5-17-2014

The Renaissance of Isothermal Titration Calorimetry

Vu Hoang Le

The renaissance of isothermal titration calorimetry

By

Vu Hoang Le

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Chemistry
in the Department of Chemistry

Mississippi State, Mississippi

May 2014

The renaissance of isothermal titration calorimetry

By

Vu Hoang Le

Approved:

_____

Edwin A. Lewis
(Major Professor)

_____

Joseph P. Emerson
(Committee Member)

_____

Nicholas C. Fitzkee
(Committee Member)

_____

Steven R. Gwaltney
(Committee Member)

_____

Gerald Rowland
(Committee Member)

_____

Stephen C. Foster
(Graduate Coordinator)

_____

R. Gregory Dunaway
Professor and Dean
College of Arts & Sciences

Name: Vu Hoang Le

Date of Degree: May 17, 2014

Institution: Mississippi State University

Major Field: Chemistry

Major Professor: Edwin A. Lewis

Title of Study: The renaissance of isothermal titration calorimetry

Pages in Study: 201

Candidate for Degree of Doctor of Philosophy

This dissertation is a composite of some of the research that I have conducted during the course of my PhD study. The larger goal of this dissertation is to renew the interests among the scientific community for an otherwise under-appreciated technique called Isothermal Titration Calorimetry. The resurgence of calorimetry in the biophysical community and the shift to investigations of more complex biological systems signal a real need for more sophisticated analysis techniques. This dissertation expounds on new ITC analysis methods that we have developed as well as results from the study of thermodynamic properties of higher order DNA structures.

In 1978, Peter Privalov described the first use of microcalorimetry to obtain the thermodynamic properties for removing calcium from parvalbumin III protein. Fast forward 36 years: modern day electronics, highly efficient thermally conductive and chemically inert materials, in conjunction with sensitive thermal detectors, has transformed the original calorimeter into a device capable of measuring heat changes as small as 0.05 nanowatts, which is equivalent to capturing heat from an incandescent light bulb a kilometer away. However, analytical methods have not kept pace with this

technology. Commercial ITC instruments are typically supplied with software that only includes a number of simple interaction models. As a result, the lack of analysis tools for more complex models has become a limiting factor for many researchers.

We have recently developed new ITC fitting algorithms that we have incorporated into a user-friendly program (CHASM©) for the analysis of complex ITC equilibria. In a little over a year, CHASM© has been downloaded by over 370 unique users. Several chapters in this dissertation demonstrate this software's power and versatility in the thermodynamic investigations of two model systems in both aqueous and non-aqueous media. In chapter VI, we assembled a model NHE-III_1 : a novel structure of G-quadruplex in a double stranded form and studied its structural complexity and binding interactions with a classical G-quaduplex interactive ligand known as TMPyP4. In chapter VII, we reported the thermodynamic properties of a novel PAH system in which weak dispersion forces are solely responsible for formation of the supramolecular complexes.

Keywords: CHASM©, G-quadruplex, DNA, c-MYC, Bcl-2, NHE-III_1, TMPyP4, TMPyP3, TMPyP2, P(5), P(5,10), P(5,15), P(5,10,15), ITC, CD, DSC, Fluorescence, ESI-MS, APPI-MS, computational modeling, C_{60}, C_{70}, buckycatcher.

DEDICATION

This work is dedicated to both my birth parents, *Lê Hồng Phương &* Hoàng Thị Ánh Nguyệt, and my adoptive parents, *Cac Dao & Phuong Dao*. My achievements are without doubt solely the result of their encouragements and unfaltering support.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisor, Professor Edwin Arthur Lewis, who has been very instrumental in the successful completion of my doctoral study. I thank him for not only for sharing his knowledge but also for sharing his methodical approaches to solve problems, no matter how complex the problem is. I thank him for giving support and granting me the liberty of attending professional trainings and conferences. I thank him for invoking in me the capabilities that I did not realize that I had. Thank you for believing in me. We had a lot of fun walking down this intellectual journey together. I also thank him for all the recommendations that he sent out during my postdoctoral search. Being able to join the Lippard lab at MIT is my ultimate gift to him.

I am grateful to all of my committee members Dr. Gwaltney, Dr. Emerson, Dr. Fitzkee and Dr. Rowland. I have enjoyed learning from every one of you either from taking your classes or from our collaborations. I take to my heart all of your professional advices either regarding research or about life as a faculty member and believing that those advices will be most valuable when I have my own research group.

I want to thank all our collaborators: Dr. Dongmao Zhang (MSU), Dr. Andrzej Sygula (MSU), Dr. David Graves at the University of Alabama Birmingham, Dr. Herman Sintim at the University of Maryland, Dr. Fred MacDonnel at the University of North Texas, Dr. Jack Correia at the University of Mississippi Medical Center, and Dr. Stephen Pruett at the MSU College of Veterinary Medicine. I thank them all for allowing me the

opportunity to be involved in many very interesting projects that helped expand my understanding.

I want to thank Ms. Laura Lewis for all her kind heart and loving spirit. I would have not considered moving to Mississippi had it not been for her. She took us in and treated us like family and made us feel good and proud of our works; and that is very important !

I came to know Dr. Stephen Cottrell and his family under strange circumstances. Despite our age differences, Dr. Cottrell and I have had a lot of spirited conversation about Vietnam, about Mississippi State University, about the South, about the US, about the world and pretty much about everything philosophically. I thank him for keeping his door always open to me and his fridge always full of cold beverages.

I value both the friendship and the fellowship provided to me by the current and past members of the Lewis Lab including Mr. Matt Freyer, Dr. Joe Ramos, Dr. Jamie Dettler, Dr. Jingjing Cui, Dr. Venkata Machha, soon-to-be Dr. Clint Milek, and the two undergrad students Matthew McGuire and Kelsey Layman. Special thanks to Kate Henderson, Dr. Michael Yanney, Jesse Dupont, Kerry Barnett, and Gopalakrishna Akurathi for their help in some of the works presented in this dissertation.

Lastly and undeniably the most important person I must give thanks to is my beloved wife, Van Vo for all of her support and understanding. Because of her hours of work in loving our child Samuel Le, she has enabled the hours of research, contemplation, and writing necessary to complete my doctoral journey. She is my "*excellent wife, who is far more precious than jewels*."

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

CHAPTER I

CALORIMETRY

As previously published in eLS. John Wiley & Sons, Ltd: Chichester
DOI: 10.1002/9780470015902.a0003010.pub3

1.1 Introduction to Isothermal Titration Calorimetry

In biology, particularly in studies relating the structure of biopolymers to their functions, two of the most important questions are: 1./ how tightly does a small molecule bind to a specific interaction site and 2./ if the small molecule is a substrate and is converted to a product, how fast does the reaction take place? Since almost any chemical reaction or physical change is accompanied by a change in heat or enthalpy, an isothermal titration calorimeter (ITC) is an ideal instrument to measure either how much of a reaction has taken place or the rate at which a reaction is occurring. In contrast to optical methods, calorimetric measurements can be done with reactants that are spectroscopically silent (a chromophore or fluorophore tag is not required), can be done on opaque, turbid or heterogeneous solutions (e.g. cell suspensions), and can be done over a range of biologically relevant conditions (temperature, salt, pH, etc.)

Key Concepts:

- Almost all chemical reactions are accompanied by a change in heat or enthalpy.
- Calorimetry can determine the complete set of thermodynamic parameters that characterize binding reactions, e.g. K_a, $\Delta G°$, $\Delta H°$, $\Delta S°$, and ΔC_p.

- Binding reactions are typically pH, salt and temperature dependent.

The binding of a small molecule to a biological macromolecule (e.g. a protein or nucleic acid) is perhaps the most fundamental reaction in biology. Such diverse processes as enzyme catalysis, gene regulation and signal transduction all rely on the binding of one molecule, the ligand (L), to another, the macromolecule (M). In biological terms, the ligand could be a substrate, inhibitor, drug, cofactor, coenzyme, prosthetic group, metal ion, polypeptide, protein, oligonucleotide, nucleic acid or any one of a number of molecules thought (or known) to non-covalently interact with a specific site on a second molecule (typically a protein or nucleic acid). A fundamental understanding of a binding interaction requires at a minimum, knowledge of the equilibrium constant, K, and the binding stoichiometry, n (how many ligands are bound to the macromolecule at saturation). A richer understanding of the ligand macromolecule interaction is established if the heat of the reaction, or enthalpy change ($\Delta H°$) is also known.

The importance of having the complete thermodynamic data set (K (or $\Delta G°$), $\Delta H°$, $-T\Delta S°$, ΔC_p) versus having only a value for K (or $\Delta G°$), and the ability to parse the free energy change into its enthalpy and entropy change components, is illustrated for the binding of TMPyP4 to a G-quadruplex DNA construct (1,2). Without knowing the different $\Delta G°$ and $\Delta H°$ values observed for binding the four moles of TMPyP4 to the DNA, it would be impossible to know the number of different complex species and complex structures formed in these systems. In fact, without the ITC data, the complexity of this system would be seriously underestimated. Figure 1.1 presents two different complex structures that are proposed to form from the addition of the ligand TMPyP4 to a c-MYC promoter sequence G-quadruplex (3).

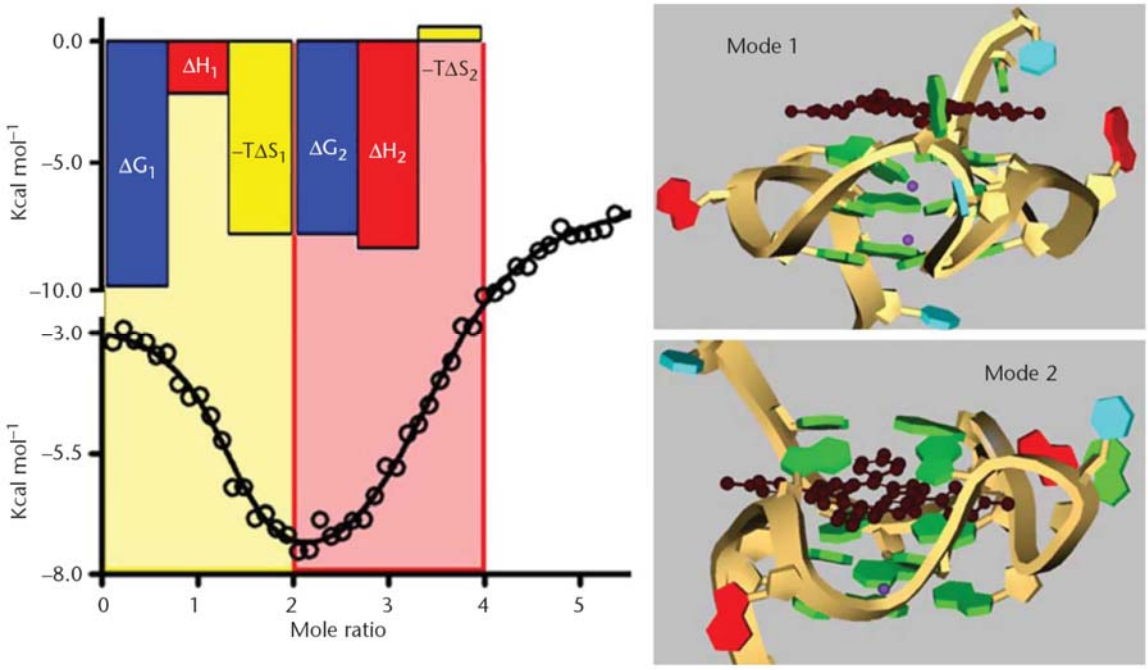


Figure 1.1 Proposed structures are shown for the two 1:1 TMPyP4/c-MYC promoter sequence G-quadruplex complexes.

The ITC binding isotherm indicates that there are at least two binding processes or modes, that the saturation stoichiometry is 4:1, and that the heat change data are well fit with a binding model that includes two overlapping equilibria, the first being entropy driven and the second being enthalpy driven. The higher affinity TMPyP4 binding site (Mode 1) is attributed to end-binding whereas the lower affinity-binding site (Mode 2) has been attributed to intercalation. Two moles of TMPyP4 are bound via each mode.

Although both complexes have favorable free energies of formation ($\Delta G° \approx -10$ and -8 kcal/mol for the end-stacked and intercalated complexes, respectively), the end stacked complex exhibits slightly higher affinity with a *K* that is 30 times larger than the *K* for the intercalation. The favorable $\Delta G°$ for the end-stacking reaction is largely the result of a very favorable entropy change for formation ($-T\Delta S° \approx -8$ kcal/mol) while the favorable $\Delta G°$ for formation of the intercalated complex is the result of a very favorable enthalpy change ($\Delta H° \approx -8.2$ kcal/mol). Parsing the two free energy changes into the enthalpy and entropy change components leads to the model presented in Figure 1.1. It

is very clear that two different binding modes with different thermodynamic signatures are in play for this ligand DNA interaction.

1.2 Mathematical and chemical description

Isothermal titration calorimetry (ITC) is a technique used to measure the heat associated with ligand binding. 'Isothermal' indicates that the experiments are performed at constant temperature, while 'titration' indicates that the ligand is added incrementally to the macromolecule. To understand how calorimetry can be used to study a binding reaction, it is necessary to review the basic thermodynamics of ligand binding (a detailed treatment of this topic is found in Wyman and Gill (4).

The simplest case of ligand binding occurs when the macromolecule can only bind to a single ligand. The binding constant, K, is defined as the ratio of the concentration of the bound complex, [ML], to the product of the concentrations of the unbound ligand, [L] and unbound macromolecule, [M] (eqn (1.1)).

$$K = \frac{[ML]}{[M][L]} \tag{1.1}$$

The value of the binding constant defines the difference in the standard state Gibbs energy, $\Delta G°$ given by eqn (1.2), where R is the gas constant and T the absolute temperature in Kelvin.

$$\Delta G^{\circ} = -RTln(K) \tag{1.2}$$

The Gibbs energy in turn is expressed in terms of the enthalpy, $\Delta H°$, and entropy, $\Delta S°$ changes (eqn (1.3)), so that $\Delta G°$ depends on temperature through $\Delta S°$.

$$\Delta G^{\circ} = \Delta H^{\circ} - T\Delta S \tag{1.3}$$

One of the driving forces for ligand binding is the hydrophobic effect. A hallmark of the hydrophobic effect is a strong temperature dependence of both $\Delta H°$ and $\Delta S°$. Consequently, a more complete expression of the Gibbs energy is required that reflects this (eqn (1.4)).

$$\Delta G = \left[\Delta H_R^\circ + \Delta C_p(T - T_R)\right] - T \times (\Delta S_R^\circ + \Delta C_p ln\frac{T}{T_R}) \quad (1.4)$$

ΔC_P is the change in heat capacity, which defines the temperature dependences of $\Delta H°$ and $\Delta S°$ and the subscript R indicates the value of $\Delta H°$ and $\Delta S°$ at the reference temperature, T_R. The dependence of K on temperature is given from eqns (1.2) and (1.4) as eqn (1.5).

$$K = \exp\left(\frac{-\Delta G^\circ}{RT}\right) \quad (1.5)$$

Finally, the fraction θ of binding sites that have a ligand bound is given by eqn (1.6).

$$\theta = \frac{K[L]}{1+K[L]} \quad (1.6)$$

The heat of reaction when a ligand binds to a macromolecule is the product of the change in θ, the total number of moles of macromolecule and $\Delta H°$. The heat measured in the ITC includes the heat of reaction, heat arising from any linked equilibria, and heats of dilution for both the titrant (dilute ligand solution) and the macromolecule. Using ITC, we can determine K and $\Delta H°$, and knowing these, can calculate $\Delta G°$ and $\Delta S°$ for complex formation. By performing experiments at different temperatures, we can also determine ΔC_P.

1.3 Detection of heat

Commercially available ITC instruments are available from GE Healthcare (Northampton, MA) and TA Instruments (New Castle, DE). Instruments are available with cell volumes ranging from 0.25 to 1.5 ml, and can be equipped with auto samplers. All of the ITC instruments have the same inherent sensitivity in terms of heat measurement, so the advantage of the smaller cell volumes is lost since the samples will need to be correspondingly more concentrated. To simplify the following discussion, we have used volumes and concentrations that are typically used in instruments having a nominal cell volume of 1.5 ml and using a 250 µl buret.

