scientific DXR SmartRaman spectrometer with a charge coupled device (CCD) camera and a continuous wave (CW) laser of 633 nm excitation wavelength and 6–8 mW output power. Due to extreme background fluorescence, some proteins were externally photobleached with a homemade photobleaching unit consisting of CW diode lasers of 532 and 635 nm excitation wavelength and 200 mW output power previous to the measurement. The number of accumulations ranged from 1000–9999 depending on the amount of fluorescence found and the spectral region from 50–3500 cm^{-1}.

Although the smart background option was selected, a further background subtraction was performed to eliminate artefacts in some regions of the spectrum for which an empty cuvette spectrum was previously collected. Moreover, a cubic spline fit method was applied to baseline the spectra and after scaled by the interval method.

Conflicts of interest

There are no conflicts to declare.

References

1 P. W. Rose, B. Beran, C. Bi, W. F. Bluhm, D. Dimitropoulos, D. S. Goodsell, A. Prlić, M. Quesada, G. B. Quinn, J. D. Westbrook, J. Young, B. Yukich, C. Zardecki, H. M. Berman and P. E. Bourne, *Nucleic Acids Research*, 2011, **39**, D392–D401.

2 V. Hall, A. Nash and A. Rodger, *Analytical Methods*, 2014, **6**, 6721–6726.

3 M. A. Andrade, P. Chacón, J. J. Merelo and F. Morán, *Protein Engineering*, 1993, **6**, 383–390.

4 P. Unneberg, J. J. Merelo, P. Chacón and F. Morán, *Proteins: Structure, Function, and Bioinformatics*, 2001, **42**, 460–470.

5 T. Kohonen, *Biological Cybernetics*, 1982, **43**, 59–69.

Chapter 5

Other outputs

This chapter highlights additional research that was not the major focus of my PhD but where my contributions were sought from collaborators. These include published articles, journal covers, conference abstracts, posters, and grants and scholarships received. All are presented in chronological order with most recent first and are accompanied by an explanation of relevance to my main research.

5.1 Additional refereed articles

The following is a list of additional published papers that I have co-authored during my PhD as part of ongoing research undertaken within my research group and broader collaborators. The paper reference, abstract, impact and my contribution are provided. Citations were obtained from Google Scholar and are current as of 20 November 2019.

Paper 1 - Review: Deo, K. M.; **Ang, D. L.**; Gordon, C. P.; Aldrich-Wright, J. (Nov 2018). *Platinum coordination compounds with potent anticancer activity*. Coordination Chemistry Reviews. DOI: 10.1016/j.ccr.2017.11.014. (Journal Impact Factor: 14.499, Citations = 29).[24]

Abstract: Many international researchers have strived to understand the mechanism of action or improve the efficacy of inorganic coordination compounds

that have been identified to exhibit anticancer activity. The inherent challenges of chemotherapy demand that new strategies be developed utilising different mechanisms of action to interrupt the cellular machinery of cancer cells. In Australia, we have benefited from the research of colleagues who have influenced modern platinum chemistry by contributing to our understanding of platinum oxidation and reduction, the mechanism of action of cisplatin, and unique design strategies for new platinum complexes. The purpose of this review is to provide some background in the history and development of platinum(II) and platinum(IV) complexes.

Impact: Research for this review provided a way for me to keep abreast of the current research on this important category of biologically active compounds. The recent advances in platinum(IV) complexes yields an interesting and functionally convenient way to design these complexes with additional payloads and opens many new avenues to making more selective and effective compounds.

Contribution: All figures and illustrations, significant writing, and final editing.

Paper 2 - Original research: Pinto-Corujo, M.; Sklepari, M.; **Ang, D. L.**; Millichip, M.; Reason, A.; Goodchild, S. C.; Wormell, P.; Amarasinghe, D. P.; Lindo, V.; Chmel, N. P.; Rodger, A. (Aug 2018). *Infrared absorbance spectroscopy of aqueous proteins: comparison of transmission and ATR data collection and analysis for secondary structure fitting*. Chirality. DOI: 10.1002/chir.23002. (IF: 1.833, Citations = 1).[25]

Abstract: Attenuated total reflectance (ATR) infrared absorbance spectroscopy of proteins in aqueous solution is much easier to perform than transmission spectroscopy, where short path-length cells need to be assembled reproducibly. However, the shape of the resulting ATR infrared spectrum varies with the refractive index of the sample and the instrument configuration. Refractive index in turn depends on the absorbance of the sample. In this work, it is shown that a room temperature triglycine sulfate detector and

a ZnSe ATR unit can be used to collect reproducible spectra of proteins. A simple method for transforming the protein ATR spectrum into the shape of the transmission spectrum is also given, which proceeds by approximating a Kramers-Krönig–determined refractive index of water as a sum of four linear components across the amide I and II regions. The light intensity at the crystal surface (with 45° incidence) and its rate of decay away from the surface is determined as a function of the wave number–dependent refractive index as well as the decay of the evanescent wave from the surface. The result is a single correction factor at each wave number. The spectra were normalized to a maximum of 1 between 1600 cm^{-1} and 1700 cm^{-1} and a self-organizing map secondary structure fitting algorithm, SOMSpec, applied using the BioTools reference set. The resulting secondary structure estimates are encouraging for the future of ATR spectroscopy for biopharmaceutical characterization and quality control applications.

Impact: Machine learning techniques are entirely dependent on high quality training sets, and SOMSpec is no different. Given that there is a dearth of quality IR spectral training sets of proteins we set out to satiate this need and provide a useful, and freely available, training set for such an application. Transmission IR is a convenient way of obtaining a spectrum but is limiting in several ways including path length, solubility of the analyte of interest, and other factors. Attenuated Total Reflectance (ATR) IR spectroscopy is a related technique that alleviates some of the problems, but at the cost that the spectrum is slightly different to a transmission IR spectrum. This paper set out to compare and contrast these differences. A future paper, soon to be published, actually deals with the theoretical aspects of converting an ATR IR spectrum to a transmission spectrum.

Contribution: Wrote all code and processed spectra (see SOMSpec, Chapter 4), minor writing.

Paper 3 - Original research: Šmidlehner, T.; Kurutos , A.; Slade, J.; Belužić, R.;

Ang, D. L.; Rodger, A.; Piantanida, I. (Feb 2018). *Versatile cyanine-click-amino acid conjugates showing one-atom-influenced recognition of DNA/RNA secondary structure and mitochondrial localization in living cells*. European Journal of Organic Chemistry. DOI: 10.1002/ejoc.201701765. (IF: 2.834, Citations = 5).[26]

Abstract: By a simple click CuAAC (copper(I)-catalysed azide alkyne cycloaddition) procedure several cyanine dye analogues have been attached to the side-chain of an amino acid to yield chromophore amino acid conjugates with the potential to fluoresce upon binding to a target. Due to the availabiltiy of the amino acid C and N termini for peptide coupling, these conjugates are suitable for easy incorporation into the backbone of peptides. The novel amino acid dyes prepared in this work, although intrinsically non-fluorescent, gave rise to strong fluorimetric responses upon binding to double-stranded (ds) DNA or RNA, the emission response to various polynucleotide secondary structures being controlled either by linker length or a halogen atom located on the cyanine part of the molecule. Molecular modelling confirmed the mode of binding to different polynucleotides, which was responsible for the recognition. Interestingly, cell localisation experiments showed that the dyes were specifically localised in mitochondria at variance with the localisation of the parent dyes, which accumulate in cell nuclei, which suggests that the amino acid tail (containing a triazole ring) might function as a novel mitochondria-directing appendage.

Impact: The specific DNA interactions these compounds underwent was visualised using molecular dynamics simulations. Additionally, the aggregation behaviour was simulated to determine the primary form of aggregate, being a H-type head-to-tail aggregate.

Contribution: Molecular dynamics simulations and analysis of the molecule with DNA including generation of final illustrations, and significant writing of computational components. Assistance with spectroscopic experiments.

Paper 4 - Review: Bjelosevic, A.; Pages, B. J.; Spare, L. K.; Deo, K. M.; **Ang, D. L.**; Aldrich-Wright, J. R.. (Jan 2018). *Exposing "bright" metals: promising advances in photoactivated anticancer transition metal complexes*. Current Medicinal Chemistry. DOI: 10.2174/0929867324666170530085123. (IF: 3.249, Citations = 6).[27]

Abstract: Photodynamic therapy (PDT) is an increasingly prominent field in anti-cancer research. PDT agents are typically nontoxic in the absence of light and can be stimulated with nonionising irradiation to "activate" their cytotoxic effect. Photosensitzers are not classified as chemotherapy drugs although it is advantageous to control the toxicity of a drug through localised irradiation allowing for selective treatment. Transition metals are an extremely versatile class of compounds with various unique properties such as oxidation state, coordination number, redox potential and molecular geometry that can be tailored for specific uses. This makes them excellent PDT candidates as their properties can be manipulated to absorb a specific range of light wavelengths, cross cellular membranes or target specific sites *in vitro*. This article reviews recent advances in transition metal PDT agents, with a focus on structural scaffolds from which several metal complexes in a series are synthesised, as well as their *in vitro* cytotoxicity in the presence or absence of irradiation.

Impact: As the title suggests, this review focussed on photoactivated anticancer metal complexes. This is a dynamic and innovative area of research that provides some context on other areas of metals research.

Contribution: All authors contributed equally to the writing and editing where each section of the document was allocated to a different author. Additional contributions by me included final editing and reviewing of the document.

Paper 5 - Review: Rodger, A.; Dorrington, G.; **Ang, D. L.**, (Nov 2016). *Linear dichroism as a probe of molecular structure and interactions*. Analyst. DOI: 10.1039/C6AN01771A. (IF: 4.107, Citations = 9).[28]

Abstract: Linear dichroism (LD) spectroscopy involves measuring the wavelength

(or energy) dependence of the difference in absorption of light parallel and perpendicular to an orientation direction. It requires samples to have a net orientation. The aim of this review is to summarise some UV–visible linear dichroism (LD) methods that can be usefully applied to increase our understanding of biomacromolecules and their complexes that have a high aspect ratio. LD shares the advantages of most spectroscopic techniques including the fact that data collection is fairly straightforward and many sample types can be investigated. Conversely, LD shares the disadvantage that the measured signal is an average over all species in the sample on which the light beam is incident. LD mitigates this disadvantage somewhat in that only species which are oriented give a net signal. How the data can be analysed to give structural information about small molecules in stretched films and membrane systems or bound to biomacromolecules and directly about biomacromolecules such as DNA and protein fibres forms part of this review. In the UV-visible region LD often suffers noticeably from light scattering since the samples tend to be large relative to the wavelength of the incident light, so consideration is also given to data analysis challenges including removal of scattering contributions to an observed signal. Brief mention is made of fluorescence detected LD.

Impact: Somewhat like circular dichroism, linear dichroism is often applied to the analysis of small molecule binding. The speciality of linear dichroism is that it can observe orientable molecules, those comprised of long chain, such as DNA. This review was important to summarise and provide some background context on the theory and applications of linear dichroism.

Contribution: Writing and significant editing, along with generation of illustrations.

Paper 6 - Review: Deo, K. M.; Pages, B. J.; **Ang, D. L.**; Gordon, C. P.; Aldrich-Wright, J. (Oct 2016). Transition Metal Intercalators as Anticancer Agents—Recent Advances. International Journal of Molecular Sciences.

DOI:10.3390/ijms17111818. (IF: 3.687, Citations = 37).[29]

Abstract: The diverse anticancer utility of cisplatin has stimulated significant interest in the development of additional platinum-based therapies, resulting in several analogues receiving clinical approval worldwide. However, due to structural and mechanistic similarities, the effectiveness of platinum-based therapies is countered by severe side-effects, narrow spectrum of activity and the development of resistance. Nonetheless, metal complexes offer unique characteristics and exceptional versatility, with the ability to alter their pharmacology through facile modifications of geometry and coordination number. This has prompted the search for metal-based complexes with distinctly different structural motifs and non-covalent modes of binding with a primary aim of circumventing current clinical limitations. This review discusses recent advances in platinum and other transition metal-based complexes with mechanisms of action involving intercalation. This mode of DNA binding is distinct from cisplatin and its derivatives. The metals focused on in this review include Pt, Ru and Cu along with examples of Au, Ni, Zn and Fe complexes; these complexes are capable of DNA intercalation and are highly biologically active.

Impact: This review focussed on the larger family of anticancer metal complexes. In addition to platinum we also reviewed recent advances in ruthenium, copper, gold, nickel, zinc and iron complexes.

Contribution: All authors contributed equally to the writing where each was allocated a range of metals to review. I was also involved with the creation of all figures and illustrations, and final editing.

Paper 7 - Original research: Pages, B. J.; Sakoff, J.; Gilbert, J.; Rodger, A.; Chmel, N. P.; Jones, N. C.; Kelly, S. M.; **Ang, D. L.**; Aldrich-Wright, J. R. (May 2016). Multifaceted Studies of the DNA Interactions and *In Vitro* Cytotoxicity of Anticancer Polyaromatic Platinum(II) Complexes, Chemistry A European Journal. DOI: 10.1002/chem.201601221. (IF = 5.731, Citations = 14).[30]

Abstract: This study reports a detailed biophysical analysis of the DNA binding and cytotoxicity of six platinum complexes (PCs). They are of the type $[Pt(P_L)(SS\text{-dach})]Cl_2$, where P_L is a polyaromatic ligand and *SS*-dach is 1*S*,2*S*-diaminocyclohexane. The DNA binding of these complexes was investigated using six techniques including ultraviolet and fluorescence spectroscopy, linear dichroism, synchrotron radiation circular dichroism, isothermal titration calorimetry and mass spectrometry. This portfolio of techniques has not been extensively used to study the interactions of such complexes previously; each assay provided unique insight. The *in vitro* cytotoxicity of these compounds was studied in ten cell lines and compared to the effects of their *R,R* enantiomers; activity was very high in Du145 and SJ-G2 cells, with some submicromolar IC_{50} values. In terms of both DNA affinity and cytotoxicity, complexes of 5,6-dimethyl-1,10-phenanthroline and 2,2′-bipyridine exhibited the greatest and least activity, respectively, suggesting that there is some correlation between DNA binding and cytotoxicity.

Impact: These platinum complexes are a variation on the usual structure investigation and were important to assess. Some of these complexes have formed the basis for my own, so far unpublished research on larger aromatic platinum complexes.

Contribution: Assistance with spectroscopic experiments and interpretation, along with minor editing.

Paper 8 - Review: Pages, B.; **Ang, D. L.**; Wright, E.; Aldrich-Wright, J. R., (Oct 2014). Metal Complex Interactions with DNA. Dalton Transactions. DOI: 10.1039/C4DT02700K. (IF: 4.197, Citations = 143).[31] Inside cover.

Abstract: Increasing numbers of DNA structures are being revealed using biophysical, spectroscopic and genomic methods. The diversity of transition metal complexes is also growing, as the unique contributions that transition metals bring to the overall structure of metal complexes depend on the various coordination numbers, geometries, physiologically relevant redox potentials,

as well as kinetic and thermodynamic characteristics. The vast range of ligands that can be utilised must also be considered. Given this diversity, a variety of biological interactions is not unexpected. Specifically, interactions with negatively-charged DNA can arise due to covalent/coordinate or subtle non-coordinate interactions such as electrostatic attraction, groove binding and intercalation as well as combinations of all of these modes. The potential of metal complexes as therapeutic agents is but one aspect of their utility. Complexes, both new and old, are currently being utilised in conjunction with spectroscopic and biological techniques to probe the interactions of DNA and its many structural forms. Here we present a review of metal complex-DNA interactions in which several binding modes and DNA structural forms are explored.

Impact: This review was a significant article in so much as it comprehensively reviewed the existing arena of biologically active metal complexes as can be attested by the large number of citations.

Contribution: All illustrations, significant editing, minor writing, design of inside cover.

Paper 9 - Original research: Pages, B.; Li, F.; Wormell, P.; **Ang, D. L.**; Clegg, J.; Kepert, C.; Spare, L.; Supawich Danchaiwijit, S.; Aldrich-Wright, J., (Nov 2014). Synthesis and analysis of the anticancer activity of platinum(II) complexes incorporating dipyridoquinoxaline variants. Dalton Transactions. DOI: 10.1039/ C4DT02133A. (IF: 4.097, Citations = 24).[32]

Abstract: Eight platinum(II) complexes with anticancer potential have been synthesised and characterised. These complexes are of the type $[Pt(I_L)(A_L)]^{2+}$, where I_L is either dipyrido[3,2-f:2′,3′-h]quinoxaline (dpq) or 2,3-dimethyl-dpq ($23Me_2dpq$) and A_L is one of the *R,R* or *S,S* isomers of either 1,2-diaminocyclohexane (*SS*-dach or *RR*-dach) or 1,2-diaminocyclopentane (*SS*-dacp or *RR*-dacp). The CT-DNA binding of these complexes and a series of other complexes were assessed using fluorescent intercalator displacement

assays, resulting in unexpected trends in DNA binding affinity. The cytotoxicity of the eight synthesised compounds was determined in the L1210 cell line; the most cytotoxic of these were [Pt(dpq)(*SS*-dach)]Cl_2 and [Pt(dpq)(*RR*-dach)]Cl_2, with IC_{50} values of 0.19 and 0.80 μM, respectively. The X-ray crystal structure of the complex [Pt(dpq)(*SS*-dach)]$(ClO_4)_2 \cdot 1.75H_2O$ is also reported.

Impact: My experience was in synthesising achiral dinuclear platinum(II) terpyridine complexes so this was a departure from my normal protocols and laid the groundwork for the synthesis of the complexes used in Chapters 1 and 3.

Contribution: Synthesis and characterisation of selected chiral platinum(II) complexes.

5.2 Journal covers

Two of my publications were selected to be published with an accompanying journal cover in recognition of the significance of the works. These published artworks, for which I am wholly attributable to the design, are presented in the following pages.

Cover 1: This artwork was published as an inside cover of RSC Dalton Transactions, Volume 44, Number 8, Feb 2015, accompanying the article titled "*Metal Complex Interactions with DNA*", DOI: 10.1039/C4DT02700K.[31]

Description: This visually represents the many different interactions of various metal complexes with a large loop of double stranded DNA. The central image is a short section of double stranded DNA complexed with cisplatin. The metal complex:DNA models were sourced from the Protein Data Bank and generated with UCSF Chimera[13] while the long DNA strands were built in and rendered using Blender.

Cover 2: This was published as the outer back cover of Chemistry–A European Journal and accompanied the article presented in Chapter 2 titled "*Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers*", DOI: 10.1002/chem.201503663.[1]

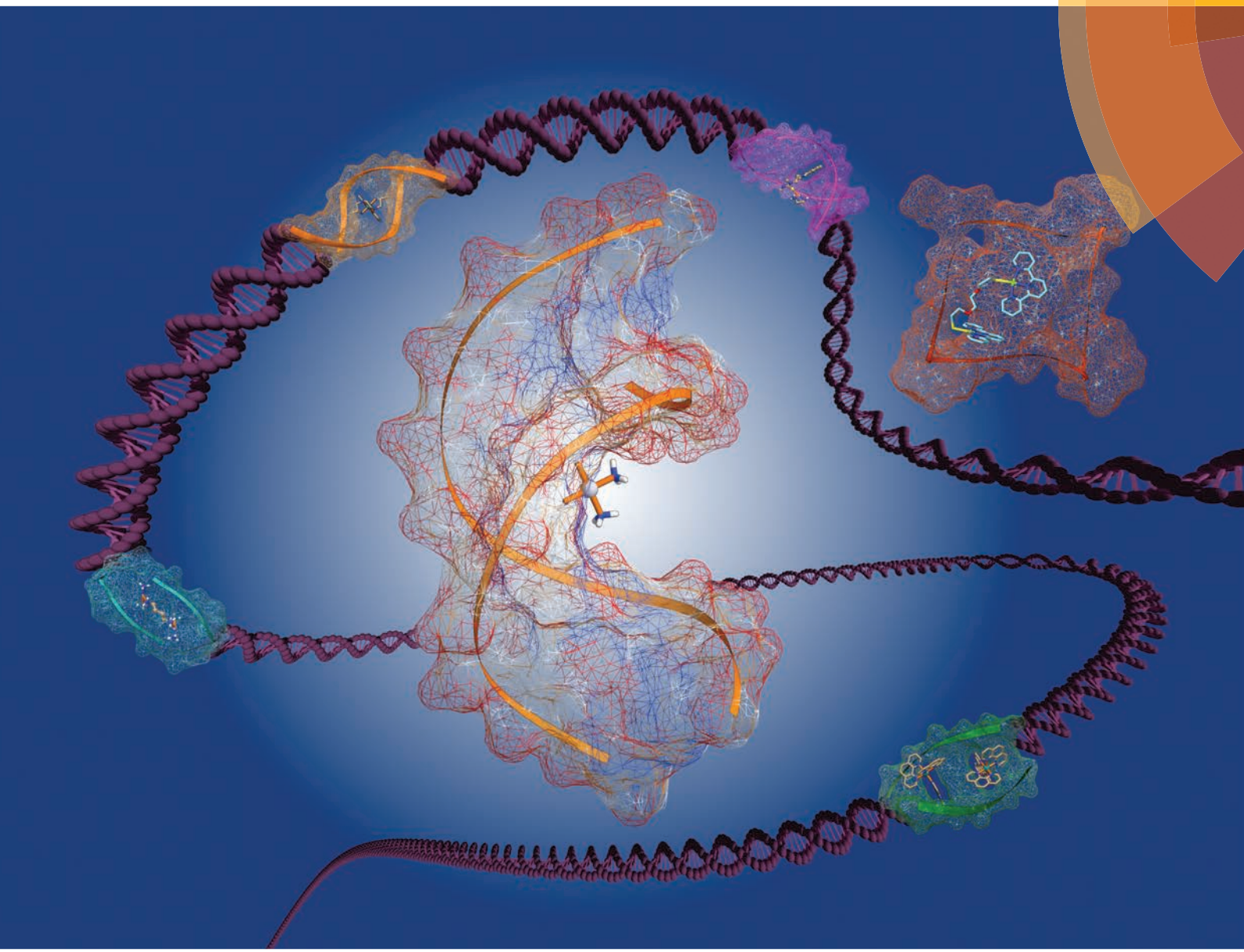
Description: This image represents the binding and stabilisation effects of a dinuclear platinum complex. The ball and stick central molecule model was built in and visualised using Avogadro.[12] The DNA model was created with UCSF Chimera.[13] Upper left is a QDNA structure, with the accompanying CD melt spectra in the lower right. The Southern Cross in the lower left corner was inserted to symbolise the Australian contribution to a European journal.

Volume 44 | Number 8 | 28 February 2015 | Pages 3489–3878

Dalton Transactions

An international journal of inorganic chemistry

www.rsc.org/dalton



ISSN 1477-9226

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

Janice R. Aldrich-Wright *et al.*

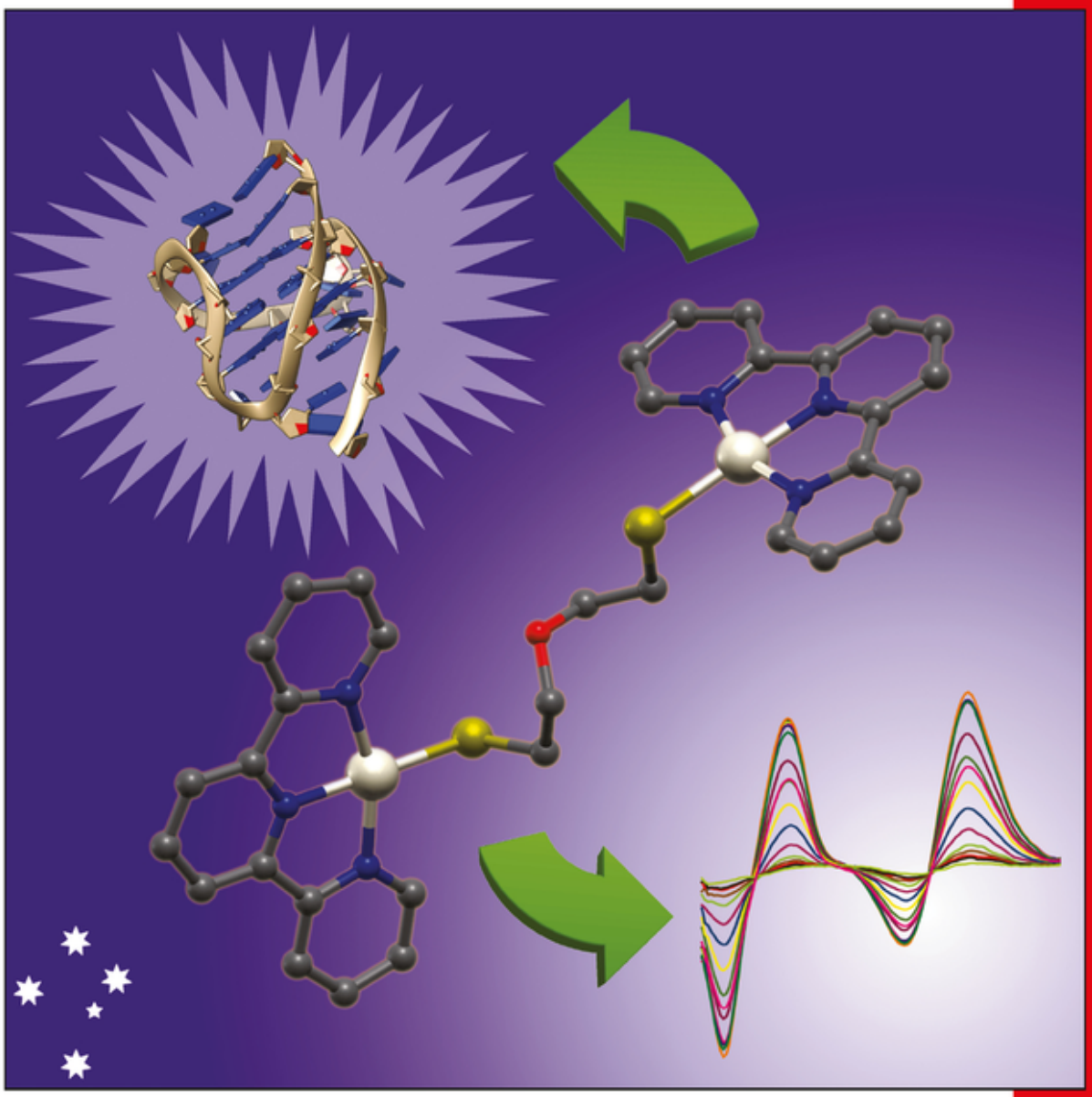
Metal complex interactions with DNA

CHEMISTRY
A European Journal

www.chemeurj.org

A Journal of ChemPubSoc Europe

2016-22/7



Cover Picture:
R. Vilar, J. Aldrich-Wright et al.
Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers

Supported by ACES Asian Chemical Editorial Society

WILEY-VCH

5.3 Refereed conference abstracts in conference proceedings

Abstract 1: Šmidlehner, T.; Kurutos , A.; Slade, J.; Belužić, R.; **Ang, D. L.**; Rodger, A.; Piantanida, I. (2017) *Fluorophore one-atom-controlled recognition of DNA/RNA secondary structures*. EFMC International Symposium on Advances in Synthetic and Medicinal Chemistry, August 2017, Vienna, Austria.