An isothermal titration calorimeter typically contains two cells: a sample (S) cell and a reference (R) cell. A dilute solution of the macromolecule is loaded into S and an equal volume of buffer is loaded into R. The dilute ligand solution is loaded into the syringe and injected into S by a stepping motor. A typical ITC sample cell holds approximately 1.5 mL of solution. In the typical ITC experiment from 10 to 50 injections of 1–10 µL of ligand solution are made at intervals of from 3 to 5 min. As the ligand binds, heat is either generated or absorbed, thus creating a transient temperature difference between the sample and reference cells. The difference in temperature results in a change in the heating power being supplied to the sample cell (an increase in heating power for an endothermic reaction and a decrease in heating power for an exothermic reaction) bringing the two cells back to the same temperature.

After the power signal returns to baseline (the reference power), another injection of ligand solution is made. Injections are continued until subsequent injections result in a small constant heat effect, indicating that all of the binding sites have ligand bound to

them and that only the heat of dilution of the titrant is being detected. In some instances the heat of dilution may be large making accurate heat of dilution corrections even more important. In these cases blank titrations are typically required to determine accurate values for the heat of dilution and the dependence of the heat of dilution on the ligand concentration in the calorimeter cell. The thermogram generated in an ITC experiment is a simple summation of all of the heat-producing reactions that occur as an aliquot of titrant is added. The initial injection heats are larger than the heats for subsequent additions since at the beginning of the titration there is a large excess of empty or unpopulated binding sites. Initial heats most typically are the result of the almost complete reaction of the added ligand. As the titration proceeds, less and less of the added ligand is bound, and there are three species existing in solution: free ligand, [L], free macromolecule, [M], and the ligand/macromolecule complex, [ML].

The results from an ITC experiment present heat as a function of the number of injections (or as a function of the number of moles of added ligand), which can be analyzed to determine K, $\Delta H°$ and n, the number of binding sites on the macromolecule The heat produced in the ITC experiment is linearly dependent on the enthalpy change, $\Delta H°$, for the binding reaction, and nonlinearly dependent on the equilibrium constant, K.

1.4 Determination of Binding Constants, Enthalpy Changes and Heat Capacity Changes

The analysis of ITC data has been discussed by Eatough *et al.* (5), Wiseman *et al.* (6), Freire *et al.* (7), and Freyer and Lewis (8) among others. When the ligand binds to the macromolecule with a high binding constant ($>10^8$), approximately all of the ligand injected into the cell binds to the macromolecule until all of the ligand binding sites are

filled. In effect when $K>10^8$, $[L]_{free} <<< [L]_{bound} \approx [L]_{total}$. Under these conditions, called stoichiometric binding, the heat evolved for a presaturation injection can be divided by the number of moles of ligand injected to obtain the enthalpy of binding, $\Delta H°$; however, a binding constant cannot be determined under these conditions.

When the binding constant is smaller (typically less than approximately 10^8), there will be injections for which both free and bound sites and free ligand are in equilibrium. With increasing concentration of the ligand, a higher fraction of the binding sites become occupied until they are all filled at some high concentration. The range of concentrations required to saturate the binding sites is determined by the binding constant and the concentration of the macromolecule. To determine a binding constant accurately, the product of the binding constant and the macromolecule concentration should be between 10 and 1,000 (6).

Figure 1.2 shows simulated ITC experiments for ligand binding to a macromolecule with K ranging from 10^4 to 10^9. In the simulation, it was assumed that the concentration of macromolecule was 0.1 mM and the concentration of ligand was 1.0 mM. The $\Delta H°$ of binding was set at –50 kJ/mol, the cell volume was set to 1.5 mL and the injection volume was 10 μL.

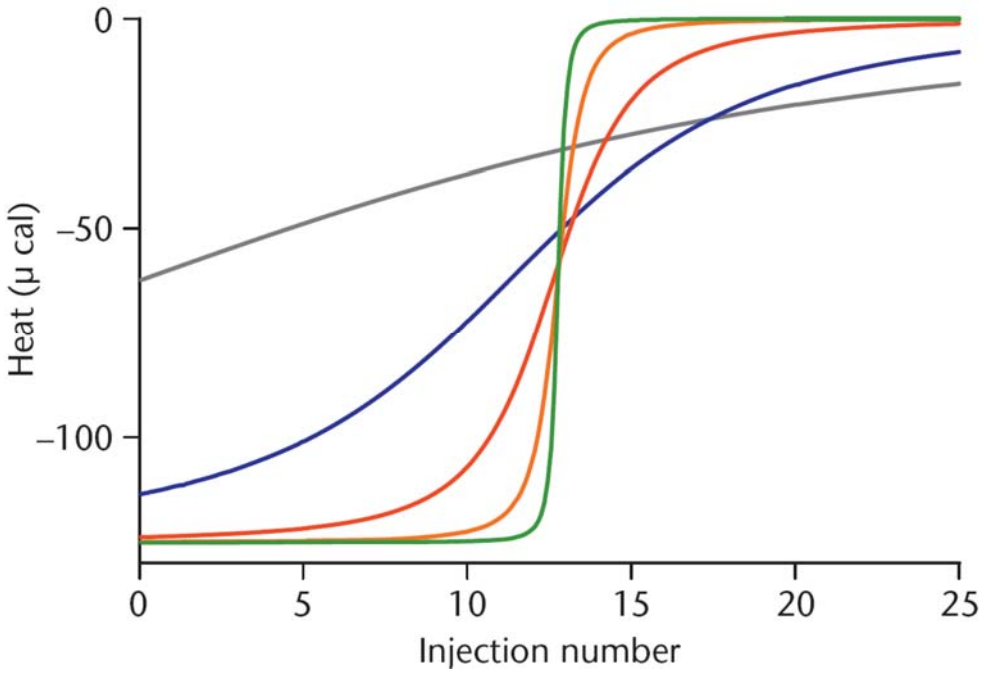


Figure 1.2 Simulated results from isothermal titration calorimetry experiments showing the effect of different binding constants on the heat evolved for an injection versus injection number.

In the simulation, it was assumed that the concentration of M was 0.1 M and the concentration of L was 1.0 M. The $\Delta H°$ of binding was set at -50 kJ, the cell volume was set to 1.5mL and the injection volume was 10 μL. The different colors correspond to different binding constants as follows: — 10^4; —10^5; —10^6; —10^7 and —10^8.

The curves simulated for K of 10^4 and 10^5 are too shallow for an accurate estimate of the binding constant and do not show saturation of the binding sites over the concentrations used. (Accurate estimates of the binding constant would require using higher concentrations of the macromolecule.) The curves simulated for K of 10^6 and 10^7 show an initial plateau, good curvature in between, and saturation towards the end of the experiment, indicating that the product of the macromolecule concentration and the binding constant is good for determining the binding constant accurately. At higher values of the K (10^8 and 10^9), the experiment is under stoichiometric conditions so there is no curvature in the thermogram. Even though K is inaccessible directly, estimates of

these large K values may be made in competition or ligand exchange experiments. However, even for the case where K is too large, the value of $\Delta H°$ for the binding reaction is readily determined.

Note that the heat for the initial injections under stoichiometric conditions is –500 μJ. The amount of ligand injected is 10^{-8} mol (10 μL×1 mM). Thus the $\Delta H°$ is calculated to be –500 μJ/10^{-8} mol or –50 kJ mol^{-1}. Sometimes macromolecules have more than one binding site for the ligand. If these sites are identical, i.e. have the same binding constant and $\Delta H°$, then the curves in Figure 1.2 would be shifted to the right; it takes more ligand to saturate the binding sites on the macromolecule. If the sites have different values of K and $\Delta H°$, then more complicated results are observed, but the values of the binding constant and $\Delta H°$ for each site can still be determined.

Finally, experiments can be performed under different conditions of temperature, pH, salt, etc. When the temperature is varied, $\Delta H°$ is usually a linear function of temperature (eqn (1.7)), where T is the experimental temperature, T_R a reference temperature and ΔC_P the change in heat capacity.

$$\Delta H_T^\circ = \Delta H_R^\circ + \Delta C_p(T - T_R) \tag{1.7}$$

A plot of the experimental $\Delta H°$ versus temperature will yield a straight line with a slope equal to ΔC_P.

1.5 Calorimetry Case Studies

As stated earlier, the binding of small molecules to biological macromolecules (e.g. a protein or nucleic acid) is a fundamental part of such diverse processes as enzyme catalysis, gene regulation and signal transduction. A richer understanding of a ligand

macromolecule interaction is established if the enthalpy ($\Delta H°$) and entropy ($\Delta S°$) change contributions to the free energy ($\Delta G°$) for formation of the complex are also known. In some cases, ITC seems uniquely suited to deconvolute the complicated or overlapping equilibria often seen in biological systems. In this section we describe the use of ITC in discovering a previously unknown process in minor groove binding, and the use of ITC in probing the Zn^{2+} binding site(s) of Carbonic anhydrase. These two case studies demonstrate some capabilities of ITC that have been descried in the earlier sections.

1.5.1 Case study 1: Complexity in the Binding of Minor Groove Agents to a Single A_2T_2 Site: Netropsin Binding to Four Different Hairpin DNA Constructs

Compounds that bind in the minor groove of duplex DNA are of interest for their therapeutic potential against diseases from cancer to infectious diseases caused by eukaryotic parasites. Minor groove-binders typically have a number of common structural features; a curved shape or the ability to adopt such a shape, groups with positive charges, H-bond donating ability, and a relatively flat conformation that has some flexibility in adjustment of intramolecular dihedral angles. Many of these compounds exhibit a preference for binding to AT sequences of at least four consecutive AT base pairs in the binding site. Our knowledge of these small molecule DNA interactions has been enhanced by both X-ray and NMR studies. Crystal structures illustrate excellent minor groove recognition in AT sequences. Minor grove binders form H-bonds with acceptors on the base pair edges at the floor of the groove, they effectively stack with the walls of the narrow minor groove in AT sequences and they have a positive charge to offset the phosphate charges and negative electrostatic potential of the groove. The steric bulk of the third H-bond in GC base pair containing sequences, as

well as the generally larger width of the minor groove in such sequences, prevents similar molecular recognition between the compounds and the minor groove in GC sequences.

Given the previous success of these structural models for minor-groove complex formation, we were surprised to find in high-resolution ITC studies that a classical polyamide minor-groove binding agent, netropsin, exhibits two quite different binding enthalpies in a 1:1 complex with a single AATT site. The unexpected and complicated binding profile exhibiting two binding equilibria having two different $\Delta G°$ values with distinctly different and compensating $\Delta H°$ and $\Delta S°$ values for formation of the netropsin AATT complex at a single AATT binding site clearly implies that there are two different complexes for netropsin. We had proposed that this observation was due to netropsin bound in different modes in the 1:1 AATT complex but also noted that other models, for example, different DNA conformations, could explain the ITC results. Other groups subsequently took up the study of the two-complex observation at a single binding site. Petty and co-workers confirmed our two-site observation with netropsin and showed that the magnitude of the thermodynamic differences determined in ITC experiments is dependent on the GC sequences flanking the AATT binding site (9). It had also been proposed that the two observed complexes were not due to netropsin binding modes in the hairpin duplex but were due to a conversion from a hairpin duplex to a two-stranded duplex with an internal bulge in place of the hairpin loop (10).

Given the importance of minor-groove binding compounds to our understanding of small molecule DNA interactions, we thought it essential to evaluate the different proposals for the molecular basis of the two binding modes at a single site in more detail and establish an explanation for the observation. Sequences with G or C bases flanking

the AATT binding site were included in the experiments. In addition to high-resolution ITC, the powerful methods of mass spectral (MS) analysis, native polyacrylamide gel electrophoresis (PAGE), and 2D NMR were used to investigate netropsin/DNA complexes.

All of the netropsin ITC titrations of the target hairpin DNAs exhibit a prominent dip in the titration curve immediately prior to the titration endpoint (see Figure 1.3). Two sets of binding parameters are needed to fit these data within experimental error, a clear indication that two unique netropsin complexes are being formed in these experiments.

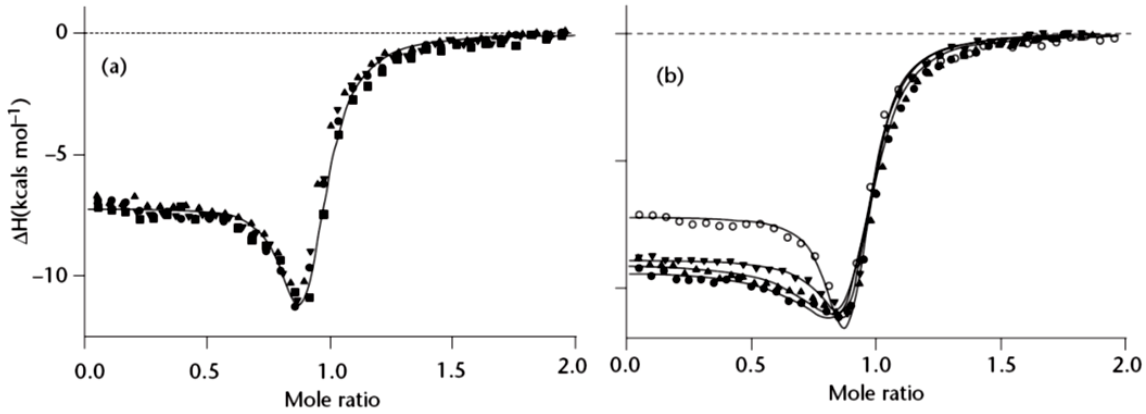


Figure 1.3 ITC titrations for the addition of Netropsin to the target hairpin DNAs.

The data for the titration of the GCG target hairpin are shown in panel (a). The symbols – filled circle, filled triangle, filled square and filled inverted triangle indicate data points taken in each of the four separate Netropsin titrations of the GCG hairpin. The line is calculated from the average of the best-fit parameters (K_1, K_2, ΔH_1, ΔH_2, n_1 and n_2) shown for the GCG oligo in Table 1.1. All titrations were conducted at pH 6.5 in 10mM cacodylic acid buffer, 100mM NaCl and 1mM EDTA with 25 μM DNA hairpin concentration. Typical ITC titrations are shown for each of the four different target hairpins in panel (b). The symbols, open circle, filled inverted triangle, filled triangle and filled circle are for the GCG, CGC, CG and GC target hairpins, respectively. The line through each set of data is calculated from the best-fit parameters (K_1, K_2, ΔH_1, ΔH_2, n_1 and n_2) for these individual experiments. The average best-fit parameters for each of the four target hairpins are listed in Table 1.1. The experimental conditions are the same as above.

The data shown in Figure 1.3A are for four different netropsin titrations of the GCG target hairpin. The four data sets were fit independently, and the average fit parameters used to produce the line in the figure. The data shown in Figure 1.3B are for a single netropsin titration done on each of the four target hairpins. Obviously all four target hairpins exhibit two binding events and all of the titration data are well fit with our two overlapping site model with n_{total} constrained to a value of 1.0. The results of the netropsin/hairpin ITC titrations are summarized in Table 1.1.

Table 1.1 Best-fit thermodynamic parameters for a two competing site model for binding netropsin to four different hairpin DNAs

Oligo	K_1 ($\times 10^7\ M^{-1}$)	ΔH_1 (kcal mol^{-1})	$-T\Delta S_1$ (kcal mol^{-1})	K_2 ($\times 10^5\ M^{-1}$)	ΔH_2 (kcal mol^{-1})	$-T\Delta S_2$ (kcal mol^{-1})	n_1/n_2
GCG	4.5 ± 1.2	-7.2 ± 0.2	-3.2 ± 0.9	8.8 ± 1.2	-21.5 ± 0.6	13.5 ± 1.3	6.3
CGC	2.9 ± 0.1	-8.9 ± 0.1	-1.3 ± 0.2	8.8 ± 1.6	-20.7 ± 0.1	12.6 ± 1.0	6.8
CG	1.8 ± 0.8	-9.1 ± 0.1	-0.7 ± 0.6	9.1 ± 1.1	-21.2 ± 0.6	15.2 ± 1.3	6.7
GC	1.3 ± 0.1	-9.4 ± 0.1	-0.2 ± 0.2	11.1 ± 1.7	-21.2 ± 0.7	13.0 ± 1.7	5.2

Three replicate titrations were performed on each oligonucleotide. The error limits in the table are the standard deviation of the mean parameters for the three titrations fit individually.