Abstract 2: **Ang, D. L.**; Jones, N. C.; Stootman, F.;Ghadirian, B.; Aldrich-Wright, J.R. (2015). *DNA Binding Studies using Synchrotron Radiation Circular Dichroism and Mathematica*. Proceedings of the Australian Synchrotron User Meeting, 19th November 2015, Melbourne, Australia.

Abstract 3: **Ang, D.L.**; Jones, N.; Stootman, F.; Ghadirian, B.; Aldrich-Wright, J.R. (2015). *Improved DNA equilibrium binding affinity determinations of platinum(II) complexes using synchrotron radiation circular dichroism*. Pacifichem, Hawaii, 2015. See Chapter 1.

Abstract 4: **Ang, D.L.**; Craze, A.; Yuliani, Y.; Kelso, C.; Beck, J.L; Ralph, S.F.; Harman, D.G.; Aldrich-Wright, J.R. (2015). *ESI-MS probing of G-quadruplex DNA platinum(II) complex interactions*. December 2015, Pacifichem, Hawaii.

Abstract 5: **Ang, D.L.**; Harman, D.G.; Ralph, S.F.; Aldrich-Wright, J.R. (2015). *Quadruplex-DNA stabilising dinuclear platinum(II) complexes*. December 2015, Pacifichem, Hawaii.

Abstract 6: **Ang, D. L.**; Jones, N. C.; Stootman, F.;Ghadirian, B.; Aldrich-Wright, J.R. (2015) *DNA Binding Studies using Synchrotron Radiation Circular Dichroism and Mathematica*. Proceedings of the Australian Synchrotron User Meeting, 19th November 2015, Melbourne, Australia.

5.4 Posters

The following are selected posters most relevant to the work presented in this thesis. These were presented at various conference throughout the world.

Poster 1: SOMSpec – a machine learning approach to predicting protein secondary structure from spectra. **Ang, D. L.**; Rodger, A.; Aldrich-Wright, J. R.. Western Sydney University HDR showcase, December 2017.

Relevance: This poster presented my work on SOMSpec (Chapter 4) at the annual WSU Higher Degree Research showcase. It gives a broad overview of what SOMSpec does, how it's designed, and a brief overview of the underlying theory.

Poster 2: Fluorophore one-atom-controlled recognition of DNA/RNA secondary structure. Šmidlehner, T.; Kurutos, A.; **Ang, D. L.**; Belužić, R.; Rodger, A.; Piantanida, I.. Circular Dichroism summer school, Warwick University, UK, July 2017.

Relevance: This poster presented the results of my work[26] on simulating the DNA interactions of these compounds.

Poster 3: Automated DNA equilibrium binding affinity calculations of platinum(II) complexes using Synchrotron Radiation Circular Dichroism. **Ang, D. L.**; Jones, N. C.; Stootman, F.; Ghadirian, B.; Aldrich-Wright, J.. Nucleic Acids Forum, London, UK, July 2016.

Relevance: This poster presented my work as presented in Chapter 1 on DNA binding affinity calculations using SRCD.[8]

Poster 4: Molecular dynamics studies of the *E. coli* division protein FtsZ. **Ang, D. L.**; Broughton, C.; Khalid, S.; Rodger, A.. Circular Dichroism summer school, Warwick University, UK, July 2016.

Relevance: My molecular dynamics simulation work performed at Warwick University on proteins is summarised in this poster. It formed the basis of my further expansion into complex simulation and visualisation. This is unpublished work.

Poster 5: Probing QDNA platinum(II) complex interactions with ESI-MS. **Ang, D. L.**; Craze, A.; Yuliani, Y.; Kelso, C.; Beck, J.; Harman, D. G.; Ralph, S. F.; Aldrich-Wright, J. R.. Pacifichem, Hawaii, Dec 2015.

Relevance: This poster was a summary of the work contained in the paper presented in Chapter 3.

5.5 Scholarships and grants awarded

During my candidature I have been awarded several scholarships and grants which enabled me to explore and expand on my research. The details of these are listed here in chronological order.

Grant 1: Marie Curie Fellowship, 2016. The Centre for Analytical Science Innovative Doctoral Programme (CAS-IDP) was funded by the EU under FP7 Marie Curie Actions to train an international group of early stage researchers. This project received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number 316630.

Outcome: My Fellowship involved 12 months at Warwick University, Coventry, UK, under the supervision of Professor Alison Rodger. During this time I furthered my understanding of CD and LD spectroscopy, began development of SOMSpec, and gained significant molecular dynamics expertise simulating proteins and DNA. Several other projects were started here and will be published in the near future.

Grant 2: Endeavour Research Fellowship, 2016. This highly competitive scholarship is offered by the Australian Government to participate in short-term (4–6 months) research with an international University.

Outcome: I declined this offer as I was also offered the Marie Curie Fellowship and the placements clashed.

Grant 3: Diamond Light Source, UK, September 2016. Reference: SM14808-1. Title: *SRCD Characterisation of unusual DNA structures*. Principal investigator: Janice Aldrich-Wright.

Outcome: This funded a 24 h window of experiments on the synchrotron circular dichroism beamline at Diamond Light Source, UK. Several QDNA CD spectra

were obtained and melt temperatures measured. These will contribute to the library of QDNA CD spectra for the SOMSpec training sets. This project was designed and performed solely by myself.

Grant 4: Australian Synchrotron Round ISAP 2015/1, ID: ISP9258. Title: *Characterisation of quadruplex DNA motifs by Synchrotron Radiation Circular Dichroism (SRCD)*. Type: International. Principal scientist: Janice Aldrich-Wright.

Outcome: This funded 6.5 days of access on the synchrotron circular dichroism beamline at ISA, Aarhus University, Denmark. This work laid the foundation for investigating QDNA binding using synchrotron radiation circular dichroism (see Chapter 2). These experiments were all designed and optimised by myself and most were performed by myself with assistance from my supervisor. All post-processing of data was performed by myself.

Chapter 6

Conclusions and future work

A key goal of this work was to gain a deeper understanding of the binding of small molecules with biomacromolecular structures with a specific emphasis on QDNA, leading to the design of improved biologically active platinum(II) complexes. Towards this goal I was motivated to develop new tools to automate and assess the spectral data obtained from the various experimental methods. In Chapter 1 a new method for determining binding constants was presented and implemented.[8] This method allows near real time processing of spectra as it is acquired. As a consequence, the prompt calculation of binding constants and critical feedback on the quality of acquired spectra, reducing potential acquisition errors, is made possible. The results also shed light on possible wavelength specific methodology considerations when considering the design of binding experiments.

In Chapter 2, QDNA binding of a series of dinuclear platinum(II)-terpyridine complexes was assessed. These were shown to exhibit strong stabilising effects and specificity towards QDNA, increasing melting temperatures by up to 17 °C, while not significantly affecting the conformation.[1] The SOS-linked complex favoured the parallel topologies of QDNA over dsDNA. Competitive FRET experiments revealed that the dipyridal-linked complex exhibited especially high selectivity towards QDNA in the presence of up to a 600-fold concentration of dsDNA, with no change in the increased melt temperatures of the dsDNA. These are significant findings and suggest this family of complexes should be pursued for further development and

understand why dsDNA precipitation occurred.

The effect of methylation on the phenanthroline moiety of a series of platinum(II)-phenanthroline complexes was then investigated by ESI-MS, molecular docking simulations and associated techniques, in regard to binding with QDNA and dsDNA in Chapter 3. Additionally, an improved method for the preparation of oligonucleotides for ESI-MS experimentation was presented. The results highlighted some important structural characteristics to consider when designing bioactive platinum(II)-phenanthroline complexes specifically targeting QDNA.

Chapter 4 presents SOMSpec, a software application I began developing in my time at the University of Warwick for the elucidation of secondary structure information from spectra. What started as a request from my supervisor to "*Take a look and see what you can do*" morphed into a fully fledged stand alone machine learning program, currently at around ten thousand lines of code. This substantially increased my understanding of the informational content obtained through spectrophotometric methods, and served as an introduction to machine learning algorithms. SOMSpec is in active development with the ultimate aim to create a suite of integrated tools allowing the experimentalist to process, manipulate, predict and present spectra from a variety of sources including CD, UV, IR, and RAMAN. SOMSpec is in active development and the implementation of automated binding calculations as presented in Chapter 1 is a current priority, which should also increase the uptake of this method. Finally, the implementation of more advanced machine learning algorithms within SOMSpec is in the pipeline and should yield improved prediction accuracy and training speed.

Future directions of this research include the further development of the dinuclear platinum(II)-terpyridine complexes, investigated in chapter 2, to improve the stabilisation effects of QDNA and to further increase specificity. Alongside this, a family of symmetric platinum(II)-phenanthroline complexes have been synthesised, based on the findings of Chapter 3, with initial QDNA melt data showing promising stabilisation properties. This data will be the subject of a paper in the near future.

Also critical is the parameterisation of these platinum(II) complexes to enable full molecular dynamics simulations in the presence of DNA as an improved way of exploring these interactions over rudimentary docking methodologies.

As part of the experiments performed at the Diamond Light Source, UK, and the ISA, Denmark, synchrotron CD beamlines I have compiled a library of synchrotron radiation CD (SRCD) melt spectra of various DNA topologies. The aim is for this data to become a useful reference set complementing the Protein Circular Dichroism Data Bank[33] and as a training set for SOMSpec to enable the prediction of DNA motifs.

Finally, but crucially related this work, is the pressing need to adopt a better CD instrument calibration standard than camphorsulfonic acid (CSA), especially for benchtop instruments. CSA is considered the primary standard for CD calibration, but is highly hygroscopic.[34] The related, less hygroscopic secondary standard ammonium camphor sulfonate (ACS) is sometimes used as it is easier to measure, but less stable in storage. For both, there is only one hand readily available with a (very broad) positive peak at 290 and negative peak at 192.5 nm. A machine is considered within 'spec' if the ratio of ellipticities of the two peaks is ~1.90–2.20.[35] Difficulties arise for most benchtop in this situation as it is difficult to adequately resolve the 192.5 nm peak due to atmospheric absorption and poor lamp intensity. This also doesn't provide any information on machine performance at longer wavelengths. As part of an ongoing collaboration to create a new CD standard for calibrating CD machines based on the chiral Co[EDDS] complex,[36] I have been developing a piece of software and associated methodology for the automated interpretation of the CD spectrum calibration results. The aim is to develop a robust method for the calibration and validation of circular dichroism spectrophotometers.

References

[1] D. L. Ang, B. W. J. Harper, L. Cubo, O. Mendoza, R. Vilar and J. Aldrich-Wright, *Chemistry – A European Journal*, 2016, **22**, 2317–2325.

[2] Q. Wang, J.-q. Liu, Z. Chen, K.-w. Zheng, C.-y. Chen, Y.-h. Hao and Z. Tan, *Nucleic Acids Research*, 2011, **39**, 6229–6237.

[3] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Research*, 2006, **34**, 5402–5415.

[4] J. Bidzinska, G. Cimino-Reale, N. Zaffaroni and M. Folini, *Molecules*, 2013, **18**, 12368–12395.

[5] D. Dutta, M. Debnath, D. Müller, R. Paul, T. Das, I. Bessi, H. Schwalbe and J. Dash, *Nucleic Acids Research*, 2018, **46**, 5355–5365.

[6] Q. Zhai, C. Gao, J. Ding, Y. Zhang, B. Islam, W. Lan, H. Hou, H. Deng, J. Li, Z. Hu, H. I. Mohamed, S. Xu, C. Cao, S. M. Haider and D. Wei, *Nucleic Acids Research*, 2019, **47**, 2190–2204.

[7] S. Bandeira, J. Gonzalez-Garcia, E. Pensa, T. Albrecht and R. Vilar, *Angewandte Chemie International Edition*, 2018, 310–313.

[8] D. L. Ang, N. C. Jones, F. Stootman, B. Ghadirian and J. R. Aldrich-Wright, *Analyst*, 2015, **140**, 4162–4169.

[9] D. L. Ang, *Honours Thesis*, University of Western Sydney, 2014.

[10] O. Trott and A. J. Olson, *Journal of Computational Chemistry*, 2010, **31**, 455–461.

[11] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.

[12] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Journal of Cheminformatics*, 2012, **4**, 17–33.

[13] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *Journal of Computational Chemistry*, 2004, **25**, 1605–1612.

[14] W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics*, 1996, **14**, 33–38.

[15] *Zotero*, Center for History and New Media, George Mason University, 2006.

[16] G. Scatchard, *Annals of the New York Academy of Sciences*, 1949, **51**, 660–672.

[17] D. E. V. Schmechel and D. M. Crothers, *Biopolymers*, 1971, **10**, 465–480.

[18] C. G. Hartinger, W. H. Ang, A. Casini, L. Messori, B. K. Keppler and P. J. Dyson, *Journal of Analytical Atomic Spectrometry*, 2007, **22**, 960–967.

[19] D. L. Ang, C. Kelso, J. L. Beck, S. F. Ralph, D. G. Harman and J. R. Aldrich-Wright, *JBIC Journal of Biological Inorganic Chemistry*, 2020, **25**, 429–440.

[20] V. Hall, A. Nash and A. Rodger, *Analytical Methods*, 2014, **6**, 6721–6726.

[21] T. Kohonen, *Biological Cybernetics*, 1982, **43**, 59–69.

[22] J. G. Lees, A. J. Miles, F. Wien and B. A. Wallace, *Bioinformatics*, 2006, **22**, 1955–1962.

[23] W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577–2637.

[24] K. M. Deo, D. L. Ang, B. McGhie, A. Rajamanickam, A. Dhiman, A. Khoury, J. Holland, A. Bjelosevic, B. Pages, C. Gordon and J. R. Aldrich-Wright, *Coordination Chemistry Reviews*, 2018, **375**, 148–163.

[25] M. P. Corujo, M. Sklepari, D. L. Ang, M. Millichip, A. Reason, S. C. Goodchild, P. Wormell, D. P. Amarasinghe, V. Lindo, N. P. Chmel and A. Rodger, *Chirality*, 2018, **30**, 957–965.

[26] T. Šmidlehner, A. Kurutos, J. Slade, R. Belužić, D. L. Ang, A. Rodger and I. Piantanida, *European Journal of Organic Chemistry*, 2018, **2018**, 1682–1692.

[27] A. Bjelosevic, B. J. Pages, L. K. Spare, K. M. Deo, D. L. Ang and J. R. Aldrich-Wright, *Current Medicinal Chemistry*, 2018, **25**, 478–492.

[28] A. Rodger, G. Dorrington and D. L. Ang, *Analyst*, 2016, **141**, 6490–6498.

[29] K. M. Deo, B. J. Pages, D. L. Ang, C. P. Gordon and J. R. Aldrich-Wright, *International Journal of Molecular Sciences*, 2016, **17**, 1818.

[30] B. J. Pages, J. Sakoff, J. Gilbert, A. Rodger, N. P. Chmel, N. C. Jones, S. M. Kelly, D. L. Ang and J. R. Aldrich-Wright, *Chemistry – A European Journal*, 2016, **22**, 8943–8954.

[31] B. J. Pages, D. L. Ang, E. P. Wright and J. R. Aldrich-Wright, *Dalton Transactions*, 2015, **44**, 3505–3526.

[32] B. J. Pages, F. Li, P. Wormell, D. L. Ang, J. K. Clegg, C. J. Kepert, L. K. Spare, S. Danchaiwijit and J. R. Aldrich-Wright, *Dalton Transactions*, 2014, **43**, 15566–15575.

[33] L. Whitmore, A. J. Miles, L. Mavridis, R. W. Janes and B. A. Wallace, *Nucleic Acids Research*, 2017, **45**, D303–D307.

[34] G. C. Chen and J. T. Yang, *Analytical Letters*, 1977, **10**, 1195–1207.

[35] A. J. Miles, F. Wien, J. G. Lees, A. Rodger, R. W. Janes and B. A. Wallace, *Journal of Spectroscopy*, 2003, **17**, 653–661.

[36] A. Damianoglou, E. J. Crust, M. R. Hicks, S. E. Howson, A. E. Knight, J. Ravi, P. Scott and A. Rodger, *Chirality*, 2008, **20**, 1029–1038.

Appendix A

Supplementary information.

The following pages contain the supplementary and supporting information for the published articles from Chapters 1–3. For Chapter 4, the SOMSpec user guide is instead included.

1. Supplementary information for the Chapter 1 paper "Improved DNA Equilibrium Binding Affinity Determinations of Platinum(II) Complexes using Synchrotron Radiation Circular Dichroism". Includes description of input data formats, derivation of the binding equation, and SRCD data from binding experiments.

2. Supporting information for the Chapter 2 paper "Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers". Includes the SRCD melt spectra and curves, FID competition spectra, FRET melting results, DFT calculations, and docking simulation details.

3. Supplementary information for the Chapter 3 paper "A study of Pt(II)-phenanthroline complex interactions with double-stranded and G-quadruplex DNA by ESI-MS, circular dichroism, and computational docking". Includes ESI-MS spectra from the quadruplex and double stranded binding experiments, HPLC traces, calculated partial atomic charges, electrostatic potential maps, and Gaussian input files.

4. The presentation and training guide for SOMSpec (Chapter 4).

SUPPLEMENTARY INFORMATION

Improved DNA Equilibrium Binding Affinity Determinations of Platinum(II) Complexes using Synchrotron Radiation Circular Dichroism

Dale L. Ang,[a] Nykola C. Jones,[b] Frank Stootman,[a] Bahman Ghadirian[a] and Janice R. Aldrich-Wright[a*]

[a]*Nanoscale Organisation and Dynamics Group, School of Science and Health, University of Western Sydney, Penrith, NSW, 2751, Australia. Fax: (+61) 246203025; Tel: (+61) 246203218; E-mail: J.Aldrich-Wright@uws.edu.au*

[b]ISA, *Department of Physics and Astronomy, Aarhus University, DK 8000 Aarhus C, Denmark.*

* Please address editorial correspondence to Prof. Janice R. Aldrich-Wright via e-mail: j.aldrich-wright@uws.edu.au or via facsimile: +61 2 4620 3025.

Table of Contents

Input Data File CSV format

The *Mathematica* notebook accepts a standard CSV file; the path of which must be specified within the notebook. The format is as follows. Cell A1 contains the DNA concentration [M]. The first column contains the wavelengths. Subsequent columns contain the blank/buffer, DNA, followed by the DNA + MC data at the various concentrations.

An example is shown below.

[DNA conc]				
	Blank/buffer	DNA, no MC	DNA + MC conc 1	DNA + MC conc 2
Wavelength 1				
Wavelength 2				
Wavelength 3				
Wavelength 4				

0.000035				
	0	0	0.00000166	0.00000332
400	0.366	0.475	0.397	0.415
399	0.355	0.442	0.426	0.467
398	0.346	0.455	0.412	0.437
397	0.345	0.457	0.443	0.441
396	0.384	0.511	0.418	0.416

Equation Derivation

In what follows we use the following labels for description: L_F (free metal complex concentration), S_F (free binding site concentration), L_B (bound metal complex concentration), S_B (bound/occupied binding site concentration), L_T (total metal complex concentration) and S_T (total binding site concentration).

In the original development of the theory the binding equilibrium rate equation is given by,

$$L_F + S_F \underset{K}{\leftrightarrow} L_B$$

Giving generally,

$$K = \frac{L_B}{L_F \times S_F} \qquad (1)$$

The value of K is assumed to be a constant at a given wavelength and independent of L_B. Substituting for $L_F = L_T - L_B$ and $S_F = S_T - S_B$ into Eq. (1) and making a reasonable assumption that $S_B = L_B$, results in a quadratic equation for L_B which has a solution of

$$L_B = 0.5R\left(\frac{1}{K} + L_T + S_T - \sqrt{\left(\frac{1}{K} + L_T + S_T\right)^2 - 4S_T L_T}\right) \qquad (2)$$

where R is a scaling constant, and L_B is directly proportional to the measured normalized molar absorption coefficient, ε_M . The value of K can be found by fitting the titration data directly in Eq. (2), which is the preferred approach.[1] The number of DNA binding sites, n, is related to the total DNA concentration and calculated where n = [DNA]/S_T.

In the literature, however, Scatchard,[2] Schmechel and Crothers[3] and others[4-6] have attempted to linearize Eq. (1) by approximation. Substituting the above values in equation (1) gives

$$K = \frac{L_B}{(L_T - L_B)(S_T - L_B)} \qquad (3)$$

A rearrangement results in

$$\frac{1}{L_B} = \frac{1}{K(L_T - L_B)L_T} + \frac{1}{L_T} \qquad (4)$$

At this point a further assumption is made that at low total ligand concentration $L_B \rightarrow L_T$ on the RHS of Eq. (4). This implies that by this method K is an extrapolation as $L_B \rightarrow 0$; this is at best artificial because K is determined when there is effectively no binding. Such an assumption is not made by Eq. (2). In the literature

Eq. (4) leads to a more traditional form in order to determine K by using ε_M (measured absorbance), ε_B (bound absorbance), ε_F (free absorbance), and the identity

$$\frac{1}{L_B} = \frac{(\varepsilon_B - \varepsilon_F)}{(\varepsilon_M - \varepsilon_F)} \frac{1}{L_T}$$

To obtain

$$\frac{1}{\varepsilon_M - \varepsilon_F} = \frac{1}{K(\varepsilon_B - \varepsilon_F)} \times \frac{1}{(S_T - L_B)} + \frac{1}{(\varepsilon_B - \varepsilon_F)} \quad (5)$$

Plotting the experimentally derived $\varepsilon_M - \varepsilon_B$ versus $S_T - L_B$ yields a value of slope and intercept from which K, again extrapolated to $L_B \rightarrow 0$, is found. If the plot is not a straight line (as mathematically it is not), it is argued that a value of K can be found from the initial slope of the graph. Both approaches determine artificially the binding constant at very low bound ligand concentrations.

Binding Data obtained from SRCD experiments

Table S1 Complex **1**, $[Pt(phen)(en)]^{2+}$ binding data, experiment B.

Wavelength nm	Binding Constant $K \times 10^5$	Estimated Binding sites per complex, *n*
297	1.1 ± 0.54	2.4 ± 0.1
320	5.8 ± 0.40	4.3 ± 0.2

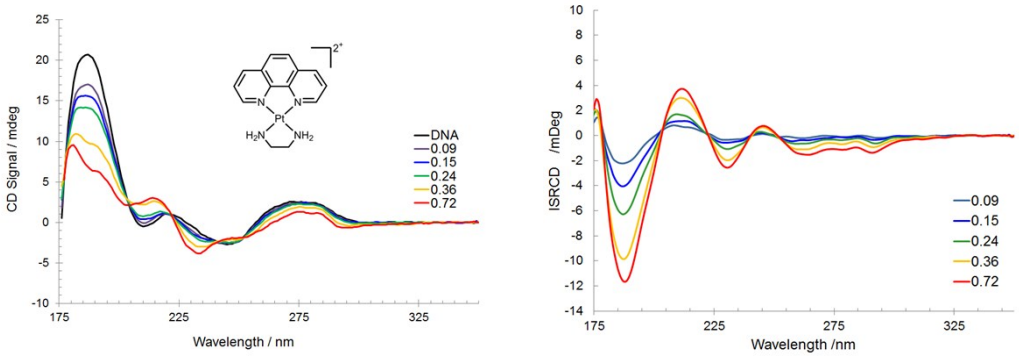


Figure S1 **Expt B –** SRCD and ISRCD spectra at different concentrations of metal complex **1**, into calf thymus DNA in PS buffer.

Table S2 Complex **2**, $[Pt(4\text{-Mephen})(en)]^{2+}$ binding data, experiment A.

Wavelength nm	Binding Constant $K \times 10^4$	Estimated Binding sites per complex, n
181	2.14 ± 0.84	2.6 ± 0.1
185	0.76 ± 0.09	3.6 ± 0.1
186	1.08 ± 0.15	3.3 ± 0.1
190	1.75 ± 0.25	2.8 ± 0.1
192	1.79 ± 0.14	2.4 ± 0.1
194	3.37 ± 0.95	2.1 ± 0.1
195	1.46 ± 0.21	2.1 ± 0.5

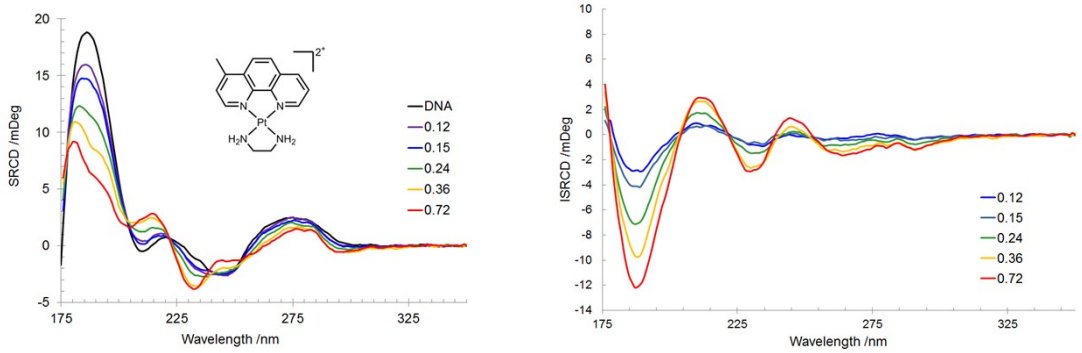


Figure S2 Expt A – SRCD and ISRCD spectra at different concentrations of metal complex **2**, into ct-DNA in PS buffer.

Table S3 Complex **2**, $[Pt(4\text{-Mephen})(en)]^{2+}$ binding data, experiment B.

Wavelength nm	Binding Constant $K \times 10^4$	Estimated Binding sites per complex, n
177	11.6 ± 3.0	2.4 ± 0.2
184	0.95 ± 0.52	1.6 ± 0.0
186	1.8 ± 0.74	1.7 ± 0.1
192	3.4 ± 0.52	1.6 ± 0.0
194	6.3 ± 3.5	1.5 ± 0.0

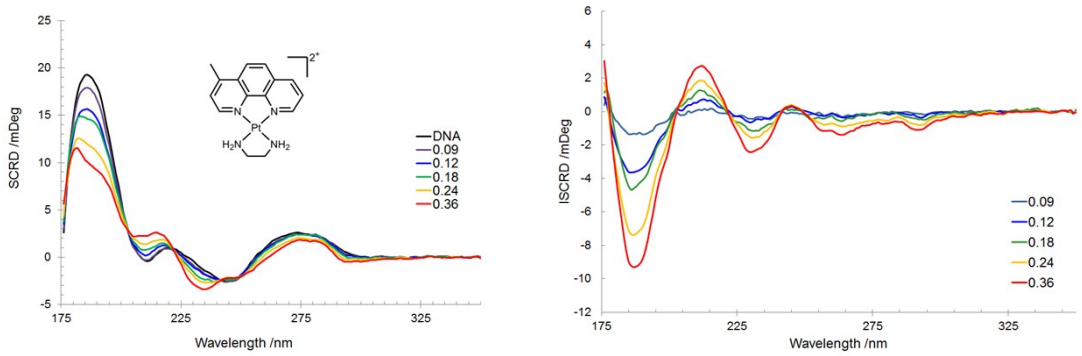


Figure S3 Expt B - SRCD and ISRCD spectra at different concentrations of metal complex **2**, into ct-DNA in PS buffer.

Table S4 Complex **3**, $[Pt(5\text{-Mephen})(en)]^{2+}$ binding data, experiment A.

Wavelength nm	Binding Constant $K \times 10^5$	Estimated Binding sites per complex, n
192	2.33 ± 0.18	2.2 ± 0.0
193	3.34 ± 0.22	2.2 ± 0.0
221	1.32 ± 0.78	6.5 ± 1.0

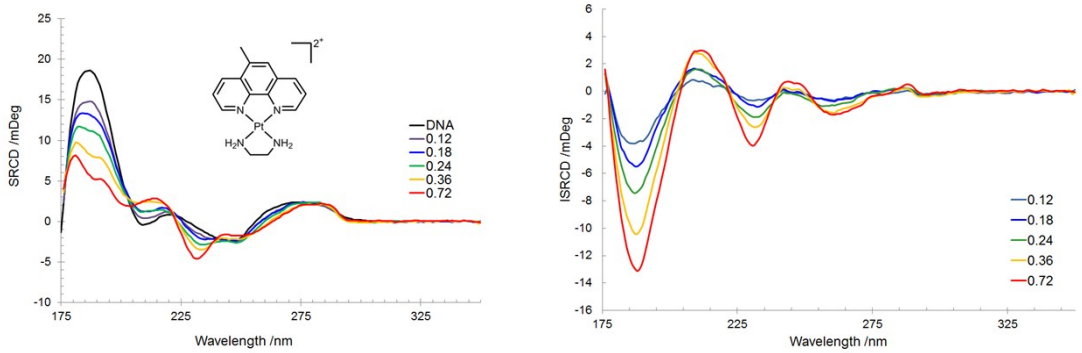


Figure S4 Expt A - SRCD and ISRCD spectra at different concentrations of metal complex **3**, into ct-DNA in PS buffer.

Table S5 Complex **3**, $[Pt(5\text{-Mephen})(en)]^{2+}$ binding data, experiment B.

Wavelength nm	Binding Constant $K \times 10^4$	Estimated Binding sites per complex, n
181	4.5 ± 1.9	2.5 ± 0.1
185	5.8 ± 1.2	2.5 ± 0.1
191	8.5 ± 0.73	2.4 ± 0.1
208	5.2 ± 1.2	4.0 ± 0.2
210	5.6 ± 0.87	3.6 ± 0.2
213	2.7 ± 0.86	3.9 ± 0.1
214	2.2 ± 0.42	4.0 ± 0.2
215	2.4 ± 0.75	3.8 ± 0.1
218	1.2 ± 0.47	6.0 ± 0.2
279	0.46 ± 0.10	1.9 ± 0.0
330	7.0 ± 0.78	3.9 ± 0.3

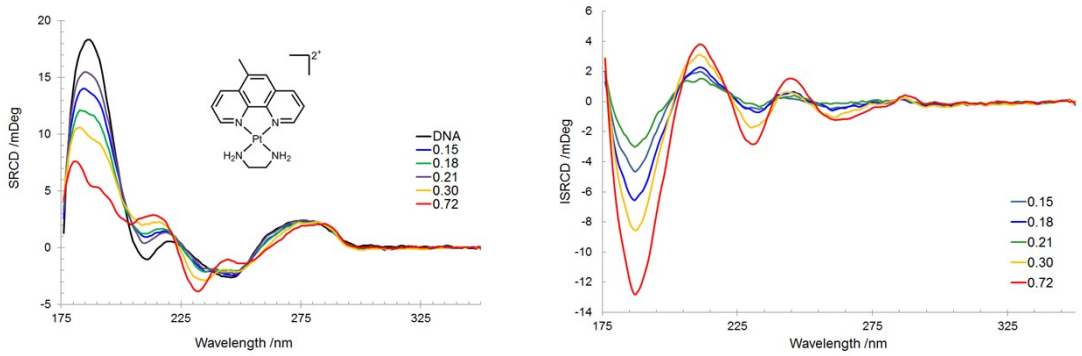


Figure S5 Expt B - SRCD and ISRCD spectra at different concentrations of metal complex **3**, into ct-DNA in PS buffer.

Table S6 Complex **4**, $[Pt(4,7\text{-}Me_2phen)(en)]^{2+}$ binding data, experiment A.

Wavelength nm	Binding Constant $K \times 10^4$	Estimated Binding sites per complex, *n*
194	4.7 ± 0.61	2.6 ± 0.2
195	3.4 ± 0.70	2.4 ± 0.1
199	0.35 ± 0.09	1.9 ± 0.0
307	33 ± 1.5	2.4 ± 0.1
312	24 ± 1.4	2.2 ± 0.1
336	1.4 ± 0.2	2.1 ± 0.1

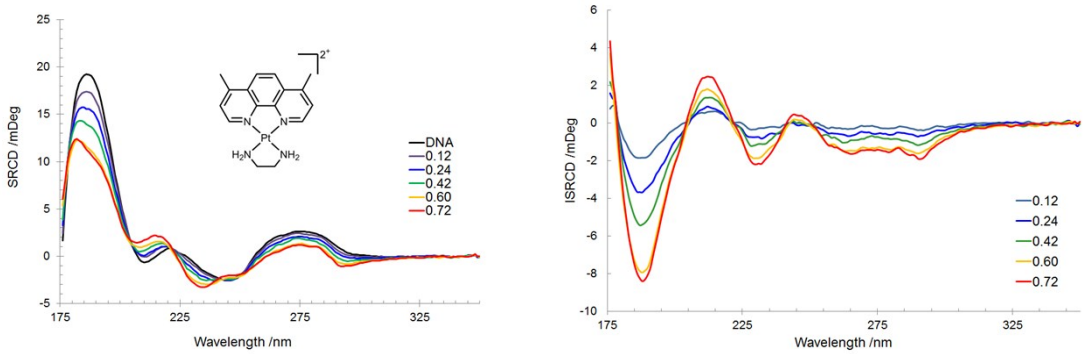


Figure S6 Expt A - SRCD and ISRCD spectra at different concentrations of metal complex 4, into ct-DNA in PS buffer.

Table S7 Complex **4**, $[Pt(4,7\text{-}Me_2phen)(en)]^{2+}$ binding data, experiment B.

Wavelength nm	Binding Constant $K \times 10^5$	Estimated Binding sites per complex, n
184	3.5 ± 0.28	2.0 ± 0.1
209	1.9 ± 0.12	3.2 ± 0.1
311	8.6 ± 0.76	3.6 ± 0.2

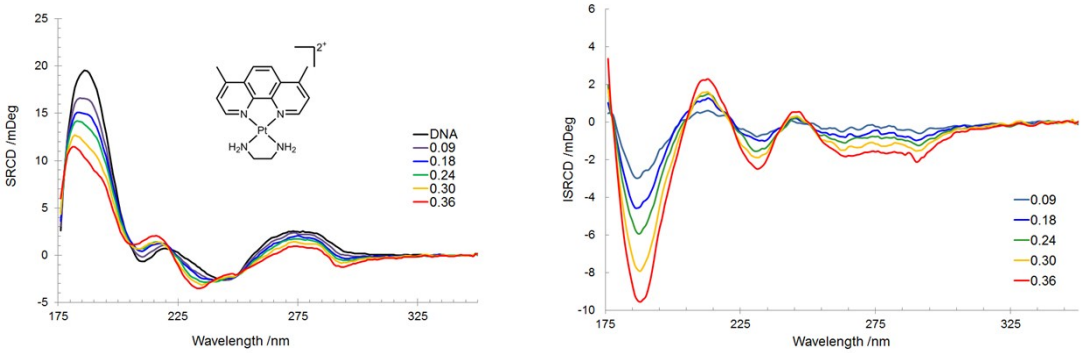


Figure S7 Expt B - SRCD and ISRCD spectra at different concentrations of metal complex 4, into calf thymus DNA in PS buffer.

Table S8 Complex **5**, $[Pt(5,6\text{-}Me_2phen)(en)]^{2+}$ binding data, experiment A.

Wavelength nm	Binding Constant $K \times 10^4$	Estimated Binding sites per complex, *n*
207	13.4 ± 0.96	3.3 ± 0.1
210	6.9 ± 0.41	3.2 ± 0.0
212	5.6 ± 0.33	3.1 ± 0.0
214	7.3 ± 0.27	3.3 ± 0.0
215	9.1 ± 0.52	3.3 ± 0.0
216	9.8 ± 0.34	3.2 ± 0.0
220	3.3 ± 0.21	4.8 ± 0.1
241	0.53 ± 0.10	1.6 ± 0.0
242	2.0 ± 0.35	2.0 ± 0.1
298	2.2 ± 0.22	2.5 ± 0.1
300	0.81 ± 0.20	2.8 ± 0.1

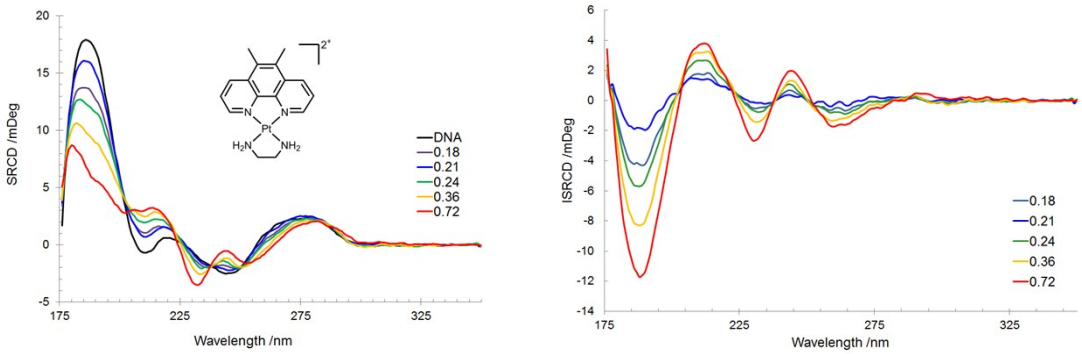


Figure S8 Expt A - SRCD and ISRCD spectra at different concentrations of metal complex 5, into ct-DNA in PS buffer.

Table S9 Complex **5**, $[Pt(5,6\text{-}Me_2phen)(en)]^{2+}$ binding data, experiment B.

Wavelength nm	Binding Constant $K \times 10^4$	Estimated Binding sites per complex, *n*
176	6.2 ± 0.98	1.4 ± 0.1
181	7.4 ± 1.5	2.3 ± 0.1
183	2.9 ± 0.37	2.4 ± 0.1
185	2.5 ± 0.13	2.4 ± 0.1
186	3.5 ± 0.35	2.3 ± 0.1
191	4.0 ± 1.4	2.2 ± 0.1
193	5.1 ± 0.39	2.0 ± 0.1
235	10 ± 0.73	2.7 ± 0.1
259	5.3 ± 0.54	2.8 ± 0.1
260	1.7 ± 0.20	3.2 ± 0.1
261	1.8 ± 0.16	3.0 ± 0.1
262	3.8 ± 0.73	2.7 ± 0.1
269	8.5 ± 0.43	2.6 ± 0.0
304	2.4 ± 0.23	2.1 ± 0.0
329	3.6 ± 1.9	3.0 ± 0.2

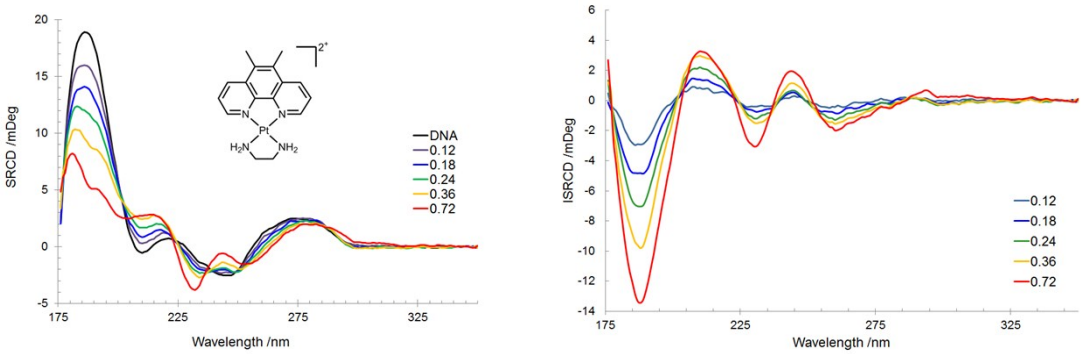


Figure S9 Expt B - SRCD and ISRCD spectra at different concentrations of metal complex 5, into ct-DNA in PS buffer.

Table S10 Complex **6**, $[Pt(3478\text{-}Me_4phen)(en)]^{2+}$ binding data, experiment A.

Wavelength nm	Binding Constant $K \times 10^5$	Estimated Binding sites per complex, *n*
208	0.77 ± 0.13	4.9 ± 0.4
254	0.71 ± 0.07	6.0 ± 0.4
261	0.10 ± 0.01	5.2 ± 0.1
323	2.5 ± 0.27	4.5 ± 0.2
330	1.1 ± 0.10	9.8 ± 0.5

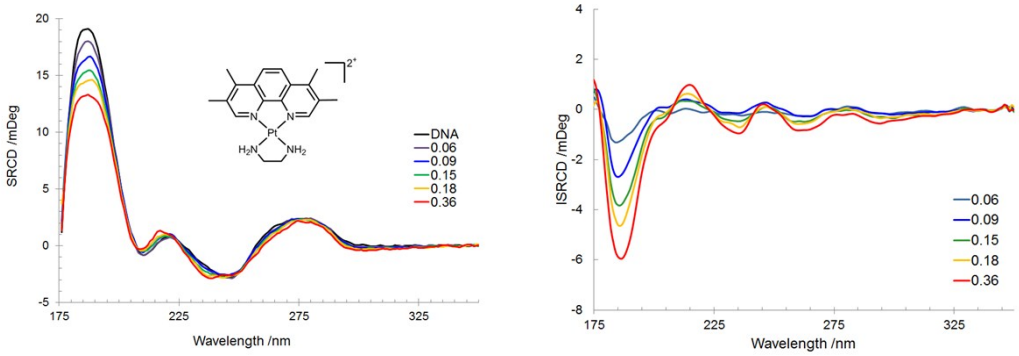


Figure S10 Expt A - SRCD and ISRCD spectra at different concentrations of metal complex 6, into ct-DNA in PS buffer.

Table S11 Complex **6**, $[Pt(3478\text{-}Mephen)(en)]^{2+}$ binding data, experiment B

Wavelength nm	Binding Constant $K \times 10^4$	Estimated Binding sites per complex, *n*
300	1.6 ± 0.15	3.5 ± 0.1
301	2.4 ± 0.20	4.7 ± 0.1
323	0.5 ± 0.19	5.4 ± 0.2

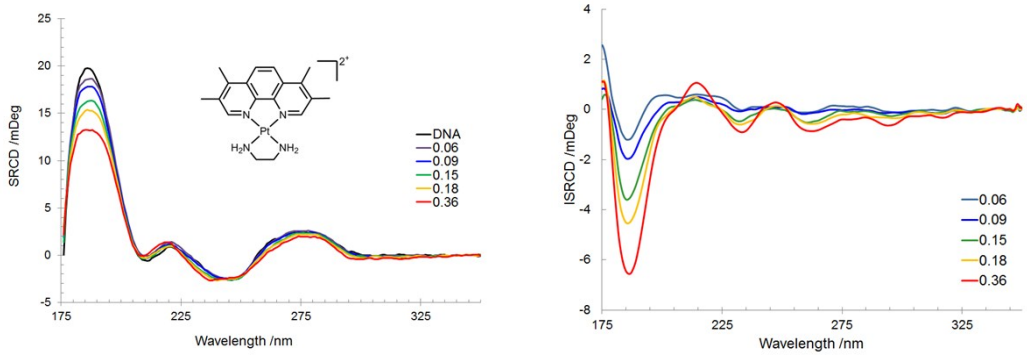


Figure S11 Expt B - SRCD and ISRCD spectra at different concentrations of metal complex 6, into calf thymus DNA in PS buffer.

References

1. F. H. Stootman, D. M. Fisher, A. Rodger and J. R. Aldrich-Wright, *The Analyst*, 2006, 131, 1145-1151.
2. G. Scatchard, *Ann. N. Y. Acad. Sci.*, 1949, 51, 660-672.
3. D. E. Schmechel and D. M. Crothers, *Biopolymers*, 1971, 10, 465-480.
4. S. Sarkar and B. Das, *Complex Metals*, 2014, DOI: 10.1080/2164232X.2014.887448, 80-87.
5. H. Wu, F. Kou, F. Jia, B. Liu, J. Yuan and Y. Bai, *Bioinorganic Chemistry and Applications*, 2011, 2011.
6. A. M. Pyle, J. P. Rehmann, R. Meshoyrer, C. V. Kumar, N. J. Turro and J. K. Barton, *Journal of the American Chemical Society*, 1989, 111, 3051-3058.

CHEMISTRY
A European Journal

Supporting Information

Quadruplex DNA-Stabilising Dinuclear Platinum(II) Terpyridine Complexes with Flexible Linkers

Dale L. Ang,[a] Benjamin W. J. Harper,[a] Leticia Cubo,[b] Oscar Mendoza,[b] Ramon Vilar,*[b, c] and Janice Aldrich-Wright*[a]

chem_201503663_sm_miscellaneous_information.pdf

Table of contents

SRCD Acquisition

Specifications of ASTRID include a maximum energy of 580 MeV, a wavelength range of 125–600 nm, maximum flux (180 and 240 nm) of 10^{12} photons/s, a TGM monochromator with a CaF_2 window and a variable sample beam size between 2 × 1 mm to 6 × 2 mm, a sample detector distance of 25 mm and a bandwidth of 0.6 mm. Daily calibration was achieved by determination of the ratio of D-10-camphorsulfonic acid (CSA) peaks at 192 and 290 nm, (2.05–2.07), which was consistent with the expected ratio of 1.96–2.15.[1]

SRCD Melt Spectra

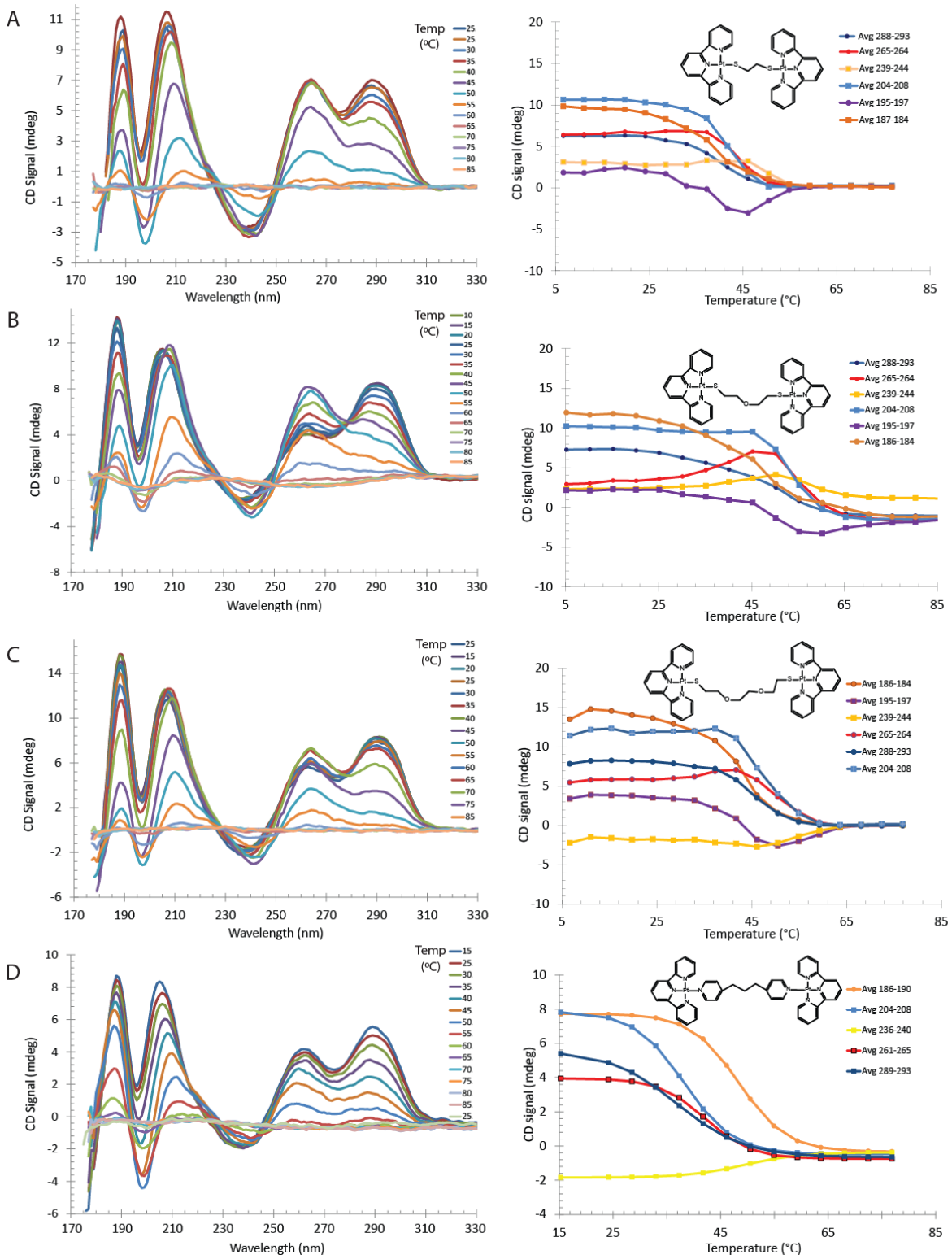


Figure S1. SRCD melt spectra of **1** (A), **2** (B), **3** (C) and **4** (D) with Q1 at a 1:1 MC:QDNA ratio (left). Thermally induced change in SRCD signal as a function of temperature with **1**, **2**, **3** and **4** respectively (right).