In general, all four target hairpin DNAs behave similarly in that all oligos appear to form two unique netropsin/DNA complexes. Subtle differences in binding energetics with respect to hairpin length and sequence were observed. It appears that the shorter 20-nt hairpins exhibit marginally higher type I binding affinities and less favorable enthalpy changes than the longer 24-nt hairpins. Type II binding trends are reversed in that the longer hairpins exhibit slightly higher affinities than the shorter hairpins.

The ITC studies of netropsin binding to the four different AATT containing duplexes present a consistent picture in which the thermodynamic signatures for complex formation is comprised of two thermodynamically different fractional binding events at a

1:1 ratio. In these studies, we confirmed the presence of the two binding modes at a single binding site and the dependence of the thermodynamic profile on the sequences of the flanking base pairs. ESI-MS and PAGE results clearly showed that the proposed netropsin induced hairpin to duplex transition is not correct.

In this study, ITC and NMR results clearly show that there are at least two netropsin complexes present in the 1:1 total complex of netropsin at AATT. PAGE, MS and NMR conclusively show that the two complexes are not due to a hairpin DNA to duplex transition. Imino proton NMR titrations strongly support the presence of two different modes of netropsin binding to AATT. Without the initial ITC results, the complexity of the netropsin DNA minor groove interaction would have been overlooked (11).

1.5.2 Case study 2: Zinc binding in human Carbonic anhydrase II

Approximately one-third of all enzymes contain a metal center that is required for structural stability and/or biological reactivity. A large portion of these enzymes are known as hydrolytic enzymes, which activate water molecules to hydrolyze specific substrates. Carbonic anhydrases (CAs) are ubiquitous hydrolytic enzymes that use a divalent zinc center to reversibly activate water to hydrolyze carbon dioxide using a ping-pong type mechanism (eqn (1.8)) (12).

$$CO_2 + [\text{HO-ZnCA}]^{2+} \rightleftharpoons [\text{HO}_2\text{CO-ZnCA}]^{+} \underset{\text{H}_2\text{O}}{\longleftrightarrow} [\text{H}_2\text{O-ZnCA}]^{2+} + \text{HCO}_3^{-} \qquad (1.8)$$

$$[\text{H}_2\text{O-ZnCA}]^{2+} \rightleftharpoons [\text{HO-ZnCA}]^{+} + \text{H}^{+}$$

This chemistry is catalyzed by a mononuclear Zn^{2+} center that is coordinated to the protein through three protein-derived histidine residues (13). A fourth metal coordination site is occupied by a labile solvent molecule to form a tetrahedral coordination as shown in Figure 1.4.

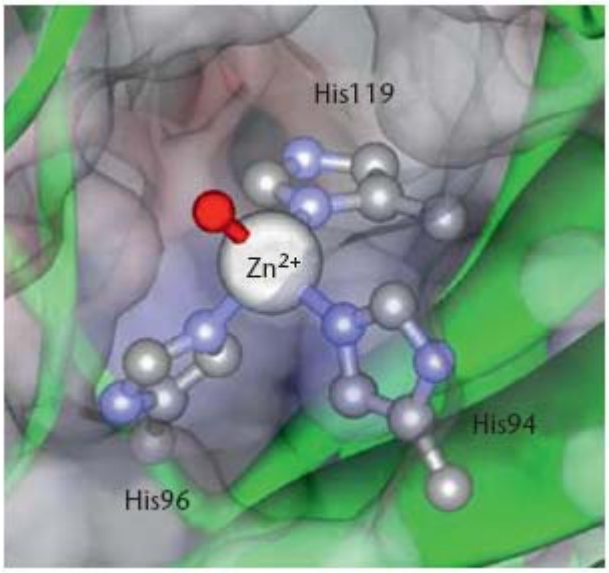


Figure 1.4 Tetrahedral zinc coordination in the mononuclear metal binding site in human carbonic anhydrase II.

The image shown above is based on the high resolution (0.9 Å) crystallographic structure (PDB accession code: 3KS3)

Beyond the zinc coordination environment, there is thought to be a number of other stabilizing interactions, such as a strong hydrogen bonding system and a series of hydrophobic interactions that lead to a strong metal association (14). The pK_D associated for Zn^{2+} in this system has been estimated to be on the order of 12 (15). However, the metal can be easily removed forming the metal-free or apo-CA, and then be reconstituted to regenerate a protein with the same reactivity (16,17) This property allowed for the

thermodynamics associated with Zn^{2+} binding to be interrogated using ITC techniques (18).

Zinc ion binds to the apo-enzyme through a complex process; this is most evident in PIPES buffer. However this binding event appears to be more consistent with a one-site binding process in other biological buffers. Upon correcting the data for a number of dilution effects with this Zn^{2+} titration, the Zn^{2+} binding data can be fit with a one-site binding model, where n = ~ 1 and the equilibrium constant (K_{ITC}) is measured in the ITC to be on the order of 2 x 10^6 (18). The buffers used in this study have different ionization energies, which allows us to assess if the metal binding equilibria are linked proton release or absorption event (19). In PIPES, MOPS, ACES, and Tris buffers, the heats associated with ΔH_{obs} varies as a function of the ionization energy associated with the buffer system. This is indicative of the equilibrium being linked to proton dependent processes. When the ΔH_{obs} values are plotted versus the enthalpy of ionization (ΔH_{ion}) associated with each buffer, there is a clear linear dependence on both observed enthalpy terms as observed in Figure 1.5.

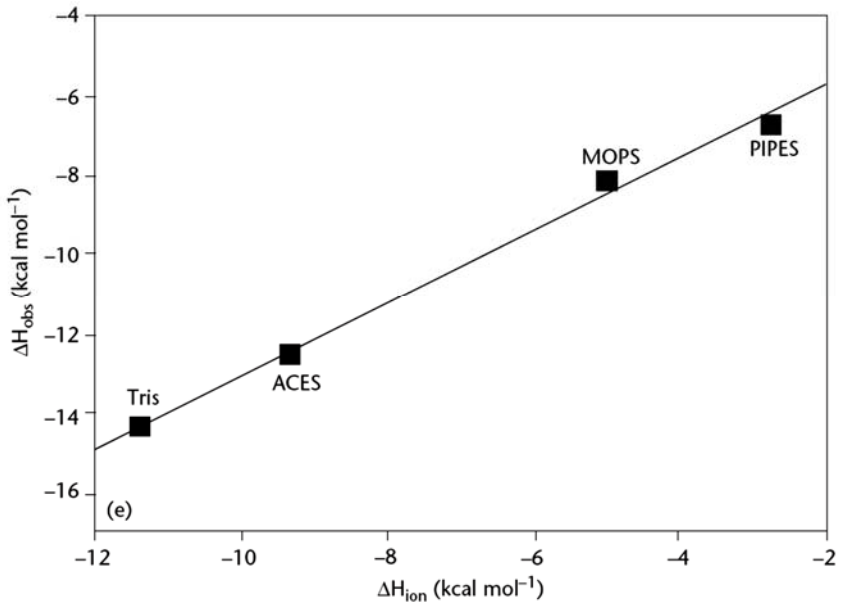


Figure 1.5 Plot of experimental enthalpies versus enthalpies of ionization for four common biological buffers.

The ΔH_{obs} measured for both processes is linearly dependent on the ΔH_{ion}. Eqn (1.9) can be used to estimate the thermodynamic parameters associated with the complex Zn^{2+} binding process to CA, where the y-intercepts (y_0) are the corrected enthalpy of metal binding (ΔH) and slopes (n_{H^+}) of these lines yields the number of protons released or absorbed by the respective systems.

$$\Delta H_{obs} = \Delta H^* + (n_{H^+} \times \Delta H_{ion}) \qquad (1.9)$$

Using the corrected enthalpies of metal binding (ΔH^*) with eqns (1.2) and (1.3), the changes in free energy (ΔG) and entropy (ΔS) were calculated to yield data in each of the buffer systems shown in Table 1.2. The metal binding equilibrium has a favorable ΔG value, which are driven by a favorable enthalpy term. Additionally 0.9 protons are released as a function of Zn^{2+} binding to CA.

Table 1.2 Average ITC determined thermodynamic parameters for zinc-binding to apo-CA in four buffers

Buffer	n	K_{ITC} ($\times 10^6$)	ΔH^* (kcal mol^{-1})	ΔG (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})
PIPES	0.97	3.3 ± 0.3	-7.2 ± 0.1	-8.9 ± 0.1	-6.4 ± 0.6
MOPS	1.03	2.2 ± 0.3	-8.1 ± 0.1	-8.6 ± 0.1	-0.6 ± 0.3
ACES	1.02	1.1 ± 0.1	-7.4 ± 0.0	-8.2 ± 0.0	-0.8 ± 0.0
MOPS	1.05	7.3 ± 0.6	-12.3 ± 0.3	-9.4 ± 0.1	3.0 ± 0.3

Average thermodynamic constants associated with Zn^{2+} binding to carbonic anhydrase. Equilibrium constants (K), changes in enthalpy (ΔH^*), change in Gibbs Free energy (ΔG), and change in entropy are the average of multiple titrations and corrected for proton release or absorption. ITC data were adjusted based on zinc concentrations.

Clearly these metal binding events are complex systems, where the buffer (a Lewis base) is competing with the CA for Zn^{2+}. The equilibrium constant (K_{ITC}) measured in the ITC is truly the product of the dissociation constant associated with the buffer-Zn^{2+} complexes (K_1) and the association constant for Zn^{2+} with apo-CA (K_2) as shown in eqn (1.10).

$$[\mathrm{Zn(Buffer)}]^{2+} + \mathrm{CA} \overset{K_1}{\leftrightarrow} \mathrm{Zn}^{2+} + \mathrm{Buffer} + \mathrm{CA} \overset{K_2}{\leftrightarrow} \mathrm{ZnCA} + \mathrm{Buffer} \quad (1.10)$$

The equilibria associated with Zn^{2+} coordination to many biological buffers have been well-studied, (these values are listed in NIST Standard Reference Database 46) which allow us to estimate the association constant for Zn^{2+} binding apo-CA (K_2) based on our ITC data. Zn^{2+} has an association constant (K_2) for CA on the order of 2×10^9, and this is an enthalpy driven process (- 16.2 kcal/mol), with only a slight increase in entropy ($-T\Delta S$ = + 3.5 kcal/mol).

1.6 Summary

Isothermal titration calorimetry is a powerful technique for studying the interactions of macromolecules with ligands. All of the thermodynamic quantities that

characterize the binding reaction, including $\Delta G°$, $\Delta H°$, $\Delta S°$ and ΔC_P, can be determined. Additional information regarding changes in protonation, as well as changes in solvation, can also be studied using ITC.

Biologists need to better understand the recognition of small molecules for specific interaction sites on larger molecules and the nature of the weak individual interactions that can result in very high affinity. Dr. F. Raymond Salemme, chief scientific officer of 3-D Pharmaceuticals, was quoted in a *C&E News* feature article as saying, "The initial expectation of structure-based drug design, that you were going to be able to design molecules and they were going to work right out of the box, was unrealistic. We didn't understand the thermodynamics well enough" (20). The use of ITC methods to probe the energetics of biologically relevant binding interactions is somewhat underappreciated (21,22). ITC provides a universal approach in determining the molecular nature of non-covalent interactions involved in the binding of small molecules to biopolymers (and even biopolymers to other biopolymers). In particular, ITC binding experiments can yield a complete set of thermodynamic parameters for complex formation in a single experiment (5,6,23,24). The parsing of the Gibbs free energy change into the enthalpy and entropy contributions can provide new insight into the molecular nature of the binding interaction being studied (22,25). The energetic information is fundamental not only in understanding naturally occurring binding interactions but also is particularly useful in drug discovery studies (1,2,9,22,26-28).

The remaining chapters in this dissertation were selected to demonstrate the power and versatility of ITC technique in the thermodynamic investigations of two model systems in aqueous and non-aqueous media.

1.7 References

1. Freyer, M.W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L.H. and Lewis, E.A. (2007) Biophysical studies of the c-MYC NHE III1 promoter: model quadruplex interactions with a cationic porphyrin. *Biophysical journal*, **92**, 2007-2015.

2. Dettler, J. and Lewis, E.A. (2011) In Sheardy, R. D. and Winkle, S. A. (eds.), *Frontiers in Nucleic Acids*, pp. 33-50

3. Cashman, D.J., Buscaglia, R., Freyer, M.W., Dettler, J., Hurley, L.H. and Lewis, E.A. (2008) Molecular modeling and biophysical analysis of the c-MYC NHE-III1 silencer element. *J Mol Model*, **14**, 93-9101.

4. Wyman, J. and Gill, S.J. (1990) *Binding and Linkage: The Functional Chemistry of Biological Macromolecules*. University Science Books, Mill Valley, CA.

5. Eatough, D.J., Lewis, E.A. and D., H.L. (1985) In Grime, K. (ed.), *Analytical Solution Calorimetry*, New York: Wiley, pp. 137–161.

6. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, **179**, 131-137.

7. Freire, E., Mayorga, O.L. and Straume, M. (1990) Isothermal titration calorimetry. *Analytical Chemistry*, **62**, 950A–959A.

8. Freyer, M.W. and Lewis, E.A. (2008) In Correia, J. J. and Detrich, H. W. (eds.), *Biophysical Tools for Biologists*. Academic Press. , San Diego, CA, pp. 79–113.

9. Degtyareva, N.N., Fresia, M.J. and Petty, J.T. (2007) DNA conformational effects on the interaction of netropsin with A-tract sequences. *Biochemistry*, **46**, 15136-15143.

10. Lah, J., Drobnak, I., Dolinar, M. and Vesnaver, G. (2008) What drives the binding of minor groove-directed ligands to DNA hairpins? *Nucleic Acids Res*, **36**, 897-904.

11. Lewis, E.A., Munde, M., Wang, S., Rettig, M., Le, V., Machha, V. and Wilson, W.D. (2011) Complexity in the binding of minor groove agents: netropsin has two thermodynamically different DNA binding modes at a single site. *Nucleic Acids Res*, **39**, 9649-9658.

12. Krishnamurthy, V.M., Kaufman, G.K., Urbach, A.R., Gitlin, I., Gudiksen, K.L., Weibel, D.B. and Whitesides, G.M. (2008) Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem Rev*, **108**, 946-941051.

13. Awaru, B.S., Kim, C.U., Sippel, K.H., Gruner, S.M., Agbandje-McKenna, M., Silverman, D.N. and McKenna, R. (2010) A Short, Strong Hydrogen Bond in the Active Site of Human Carbonic Anhydrase II. *Biochemistry*, **49**, 249-251.

14. Cox, J.D., Hunt, J.A., Compher, K.M., Fierke, C.A. and Christianson, D.W. (2000) Structural Influence of Hydrophobic Core Residues on Metal Binding and Specificity in Carbonic Anhydrase II. *Biochemistry*, **39**, 13687-13694.

15. McCall, K.A. and Fierke, C.A. (2004) Probing Determinants of the Metal Ion Selectivity in Carbonic Anhydrase Using Mutagenesis. *Biochemistry*, **43**, 3979-3986.

16. DiTusa, C.A., Christensen, T., McCall, K.A., Fierke, C.A. and Toone, E.J. (2001) Thermodynamics of Metal Ion Binding. 1. Metal Ion Binding by Wild-type Carbonic Anhydrase. *Biochemistry*, **40**, 5338-5344.

17. Lindskog, S. (1963) Effects of pH and Inhibitors on Some Properties Related to Metal Binding in Bovine Carbonic Anhydrase. *J Biol Chem*, **238**, 945-951.

18. Song, H., Wilson, D.L., Farquhar, E.R., Lewis, E.A. and Emerson, J.P. (2012) Revisiting zinc coordination in human carbonic anhydrase II. *Inorg Chem*, **51**, 11098-11105.

19. Grossoehme, N.E., Spuches, A.M. and Wilcox, D.E. (2010) Application of isothermal titration calorimetry in bioinorganic chemistry. *J Biol Inorg Chem*, **15**, 1183-1191.

20. Henry, C.M. (2001), *C&E News*, pp. 69-74.

21. Chaires, J.B. (2006) A thermodynamic signature for drug-DNA binding mode. *Arch Biochem Biophys*, **453**, 26-31.

22. Freyer, M.W., Buscaglia, R., Nguyen, B., Wilson, W.D. and Lewis, E.A. (2006) Binding of netropsin and 4,6-diamidino-2-phenylindole to an A2T2 DNA hairpin: a comparison of biophysical techniques. *Anal Biochem*, **355**, 259-266.

23. Christensen, J.J., Izatt, R.M., Hansen, L.D. and Partridge, J.A. (1966) Entropy titration. A calorimetric method for the determination of ΔG, ΔH and ΔS from a single thermometric titration. *Journal of Physical Chemistry*, **70**, 2003-2010.

24. Doyle, M.L. (1997) Characterization of binding interactions by isothermal titration calorimetry. *Curr Opin Biotechnol*, **8**, 31-35.

25. Ladbury, J.E. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem Biol*, **3**, 973-980.

26. Ladbury, J.E. (2004) Application of isothermal titration calorimetry in the biological sciences: things are heating up! *Biotechniques*, **37**, 885-887.

27. Ladbury, J.E. (2001) Isothermal titration calorimetry: application to structure-based drug design. *Thermochimica Acta*, **380**, 209-215.

28. Freyer, M.W., Buscaglia, R., Hollingsworth, A., Ramos, J., Blynn, M., Pratt, R., Wilson, W.D. and Lewis, E.A. (2007) Break in the heat capacity change at 303 K for complex binding of netropsin to AATT containing hairpin DNA constructs. *Biophysical journal*, **92**, 2516-2522.

CHAPTER II

MATERIALS AND METHODS

Beside ITC being the main technique, five additional techniques were used in this dissertation to study the structures and interactions of DNA and small molecules as well as fullerenes and their binding partner the buckycatcher.

2.1 UV-vis spectroscopy

UV-vis spectroscopy was used to determine and verify the concentrations of all samples. UV-vis experiments were performed on an Olis HP-8452 diode array (Olis, Bogart, GA). Extinction coefficient for DNA samples were estimated using the nearest neighbor calculation for single stranded DNA (1). Samples were prepared from stock solutions in appropriate buffer and organic solvent. Job plot analysis (2) is used to determine the binding stoichiometry for complex formation. The titrations were carried out by changing the amount of each species in the complex yet maintaining the total concentration of ([analyte] + [binding partner]) constant throughout the experiment.

2.2 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is routinely used to measure the thermodynamic properties of biopolymer binding interactions. Advances in sensitivity of this instrumentation has enabled ITC users to measure heat change as small as 0.1 μcal and binding affinities (K_a) as large as 10^9 M^{-1}. A schematic of an ITC instrument is

shown Figure 2.1. Prior to addition of ligand, a constant power is applied to the reference cell. This directs a feedback circuit, activating a heater located on the sample cell. During the experiment, ligand is titrated into the calorimeter cell containing the sample, causing heat to be either taken up or evolved. Measurements consist of the input of power required to maintain equal temperatures between the sample and reference cells. Analysis of ITC data involves baseline and concentration corrections, followed by a nonlinear regression process that will fit the binding curve to produce values for K_a and ΔH°. Value for ΔG is calculate from the relation ΔG° = -RTln(K_a) and –TΔS° term is calculated from ΔG° = ΔH° - TΔS°.

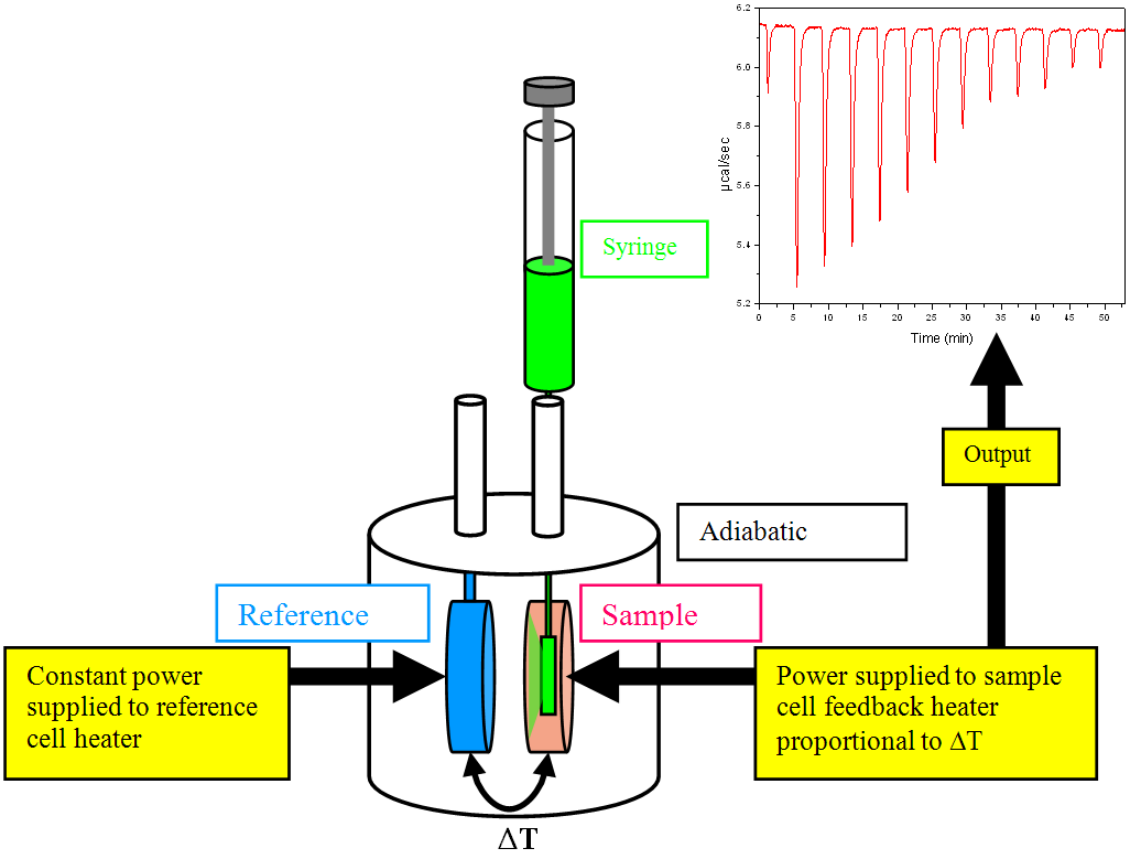


Figure 2.1 Schematic representation of an Isothermal Titration Calorimeter

ITC experiments were performed using a Microcal VP-ITC (Microcal, Northampton, MA). All titrations were performed by overfilling the cell with approximately 1.5 ml of dilute titrate solution. Typical titration involves either of 60 injections of 5 μL (or 20 injections of 14 μL for simple binding reactions) of a semi dilute titrant solution. Both titrate and titrant solutions were extensively degassed to ensure no bubbles exist in the calorimeter cell or the injection syringe. Typical experiments are carried out at room temperature but can be changed to any desired temperature . ITC experiments are routinely performed in triplicates to ensure reproducible data.

2.3 Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) is a thermoanalytical technique frequently used to measure the stability of biomacromolecules and assemblies macromolecules. The schematic of a modern DSC instrument is shown in Figure 2.2. The DSC measures the amount of power needed to maintain a constant temperature between both the sample and reference cells as the temperature rises. As a sample physically changes the amount of power used to maintain the constant temperature between the two cells is compared to that of the temperature change between the jacket and the cells. The melting temperature, T_m, is considered the temperature at which fraction of the folded biomolecule is equal to the fraction of unfolded species. Analyzing the melting curve to a specific model (e.g. two-state, non-two-state, etc.) yields the more thermodynamic properties of the unfolding process.

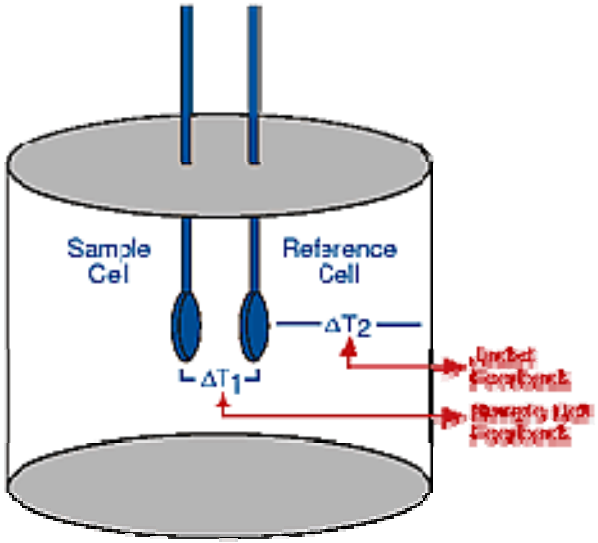


Figure 2.2 Schematic representation of a Differential Scanning Calorimeter

DSC experiments were performed using a Microcal VP-DSC (Microcal, Northampton, MA). In a typical DSC experiment, the typical scanning range is for biological sample is between 5 and 95 °C with a rate of 60 °C/hour. Samples are prepared from stock solutions and at appropriate concentrations (> 100 μM for DNA samples) to produce measureable changes in excess heat capacity. Samples and buffer solution were extensively degassed prior to loading into the sample and reference cells. The resulting melting thermograms are analyzed using Origin 7.1 software (Northampton, MA) to appropriate model. Data analysis involves deconvolution of the melting curve to give the melting temperature T_m , integration of the excess ΔC_p curve against temperature to give calorimetric enthalpy change, ΔH_{cal}, and the fitting the shape of the melting curve to yield van't Hoff enthalpy change, ΔH_{VH}, for the unfolding process.

2.4 Circular Dichroism

Circular Dichroism (CD) measures the differential absorbance of left-circularly polarized light and right-circularly polarized light after being absorbed by the sample (see Figure 2.3). This technique was used to determine the secondary structure of the DNA as well as any conformational change brought about by changes in temperature, addition of denaturants, or upon binding to a small molecule or protein. Specifically, in this study, CD was able to show changes in the equilibriums between anti-parallel and parallel quadruplex as well as the equilibriums between the different folded species of the G-quadruplex.

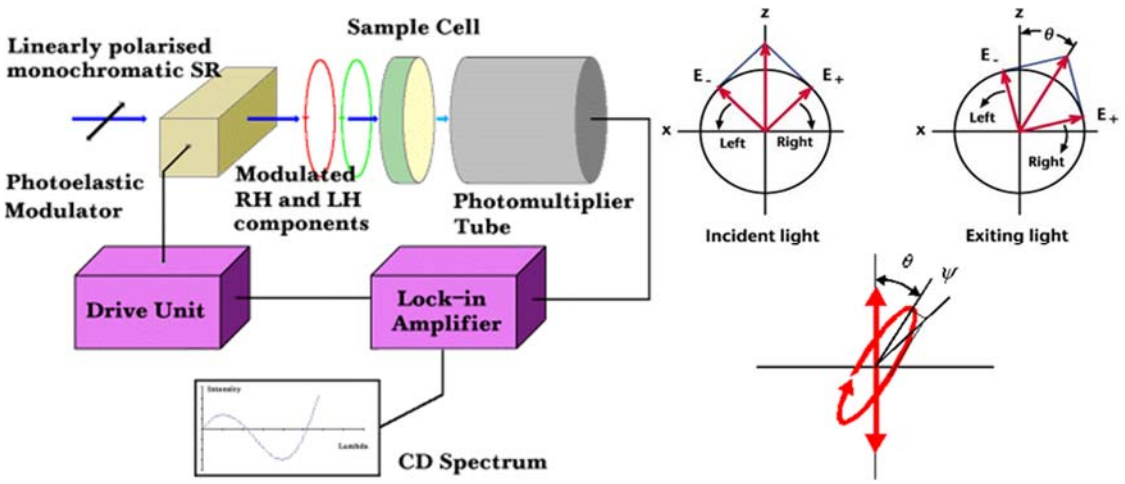


Figure 2.3 Schematic representation of a Circular Dichroism Spectropolarimeter

CD experiments were performed using an Olis DSM 20 Spectropolarimeter (Olis, Bogart, GA). Samples were prepared from stock solutions to produce a sample solution having a nominal absorbance of 0.9 AU. Prior to data collection, the instrument was

flushed with at least 10 minutes of Nitrogen gas flowing at a rate of 2 L/min to eliminate formation of ozone gas that can interfere with ellipticity measurements in the UV range.

2.5 Fluorescence spectroscopy

Fluorescence is yet another powerful technique to measure changes in the chemical environment of biological sample. A schematic of a spectrofluorimeter is shown in Figure 2.4. Steady-state fluorescence experiments begin by exciting a sample containing the fluorophore at a particular wavelength (usually at the maximum absorption wavelength for the fluorophore) and the emission spectra is recorded. Specifically, in this study, steady-state fluorescence was used to monitor the changes in fluorescence signal of TMPyP4 upon complexation with G-quadruplex DNA.

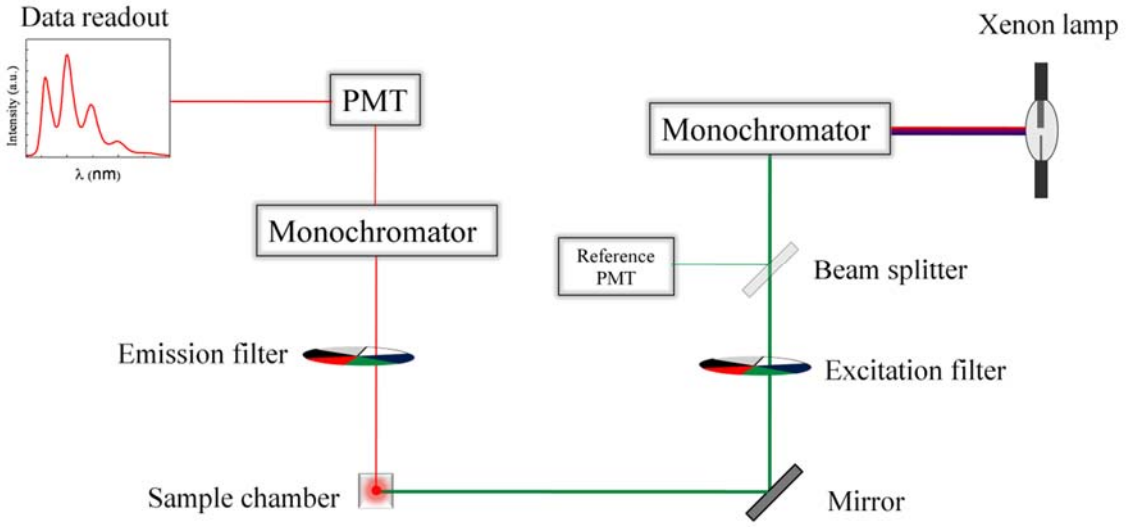


Figure 2.4 Schematic representation of a Spectrofluorimeter

Typical steady-state fluorescence experiments were performed using an Olis DM 45 spectrofluorimeter equipped with a 1 cm path length quartz cuvette. The sample turret

is temperature regulated using a Quantum Northwest (Liberty Lake, WA) Peltier control. Fluorescence is a very sensitive technique in that it can measure the fluorescence signal of analyte as low as picoMolar level. Fluorescence signal is also sensitive to thermal changes and chemical environment. Performing singular-value-decomposition on fluorescence spectra can provide rich information regarding the number of chemical species and their distributions, binding equilibrium between the different chemical species, the transition temperature, T_m, as well as enthalpy change, ΔH, for the unfolding process.

2.6 Mass spectrometry

Mass spectrometry has become very widely used for the analysis of large biomolecules ever since the improved ion source development in 2003 . Schematics of a mass spectrometer together shown with two types of sample ionization methods are shown in Figure 2.5.