FID Competition

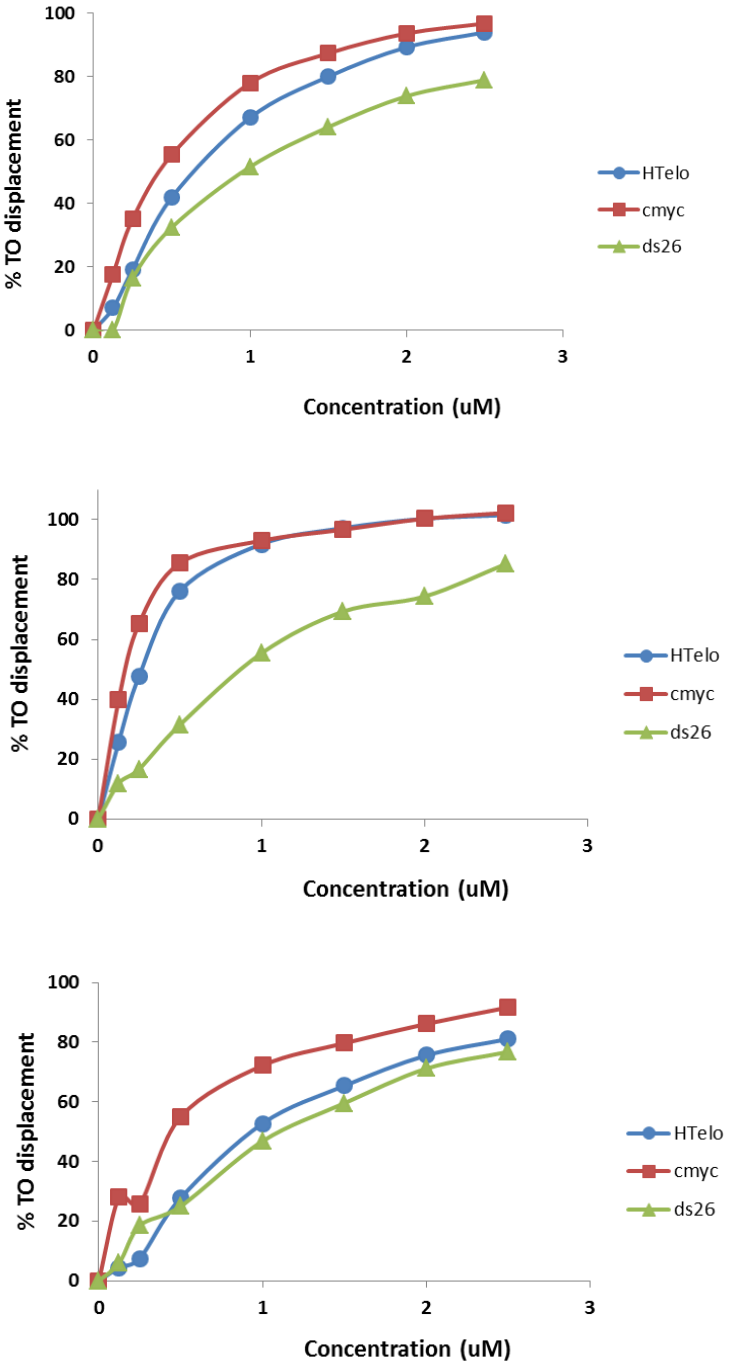


Figure S2. Graphical representation of TO displacement from *c-myc* QDNA (■), HTelo QDNA (●) and ds26 duplex DNA (▲) upon increasing concentration of complex **4** (above), **2** (centre) and **3** (below).

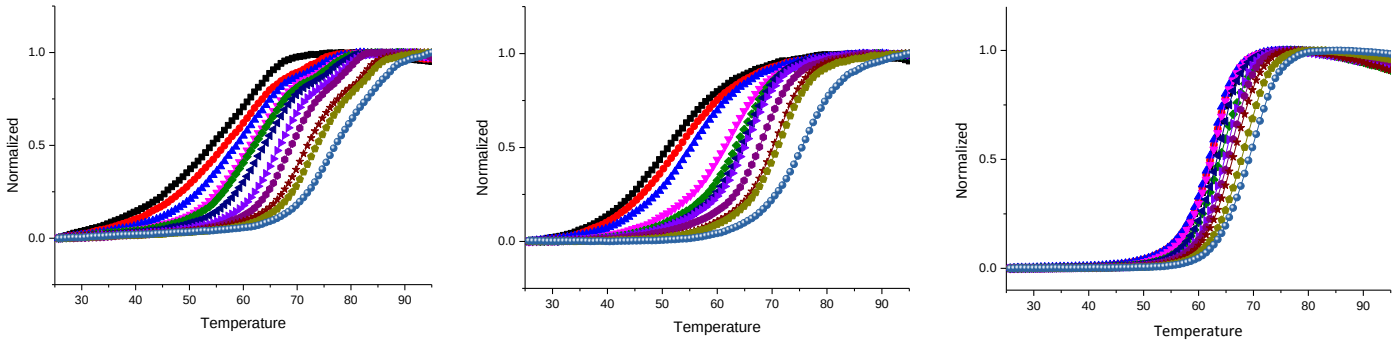


Figure S3. FRET melting results of experiments carried out with F-myc (left), F21T-HTelo (center) and DS1-T (right) with increasing concentration of **2** from 0-4 µM.

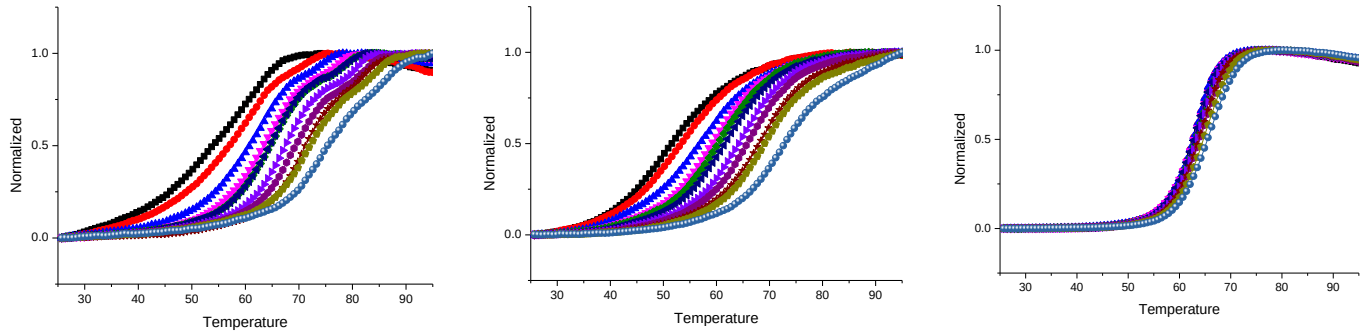


Figure S4. FRET melting results of experiments carried out with F-myc (left), F21T-HTelo (center) and DS1-T (right) with increasing concentration of **3** from 0-4 µM.

DFT Calculations

The optimised structures of **1-3** show a Pt-S distance of 2.376 Å with the Pt metal centrally placed within the ligand (Pt-N_{cis} distance 2.005 Å; Pt-N_{trans} distance 2.062 Å; N_{cis}-Pt-N_{trans} dihedral 80°). Structure **4** shows a Pt-N_{Pyr} distance of 2.107 Å with the Pt metal centrally placed within the ligand (Pt-N_{cis} distance 1.972 Å; Pt-N_{trans} distance 2.073 Å; N_{cis}-Pt-N_{trans} dihedral 87°). These values are in good agreement with experimental values previously reported for Pt-terpyridine complexes.[2] The pyridine rings of **1-4** are all coplanar; the Pt-S-C dihedral for **1-3** was ~103°, the pyridine linker of **4** perpendicular to the plane. Linker lengths from Pt–Pt were 10.2 Å (**1**), 14.6 Å (**2**), 19.1 Å (**3**) and 18.7 Å (**4**).

Docking Simulations

Preparation of the DNA substrates included removal of solvent molecules and assignment of polar hydrogens. The grid boxes were defined to encompass the entire structureof the relevant PDB entry. Each docking simulation was carried out with five repeats and an exhaustiveness value of 64.

SRCD Melt Curve Fits

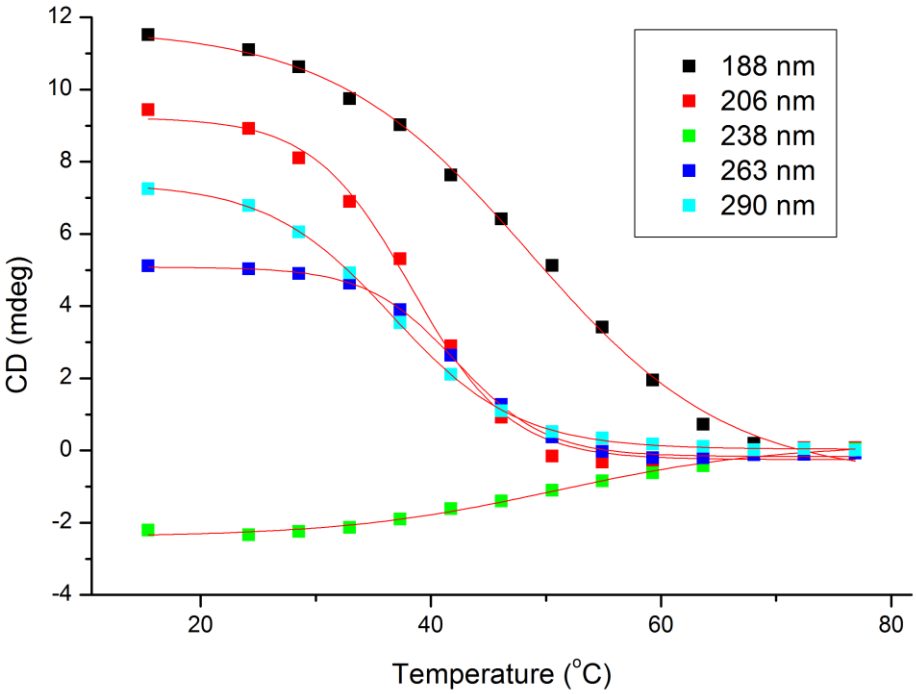


Figure S5. Melt curves of **Q1** only.

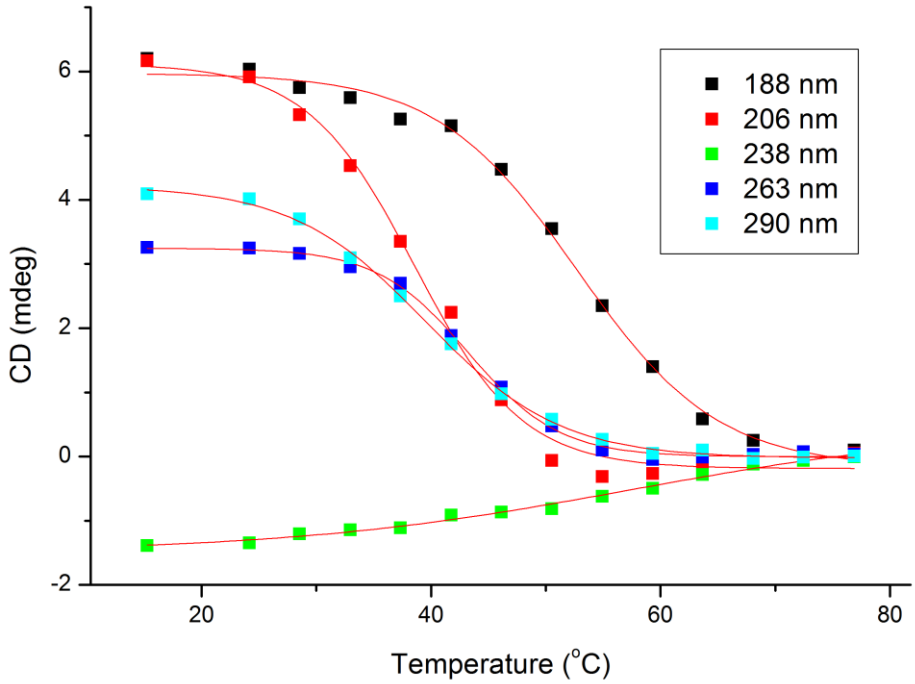


Figure S6. Melt curves of **Q1** incubated with **1**.

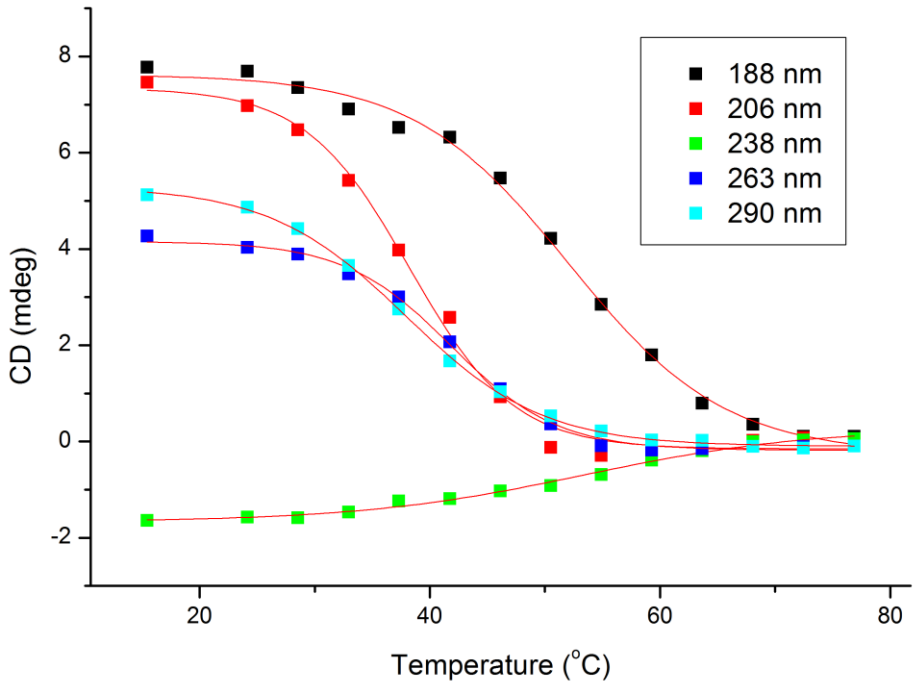


Figure S7. Melt curves of **Q1** incubated with **2**.

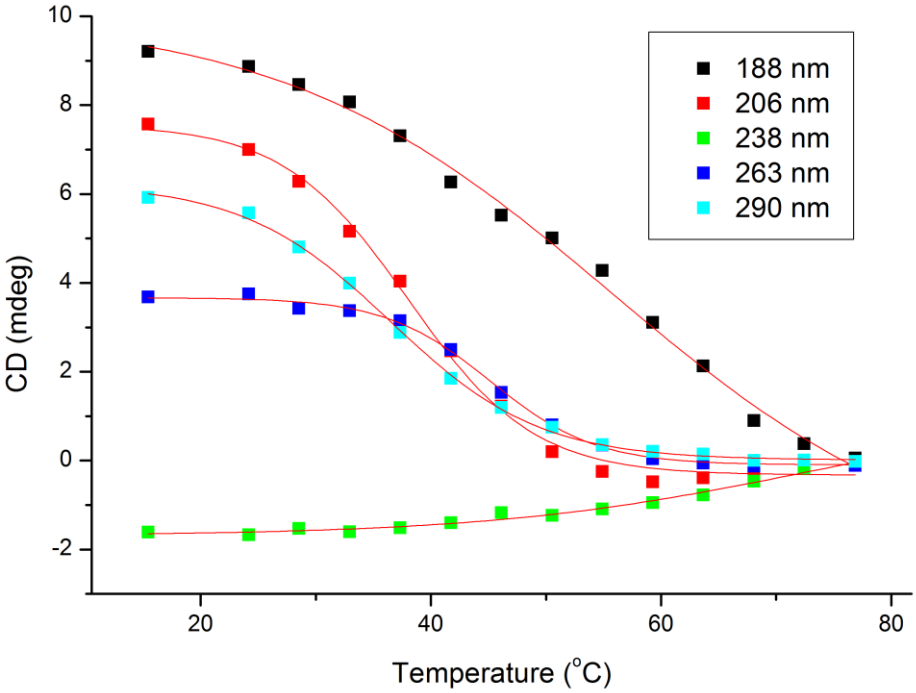


Figure S8. Melt curves of **Q1** incubated with **3**.

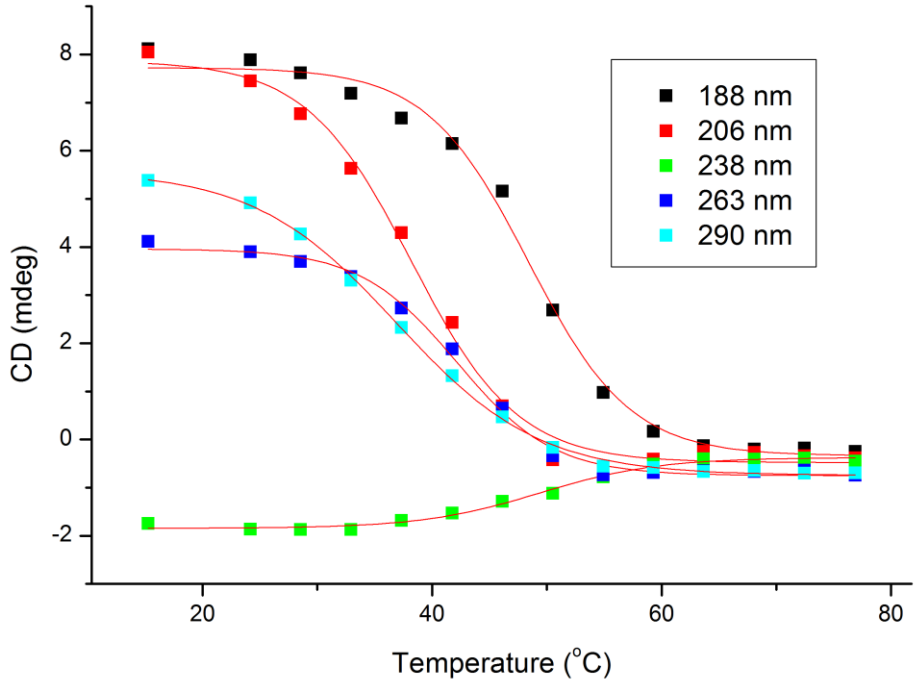


Figure S9. Melt curves of **Q1** incubated with **4**.

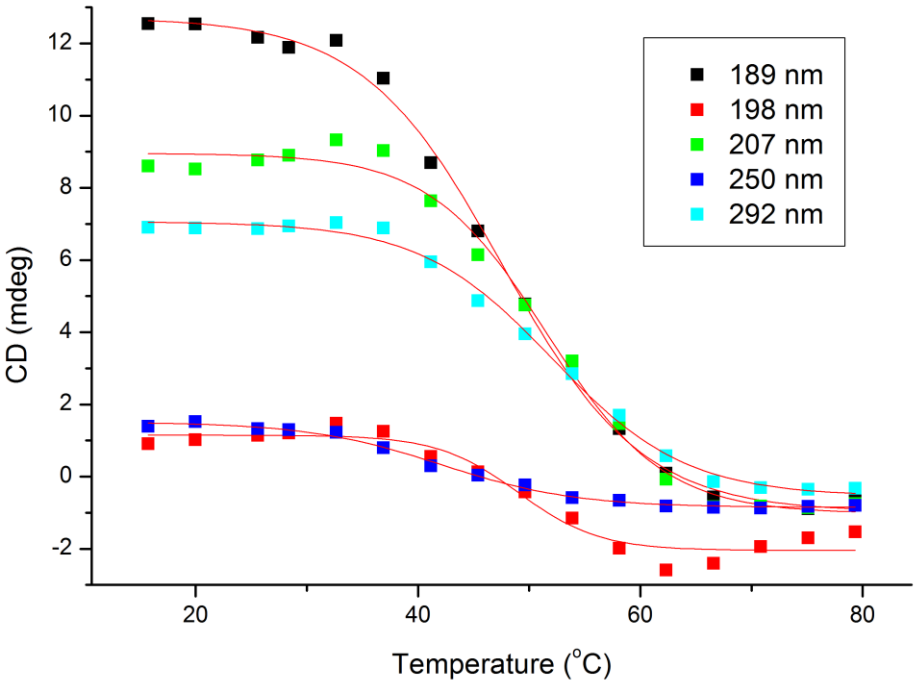


Figure S10. Melt curves of **Q2** only.

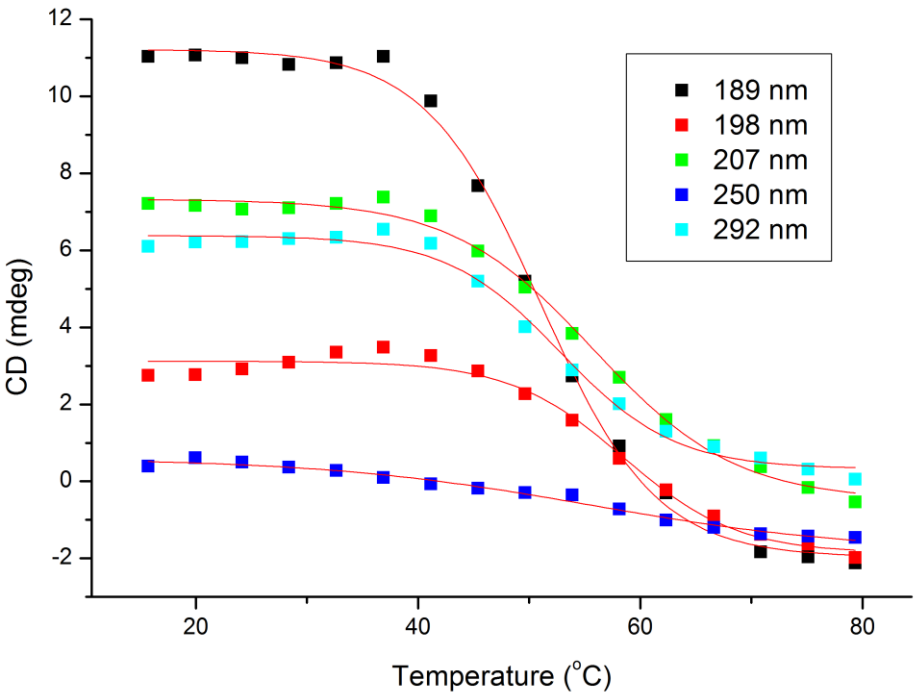


Figure S11. Melt curves of **Q2** incubated with **2**.

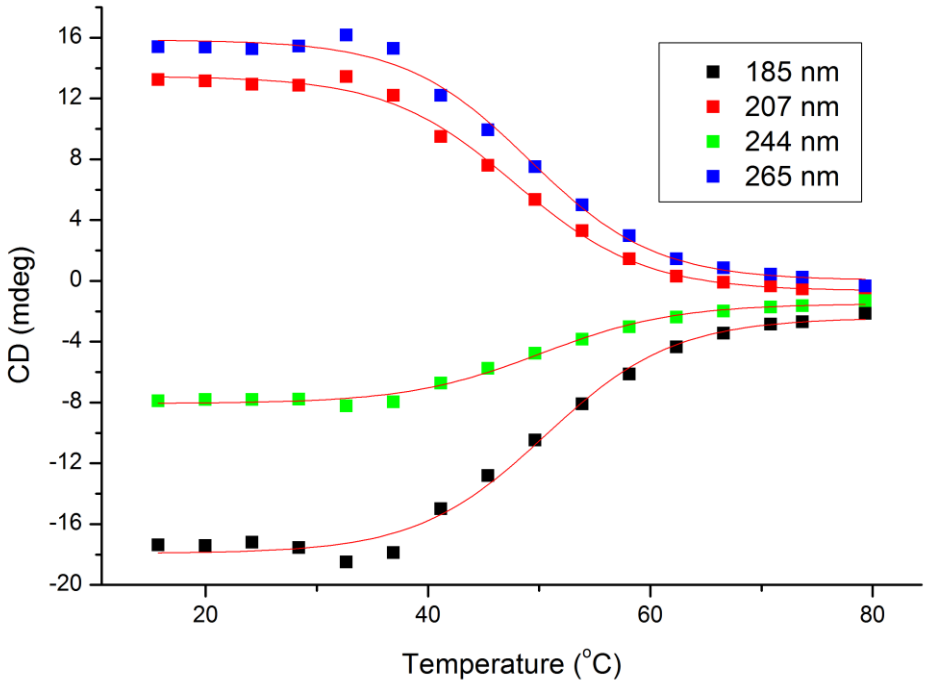


Figure S12. Melt curves of **Q3** only.

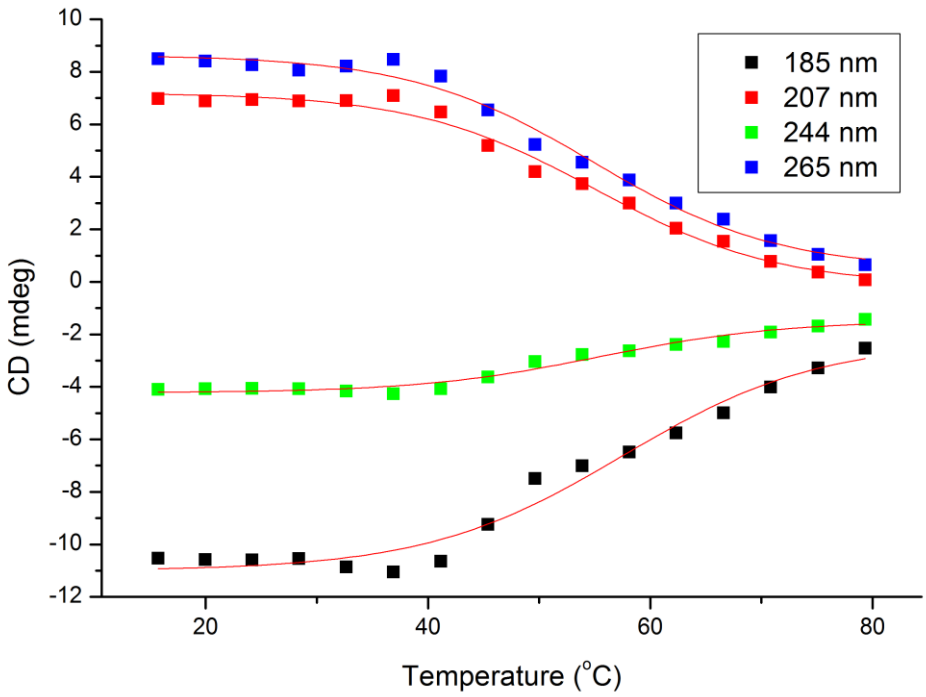


Figure S13. Melt curves of **Q3** incubated with 2.

References

1. A. J. Miles, S. V. Hoffmann, Y. Tao, R. W. Janes and B. A. Wallace, *Spectroscopy*, 2007, **21**, 245-255.
2. K. Suntharalingam, A. J. White and R. Vilar, *Inorg. Chem.*, 2009, **48**, 9427-9435.

SUPPLEMENTARY INFORMATION

A study of Pt(II)-phenanthroline complex interactions with double-stranded and G-quadruplex DNA by ESI-MS, circular dichroism, and computational docking.