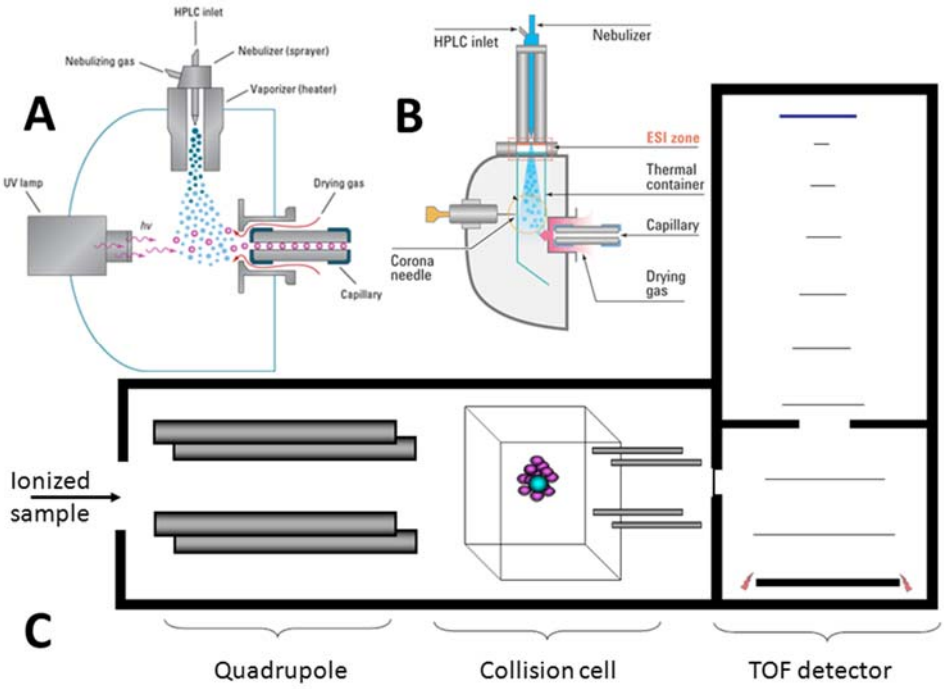


Figure 2.5 Schematic representation of a modern mass spectrometer.

Panel A shows the photoionization process for APPI. Panel B shows the electrospray ionization process for ESI. Panel C shows the mass analyzers coupled to a TOF detector.

In this dissertation, ESI-MS was used to determine the binding stoichiometry for the complex formation between porphyrin molecules and G-quadruplex DNAs. APPI was used to determine the binding stoichiometry for the complexation between buckycatcher and fullerenes.

2.7 References

1. Plum, G.E. (2000) Determination of oligonucleotide molar extinction coefficients. *Current Protocols in Nucleic Acid Chemistry*, **73**, 1-17.

2. Job, P. (1928) Formation and stability of inorganic complexes in solution. *Annali di Chimica Applicata*, **9**, 113-203.

CHAPTER III

MODELING COMPLEX EQUILIBRIA IN ISOTHERMAL TITRATION CALORIMETRY EXPERIMENTS:THERMODYNAMIC PARAMETERS ESTIMATION FOR A THREE-BINDING-SITE MODEL

As previously published in Analytical Biochemistry, 2013. 434, pp 233-241

3.1 Abstract

Isothermal titration calorimetry (ITC) is a powerful technique that can be used to estimate a complete set of thermodynamic parameters (e.g., K_{eq} (or ΔG), ΔH, ΔS, and n) for a ligand-binding interaction described by a thermodynamic model. Thermodynamic models are constructed by combining equilibrium constant, mass balance, and charge balance equations for the system under study. Commercial ITC instruments are supplied with software that includes a number of simple interaction models, for example, one binding site, two binding sites, sequential sites, and n-independent binding sites. More complex models, for example, three or more binding sites, one site with multiple binding mechanisms, linked equilibria, or equilibria involving macromolecular conformational selection through ligand binding, need to be developed on a case-by-case basis by the ITC user. In this paper we provide an algorithm (and a link to our MATLAB program) for the nonlinear regression analysis of a multiple-binding-site model with up to four overlapping binding equilibria. Error analysis demonstrates that fitting ITC data for

multiple parameters (e.g., up to nine parameters in the three-binding-site model) yields thermodynamic parameters with acceptable accuracy.

3.2 Introduction

Titration calorimetry has been used for the simultaneous determination of *K* and ΔH for more than 40 years (1-4). Isothermal titration calorimetry is now routinely used to directly characterize the thermodynamics of biopolymer binding interactions(5-13). Knowledge of the thermodynamic profiles for drug–receptor binding interactions greatly enhances drug design and development (14-17). Isothermal titration calorimetry (ITC) instruments (available from GE Healthcare (Microcal) and TA Instruments (Calorimetry Sciences)) have adequate sensitivity to measure heat effects as small as 0.1 μcal, making it possible to directly determine binding constants as large as 10^8 to 10^9 M^{-1}. Even larger values for K may be estimated from competitive binding experiments (18,19).

To take full advantage of the powerful ITC technique, the user must be able to design the optimum experiment, understand the nonlinear fitting process, and appreciate the uncertainties in the fitting parameters *K*, ΔH, and n. ITC experiment design and data analysis have been the subject of numerous publications (5,6,14-23). Recent reviews of isothermal titration calorimetry describe the ease of use of modern microcalorimeters (18,19,24). Several papers have described modern uses of isothermal titration calorimetry to study a broad range of chemical equilibria in numerous ways (25-27). For example, ITC studies are now being used to identify possible binding mechanisms for ligand–DNA complexes based on their thermodynamic signatures (28). ITC experiments exploring iron binding to Escherichia coli ferritin were accompanied by a model describing the equations for three independent binding sites (29), while ITC studies of histone

nucleoplasm interactions were best fit with a site-specific cooperative model including four equilibrium constants and four enthalpy changes (30). Examples of the use of ITC experiments to unravel the complicated binding equilibria often occurring in biology are limited since the analysis tools provided by the ITC industry cover only the simplest cases.

The improved sensitivity of the current ITC instruments has resulted in the ability to accurately estimate thermodynamic parameters for multiple overlapping binding equilibria. Figure 3.1 shows three unique thermograms that might result from the titration of a system exhibiting two overlapping binding processes.

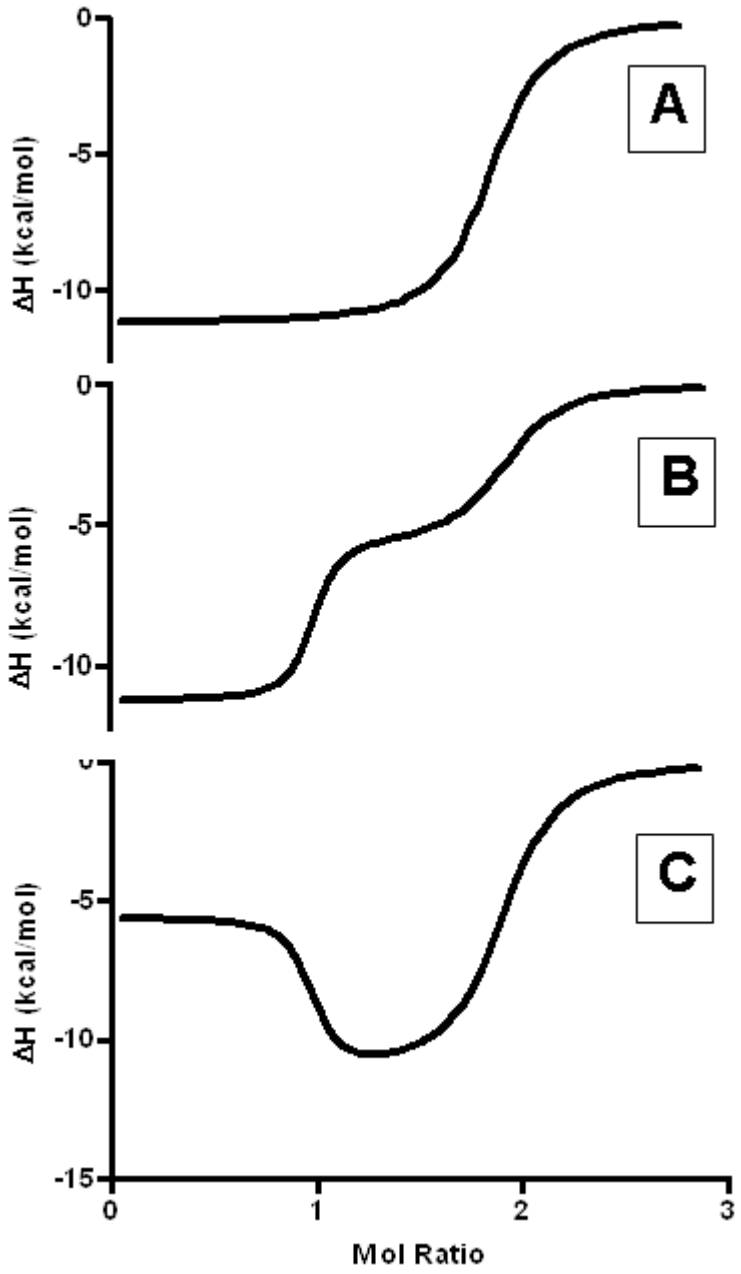


Figure 3.1 Representative two-process ITC titration experiment thermograms.

In all three cases, the stoichiometry demonstrates that two ligand molecules are binding to the receptor molecule. The top panel represents a system wherein the enthalpy change for binding of both ligands are experimentally indistinguishable. The middle panel represents a system where the binding process with higher affinity is accompanied by a more exothermic enthalpy change. The bottom panel represents a system where the higher binding affinity process demonstrates a less exothermic enthalpy change than the lower binding affinity process.

Figure 3.1B and C were simulated from a model for a system having two binding processes, wherein K_1 is much greater than K_2 and ΔH_1 is not equal to ΔH_2. These two thermograms show a clear distinction between the first and the second binding process, which can be differentiated when fitting data based on the differences in enthalpy change. However, Figure 3.1A also demonstrates a simulation for a system exhibiting two binding sites in which the binding affinities have the same relative magnitude as the previous cases, $K_1 \gg K_2$, but with ΔH_1 equal to ΔH_2. In this case, it is not possible to distinguish the two overlapping equilibria and only the weaker binding process (K_2) can be modeled. The equivalence of the binding enthalpy changes for the two overlapping processes causes this system to be modeled as a single binding process with $n = 2$.

The three plots shown in Figure 3.1 demonstrate the increased complexity that is often seen in ITC thermograms as the number of binding process is increased. ITC thermograms must be understood prior to modeling the data. Understanding the binding profiles ensures that the models being applied to the data give an accurate representation of the chemical equilibria being observed. Furthermore, error analysis of algorithms being applied to analyze ITC thermograms ensures best-fit solutions are accurate. Several papers have been published on methods for evaluating the error introduced into ITC results from either experimental or data fitting considerations (31-38). In general, the potential for using ITC to estimate the number of parameters needed to describe more complex thermodynamic binding models is underappreciated.

In this work, we describe the construction of algorithms that can be used to model ITC data obtained on systems exhibiting three or more binding sites described with overlapping equilibrium constants. We have also used simulated data and Monte Carlo

method (31,34,39) to evaluate the uncertainties and cross-correlation in the binding parameters (K, ΔH, and n). One- and two-site systems are used to compare our nonlinear fitting routines to those in the program commercially available from GE Healthcare (Microcal, Northampton, MA, USA; Origin 7.0). Simulated data are used to evaluate the three-binding-site algorithm implemented in MATLAB (version R2012a, 2012; The MathWorks, Natick, MA, USA) for the accuracy of multiple parameters (K_{1-3}, ΔH_{1-3}, and n_{1-3}) determined in an ITC experiment. The analysis techniques described in this paper can be extended to ITC data obtained on other complex systems if the user can construct the appropriate thermodynamic model for the binding interactions and/or linked equilibria. One novel aspect of this work is that our n-sites program for the analysis of systems having up to four binding sites is described in detail in the Supplementary material and the fully functional analysis program can be downloaded from our web site (http://lewis.chemistry.msstate.edu/download.html).

3.3 Material and methods

3.3.1 Multiple-binding-site model

Algorithms for modeling ITC thermograms demonstrating up to four overlapping binding processes were developed using MATLAB software. The algorithms model the binding affinity, K_j; the molar enthalpy change, ΔH_j; and the total stoichiometric ratio, n_j; for each of the j = 1 to 4 binding process using nonlinear regression techniques. The mass balance and equilibrium equations were manipulated to produce an $(n + 1)$th degree polynomial for the n-binding sites, with free ligand as the indeterminate.

3.3.2 General-binding-site model

The thermodynamic model algorithms were developed from a combination of the appropriate mass balance and equilibrium constant expressions. Eqs. (3.1) and (3.2) are the generalized mass balance and binding equilibrium equations, respectively. Each equation is written in a simplified form that can be expanded to include n-binding site:

$$L_{t=}[L] + M_t \sum_1^{j=n}(n_j \Theta_j) \tag{3.1}$$

$$K_j = \frac{\Theta_j}{(1-\Theta_j)[L]} \Rightarrow \Theta_j = \frac{[L]K_j}{1+[L]K_j} \tag{3.2}$$

Eq. (1) establishes that at any point in the titration, the total ligand present in the reaction vessel (L_t) must be either bound to one of the *n*-binding sites $M_t \sum_1^{j=n}(n_j \Theta_j)$ or free in the reaction vessel ($[L]$). The equilibrium constants in Eq. (3.2) have been rewritten to express the fraction of process *j* bound (Θ_j) as a function of the binding affinity, K_j, and the presence of free ligand. Substitution of Θ_j into Eq. (3.1) and expanding yields a $(n + 1)^{th}$ degree polynomial where $[L]$ is the indeterminate:

$$[L]^{k+1} + \alpha_k[L]^k + \alpha_{k-1}[L]^{k-1} + \cdots + \alpha_0 = 0 \tag{3.3}$$

Eq. (3.3) represents a simplified form of this polynomial, where α_i represents the *i*th coefficient. The coefficients, α_i, are derived from ligand concentration, macromolecule concentration, stoichiometric ratios, and binding affinities. Calculating the roots of the polynomial shown in Eq. (3.3) determines the concentration of free ligand present in the reaction vessel after the *i*th injection.

Once the concentration of free ligand is determined, substitution of $[L]$ into Eq. (2) yields the fraction of site *n* bound after the *i*th injection. The total heat produced from the start of the titration through the *i*th injection, Q_i, can be calculated using Eq. (3.4),

$$Q_i = M_t V_0 \sum_1^{i=n} (n_i \theta_i \Delta H_i) \tag{3.4}$$

where V_o is the active calorimeter cell volume and where $M_t V_o \sum_1^{j=n} (n_i \Theta_i \Delta H_i)$ is the total reaction heat, calculated as the sum of the heat produced in all of the overlapping binding reactions from the start of the titration through the *i*th injection. Eq. (3.5) is then used to calculate the differential heat produced during the *i*th injection, ΔQ_i

$$\Delta Q_i = Q_i - Q_{i-1} \tag{3.5}$$

where Q_i and Q_{i-1} represent the total heats produced from the start of the titration through injection points *i* and *i*-1 ; and M_t is the macromolecule concentration in the reaction vessel (calorimeter cell) as corrected for dilution and for displacement from the calorimeter cell as titrant is added. Eq. (3.5) assumes that all reaction heat is sensed by the calorimeter and that no heat is lost up the fill tube or owing to reactions occurring outside of the calorimeter cell and thus incompletely sensed by the calorimeter. ΔQ can be changed to include heat produced from reactions occurring outside of the reaction vessel. One such assumption is that the heat produced outside of the calorimeter cell (i.e., in the fill tube) is measured with only half the efficiency as that inside of the reaction vessel. A correction term for the heat being produced outside of the reaction vessel can be easily introduced into the fitting algorithms.

The algorithms used here introduced a correction term for the heat being produced outside of the reaction vessel by restricting the concentration of macromolecule (M_t) and

the concentration of the ligand (L_t) after each injection. Eqs. (3.6) and (3.7) are the iterative functions used to determine the concentrations of macromolecule and ligand after the *i*th injection. $M_{t,i}$ and $L_{t,i}$ are the concentrations of macromolecule and ligand, respectively, after the *i*th injection; L_i is the concentration of ligand being injected into the cell; V_i is the injection volume; V_o is the active cell volume. The initial conditions for M_t and X_t are $M_{t,o} = M_i$ and $X_{t,o} = 0$, where M_i is the concentration of macromolecule loaded into the cell initially:

$$M_{t,i} = M_{t,i-1} - (V_i M_{t,i-1}/V_0) \tag{3.6}$$

$$L_{t,i} = [L_i V_i + L_{t,i}(1 - V_i)]/V_0 \tag{3.7}$$

Eqs. (3.6) and (3.7) were constructed to accurately mimic the concentrations of ligand and macromolecule present in the active cell volume of the calorimeter. After each injection, the macromolecule concentration is lowered by the amount of macromolecule leaving the active cell volume. Similarly, the concentration of ligand is increased after each injection, but after the initial injection, a correction is also made for the amount of ligand that leaves the active cell volume. These equations assume that the stirring rate is efficient enough that the ligand is accurately mixed into the active cell volume directly after the injection.

The difference between the heats calculated for ΔQ_i (Eq. (3.5)) and those measured during an ITC experiment was used to determine the goodness of fit. A Levenberg–Marquardt nonlinear regression model common to MATLAB was used to determine best-fit parameters. (No other minimization routines were tried.) By allowing the model to iterate over the parameters ($K_1, \ldots, K_n$; $n_1, \ldots, n_n$; $\Delta H_1, \ldots, \Delta H_n$), a

solution is found such that the $\chi 2$ merit function is minimized by steepest descent and quadratic minimization.

3.3.3 Algorithm comparison

To compare the MATLAB and Origin 7.0 nonlinear regression analysis programs, the same simulated data sets were fit with both programs. Each set of simulated data was generated using 25 injections of 5 μl with a ligand concentration of 1 mM and a macromolecule concentration of 40 μM at 298 K. Randomly generated normally distributed noise, 0.1 μcal, was added to the simulated thermograms prior to fitting with either Origin 7.0 or MATLAB routines.

Solutions resulting from the MATLAB algorithms-developed two-binding-site model were compared to solutions obtained using Microcal Origin 7.0 to ensure that the resulting best-fit parameters were in agreement between the two nonlinear regression analysis programs. Simulated ITC thermograms were created for two different cases: two binding sites and three binding sites.

The parameters used to simulate ITC thermograms and the resulting best-fit parameters for the two-binding-site model are given in Table 3.1. The parameters used to simulate ITC test data for the three-binding-site case are given in Table 3.2, in which the best-fit parameters are listed only for the MATLAB program. (Origin 7.0 is limited to solving equilibria for two or fewer reactions although the sequential model in Origin could be transformed and used to simulate thermograms for three independent sites.)

Table 3.1 Parameters used for the generation of two-competitive-processes simulated ITC titration data. (A) Two-competitive-site model: ITC test data simulation parameters. (B) Two competitive- site model: best fit parameters. (C) Case 2, two-competitive-site model: average Monte Carlo parameters and standard deviations..

Parameter	Case 1	Case 2	Case 3
K_1	2.00×10^6	2.00×10^8	2.00×10^8
ΔH_1 (kcal/mol)	−9.000	−12.000	−3.000
n_1	1.000	1.000	1.000
K_2	1.00×10^5	1.00×10^6	1.00×10^6
ΔH_2 (kcal/mol)	−3.000	−5.000	−12.000
n_2	1.000	1.000	1.000

Parameter	Case 1		Case 2		Case 3	
	Origin	MATLAB	Origin	MATLAB	Origin	MATLAB
K_1	2.00×10^6	1.98×10^6	2.02×10^8	1.99×10^8	2.04×10^8	2.02×10^8
ΔH_1 (kcal/mol)	−8.99	−8.94	−12.01	−11.93	−2.99	−2.98
n_1	0.99	1.00	0.99	1.00	0.99	1.00
K_2	0.99×10^5	0.98×10^5	0.94×10^6	0.98×10^6	1.02×10^6	1.00×10^6
ΔH_2 (kcal/mol)	−3.12	−3.09	−5.01	−4.97	−12.03	−11.93
n_2	0.98	0.99	0.99	1.00	0.99	1.00

Parameter	Origin		MATLAB	
	Average	SD	Average	SD
K_1	1.991×10^8	0.068×10^8	1.944×10^8	0.066×10^8
ΔH_1 (kcal/mol)	−12.008	±0.001	−11.950	±0.001
n_1	1.005	±0.001	1.010	±0.001
K_2	0.995×10^6	$\pm 0.003 \times 10^6$	0.990×10^6	$\pm 0.003 \times 10^6$
ΔH_2 (kcal/mol)	−5.004	±0.002	−4.950	±0.002
n_2	1.005	±0.001	1.013	±0.001

Best-fit parameters (K_1, ΔH_1, n_1, K_2, ΔH_2, and n_2 values) are listed for the Origin 7.0 and MATLAB solutions for the nonlinear regression analysis of the three two-competitive processes test cases. The comparison shows that both programs result in the same best-fit parameters. Average best-fit parameters are also listed for 100 Monte Carlo analyses of the Case 2 two-site test data using both algorithms. The standard deviations (SD) are also listed for the six average parameters for both methods.

Table 3.2 Parameters used for the generation of simulated ITC titration data for the three competitive- processes model. (A) Three-competitive-sites model: simulation parameters. (B) Three-binding-sites algorithm: Monte Carlo analysis ‘‘best-fit’’ parameters.

Model parameter	Site 1	Site 2	Site 3
K_i (M^{-1})	1.0×10^8	1.0×10^6	1.0×10^4
ΔH_i (kcal/mol)	−12.0	−8.0	−4.0
n_i	1.0	1.0	1.0

Model parameter	Monte Carlo mean	95% confidence interval	
		Minimum	Maximum
K_1 (M^{-1})	1.06×10^8	1.02×10^8	1.09×10^8
K_2 (M^{-1})	1.06×10^6	1.02×10^6	1.10×10^6
K_3 (M^{-1})	1.11×10^4	1.03×10^4	1.19×10^4
ΔH_1 (kcal/mol)	−12.00	−11.999	−12.001
ΔH_2 (kcal/mol)	−7.99	−7.993	−8.002
ΔH_3 (kcal/mol)	−4.24	−4.147	−4.339
n_1	0.99	0.999	1.000
n_2	0.99	0.999	1.000
n_3	1.08	1.052	1.101

The simulated data were used in the Monte Carlo analysis of the three-competitive processes fitting algorithm. Resulting Monte Carlo mean parameter values and 95% confidence intervals were determined for each of the nine parameters ($K_{1–3}$, $\Delta H_{1–3}$, and $n_{1–3}$) obtained as a solution for the three-competitive-processes model with the MATLAB code given in the Supplementary material. The Monte Carlo analysis demonstrates that all of the three-competitive-processes model parameters are well defined and have acceptable error. A correlation between n and ΔH was observed. The 95% confidence intervals show that the equilibrium constants, molar enthalpy changes, and stoichiometric ratios are well determined for all three processes. The binding affinities for each process are slightly overestimated and the third-process molar enthalpy change, ΔH_3, and stoichiometry, n_3, exhibit the largest uncertainties.

3.3.4 Monte Carlo analysis

Monte Carlo analysis (31,34,39) was used to estimate the uncertainties for each of the nine parameters ($K_{1–3}$, $\Delta H_{1–3}$, $n_{1–3}$) required to fit the three-binding-site model. The Monte Carlo analysis involved the creation of a perfect data set with the MATLAB algorithm. This perfect data set was then used to create 1000 virtual data sets by adding

random normally distributed noise to each data point in the perfect data set. The ITC instrument noise was taken to be ±0.1 μcal per injection (18).

The MATLAB algorithm was then used to fit each of the 1000 virtual ITC experiments and to obtain a set of best-fit parameters for each of the virtual experiments. The mean value for each of the nine fitting parameters was then calculated from the 1000 sets of best-fit values obtained in the Monte Carlo procedure.

3.3.5 Error analysis

Two-dimensional error plots for each of the nine parameters involved in modeling the three binding sites were generated. The error plot for parameter X was established by fixing the value of X and allowing all other parameters to be iterated until a best-fit solution was found. Fixed values for parameter X were chosen such that $\Delta\log(X) = \log(X) - \log(\text{perfect value})$ had a range of -3 to 3. The error plot was produced by plotting $\Delta\log(\text{parameter})$ versus log(error). Transformation of the data using logs condenses the data and exaggerates the local and global minima being observed. These error plots were used to evaluate the interdependence of the parameters and to establish the range of acceptable initial parameter guess values required in modeling three binding sites.

3.4 Results

3.4.1 Algorithm comparisons

ITC experiments were simulated for multiple two-site models. These simulated data (including random noise) were fit using both Origin 7.0 and the MATLAB algorithms. This comparison was done to ensure that our MATLAB algorithms yielded a solution that was in agreement with the solution obtained with Origin 7.0. The best-fit

parameters obtained with both programs, for all of the simulated ITC experiments, are in excellent agreement with one another and returned parameters close to the parameters used to create the simulated ITC data sets.

3.4.2 Monte Carlo analysis

A Monte Carlo analysis was used to estimate the uncertainty in each of the thermodynamic parameters determined in modeling simulated ITC data for a system exhibiting three binding sites. Because modeling a three-process system is not currently implemented in any of the commercially available ITC data analysis programs, a rigorous determination of the uncertainties in the best-fit parameters was conducted in lieu of a direct comparison with other approaches or solutions.

Table 3.2 gives the mean value and 95% confidence interval for each parameter estimated by the Monte Carlo method. Comparison of the mean Monte Carlo parameters with the test data simulation parameters (shown in Table 3.2) demonstrates that the MATLAB algorithm finds a solution minimum in which all nine of the binding parameters are well determined. As seen from Table 3.2, each of the binding constants (K_1, K_2, K_3) has a 95% confidence interval that is slightly skewed to higher values than the values used to simulate the ITC data. Overestimating the binding constants is typical for binding algorithms and can be attributed to the number of data points in the slope of the thermogram and the interdependence of the binding affinity on the other parameters. The molar enthalpy changes and stoichiometric ratios for processes 1 and 2 have 95% confidence intervals that are equally distributed around the noiseless values. This indicates that ΔH_1, n_1, ΔH_2, and n_2 are determined with the greatest accuracy. Because the stoichiometric ratios are well determined for the two processes with the higher affinities

and larger heats, the algorithms correctly identify the solutions for accurate estimates of ΔH_1, n_1, ΔH_2, and n_2. However, the 95% confidence interval for ΔH_3 (and/or n_3) demonstrates that these best-fit parameters contain the largest error of the nine model parameters. ΔH_3 is skewed toward more exothermic values. This error results from the inability of the binding algorithm to determine an accurate endpoint for the third process where the heat signal and integrated heats between successive injections are small.

3.4.3 Error surfaces

A two-dimensional error plot was created for each of the nine parameters required to fit the three-binding-sites model. MATLAB was used to fix the value of the parameter of interest while allowing all other parameters to be optimized by nonlinear regression. The error plots were constructed by plotting $\Delta\log(X)$ versus log(error), where $\Delta\log(X) = \log(X) - \log(\text{noiseless value})$. The error plots demonstrate the interdependence of each parameter, shown by the presence of multiple local minima in several of the parameters. However, the parameters are all found to have global minima at $\Delta\log(X)$ equal to 0. This implies that the algorithm finds the correct minima when the starting parameters are adequately set.

A typical parameter 2-D error plot is shown in Figure 3.2 This figure shows the plot of $\Delta\log(K2)$ versus log(error), where $\Delta\log(K_2) = \log(K_2) - \log(1 \times 10^6)$.

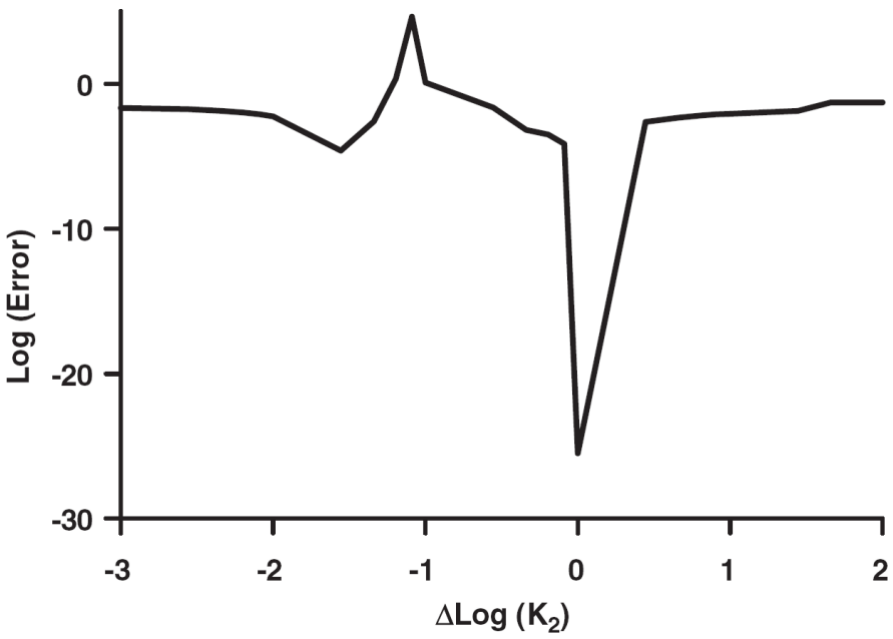


Figure 3.2 Error plot produced by holding the value of K_2 constant while all other three site parameters were optimized using nonlinear regression.

The value of K_2 ranged from 10^1 to 10^8 to cover a large enough range of values (i.e., $-3 < \Delta\log(K_2) < 2$). Error plots were produced for all nine parameters, each showing several local minima and a global minimum at $\Delta\log(X) = 0$. The error plot shown illustrates the interdependence of parameters by the local minimum found at -1.5. Error plots for n_k and ΔH_k show similar local minima due to their interdependence. The error plot demonstrates that appropriate starting positions are necessary to determine an optimal best-fit solution.

As can be seen from the K_2 error plot, at least one local minimum exists that would result in a solution not corresponding to the global minimum. Local minima may be the result of the interdependence of one or more of the fitting parameters. For K_2, the shallow local minimum occurring at binding affinities 2 orders of magnitude larger than the correct value results from the algorithm finding a two-site solution wherein K_2 and K_3 are indistinguishable. This indicates the codependence of the binding affinity values. Hence, it is up to the individual user applying these algorithms to use an appropriate number of binding processes to arrive at an optimal thermodynamic solution that is consistent with the ITC thermograms (i.e., fitting the experimental data within the

expected or observed experimental error). Furthermore, as seen from Figure 3.3, which shows a 3-D error plot for the two parameters ΔH_3 and n_3, it may be necessary for the individual user to enter appropriate starting values for each parameter being evaluated and to verify that the minimum solution found corresponds to a global minimum and represents a plausible solution for the actual chemical equilibria occurring in the system.

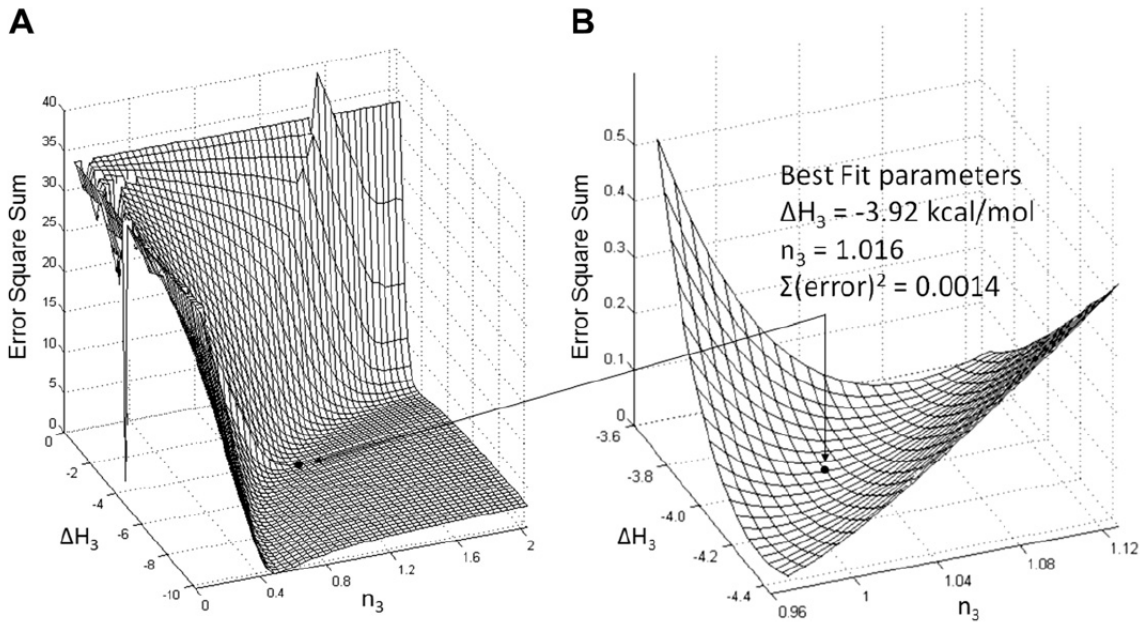


Figure 3.3 Error plot produced by holding the values of n_3 and ΔH_3 constant while all other three-site parameters were optimized using nonlinear regression.