Dale L. Ang,[a] Celine Kelso,[c] Jennifer L. Beck,[c] Stephen F. Ralph,[c] David G. Harman,[b] Janice R. Aldrich-Wright[ab*]

[a]*Nanoscale Organisation and Dynamics Group, School of Science,*
Western Sydney University, Penrith, NSW 2751, Australia
Phone: +61 2 4620 3218
Fax: +61 2 4620 3025
Email: j.aldrich-wright@westernsydney.edu.au

[b]*School of Medicine*
Western Sydney University
Locked Bag 1797 Penrith South DC, NSW 2751,
Australia

[c]*School of Chemistry and Molecular Bioscience*
University of Wollongong
Northfields Avenue, Wollongong, 2522,
Australia

* Please address editorial correspondence to Prof. Janice R. Aldrich-Wright via e-mail: j.aldrich-wright@westernsydney.edu.au or via facsimile: +61 2 4620 3025.

ORCID:

Dale L. Ang https://orcid.org/0000-0001-5142-7555

Celine Kelso https://orcid.org/0000-0003-0032-9650

Jennifer L. Beck https://orcid.org/0000-0001-6058-4886

David G. Harman https://orcid.org/0000-0003-4376-832X

Janice R. Aldrich-Wright https://orcid.org/0000-0002-6943-6908

Contents

ESI -MS data processing methodology

- Relative abundance refers to the normalised amount of DNA in complex with ligand as a percentage of all DNA (calculated from the sums of relative intensities of ions from the DNA:ligand divided by the sum of intensities of all ions in the ESI mass spectrum). The percentage bound DNA is calculated using the equation:

$$\%\,bound = \frac{[DNA:ligand]}{[DNA + DNA:ligand]}$$

- Initial processing steps within MassLynx:
    a. Identify the major abundance charge states, typically 4^-, 5^-, 6^-;
    b. Zoom into the area of interest (Display → Range → From… 1600–3000 *m/z* for G4DNA, 1100–2800 *m/z* for dsDNA);
    c. Process → Background Subtract (Polynomial order: 2, Below curve: 40, Tolerance: 0.010);
    d. Process → Spectrum Smooth (Smooth window: 20, Number of smooths: 1, Savitzky-Golay);
    e. Search the Display → List Spectrum, record, enter into spreadsheet to calculate *%bound*.
- Raw ESIMS data and calculations are contained in SUPP_ESIMS_results.pdf.

ESI-MS Spectra

Conditions: Mass spectra were obtained in ESI negative mode using a Waters XEVO QToF ESI mass spectrometer with a Z-spray ionisation source. Spectra were obtained with an MCP potential of 2.4 kV, a cone voltage of 25 V, capillary tip potential of 2.3 kV and the extraction cone at 4.0 V. The desolvation gas (nitrogen) flow was 150 L hr^{-1} at 150 °C, cone gas flow 20 L hr^{-1}, with a source temperature of 70 °C. Sample solutions were infused using a KD Scientific Model 100 syringe pump, at a flow rate of 5 µL min^{-1}. A lockspray solution of 200 pg μL^{-1} leucine enkephalin in 50% aqueous acetonitrile plus 0.1% formic acid was infused at 5 µL min^{-1} simultaneously, in order to enable mass accuracy correction. All sample spectra were obtained in ESI negative ion mode over a 50–4000 m/z range. MassLynx v4.1 was used for the processing of spectra. A 5 min window of the total ion chromatogram was combined, and the region from 1200–3000 m/z chosen so as to encompass all visible charge states. All samples were prepared at 1:1, 5:1 and 10:1 PC:DNA ratios, vortexed and adjusted to a final volume of 25 µL. The solutions were vortexed and equilibrated at room temperature for 5 min before injection. DNA concentration in each sample was 10 µM. The solution was vortexed and equilibrated at room temperature for 5 min before injection. Sequences: G4DNA = 5`-TTG GGG GT-3`, dsDNA = 5`-CGC GAA TTC GCG-3`.

PC:DNA ratios used. Lower = 1:1, Middle = 5:1, Upper = 10:1. The 10:1 ratio ESI-MS spectra for dsDNA are omitted due to excessive noise.

ESI-MS Spectra of G4DNA+PC

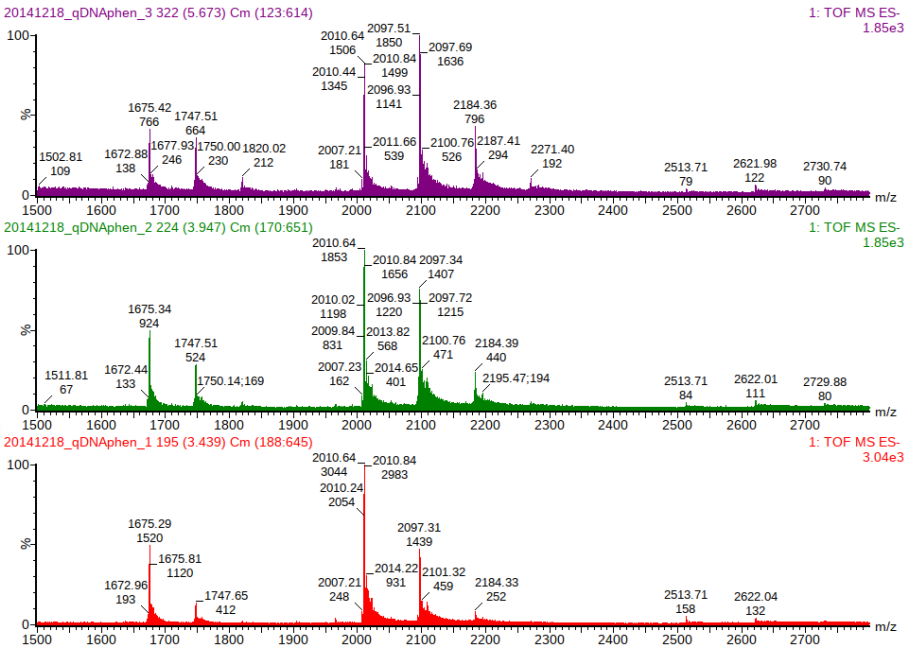


*Figure 1: Negative ion ESI mass spectra of G4DNA with **1** [Pt(en)(phen)]Cl_2*

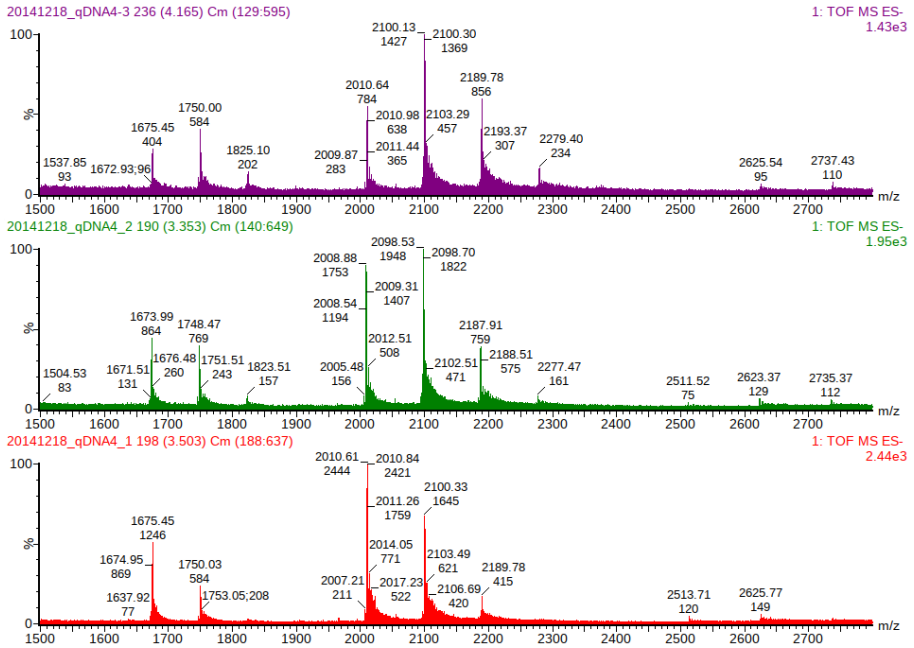


*Figure 2: Negative ion ESI mass spectra of G4DNA with **2** [Pt(en)(4-mephen)]Cl_2*

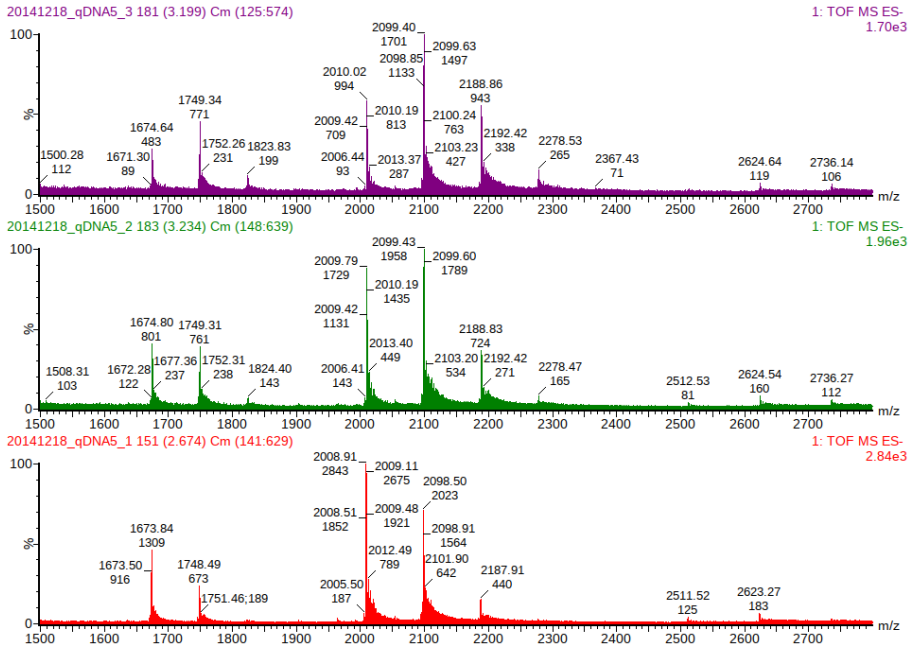


*Figure 3: Negative ion ESI mass spectra of G4DNA with **3** [Pt(en)(5-mephen)]Cl_2*

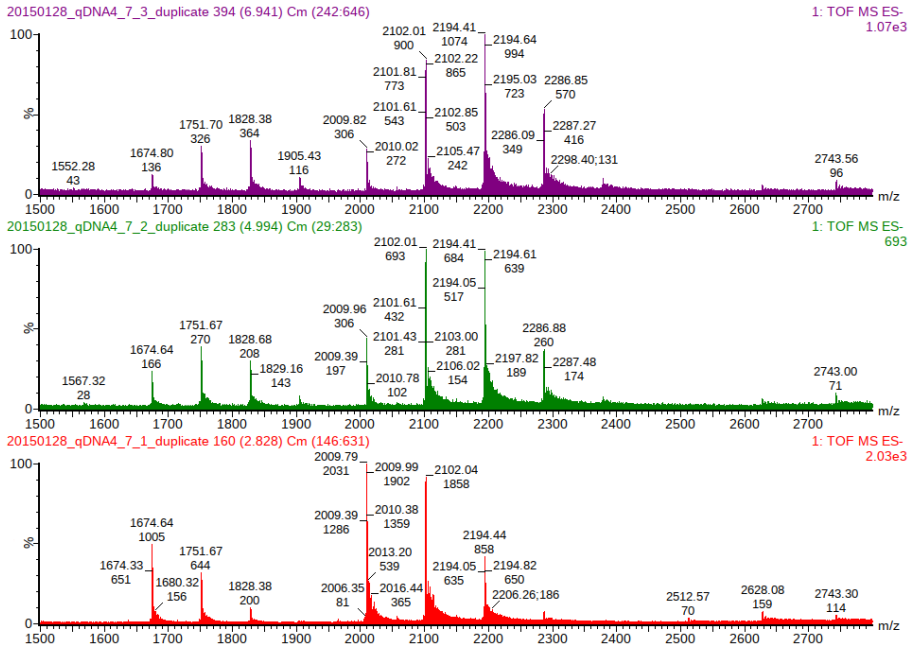


*Figure 4: Negative ion ESI mass spectra of G4DNA with **4** [Pt(en)(4,7-me$_2$phen)]Cl$_2$*

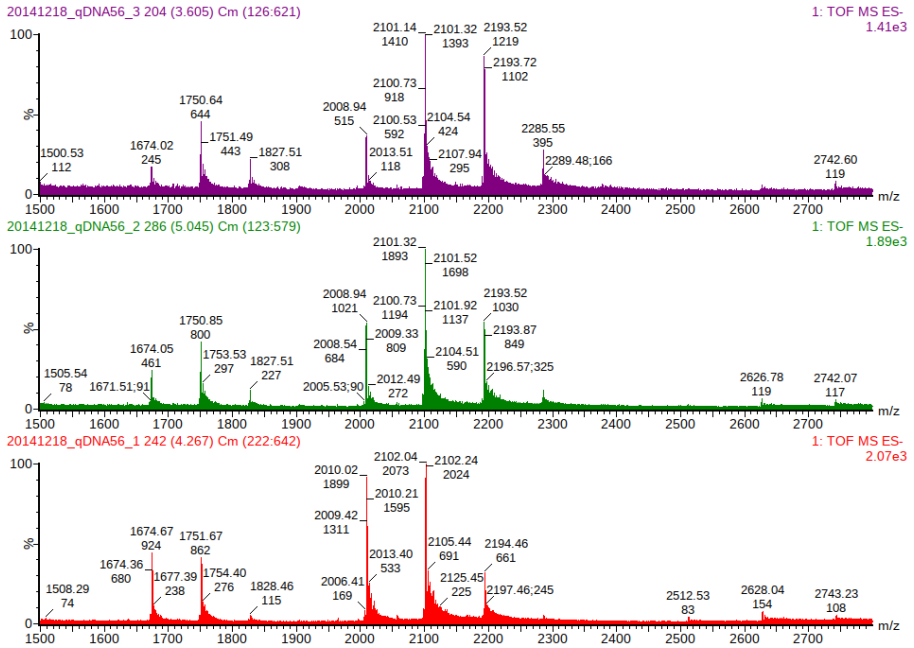


*Figure 5: Negative ion ESI mass spectra of G4DNA with **5** [Pt(en)(5,6-me$_2$phen)]Cl$_2$*

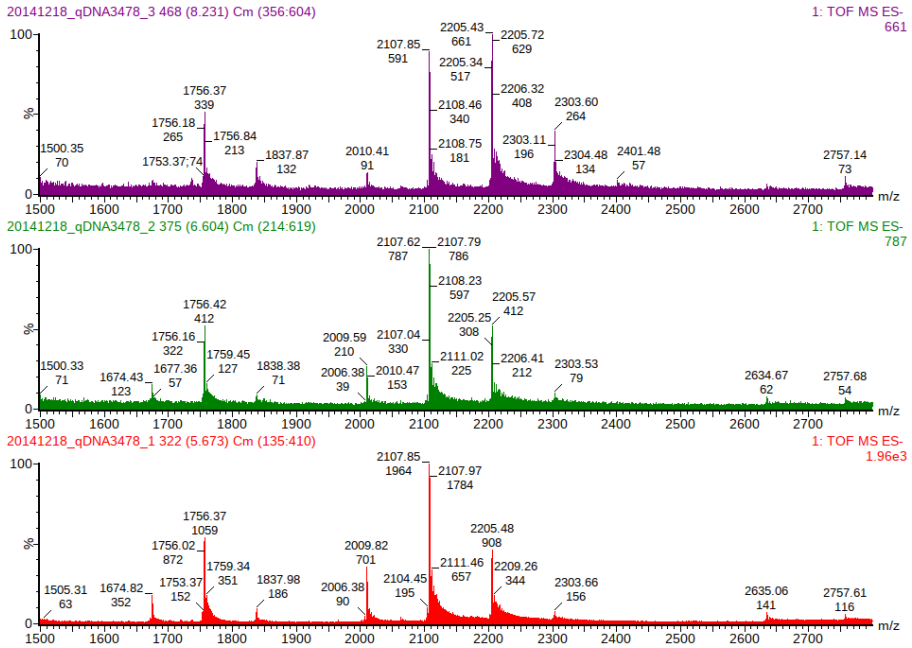


*Figure 6: Negative ion ESI mass spectra of G4DNA with **6** [Pt(en)(3,4,7,8-me$_4$phen)]Cl$_2$*

ESI-MS Spectra of dsDNA+PC

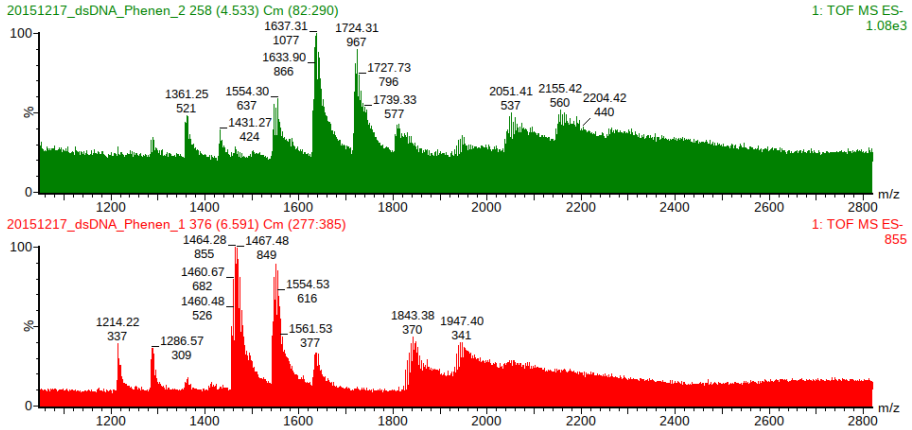


*Figure 7: Negative ion ESI mass spectra of dsDNA with **1** [Pt(en)(phen)]Cl$_2$*

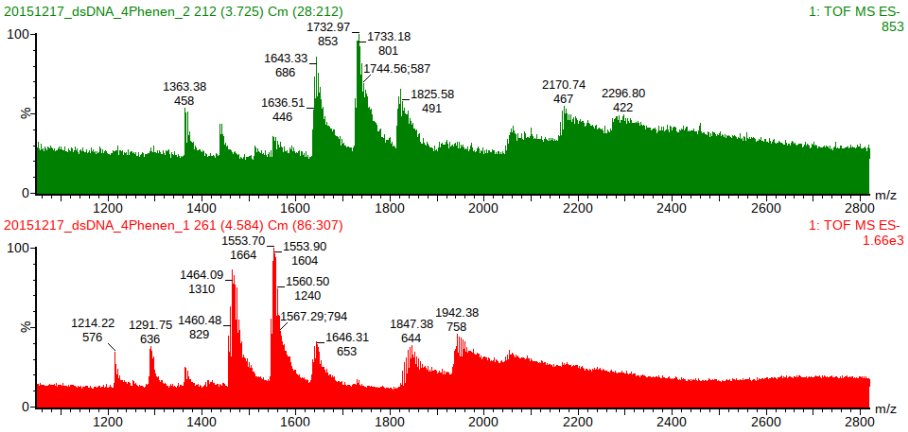


*Figure 8: Negative ion ESI mass spectra of dsDNA with **2** [Pt(en)(4-mephen)]Cl_2*

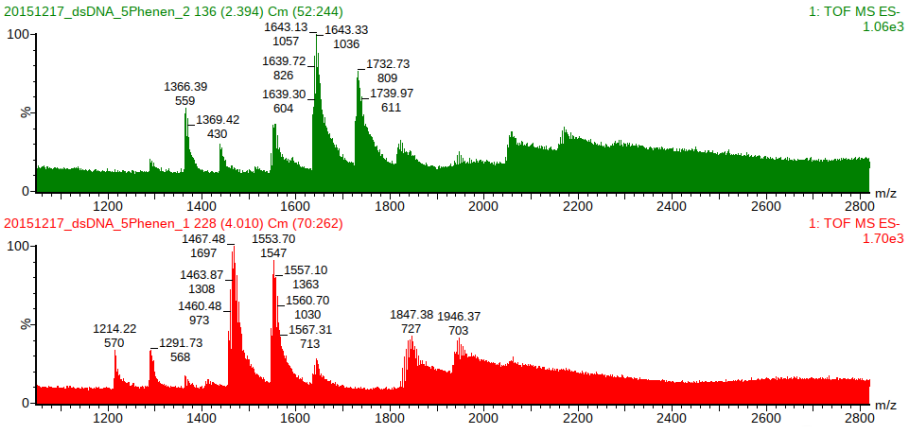


*Figure 9: Negative ion ESI mass spectra of dsDNA with **3** [Pt(en)(5-mephen)]Cl_2*

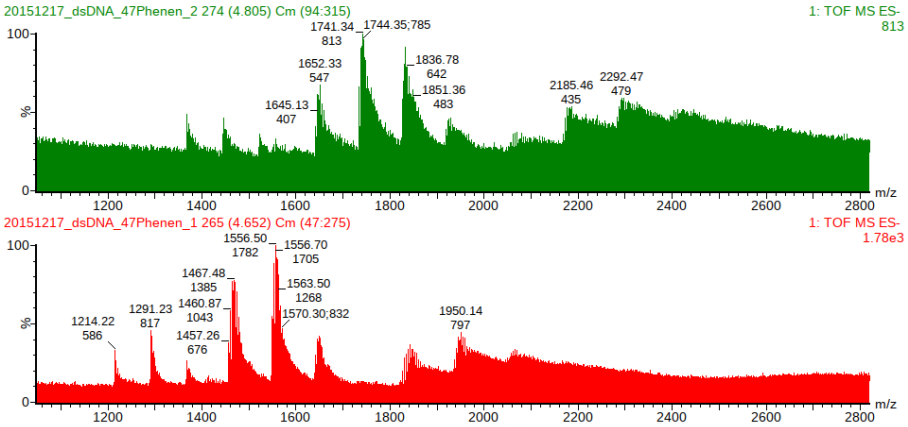


*Figure 10: Negative ion ESI mass spectra of dsDNA with **4** [Pt(en)(4,7-me$_2$phen)]Cl$_2$*

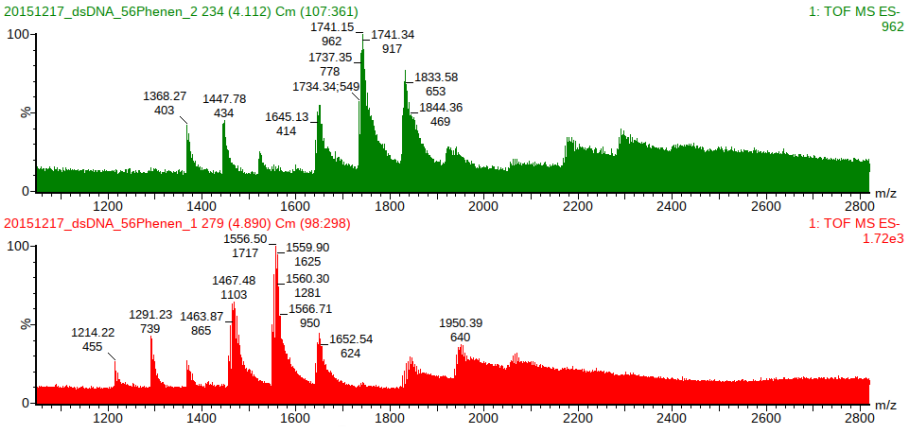


*Figure 11: Negative ion ESI mass spectra of dsDNA with **5** [Pt(en)(5,6-me$_2$phen)]Cl$_2$*

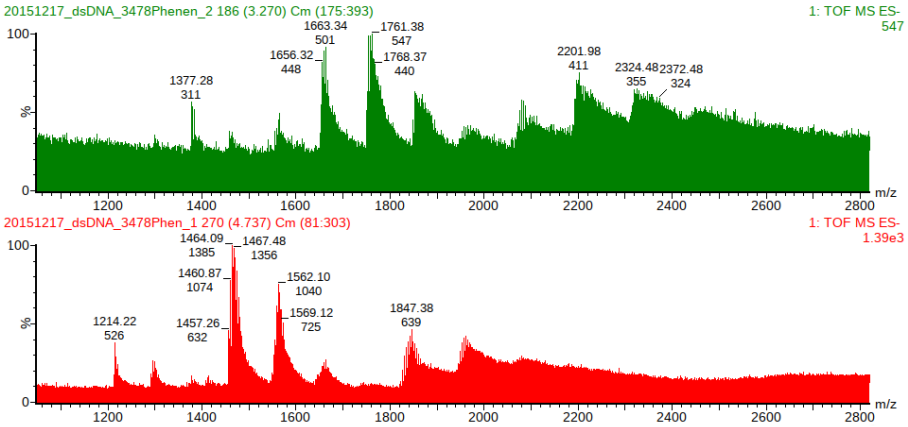


*Figure 12: Negative ion ESI mass spectra of dsDNA with **6** [Pt(en)(3,4,7,8-me$_4$phen)]Cl$_2$*

HPLC retention times

Table S1: Retention times and corrected retention times (k`) for each complex.

Compound	Name	t_0 (min)	Retention time (t_R min)	k` ($t_R - t_0$)
1	[Pt(en)(phen)]Cl_2	1.487	6.223	4.736
2	[Pt(en)(4-Mephen)]Cl_2	1.485	7.273	5.788
3	[Pt(en)(5-Mephen)]Cl_2	1.485	7.423	5.938
4	[Pt(en)(4,7-Me_2phen)]Cl_2	1.489	8.317	6.828
5	[Pt(en)(5,6-Me_2phen)]Cl_2	1.490	8.477	6.987
6	[Pt(en)(3,4,7,8-Me_4phen)]Cl_2	1.586	10.020	8.434

HPLC traces

Conditions: Analytical reverse-phase HPLC (RP-HPLC) of the PCs was performed on an Agilent Technologies 1260 Infinity, equipped with a Phenomenex C18 reverse phase column (4.6 × 150 mm, 5 µm). The mobile phase consisted of solvent A (0.06% TFA in water) and solvent B (0.06% TFA in ACN:H_2O) (90:10). Samples were injected and eluted with a gradient of 0–40 (% B) over 15 min at a flow rate of 1 mL min^{-1}. Detection was by photodiode array at 254 nm and 214 nm. Reported retention times were corrected by subtracting the elution time of a non-retained compound.