The value of ΔH_3 ranged from 0 to -10 kcal/mol and the value of n_3 ranged from 0 to 2, to include the best-fit values for each of these parameters. (A) The error plot for the whole range of values for the two parameters. (B) The magnified error surface near the best-fit values for the parameters ΔH_3 and n_3 at the global minimum.

3.5 Discussion

3.5.1 Multiple-binding-site models

The theory used for developing the three-process binding model and fitting algorithm can be extended to model multiple-binding-site models of almost any order.

The challenge in modeling such complex systems can be attributed to two factors. The first is the ability to determine the roots of the polynomial (Eq.(3.3)) used for calculating the concentration of free ligand within the reaction vessel. For modeling simpler binding models, including one or two binding sites, the polynomial expands only to a quadratic or cubic equation, respectively. Although closed-form solutions can be obtained for both quadratic and cubic equations, the solution for the cubic equation can be quite cumbersome. For more complex cases, such as three or more binding sites, a closed-form solution for the polynomial does not exist. Software such as MATLAB can provide numerical methods for efficient and accurate means of determining the correct root of Eq. (3.3). It is possible to simplify the determination of roots from Eq. (3.3) based on chemical restrictions. The restriction that the concentration of free ligand in the cell cannot fall below 0 or exceed the injected ligand concentration allows for the root-finding processes to efficiently determine the correct root of the polynomial.

Second, using the smallest correct number of reactions and introducing restrictions to the binding algorithm can greatly increase the efficiency of complex algorithms. Knowing that the thermodynamic model accurately describes the system under study (i.e., includes only chemical equilibria actually observed in the system) allows for the minimum number of fitting parameters to be determined by the nonlinear regression analysis. Similarly, if any restrictions can be placed on the binding algorithm to reduce the number of modeled parameters (e.g., fixing any of the parameters using experimentally determined values), the uncertainty in the best-fit parameters can be dramatically reduced. In particular, restrictions placed on the binding stoichiometry, n, are generally helpful for improving fitting efficiency.

The stoichiometric ratio, *n*, can also greatly influence the enthalpy change, ΔH, of the reaction because these two variables are inversely correlated. That is, if the ΔH of the reaction is increased by an order of magnitude the stoichiometric ratio will decrease proportionally. Caution must be taken when using these complex algorithms to ensure that the proper minimum is being approached. Because of the compensation of stoichiometries and enthalpy changes, it is possible for the algorithm to enter a local minimum at which the enthalpy change becomes increasingly large to compensate for underestimated binding affinities and/or stoichiometric ratios. The relationship between ΔH and n can be seen in Eq. (3.4), which is the only equation in which ΔH appears. Because ΔH enters linearly into the binding models, it is possible to further improve the binding algorithms through linearization of the ΔH parameter.

3.5.2 Monte Carlo evaluation

Monte Carlo evaluation is a valuable technique for determining the statistical uncertainties in best-fit binding parameters. The uncertainties estimated from a Monte Carlo analysis for one possible three-binding-site model in which $\Delta H_1 < \Delta H_2 < \Delta H_3$ is discussed below. The Monte Carlo analysis demonstrated a strong correlation between ΔH and n when determining best-fit parameters. Other combinations of molar enthalpy changes not simulated in this work may result in different parameter uncertainties. However, the largest factor when evaluating uncertainties in the ITC method is the size of the raw heats produced from the reactions occurring in the calorimeter. Because the standard deviation in each injection interval heat, ΔQ_i, is ±0.1 μcal, larger heats are less affected by experimental error in the integrated injection heats. The results given here from the Monte Carlo analysis of the fit parameter uncertainties are expected to be typical

for modeling a system with three binding sites. The parameters with the most uncertainty are typically n_3 and ΔH_3 because these parameters are sensitive to the endpoint of the titration. Thermodynamic parameters obtained from titrations exhibiting a well-defined endpoint (i.e., having a large value for K_n) are well determined with the algorithms developed here. Furthermore, the above algorithms, if extended beyond three binding sites, can be used to evaluate even more complex systems. The Monte Carlo method provides a rigorous test of the reliability of the best-fit parameters obtained in the nonlinear regression modeling of the three-binding-site system. The Monte Carlo method can be used to estimate the uncertainties in parameters obtained in the nonlinear regression modeling of experimental data in the same manner as used here for testing the nonlinear fitting of simulated data, but should incorporate experimentally determined error.

3.5.3 Error analysis

The error plots produced for all parameters involved in modeling three binding sites clearly demonstrated the interdependence of binding parameters (Figure 3.2). Because several local minima are observed for each parameter, stress is placed on the user to accurately enter reasonable starting parameters to avoid getting trapped in a local minimum. It is clear that several parameters have strong interdependence (i.e., K_2 vs K_3, ΔH_3 vs n_3) and must be given starting values that will ensure that the true global minimum is found. The user can check the validity of the solution obtained from nonlinear regression analysis by entering several starting locations for parameters that demonstrate large variability. If the same minimum is reached regardless of starting positions, it is more likely that the solution represents a global minimum. The user must

take care to ensure that the solution both is descriptive of the system and represents a plausible thermodynamic and chemical result. However, it must be remembered that the binding model chosen strongly influences the best-fit parameters obtained in the nonlinear regression analysis of ITC data, and binding parameters will accurately reflect the system only when the correct model is chosen.

3.5.4 Example systems exhibiting three binding sites

Biological systems exhibiting multiple overlaps may be more common than previously thought. Several ITC studies of these more complex systems have been recently reported (29,30,40,41). We present three examples of quadruplex DNA binding either a small ligand or a protein that further illustrate the utility of our MATLAB n-binding sites ITC data analysis program.

ITC data for the titration of a 27-mer c-MYC promoter sequence G-quadruplex construct (having the sequence 50-TGGGGAGGGTGGGGAGGGTGGGGAAGG) at low ionic strength, 20 mM K^+, with the cationic porphyrin ligand TMPyP4 (5, 10, 15, 20-*meso*-tetra-(*N*-methyl-4-pyridyl) porphine) are shown in Figure 3.4. This wild-type (WT) c-MYC promoter sequence quadruplex binds four ligands at saturation. At physiologic ionic strength (150 mM KCl), the four binding sites in the model are represented by two sets of two sites with each set having a characteristic affinity. In effect, two ligands bind to the two high-affinity sites and two ligands bind to the low-affinity sites (40).

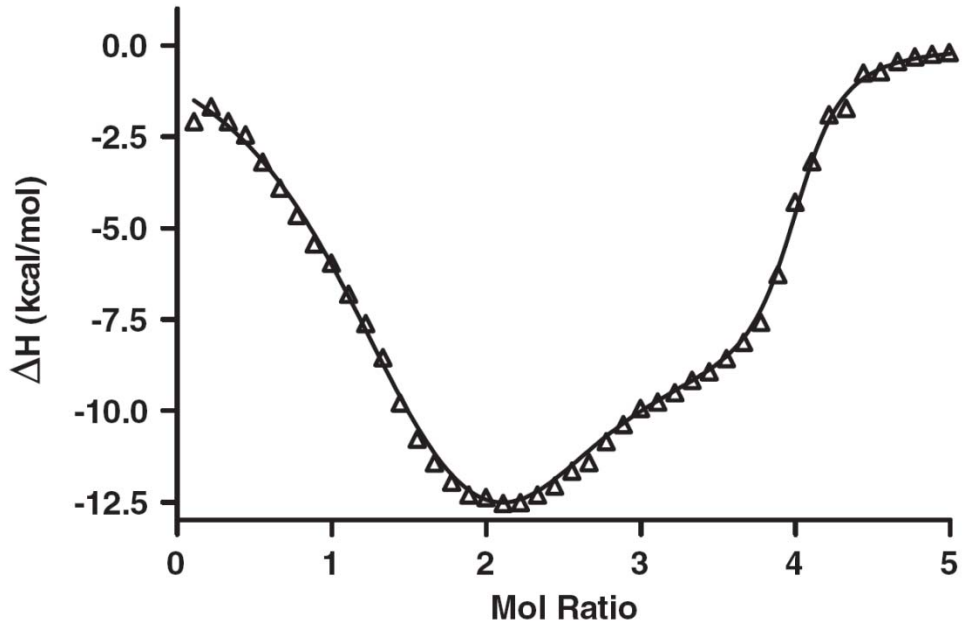


Figure 3.4 ITC data for the reaction of a cationic porphyrin, TMPyP4, with the WT c-MYC 27-mer polypurine P1 promoter sequence G-quadruplex at low ionic strength (20 mM K^+ in BPES buffer at pH 7.0).

The lower ionic strength data must be fit to the three-binding-site model. The best-fit parameters are $K = 3.9 \times 10^8$ M^{-1}, $K_2 = 4.2 \times 10^7$ M^{-1}, $K_3 = 2.9 \times 10^6$ M^{-1}, $\Delta H_1 = -0.2$ kcal/mol, $\Delta H_2 = -12.6$ kcal/mol, $\Delta H_3 = -5.9$ kcal/mol, and $n_1 = 1$, $n = 2$, $n_3 = 1$.

We have assigned the high-affinity sites to end stacking on the two ends of the G-tetrad stack, while the two lower-affinity sites were assigned to the two intercalation sites located between the first and second and the second and third G-tetrads. At low ionic strength (20 mM KCl), a more complicated model is required, having three sites exhibiting three different affinities. We have suggested that at the lower ionic strength the two ends of the G-quadruplex are no longer equivalent and the overhanging single strand is stiffened because of electrostatic repulsion. Based on the three-binding-site fit of the ITC data, the first TMPyP4 binds to one end, the next two ligands bind in the two intercalation sites, and the fourth TMPyP4 binds to the sterically hindered low-affinity

end of the G-quadruplex. Whether this structural model is correct or not, the ITC thermogram observed under low-salt conditions is well fit with the three-binding- site model and cannot be fit with any simpler model.

G-quadruplex ligand studies have typically been limited to single- stranded constructs. To better model the G-quadruplex motif in the presence of its complementary strand we have constructed capped G-quadruplex constructs in which the C-rich strand has been replaced with a shorter partially complementary strand having the ability to form short flanking duplex regions on either side of the quadruplex and with a short non-interaction region of four, five, or six T's bridging the quadruplex in the capped construct. ITC data for the TMPyP4 titration of one such construct are shown in Figure 3.5.

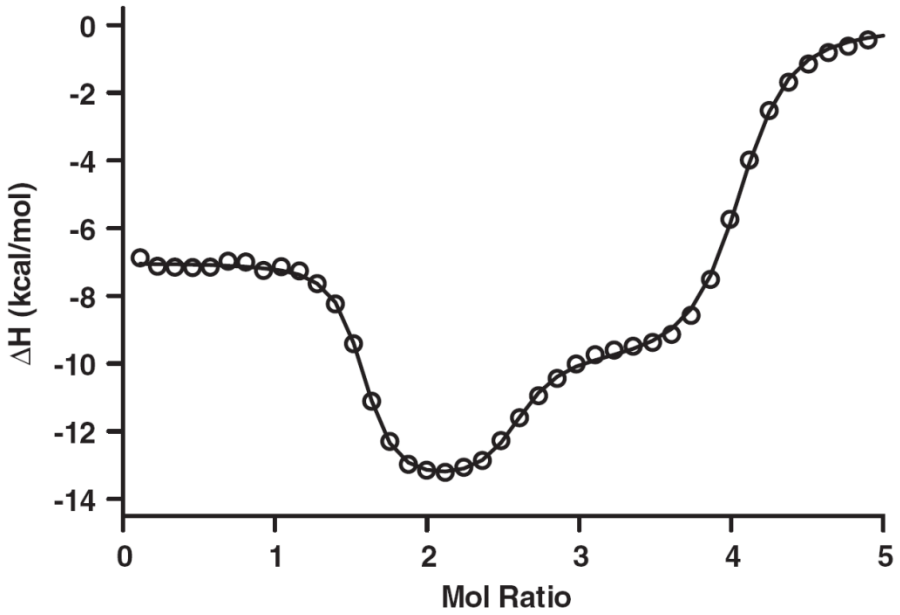


Figure 3.5 ITC data for the reaction of a cationic porphyrin, TMPyP4, with the 1:6:1 c-MYC 32-mer polypurine P1 promoter sequence G-quadruplex capped with a partially complementary 17-mer 5-T capping oligonucleotide in 100 mM K^+/BPES buffer at pH 7.0.

These ITC data have been corrected for ligand-induced folding and dilution heat effects and fit to the three-binding-site model. The best-fit parameters are $K_1 = 7.4 \times 10^9$ M^{-1}, $K_2 = 5.2 \times 10^7$ M^{-1}, $K_3 = 7.0 \times 10^5$ M^{-1}, $\Delta H_1 = -7.0$ kcal/mol, $\Delta H = = -13.7$ kcal/mol, $\Delta H_3 = -9.9$ kcal/mol, and $n_1 = 1.5$, $n_2 = 1$, $n_3 = 1.5$.

The interaction of TMPyP4 with the ‘‘capped’’ quadruplex is more complicated than the interaction with the singled-stranded quadruplex having the same sequence. In effect, the cap results in a small amount of unfolding of the G-quadruplex and in the case of the WT sequence results in a shift in the equilibrium between the 1:6:1 and the 1:2:1 foldamers. The TMPyP4 titration of the five-T capped 1:6:1 mutant c-MYC 32-mer promoter sequence is fit to a three-binding-site model after the data have been corrected for some ligand-induced quadruplex folding and dilution effects. The three binding events must be the result of mixed end binding and intercalation interactions.

We do not currently understand the stoichiometry 1.5:1.0:1.5; however, we are continuing to search for structural models that are consistent with the titration results and the best-fit parameters shown in Figure 3.5. (The ITC data shown here for the titration of the capped quadruplex with TMPyP4 are unpublished and provided here as an example of the need for the analysis of ITC data for systems exhibiting multiple overlapping equilibria.)

Human telomeric DNA contains tandem repeats of the DNA sequence (TTAGGG) and is capable of forming G-quadruplexes. Unfolding of the telomeric G-quadruplex DNA leads to elongation of the telomere by TERT in vitro. UP1 is the proteolytic fragment of the hnRNP A1 protein that is responsible for several gene-regulatory processes at the posttranscriptional level, including mRNA alternative splicing, transportation, and pre-miRNA processing. A crystal structure of UP1 complexed with a partial human telomeric sequence $d(TTAGGG)_2$ revealed that the interaction between UP1 and the telomeric DNA is established through three nucleobases, TAG. The TAG binding motif located at several corners of G-quartets was found to be critical for UP1 binding and initiating unfolding of the G-quadruplex structure. ITC experiments, including both forward and reverse titrations, and DSC experiments were used to probe how UP1 interacts with human telomeric G-quadruplex DNA in Na^+ buffer conditions. ITC data and the three binding sites fit from a forward titration of Tel-22 with UP1 are shown in Figure 3.6. These studies reveal that UP1 binds and unfolds the Tel-22 G-quadruplex DNA at a 2:1 molar ratio. The binding involves loop recognition plus unfolding plus single-strand binding for a second UP1 molecule. This binding, unfolding, binding process is described by the thermochemical cycle shown

in Figure 3.7. (The UP1/Tel-22 data are unpublished and are provided by Dr. David Graves at the University of Alabama at Birmingham, Department of Chemistry.)