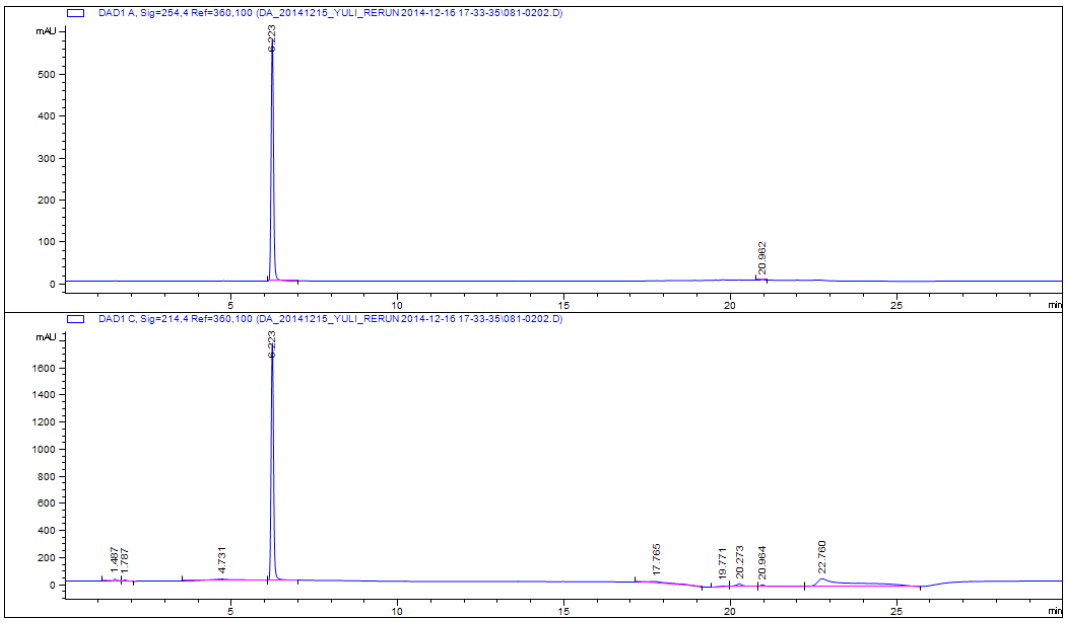


Figure S13 RP-HPLC trace of **1** [Pt(en)(phen)]Cl_2, 5 µm C18 column, 0–40 (% H_2O–ACN:H_2O) @ 1 mL min^{-1}, UV detection at 254 nm (upper) and 214 nm (lower).

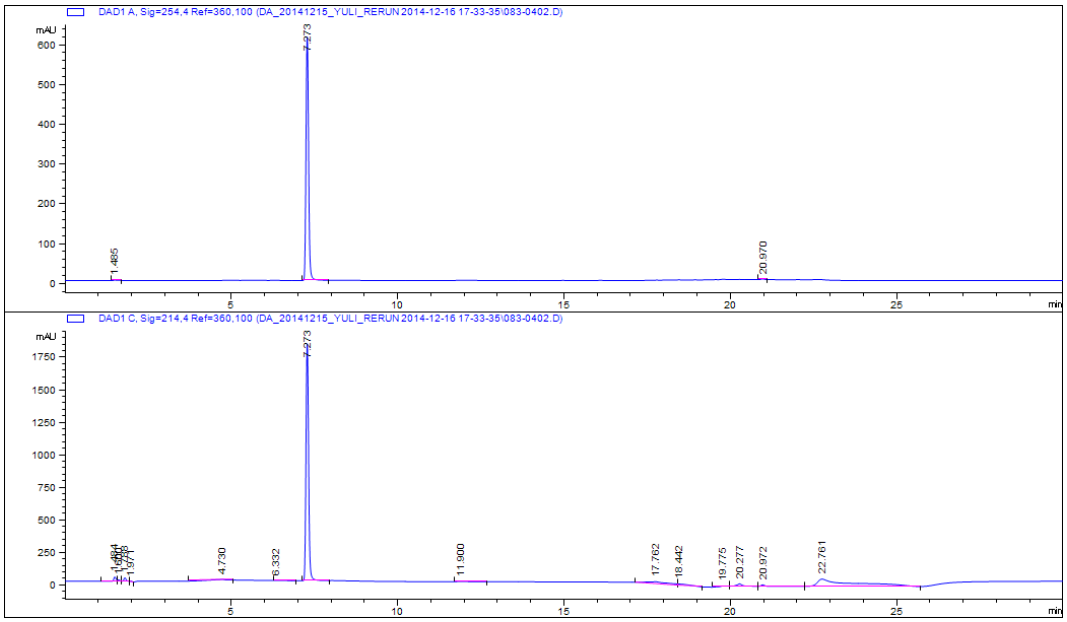


Figure S14 RP-HPLC trace of **2** [Pt(en)(4-Mephen)]Cl_2, 5 µm C18 column, 0–40 (% H_2O–ACN:H_2O) @ 1 mL min^{-1}, UV detection at 254 nm (upper) and 214 nm (lower).

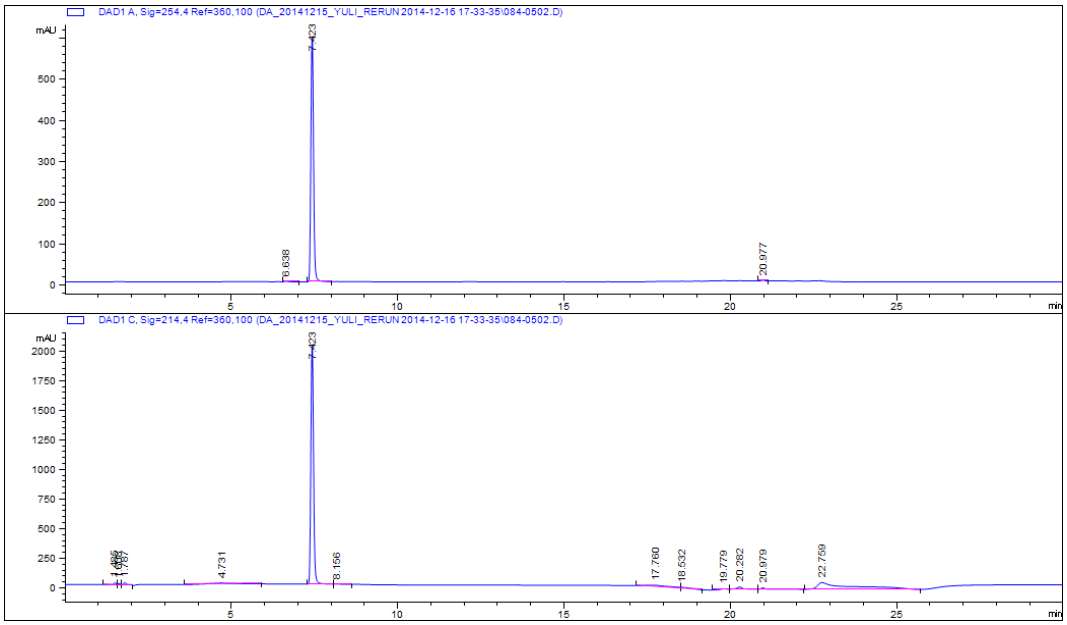


Figure S15 RP-HPLC trace of **3** [Pt(en)(5-Mephen)]Cl_2, 5 µm C18 column, 0–40 (% H_2O–ACN:H_2O) @ 1 mL min^{-1}, UV detection at 254 nm (upper) and 214 nm (lower).

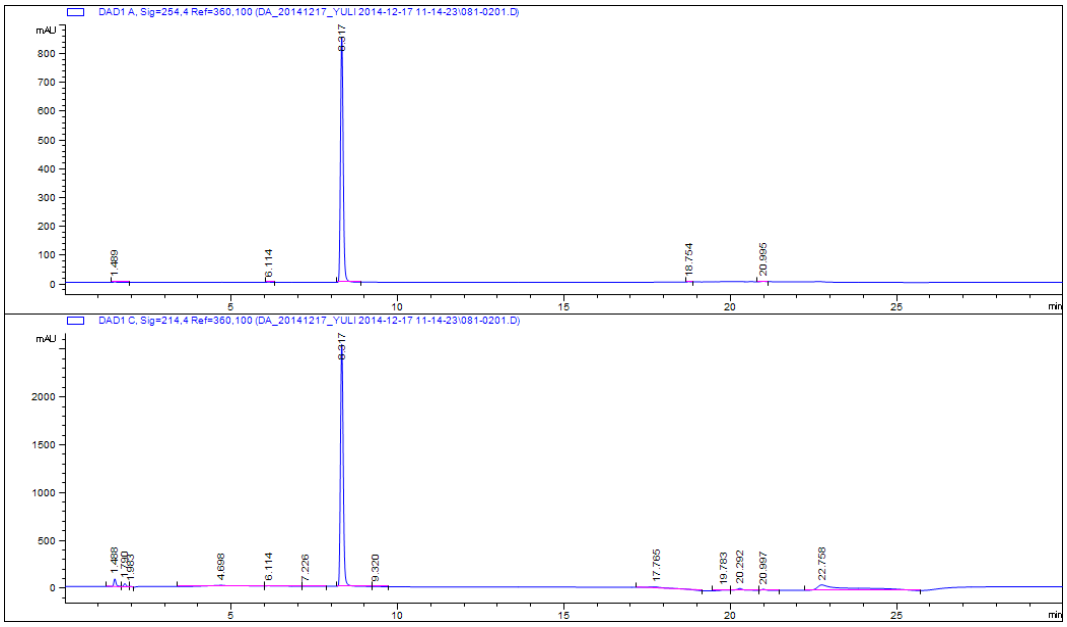


Figure S16 RP-HPLC trace of **4** [Pt(en)(4,7-Me_2phen)]Cl_2, 5 µm C18 column, 0–40 (% H_2O–ACN:H_2O) @ 1 mL min^{-1}, UV detection at 254 nm (upper) and 214 nm (lower).

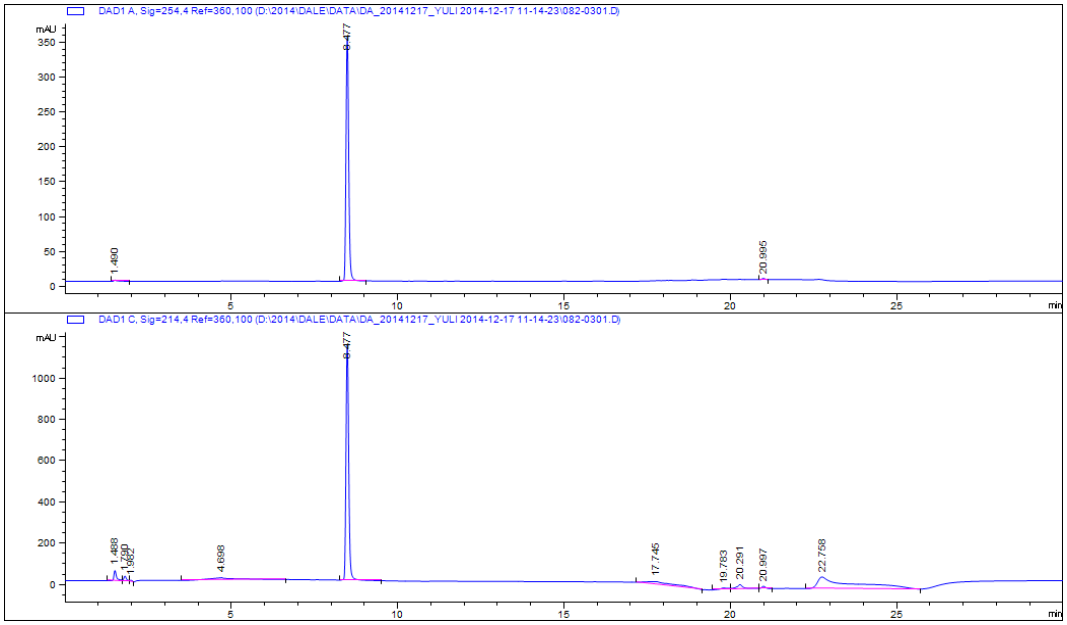


Figure S17 RP-HPLC trace of **5** [Pt(en)(5,6-Me_2phen)]Cl_2, 5 µm C18 column, 0–40 (% H_2O–ACN:H_2O) @ 1 mL min^{-1}, UV detection at 254 nm (upper) and 214 nm (lower).

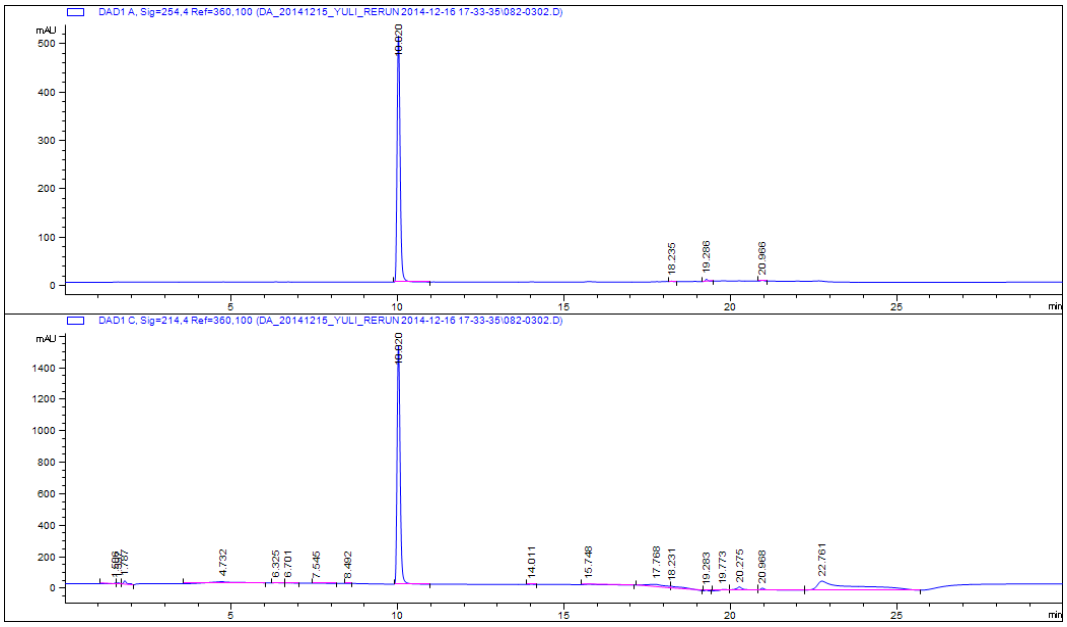


Figure S18 RP-HPLC trace of **6** [Pt(en)(3,4,7,8-Me_4phen)]Cl_2, 5 µm C18 column, 0–40 (% H_2O–ACN:H_2O) @ 1 mL min^{-1}, UV detection at 254 nm (upper) and 214 nm (lower).

Elution time versus % DNA bound

Table S2: Data for % bound DNA at 1:1 PC:DNA ratio.

PC	K` /min	% Bound G4	% Bound D2
1	4.74	32	48
2	5.79	41	66
3	5.94	45	57
4	6.83	54	67
5	6.99	56	70
6	8.43	80	51

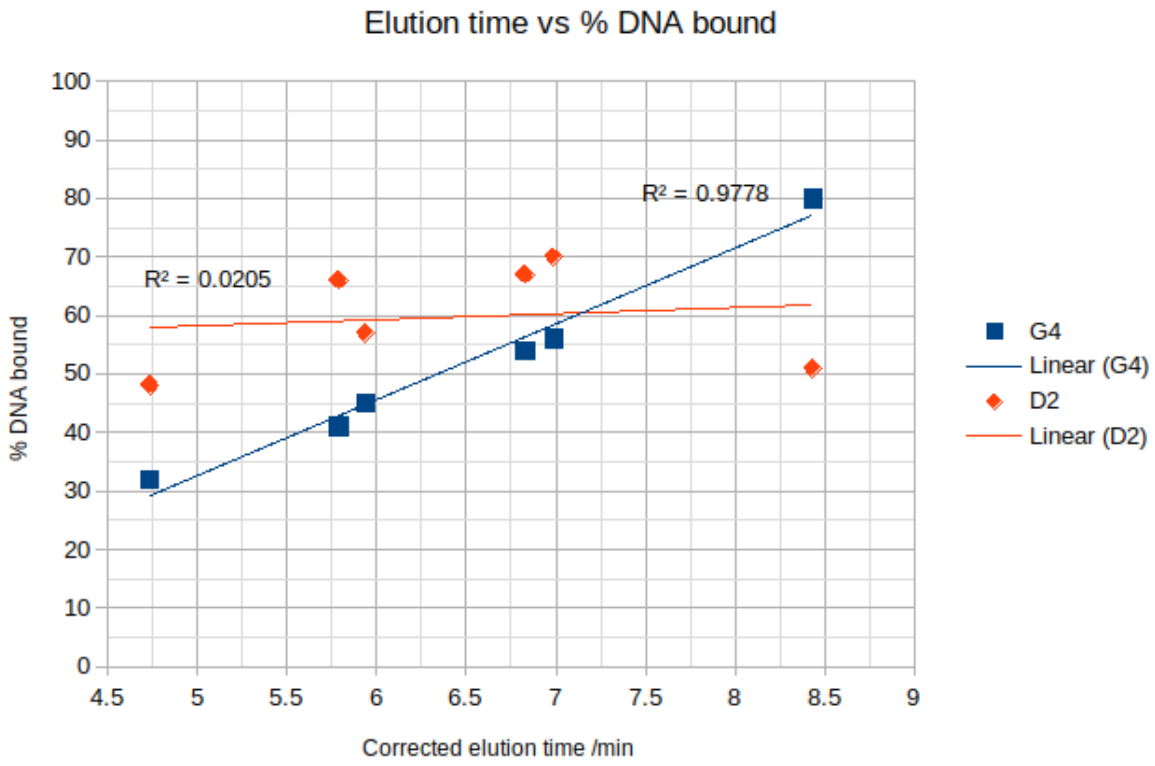


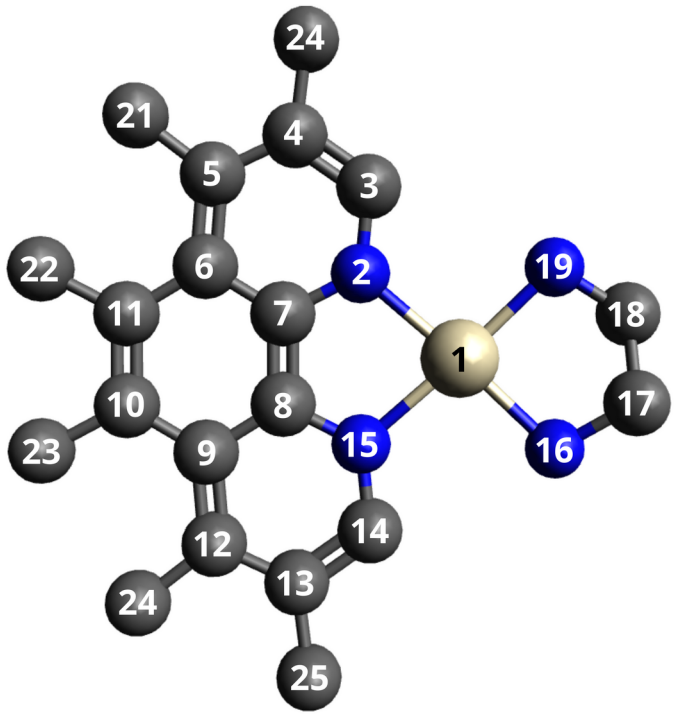
Figure S19: Plots of corrected elution time vs % bound DNA for G4 and D2 at 1:1 PC:DNA ratio.

Calculated partial atomic charges

Table S3: Calculated Mullikan charges for each complex. Atom descriptors: ar = aromatic, en = ethylenediamine.

		Complex					
Atom ID	**Atom**	**1**	**2**	**3**	**4**	**5**	**6**
1	Pt	0.7227	0.7220	0.7217	0.7188	0.7220	0.7137
2	N (ar)	-0.6662	-0.6773	-0.6698	-0.6755	-0.6748	-0.6777
3	C (ar)	0.2611	0.2571	0.2597	0.2583	0.2566	0.2059
4	C (ar)	0.0846	0.0390	0.0808	0.0263	0.0757	0.0693
5	C (ar)	0.1214	0.1167	0.1057	0.1311	0.0977	0.0856
6	C (ar)	0.1263	0.1060	0.0976	0.1062	0.1156	0.1129
7	C (ar)	0.2997	0.3023	0.3040	0.2947	0.2972	0.2939
8	C (ar)	0.2997	0.2924	0.2942	0.2947	0.2972	0.2909
9	C (ar)	0.1263	0.1300	0.1445	0.1062	0.1156	0.1110
10	C (ar)	0.0612	0.0508	-0.0162	0.0371	0.0426	0.0239
11	C (ar)	0.0612	0.0447	0.1107	0.0371	0.0425	0.0280
12	C (ar)	0.1214	0.1186	0.1151	0.1311	0.0977	0.0948
13	C (ar)	0.0846	0.0821	0.0805	0.0263	0.0757	0.0654
14	C (ar)	0.2611	0.2589	0.2556	0.2583	0.2566	0.2051
15	N (ar)	-0.6663	-0.6668	-0.6695	-0.6755	-0.6748	-0.6756
16	N (en)	0.0490	0.0475	0.0485	0.0476	0.0479	0.0471
17	C (en)	0.3017	0.2997	0.3006	0.2978	0.2989	0.2951
18	C (en)	0.3017	0.2995	0.3005	0.2978	0.2989	0.2951
19	N (en)	0.0490	0.0480	0.0479	0.0476	0.0479	0.0470
20	C (methyl)	-	-	-	-	-	0.0806
21	C (methyl)	-	0.1287	-	0.1170	-	0.1068
22	C (methyl)	-		0.0879	-	0.0815	-
23	C (methyl)	-	-	-	-	0.0815	-
24	C (methyl)	-	-	-	0.1170	-	0.1019
25	C (methyl)	-	-	-	-	-	0.0794

Atom numbering



Calculated atomic charges

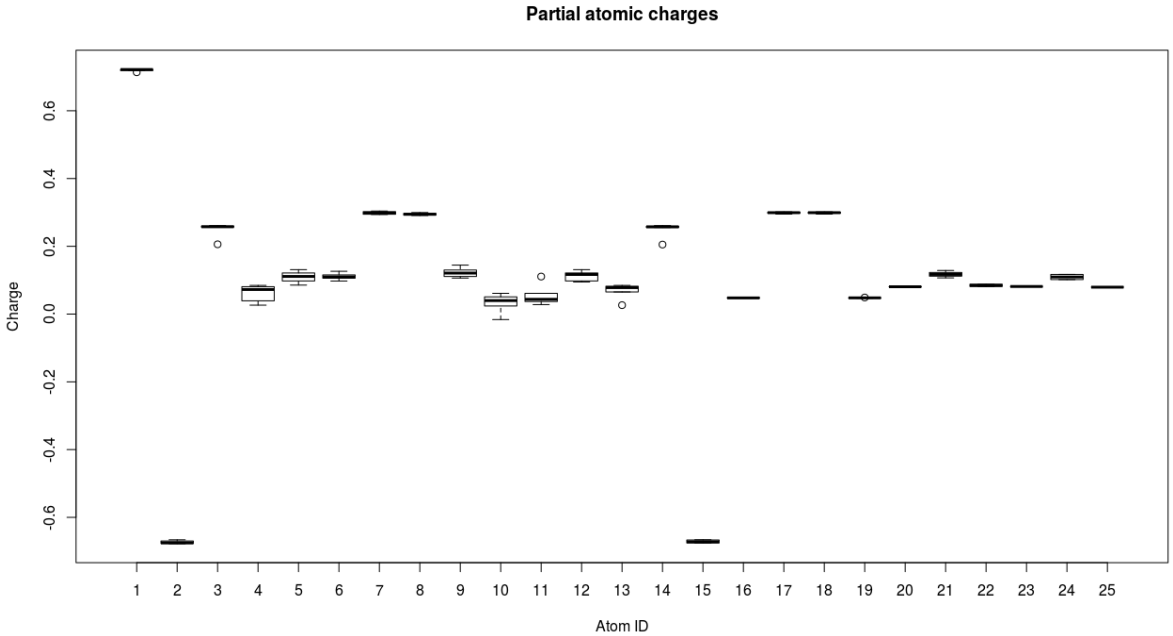


Figure S20: Boxplot of calculated Mullikan charges for each complex.

Electrostatic potential maps

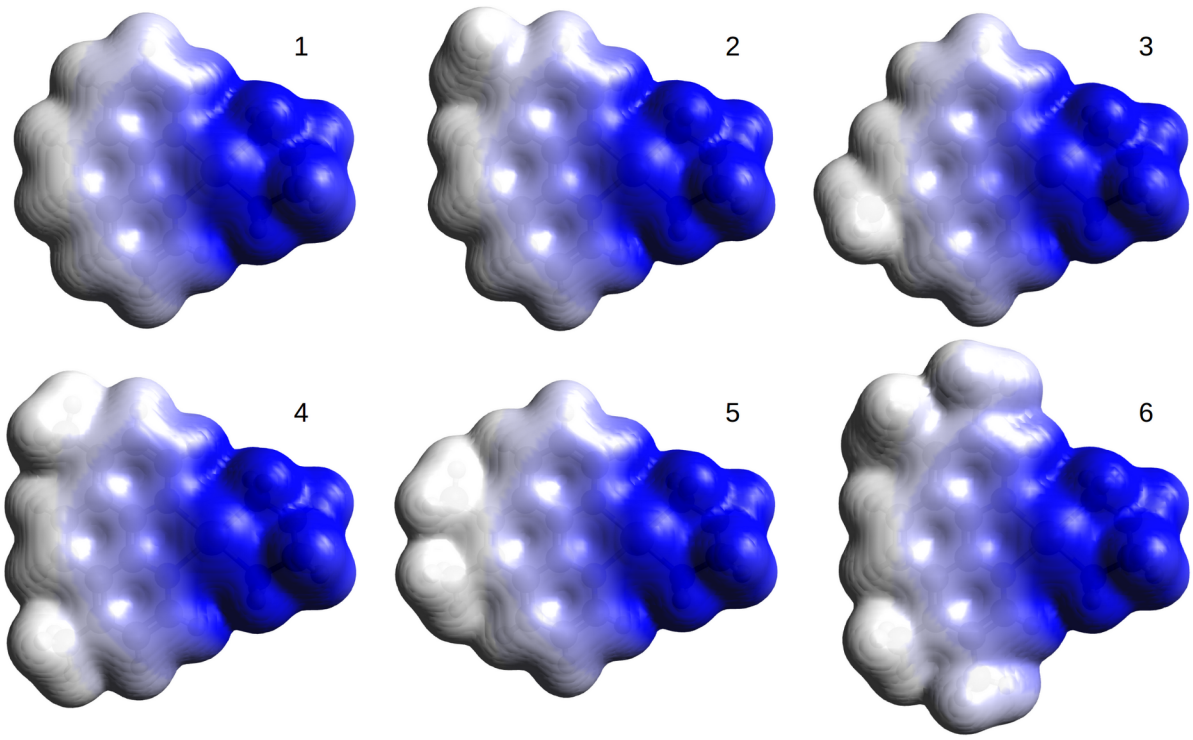


Figure S21: Electrostatic distribution. White = neutral, blue = positive.

Frontier molecular orbitals

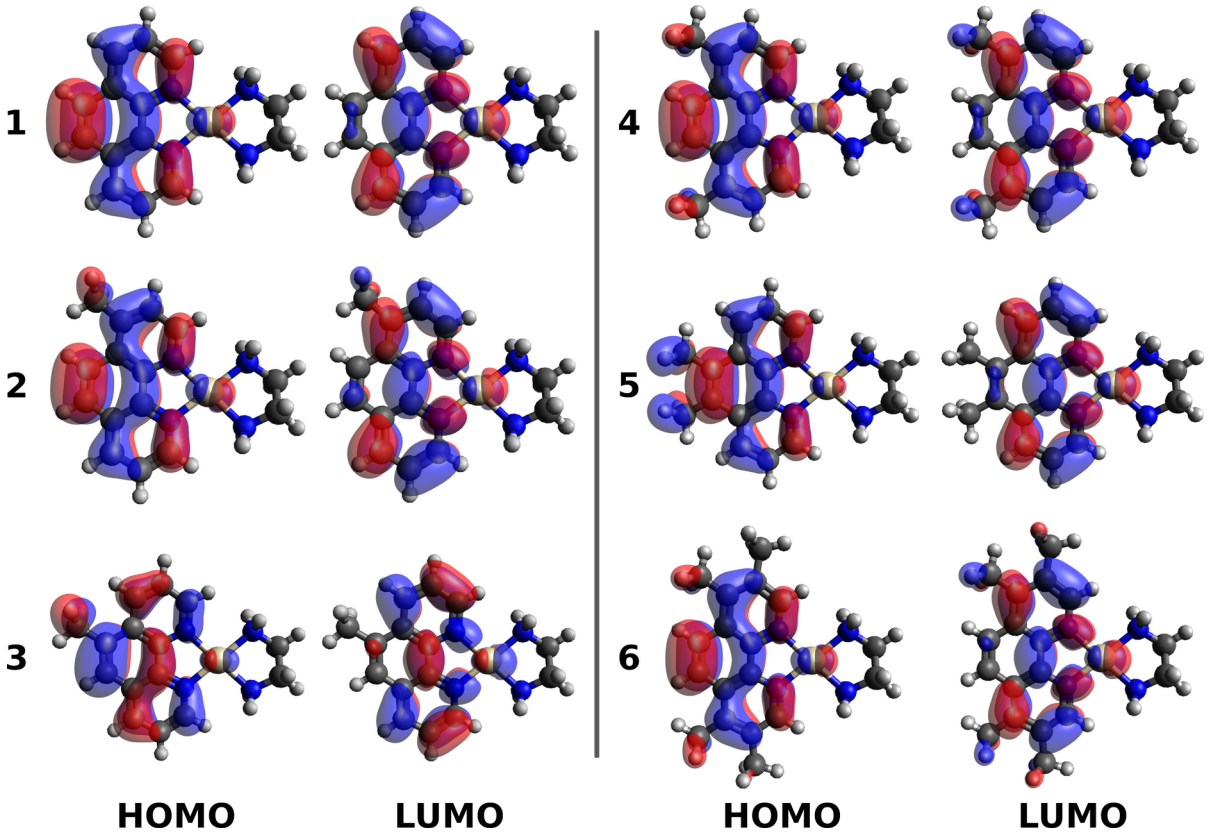


Figure S22: Frontier molecular orbitals of the Pt(II)-phenanthroline complexes studied.

Gaussian input files

```
%Mem=8GB
%NProcShared=4
%Chk=Pt_1Phenen_mixed_OPT.chk
#P B3LYP/Gen Pseudo=Read Opt

 C14H16N4Pt | Equilibrium Geometry | B3LYP/Gen

2 1
Pt     -2.208000   -0.125000    0.029000
N      -4.249000   -0.135000    0.036000
C      -3.056000    2.820000   -0.154000
C      -2.735000    4.185000   -0.220000
C      -1.390000    4.589000   -0.204000
C      -0.376000    3.623000   -0.126000
C      -0.746000    2.274000   -0.062000
N      -2.055000    1.901000   -0.072000
C       0.235000    1.267000    0.012000
```

```
C      1.595000    1.597000    0.028000
C      1.973000    2.953000   -0.034000
C      0.991000    3.962000   -0.112000
C      0.714000   -1.056000    0.154000
C      2.090000   -0.774000    0.173000
C      2.532000    0.557000    0.108000
H     -4.090000    2.514000   -0.173000
H     -3.523000    4.924000   -0.284000
H     -1.145000    5.642000   -0.256000
H      0.379000   -2.079000    0.210000
H      2.807000   -1.582000    0.240000
H      3.594000    0.772000    0.124000
N     -0.176000   -0.030000    0.069000
N     -2.278000   -2.164000    0.079000
C     -4.598000   -1.498000    0.435000
C     -3.661000   -2.483000   -0.272000
H     -4.452000   -1.611000    1.532000
H     -4.589000    0.054000   -0.935000
H     -3.801000   -2.364000   -1.369000
H     -2.071000   -2.483000    1.054000
H     -4.656000    0.542000    0.719000
H     -1.633000   -2.610000   -0.611000
H     -5.663000   -1.710000    0.196000
H     -3.897000   -3.534000    0.003000
H      3.020000    3.229000   -0.023000
H      1.297000    5.000000   -0.160000

C N H 0
6-31G(d,p)
****
Pt 0
LANL2DZ
****

Pt 0
LANL2DZ

%Mem=8GB
%NProcShared=4
%Chk=Pt_4MEEN_mixed_OPT.chk
#P B3LYP/Gen Pseudo=Read Opt

 C15H18N4Pt | Equilibrium Geometry | B3LYP/Gen

2 1
Pt    -2.284000   -0.049000    0.244000
N     -4.325000   -0.003000    0.275000
C     -3.061000    2.883000   -0.190000
```

```
C     -2.711000    4.228000   -0.380000
C     -1.358000    4.598000   -0.414000
C     -0.365000    3.619000   -0.260000
C     -0.760000    2.292000   -0.073000
N     -2.079000    1.953000   -0.036000
C      0.201000    1.269000    0.081000
C      1.579000    1.562000    0.055000
C      1.968000    2.915000   -0.134000
C      1.003000    3.928000   -0.290000
C      0.600000   -1.047000    0.425000
C      1.979000   -0.821000    0.409000
C      2.492000    0.480000    0.223000
H     -4.101000    2.602000   -0.172000
H     -3.482000    4.977000   -0.503000
H     -1.089000    5.637000   -0.562000
H      0.229000   -2.047000    0.576000
H      2.654000   -1.658000    0.543000
N     -0.250000   -0.004000    0.256000
N     -2.408000   -2.074000    0.473000
C     -4.704000   -1.315000    0.797000
C     -3.805000   -2.384000    0.168000
H     -4.545000   -1.336000    1.898000
H     -4.673000    0.110000   -0.704000
H     -3.958000   -2.357000   -0.934000
H     -2.196000   -2.313000    1.469000
H     -4.705000    0.742000    0.902000
H     -1.787000   -2.596000   -0.185000
H     -5.778000   -1.517000    0.591000
H     -4.064000   -3.400000    0.537000
C      3.994000    0.633000    0.219000
H      3.002000    3.212000   -0.165000
H      1.326000    4.952000   -0.434000
H      4.366000    1.659000    0.072000
H      4.422000    0.011000   -0.596000
H      4.400000    0.273000    1.188000

C N H 0
6-31G(d,p)
****
Pt 0
LANL2DZ
****

Pt 0
LANL2DZ

%Mem=8GB
%NProcShared=4
```

```
%Chk=Pt_5MEEN_mixed_OPT.chk
#P B3LYP/Gen Pseudo=Read Opt

 C15H18N4Pt | Equilibrium Geometry | B3LYP/Gen

2 1
Pt    -2.199000   -0.124000    0.029000
N     -4.240000   -0.127000    0.038000
C     -3.039000    2.815000   -0.154000
C     -2.719000    4.179000   -0.221000
C     -1.374000    4.581000   -0.204000
C     -0.360000    3.614000   -0.125000
C     -0.726000    2.269000   -0.061000
N     -2.037000    1.898000   -0.071000
C      0.256000    1.260000    0.013000
C      1.627000    1.585000    0.031000
C      2.005000    2.964000   -0.033000
C      1.000000    3.958000   -0.111000
C      0.706000   -1.071000    0.157000
C      2.079000   -0.811000    0.178000
C      2.544000    0.511000    0.113000
H     -4.072000    2.510000   -0.174000
H     -3.506000    4.919000   -0.286000
H     -1.129000    5.634000   -0.256000
H      0.358000   -2.090000    0.214000
H      2.783000   -1.630000    0.246000
H      3.609000    0.663000    0.133000
N     -0.167000   -0.035000    0.070000
N     -2.278000   -2.164000    0.075000
C     -4.595000   -1.488000    0.436000
C     -3.664000   -2.476000   -0.274000
H     -4.447000   -1.604000    1.533000
H     -4.581000    0.065000   -0.932000
H     -3.805000   -2.355000   -1.371000
H     -2.071000   -2.486000    1.049000
C      3.437000    3.443000   -0.023000
H      4.202000    2.653000    0.037000
H      3.591000    4.114000    0.848000
H      3.635000    4.019000   -0.952000
H     -4.644000    0.551000    0.723000
H     -1.638000   -2.612000   -0.618000
H     -5.661000   -1.695000    0.199000
H     -3.904000   -3.527000    0.000000
H      1.285000    5.003000   -0.161000

C N H 0
6-31G(d,p)
****
```

```
Pt 0
LANL2DZ
****

Pt 0
LANL2DZ

%Mem=8GB
%NProcShared=4
%Chk=Pt_47MEEN_mixed_OPT.chk
#P B3LYP/Gen Pseudo=Read Opt

 C16H20N4Pt | Equilibrium Geometry | B3LYP/Gen

2 1
Pt    -2.251000   -0.147000    0.029000
N     -4.292000   -0.189000    0.045000
C     -3.149000    2.769000   -0.151000
C     -2.867000    4.138000   -0.218000
C     -1.531000    4.588000   -0.205000
C     -0.485000    3.635000   -0.128000
C     -0.824000    2.277000   -0.063000
N     -2.127000    1.878000   -0.070000
C      0.177000    1.281000    0.009000
C      1.543000    1.627000    0.025000
C      1.880000    3.001000   -0.039000
C      0.881000    3.988000   -0.115000
C      0.667000   -1.040000    0.151000
C      2.036000   -0.760000    0.170000
C      2.497000    0.570000    0.105000
H     -4.175000    2.438000   -0.169000
H     -3.686000    4.845000   -0.281000
H      0.336000   -2.064000    0.208000
H      2.742000   -1.579000    0.236000
N     -0.223000   -0.020000    0.066000
N     -2.292000   -2.188000    0.072000
C     -4.620000   -1.558000    0.441000
C     -3.672000   -2.526000   -0.274000
H     -4.467000   -1.673000    1.537000
H     -4.640000   -0.001000   -0.924000
H     -3.819000   -2.405000   -1.370000
H     -2.076000   -2.508000    1.045000
H     -4.707000    0.478000    0.734000
H     -1.646000   -2.623000   -0.624000
H     -5.682000   -1.784000    0.206000
H     -3.891000   -3.582000   -0.001000
C      3.992000    0.780000    0.131000
H      4.326000    1.829000    0.081000
```

```
H       4.450000    0.250000   -0.731000
H       4.405000    0.351000    1.069000
C      -1.250000    6.066000   -0.278000
H      -2.186000    6.661000   -0.334000
H      -0.646000    6.290000   -1.183000
H      -0.693000    6.388000    0.628000
H       2.902000    3.339000   -0.032000
H       1.185000    5.025000   -0.162000

C N H 0
6-31G(d,p)
****
Pt 0
LANL2DZ
****

Pt 0
LANL2DZ

%Mem=8GB
%NProcShared=4
%Chk=Pt_56MEEN_mixed_OPT.chk
#P B3LYP/Gen Pseudo=Read Opt

 C16H20N4Pt | Equilibrium Geometry | B3LYP/Gen

2 1
Pt     -2.195000   -0.112000    0.028000
N      -4.237000   -0.124000    0.045000
C      -3.054000    2.803000   -0.157000
C      -2.761000    4.163000   -0.224000
C      -1.430000    4.593000   -0.209000
C      -0.370000    3.653000   -0.129000
C      -0.726000    2.295000   -0.063000
N      -2.037000    1.908000   -0.072000
C       0.255000    1.285000    0.012000
C       1.625000    1.603000    0.030000
C       2.018000    2.978000   -0.034000
C       1.017000    4.007000   -0.114000
C       0.698000   -1.053000    0.157000
C       2.067000   -0.799000    0.179000
C       2.536000    0.518000    0.114000
H      -4.081000    2.481000   -0.177000
H      -3.564000    4.886000   -0.291000
H      -1.267000    5.653000   -0.264000
H       0.346000   -2.069000    0.216000
H       2.768000   -1.622000    0.248000
H       3.601000    0.650000    0.136000
```

```
N     -0.169000   -0.012000    0.069000
N     -2.267000   -2.153000    0.070000
C     -4.585000   -1.489000    0.439000
C     -3.652000   -2.470000   -0.277000
H     -4.433000   -1.608000    1.535000
H     -4.583000    0.070000   -0.923000
H     -3.797000   -2.345000   -1.373000
H     -2.056000   -2.477000    1.041000
C      1.476000    5.458000   -0.182000
H      0.689000    6.222000   -0.242000
H      2.109000    5.608000   -1.081000
H      2.065000    5.706000    0.725000
C      3.484000    3.394000   -0.019000
H      4.226000    2.587000    0.042000
H      3.678000    4.047000    0.857000
H      3.723000    3.951000   -0.949000
H     -4.643000    0.546000    0.736000
H     -1.629000   -2.597000   -0.628000
H     -5.651000   -1.699000    0.204000
H     -3.887000   -3.522000   -0.006000

C N H 0
6-31G(d,p)
****
Pt 0
LANL2DZ
****

Pt 0
LANL2DZ

%Mem=8GB
%NProcShared=4
%Chk=Pt_3478MEEN_mixed_OPT.chk
#P B3LYP/Gen Pseudo=Read Opt

 C18H24N4Pt | Equilibrium Geometry | B3LYP/Gen

2 1
Pt    -2.247000   -0.158000    0.031000
N     -4.289000   -0.177000    0.047000
C     -3.113000    2.762000   -0.153000
C     -2.832000    4.138000   -0.221000
C     -1.475000    4.576000   -0.203000
C     -0.441000    3.604000   -0.123000
C     -0.794000    2.250000   -0.058000
N     -2.098000    1.865000   -0.068000
C      0.198000    1.241000    0.017000
```

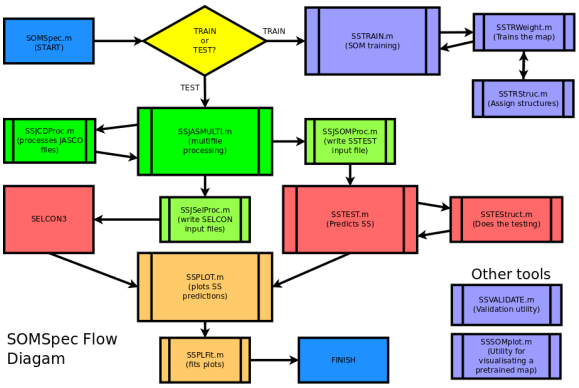
```
C      1.567000    1.568000    0.036000
C      1.917000    2.941000   -0.028000
C      0.930000    3.938000   -0.106000
C      0.654000   -1.087000    0.162000
C      2.035000   -0.839000    0.185000
C      2.515000    0.501000    0.120000
H     -4.136000    2.423000   -0.173000
C     -3.982000    5.109000   -0.311000
C     -1.125000    6.049000   -0.274000
H      0.291000   -2.100000    0.219000
C      2.983000   -2.010000    0.282000
C      4.016000    0.733000    0.147000
N     -0.217000   -0.053000    0.073000
N     -2.312000   -2.199000    0.070000
C     -4.632000   -1.543000    0.440000
C     -3.696000   -2.520000   -0.277000
H     -4.480000   -1.663000    1.536000
H     -4.635000    0.017000   -0.921000
H     -3.842000   -2.395000   -1.373000
H     -2.100000   -2.524000    1.041000
H     -4.697000    0.491000    0.739000
H     -1.673000   -2.640000   -0.629000
H     -5.697000   -1.757000    0.205000
H     -3.927000   -3.574000   -0.007000
H      1.252000    4.969000   -0.153000
H      2.939000    3.274000   -0.018000
H     -4.963000    4.589000   -0.313000
H     -3.909000    5.694000   -1.252000
H     -3.966000    5.796000    0.561000
H     -2.001000    6.721000   -0.336000
H     -0.504000    6.241000   -1.175000
H     -0.555000    6.340000    0.634000
H      2.441000   -2.978000    0.324000
H      3.597000   -1.927000    1.203000
H      3.649000   -2.034000   -0.606000
H      4.345000    1.781000    0.095000
H      4.485000    0.216000   -0.716000
H      4.437000    0.320000    1.087000

C N H O
6-31G(d,p)
****
Pt 0
LANL2DZ
****

Pt 0
LANL2DZ
```

SOMSpec

SOMSpec is a utility for predicting protein secondary structure from electronic spectra (such as circular dichroism (CD), infrared (IR/RAMAN). It does this through self organising maps – a form of machine learning.

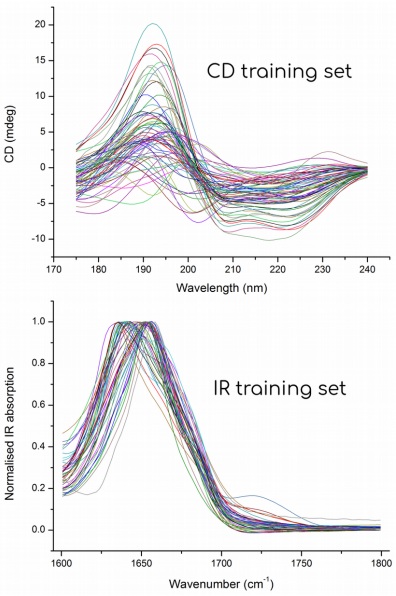
Features

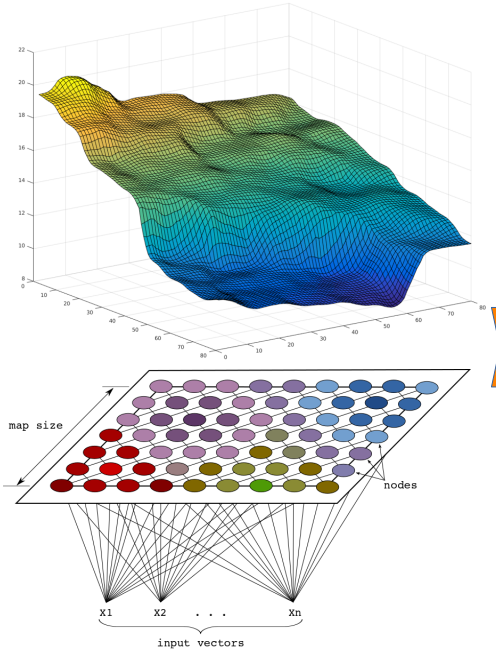
- Fast and flexible learning. A typical training run takes a few minutes on a recent desktop computer.
- Native support for Linux and Windows.
- Automated validation of training sets.
- Visualisation of the learning process.

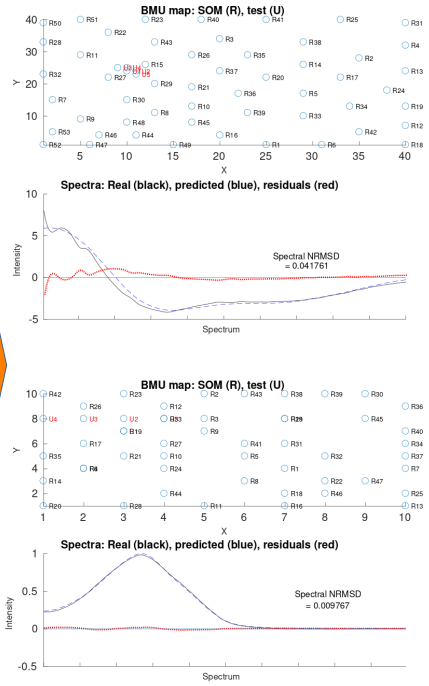


SOMSpec Flow Diagam

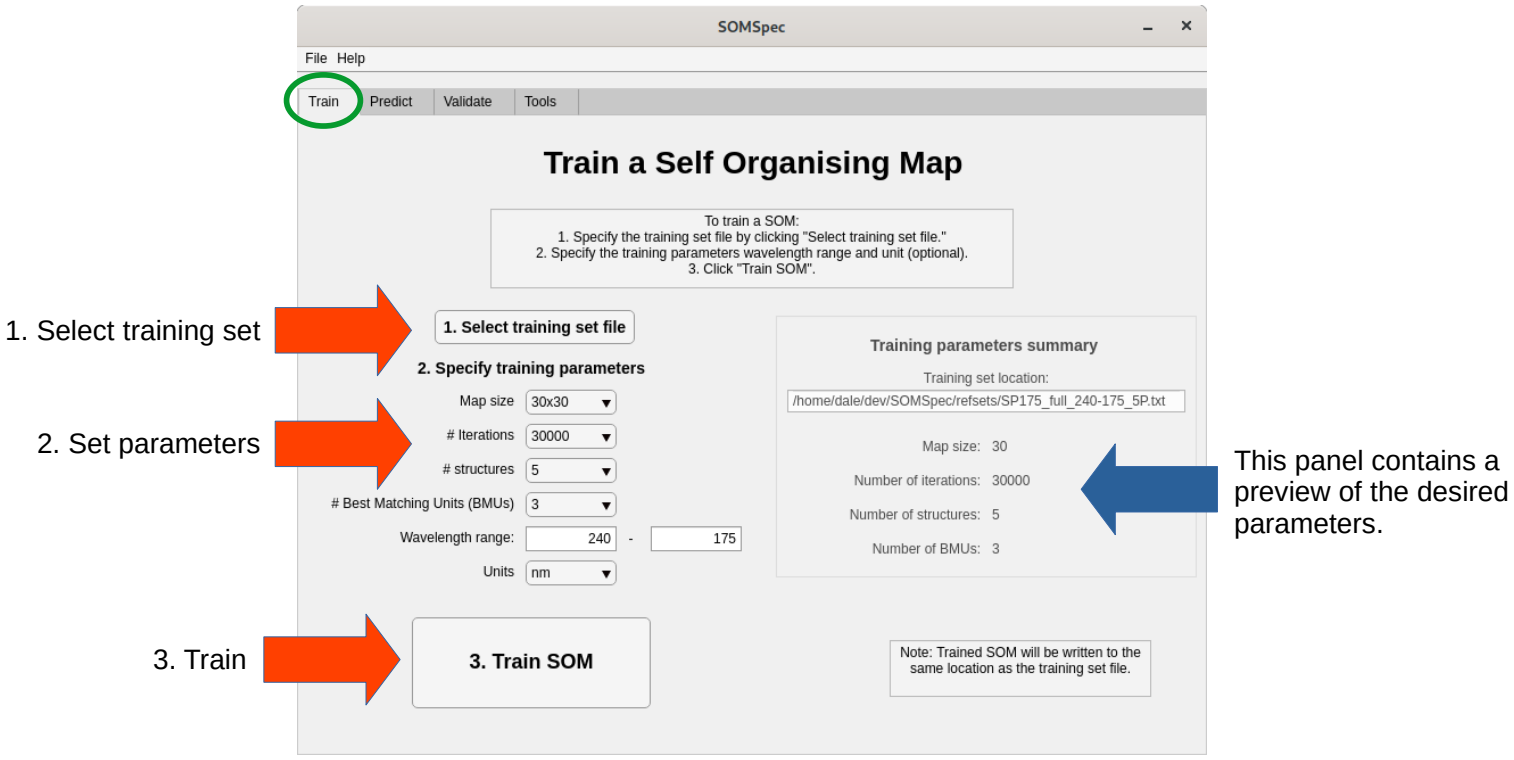
SOMSpec

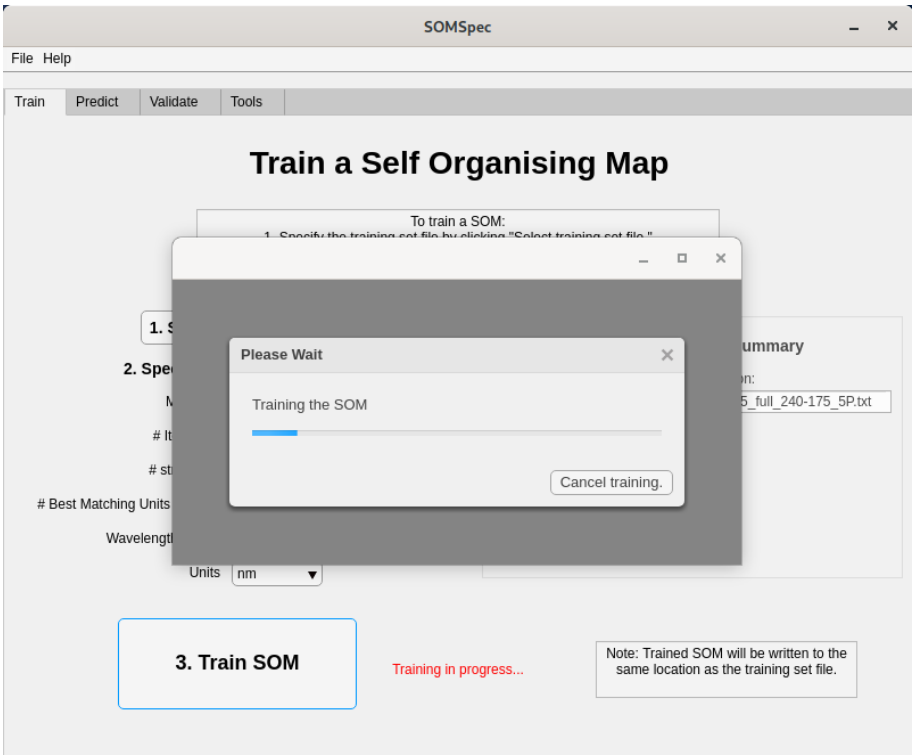






From this… to… this.

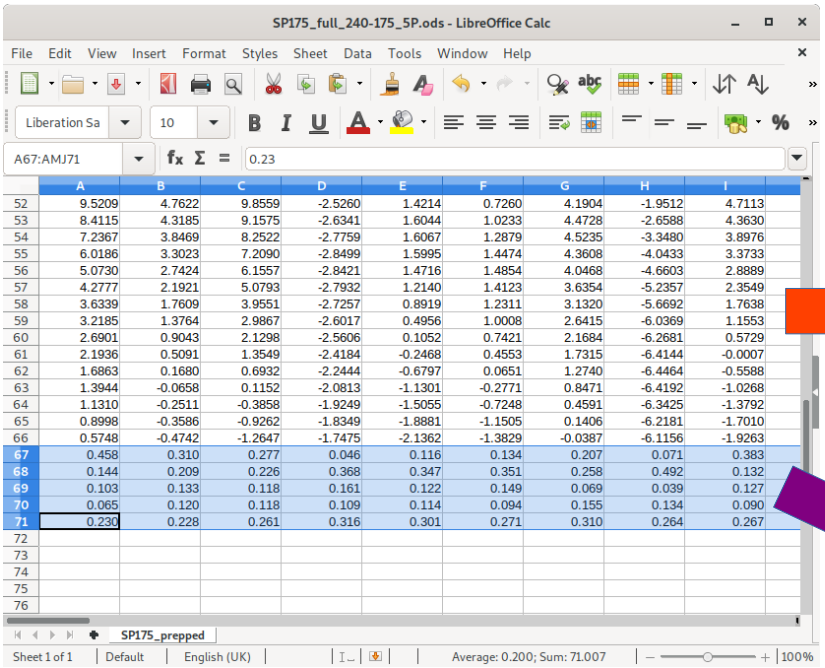

SOMSpec
File Help
Train
Predict
Validate
Tools
Train a Self Organising Map
To train a SOM:
1. Specify the training set file by clicking "Select training set file."
2. Specify the training parameters wavelength range and unit (optional).
3. Click "Train SOM".
1. Select training set
1. Select training set file
2. Specify training parameters
2. Set parameters
Map size 30x30
Iterations 30000
structures 5
Best Matching Units (BMUs) 3
Wavelength range: 240 - 175
Units nm
Training parameters summary
Training set location:
/home/dale/dev/SOMSpec/refsets/SP175_full_240-175_5P.txt
Map size: 30
Number of iterations: 30000
Number of structures: 5
Number of BMUs: 3
This panel contains a preview of the desired parameters.
3. Train
3. Train SOM
Note: Trained SOM will be written to the same location as the training set file.



- The trained map is output to the same location as the training file.
- For a training set of ~50 spectra recommended settings are anywhere from 30 to 60 map size and 50,000 iterations.
- The training step should take less that 5 minutes.

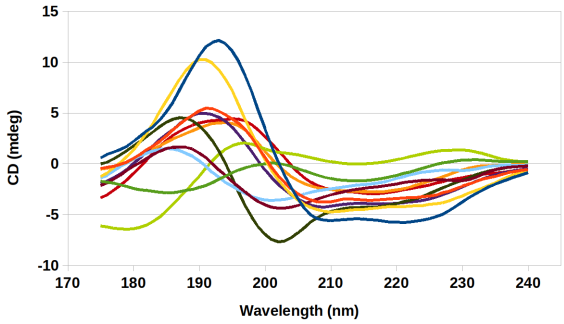
Training set data format

(the reference dataset used for training the SOM)



	A	B	C	D	E	F	G	H	I
52	9.5209	4.7622	9.8559	-2.5260	1.4214	0.7260	4.1904	-1.9512	4.7113
53	8.4115	4.3185	9.1575	-2.6341	1.6044	1.0233	4.4728	-2.6588	4.3630
54	7.2367	3.8469	8.2522	-2.7759	1.6067	1.2879	4.5235	-3.3480	3.8976
55	6.0186	3.3023	7.2090	-2.8499	1.5995	1.4474	4.3608	-4.0433	3.3733
56	5.0730	2.7424	6.1557	-2.8421	1.4716	1.4854	4.0468	-4.6603	2.8889
57	4.2777	2.1921	5.0793	-2.7932	1.2140	1.4123	3.6354	-5.2357	2.3549
58	3.6339	1.7609	3.9551	-2.7257	0.8919	1.2311	3.1320	-5.6692	1.7638
59	3.2185	1.3764	2.9867	-2.6017	0.4956	1.0008	2.6415	-6.0369	1.1553
60	2.6901	0.9043	2.1298	-2.5606	0.1052	0.7421	2.1684	-6.2681	0.5729
61	2.1936	0.5091	1.3549	-2.4184	-0.2468	0.4553	1.7315	-6.4144	-0.0007
62	1.6863	0.1680	0.6932	-2.2444	-0.6797	0.0651	1.2740	-6.4464	-0.5588
63	1.3944	-0.0658	0.1152	-2.0813	-1.1301	-0.2771	0.8471	-6.4192	-1.0268
64	1.1310	-0.2511	-0.3858	-1.9249	-1.5055	-0.7248	0.4591	-6.3425	-1.3792
65	0.8998	-0.3586	-0.9262	-1.8349	-1.8881	-1.1505	0.1406	-6.2181	-1.7010
66	0.5748	-0.4742	-1.2647	-1.7475	-2.1362	-1.3829	-0.0387	-6.1156	-1.9263
67	0.458	0.310	0.277	0.046	0.116	0.134	0.207	0.071	0.383
68	0.144	0.209	0.226	0.368	0.347	0.351	0.258	0.492	0.132
69	0.103	0.133	0.118	0.161	0.122	0.149	0.069	0.039	0.127
70	0.065	0.120	0.118	0.109	0.114	0.094	0.155	0.134	0.090
71	0.230	0.228	0.261	0.316	0.301	0.271	0.310	0.264	0.267
72									
73									
74									
75									
76									

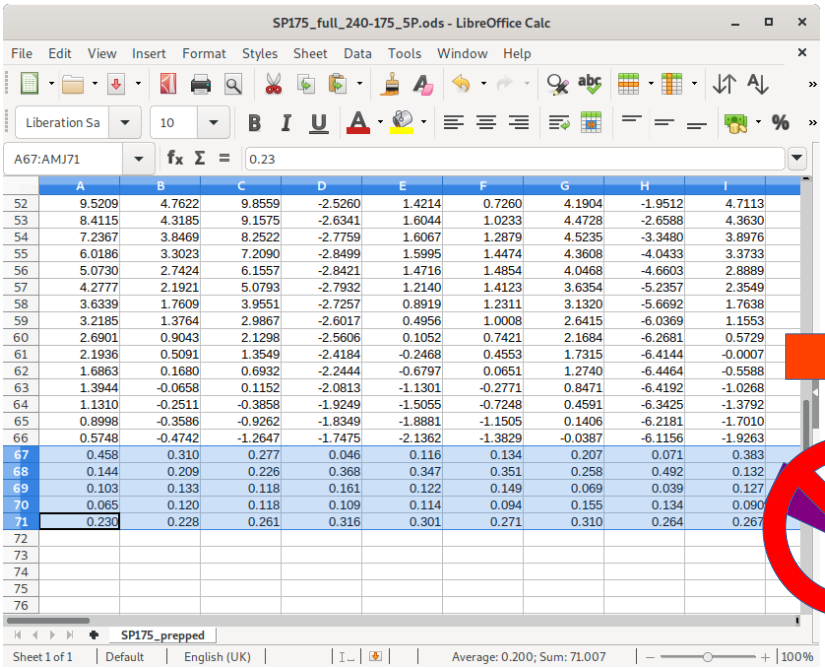
- Training set is a plain text .csv
- One column per spectrum – **wavelength and step size arbitrary**.
- Last cells of each column contain structure information.



In this example (an excerpt of the SP175 dataset) cells 67 through 71 contain the fractions of: α-helix, β-sheet, turns, bends, and other – a total of 5 structures. These values must sum to 1.

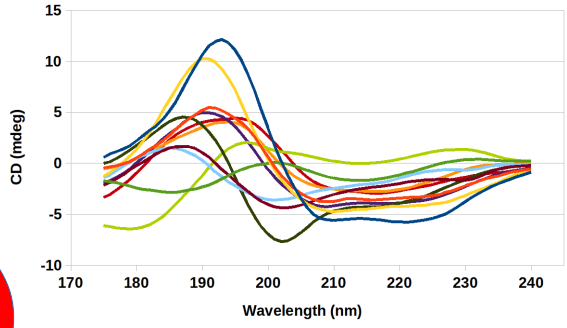
Test set data format

(the spectrum you want a prediction of)

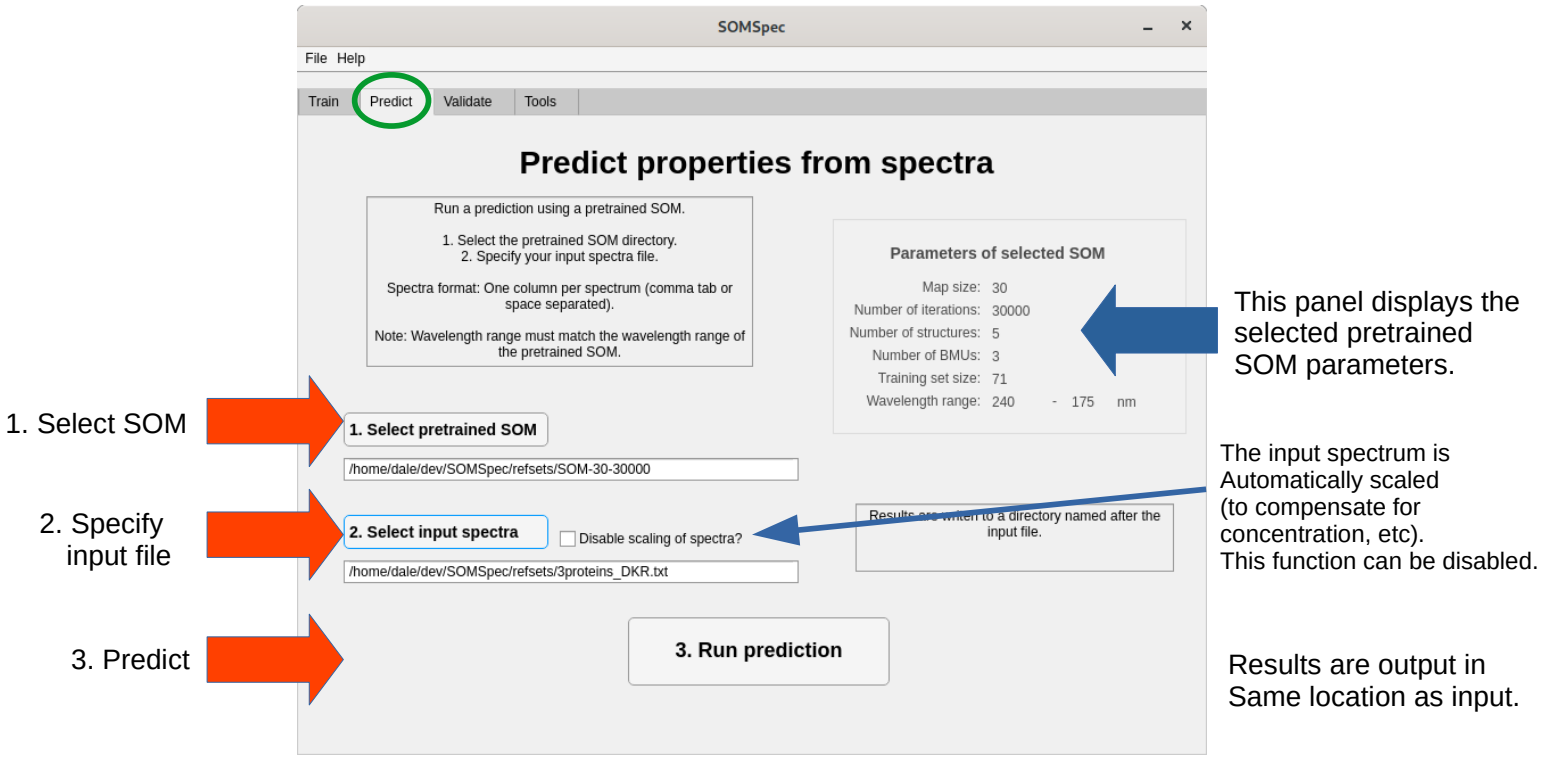


	A	B	C	D	E	F	G	H	I
52	9.5209	4.7622	9.8559	-2.5260	1.4214	0.7260	4.1904	-1.9512	4.7113
53	8.4115	4.3185	9.1575	-2.6341	1.6044	1.0233	4.4728	-2.6588	4.3630
54	7.2367	3.8469	8.2522	-2.7759	1.6067	1.2879	4.5235	-3.3480	3.8976
55	6.0186	3.3023	7.2090	-2.8499	1.5995	1.4474	4.3608	-4.0433	3.3733
56	5.0730	2.7424	6.1557	-2.8421	1.4716	1.4854	4.0468	-4.6603	2.8889
57	4.2777	2.1921	5.0793	-2.7932	1.2140	1.4123	3.6354	-5.2357	2.3549
58	3.6339	1.7609	3.9551	-2.7257	0.8919	1.2311	3.1320	-5.6692	1.7638
59	3.2185	1.3764	2.9867	-2.6017	0.4956	1.0008	2.6415	-6.0369	1.1553
60	2.6901	0.9043	2.1298	-2.5606	0.1052	0.7421	2.1684	-6.2681	0.5729
61	2.1936	0.5091	1.3549	-2.4184	-0.2468	0.4553	1.7315	-6.4144	-0.0007
62	1.6863	0.1680	0.6932	-2.2444	-0.6797	0.0651	1.2740	-6.4464	-0.5588
63	1.3944	-0.0658	0.1152	-2.0813	-1.1301	-0.2771	0.8471	-6.4192	-1.0268
64	1.1310	-0.2511	-0.3858	-1.9249	-1.5055	-0.7248	0.4591	-6.3425	-1.3792
65	0.8998	-0.3586	-0.9262	-1.8349	-1.8881	-1.1505	0.1406	-6.2181	-1.7010
66	0.5748	-0.4742	-1.2647	-1.7475	-2.1362	-1.3829	-0.0387	-6.1156	-1.9263
67	0.458	0.310	0.277	0.046	0.116	0.134	0.207	0.071	0.383
68	0.144	0.209	0.226	0.368	0.347	0.351	0.258	0.492	0.132
69	0.103	0.133	0.118	0.161	0.122	0.149	0.069	0.039	0.127
70	0.065	0.120	0.118	0.109	0.114	0.094	0.155	0.134	0.090
71	0.230	0.228	0.261	0.316	0.301	0.271	0.310	0.264	0.267
72									
73									
74									
75									
76									

- Test set is a plain text .csv
- Multiple spectra can be input and predicted at the same time. Again, one column per spectrum – wavelength and step size arbitrary but **must match training set wavelength range**.



Columns ONLY contain spectrum data, no structure values – these will be predicted.

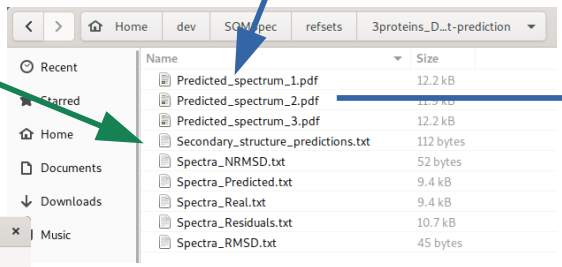

SOMSpec
File Help
Train
Predict
Validate
Tools
Predict properties from spectra
Run a prediction using a pretrained SOM.
1. Select the pretrained SOM directory.
2. Specify your input spectra file.
Spectra format: One column per spectrum (comma tab or space separated).
Note: Wavelength range must match the wavelength range of the pretrained SOM.
Parameters of selected SOM
Map size: 30
Number of iterations: 30000
Number of structures: 5
Number of BMUs: 3
Training set size: 71
Wavelength range: 240 - 175 nm
1. Select pretrained SOM
/home/dale/dev/SOMSpec/refsets/SOM-30-30000
2. Select input spectra
Disable scaling of spectra?
/home/dale/dev/SOMSpec/refsets/3proteins_DKR.txt
Results are written to a directory named after the input file.
3. Run prediction
1. Select SOM
2. Specify input file
3. Predict
This panel displays the selected pretrained SOM parameters.
The input spectrum is Automatically scaled (to compensate for concentration, etc). This function can be disabled.
Results are output in Same location as input.

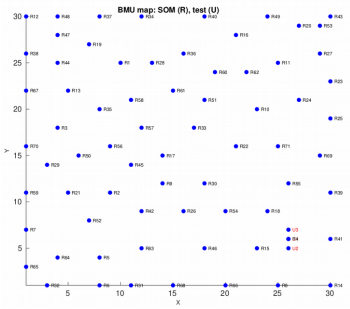
Results

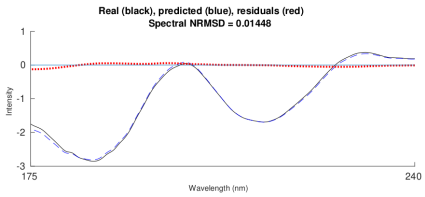
(interpreting the output of a prediction)

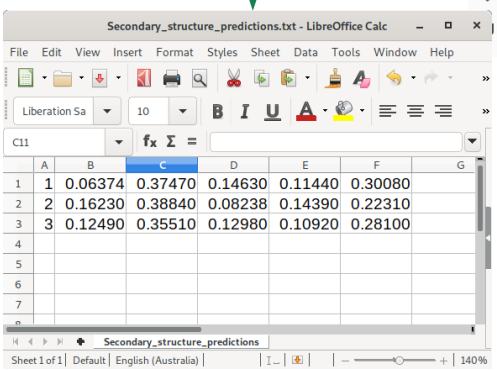
- In this example 3 protein CD spectra were input (input file contained 3 columns.) The predicted spectrum of each is output as a separate PDF.

- Secondary structure predictions and other information is contained in the other files; 1 line per spectrum.



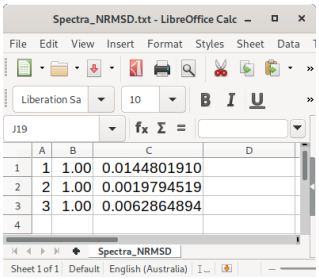






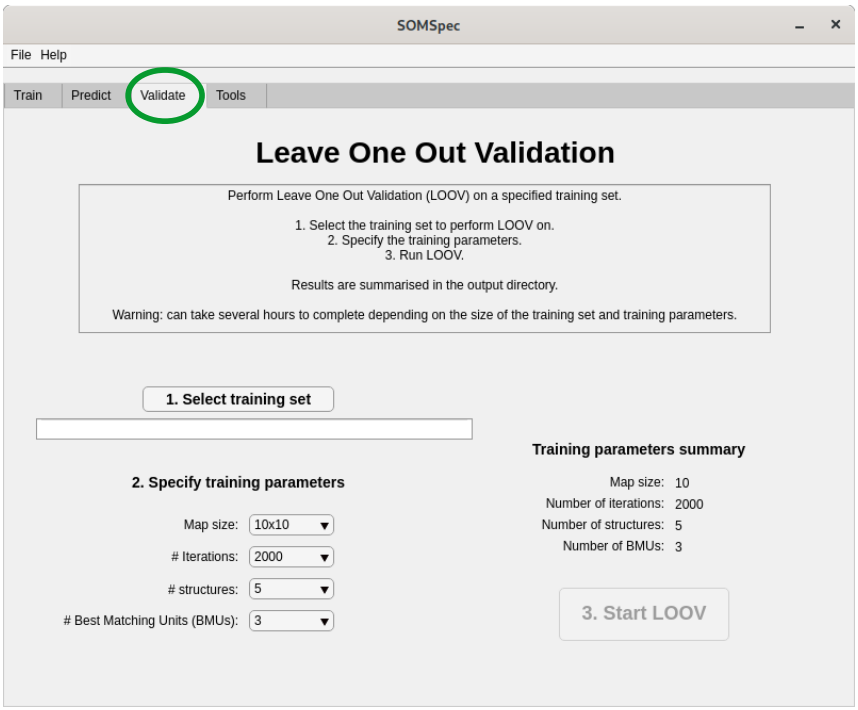
1	1	0.06374	0.37470	0.14630	0.11440	0.30080
2	2	0.16230	0.38840	0.08238	0.14390	0.22310
3	3	0.12490	0.35510	0.12980	0.10920	0.28100

- Goodness of fit (NRMSD) and scaling factors are output in Spectra_NRMSD.txt

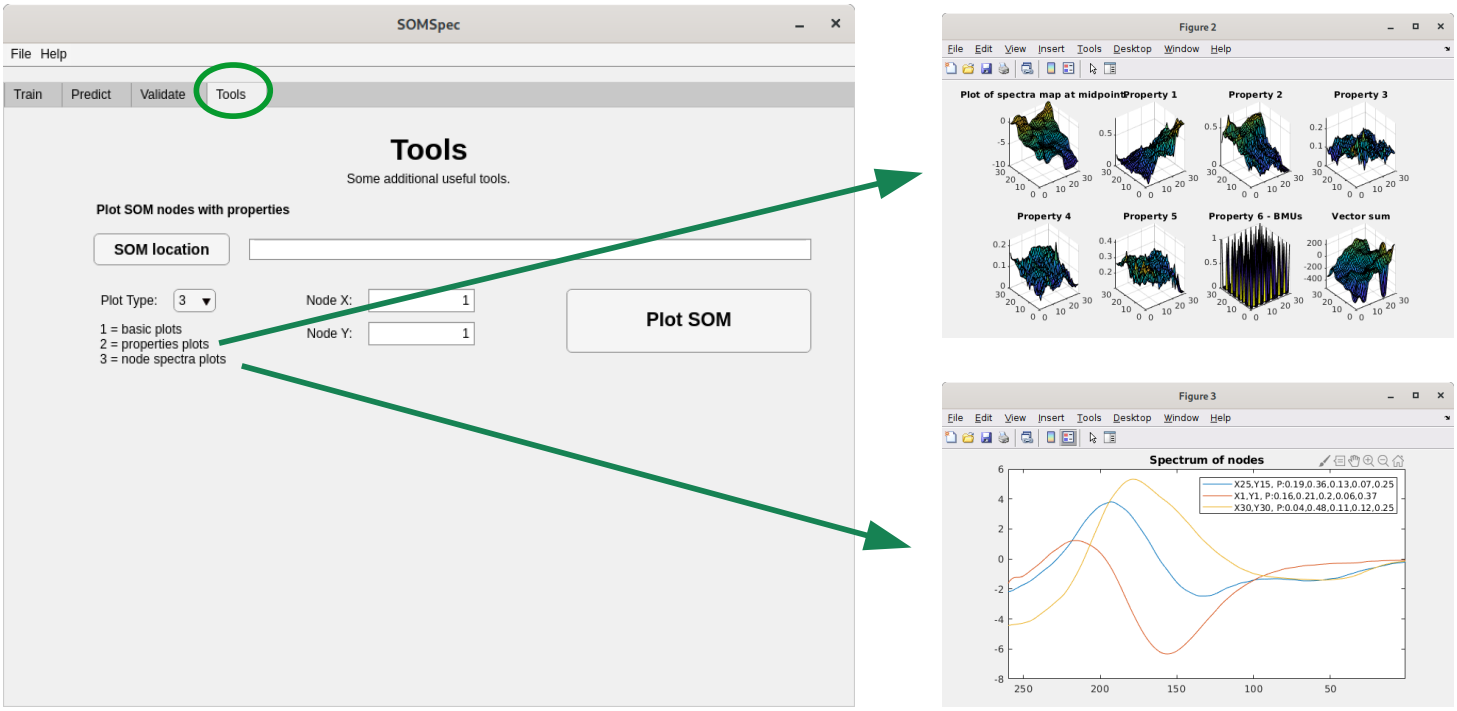


1	1	1.00	0.0144801910
2	2	1.00	0.0019794519
3	3	1.00	0.0062864894

Other features - Validation of a training set



Other features – Visualising a SOM



Availability

SOMSpec is distributed licensed under the GPL3 license.

Download the installer (Linux & Windows) from `somspec.org`

Source code hosted at `gitlab.com/daleang/somspec`

Collaborators welcome.