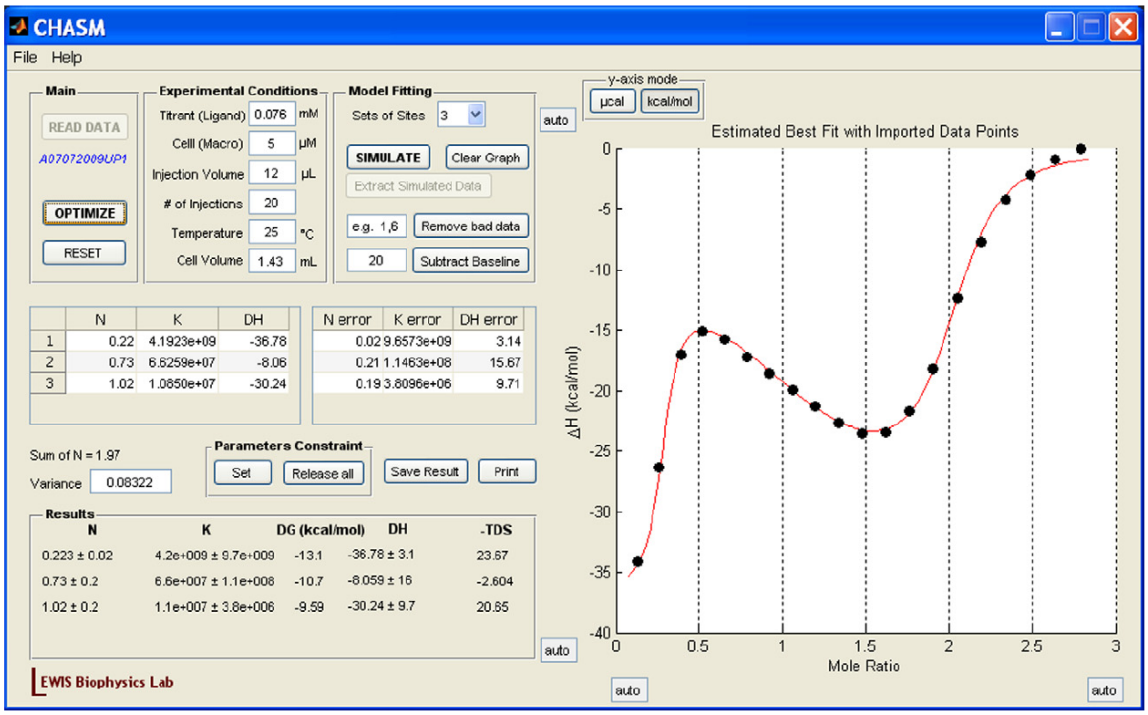


Figure 3.6 Screen capture of the MATLAB user interface showing the three-binding-sites fit of the ITC data obtained for the addition of UP1 to the human telomere quadruplex,Tel-22.

Experimental parameters, [L], [M], calorimeter cell volume, injection volume, number of injections, and temperature, are shown in the upper left quadrant of the screen. Starting parameters n_i, K_i, and ΔH_i are entered in the table shown midleft on the screen capture. Best-fit parameters and the standard deviations are shown at the bottom left of the screen. The right side of the screen shows the ITC data (corrected for heat of dilution) and the best-fit line through the data. The MATLAB program can be downloaded at http://lewis.chemistry.msstate.edu/download.html.

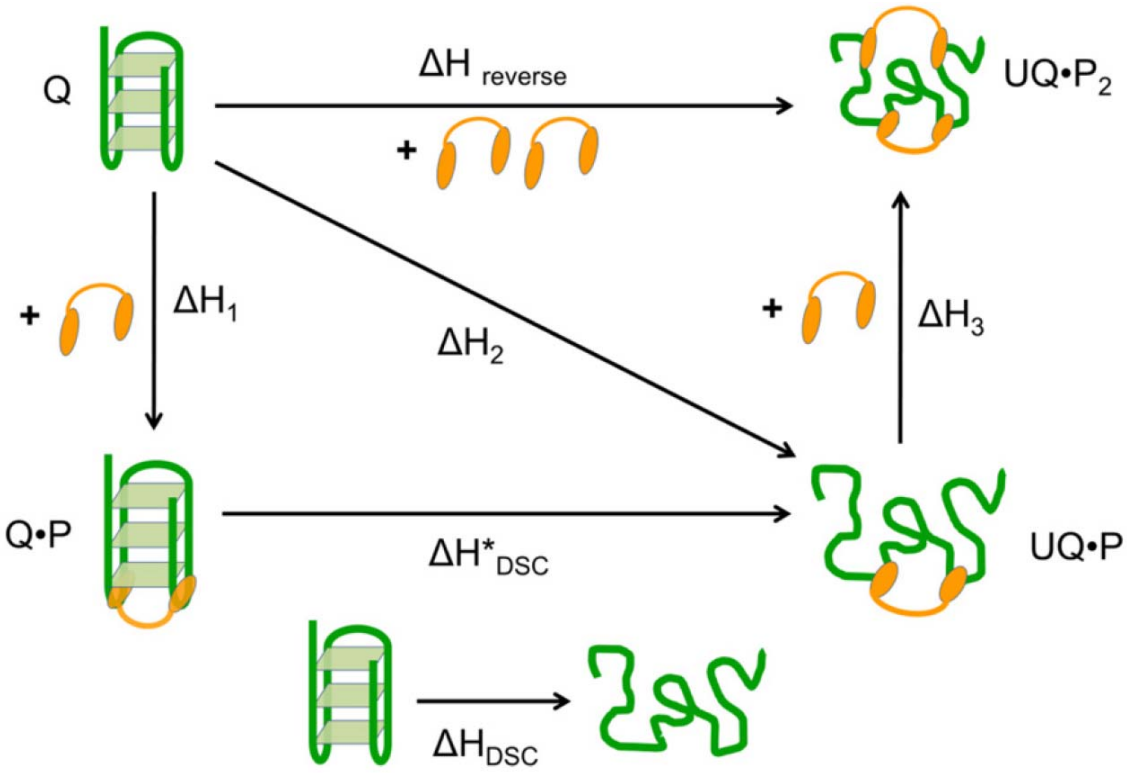


Figure 3.7 Thermochemical cycle (Hess' Law diagram) for the binding of 2 mol of UP1 to the Tel-22 G-quadruplex.

The values of ΔH_1, ΔH_2, and ΔH_3 are obtained directly from the three-binding-sites model fit to the forward titration data for the addition of UP1 to Tel-22. The value for $\Delta H_{reverse}$ is obtained from the ''model-free'' reverse titration. The value of ΔH^*_{DSC} is obtained as the difference between ΔH_1 and ΔH_2 and compared to the ΔH_{DSC} value obtained for the thermal denaturation of the naked Tel-22 G-quadruplex. The value of ΔH_3 is obtained directly from the three-binding-sites model and compared to the difference between ΔH_2 and $\Delta H_{reverse}$.

3.5.5 Further models

The models developed here assume that the only chemical equilibria being observed are reactions involving the binding of a ligand molecule to one or more specific binding sites in a macromolecule (Eq. (3.2)). In these systems, each type of binding site competes with any other binding sites present in the macromolecule and the binding at any site occurs with unitary stoichiometry.

We have shown that a large number of thermodynamic parameters may be accurately estimated from an ITC experiment, e.g., the nine parameters in the three-binding-site model were all well determined. The larger the number of parameters that need to be determined, the more uncertainty will be introduced in parameter determination. There will always be some benefit in designing the experiments or in combining data from complementary studies to limit the number of parameters that are fit. For example the enthalpy change for the strongest binding reaction may be best determined in so-called ‘‘model-free’’ experiments (42), while the thermodynamic parameters for the weakest binding process may be best determined in additional reverse titration experiments in which the low-affinity complexes are more highly populated (30). In conclusion, we would like to state that ITC data can be fit with acceptable accuracy to complex thermodynamic models having as many as three or more overlapping binding equilibria.

3.6 References

1. Hansen, L.D., Christensen, J.J. and M., I.R. (1965) Entropy titration: a calorimetric method for determination of ΔG° (k), ΔH° and ΔS°. *Chemical Communications*, 36-38.

2. Christensen, J.J., Izatt, R.M. and Hansen, L.D. (1966) Entropy titration. A calorimetric method for the determination of DG, DH and DS from a single thermometric titration. *Journal of Physical Chemistry*, **70**, 2003-2010.

3. Christensen, J.J., Wrathall, D.P., Oscarson, J.O. and Izatt, R.M. (1968) Theoretical evaluation of entropy titration method for calorimetric determination of equilibrium constants in aqueous solution. *Analytical Chemistry*, **40**, 1713-1717.

4. Eatough, D.J., Lewis, E.A. and Hansen, L.D. (1985) In Grime, K. (ed.), *Analytical Solution Calorimetry*. Wiley, New York.

5. Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical biochemistry*, **179**, 131-137.

6. Freire, E., Mayorga, O.L. and Straume, M. (1990) Isothermal titration calorimetry. *Analytical Chemistry*, **62**, 950A-959A.

7. Doyle, M.L. (1997) Characterization of binding interactions by isothermal titration calorimetry. *Curr Opin Biotechnol*, **8**, 31-35.

8. Holgate, G.A. (2001) Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics. *BioTechniques*, **30**, 164-166.

9. Ladbury, J.E. (2001) Isothermal titration calorimetry: application to structure-based drug design. *Thermochimica Acta*, **380**, 209-215.

10. Ladbury, J.E. (2004) Application of isothermal titration calorimetry in the biological sciences: things are heating up! *Biotechniques*, **37**, 885-887.

11. Ababou, A. and Ladbury, J.E. (2007) Survey of the year 2005: literature on applications of isothermal titration calorimetry. *J Mol Recognit*, **20**, 4-14.

12. Ladbury, J.E., Klebe, G. and Freire, E. (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. *Nature reviews. Drug discovery*, **9**, 23-27.

13. Ladbury, J.E. (2010) Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochem Soc Trans*, **38**, 888-893.

14. Chaires, J.B. (2008) Calorimetry and thermodynamics in drug design. *Annu Rev Biophys*, **37**, 135-151.

15. Freire, E. (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today*, **13**, 869-874.

16. Freire, E. (2009) A thermodynamic approach to the affinity optimization of drug candidates. *Chemical biology & drug design*, **74**, 468-472.

17. Garbett, N.C. and Chaires, J.B. (2012) Thermodynamic studies for drug design and screening. *Expert opinion on drug discovery*, **7**, 299-314.

18. Freyer, M.W. and Lewis, E.A. (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol*, **84**, 79-7113.

19. Velazquez-Campoy, A. and Freire, E. (2006) Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat Protoc*, **1**, 186-191.

20. Hansen, L.D., Fellingham, G.W. and Russell, D.J. (2011) Simultaneous determination of equilibrium constants and enthalpy changes by titration calorimetry: Methods, instruments, and uncertainties. *Analytical biochemistry*, **409**, 220-229.

21. Lewis, E.A. and Murphy, K.P. (2005) Isothermal titration calorimetry. *Methods Mol Biol*, **305**, 1-16.

22. Freire, E., Schon, A. and Velazquez-Campoy, A. (2009) Isothermal titration calorimetry: general formalism using binding polynomials. *Methods Enzymol*, **455**, 127-155.

23. Velazquez-Campoy, A., Ohtaka, H., Nezami, A., Muzammil, S. and Freire, E. (2004) Isothermal titration calorimetry. *Curr Protoc Cell Biol*, **Chapter 17**.

24. Lewis, E.A. (2009) Calorimetry.

25. Breslauer, K.J., Freire, E. and Straume, M. (1992) Calorimetry: a tool for DNA and ligand-DNA studies. *Methods Enzymol*, **211**, 533-567.

26. Bundle, D.R. and Sigurskjold, B.W. (1994) Determination of accurate thermodynamics of binding by titration microcalorimetry. *Methods Enzymol*, **247**, 288-305.

27. Arnaud, A. and Bouteiller, L. (2004) Isothermal titration calorimetry of supramolecular polymers. *Langmuir*, **20**, 6858-6863.

28. Chaires, J.B. (2006) A thermodynamic signature for drug-DNA binding mode. *Archives of biochemistry and biophysics*, **453**, 26-31.

29. Bou-Abdallah, F., Woodhall, M.R., Velazquez-Campoy, A., Andrews, S.C. and Chasteen, N.D. (2005) Thermodynamic analysis of ferrous ion binding to Escherichia coli ferritin EcFtnA. *Biochemistry*, **44**, 13837-13846.

30. Taneva, S.G., Banuelos, S., Falces, J., Arregi, I., Muga, A., Konarev, P.V., Svergun, D.I., Velazquez-Campoy, A. and Urbaneja, M.A. (2009) A mechanism for histone chaperoning activity of nucleoplasmin: thermodynamic and structural models. *J Mol Biol*, **393**, 448-463.

31. Saroff, H.A. (1989) Evaluation of uncertainties for parameters in binding studies: the sum-of-squares profile and Monte Carlo estimation. *Analytical biochemistry*, **176**, 161-169.

32. Correia, J.J. and Chaires, J.B. (1994) Analysis of drug-DNA binding isotherms: a Monte Carlo approach. *Methods Enzymol*, **240**, 593-614.

33. Tellinghuisen, J. (2005) Statistical error in isothermal titration calorimetry: variance function estimation from generalized least squares. *Analytical biochemistry*, **343**, 106-115.

34. Tellinghuisen, J. (2003) A study of statistical error in isothermal titration calorimetry. *Analytical biochemistry*, **321**, 79-88.

35. Straume, M. and Johnson, M.L. (1992) Analysis of residuals: criteria for determining goodness-of-fit. *Methods Enzymol*, **210**, 87-8105.

36. Horn, J.R., Russell, D., Lewis, E.A. and Murphy, K.P. (2001) Van't Hoff and calorimetric enthalpies from isothermal titration calorimetry: are there significant discrepancies? *Biochemistry*, **40**, 1774-1778.

37. Baker, B.M. and Murphy, K.P. (1996) Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophysical journal*, **71**, 2049-2055.

38. Tellinghuisen, J. (2004) Statistical error in isothermal titration calorimetry. *Methods Enzymol*, **383**, 245-282.

39. Chaires, J.B. (1997) Possible origin of differences between van't Hoff and calorimetric enthalpy estimates. *Biophys Chem*, **64**, 15-23.

40. Freyer, M.W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L.H. and Lewis, E.A. (2007) Biophysical studies of the c-MYC NHE III1 promoter: model quadruplex interactions with a cationic porphyrin. *Biophysical journal*, **92**, 2007-2015.

41. Lewis, E.A., Munde, M., Wang, S., Rettig, M., Le, V., Machha, V. and Wilson, W.D. (2011) Complexity in the binding of minor groove agents: netropsin has two thermodynamically different DNA binding modes at a single site. *Nucleic Acids Res*, **39**, 9649-9658.

42. Chaires, J.B. (2010) Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *FEBS J*, **277**, 1098-1106.

CHAPTER IV

BCL-2 PROMOTER SEQUENCE G-QUADRUPLEX INTERACTIONS WITH THREE PLANAR AND NON-PLANAR CATIONIC PORPHYRINS: TMPYP4, TMPYP3, AND TMPYP2

As previously published in PLOS ONE, 2013, 8(8), e72462:1-9

4.1 Abstract

The interactions of three related cationic porphyrins, TMPyP4, TMPyP3 and TMPyP2 with a WT 39-mer Bcl-2 promoter sequence G-quadruplex were studied using Circular Dichroism, ESI mass spectrometry, Isothermal Titration Calorimetry, and Fluorescence spectroscopy. The planar cationic porphyrin TMPyP4, (5, 10, 15, 20–meso-tetra (N-methyl-4-pyridyl) porphine), is shown to bind to a WT Bcl-2 G-quadruplex via two different binding modes, an *end binding* mode and a weaker mode attributed to *intercalation*. The related non-planar ligands, TMPyP3 and TMPyP2, are shown to bind to the Bcl-2 G-quadruplex by a single mode. ESI mass spectrometry experiments confirmed that the saturation stoichiometry is 4:1 for the TMPyP4 complex and 2:1 for the TMPyP2 and TMPyP3 complexes. ITC experiments determined that the equilibrium constant for formation of the $(TMPyP4)_1$/DNA complex ($K_1 = 3.7 \times 10^6$) is approximately two orders of magnitude greater than the equilibrium constant for the formation of the $(TMPyP2)_1$/DNA complex, ($K_1 = 7.0 \times 10^4$). Porphyrin fluorescence is

consistent with intercalation in the case of the $(TMPyP4)_3$/DNA and $(TMPyP4)_4$/DNA complexes. The non-planar shape of the TMPyP2 and TMPyP3 molecules results in both a reduced affinity for the *end binding* interaction and the elimination of the *intercalation* binding mode.

4.2 Introduction

The promoter region of the B-cell lymphoma-2 (Bcl-2) gene is guanine and cytosine rich (1-3). Bcl-2 expression at low levels is required for cell survival; however, upregulation of the Bcl-2 gene can lead to formation of many common cancers and has been reported to play a role in resistance to conventional cancer treatments (2-4). A 39 base pair region, located from19 to 58 base pairs upstream of the Bcl-2 P1 promoter, plays a crucial role in the regulation of Bcl-2 transcription (5). Dai *et al.* used NMR and CD methods to demonstrate that the Bcl-2 39-mer purine rich strand folds into multiple intramolecular G-quadruplex structures (6).

G-quadruplex structures are thought to be generally involved gene regulation, with G-rich sequences found upstream from as many as 40% of all human genes (7). Small molecules, which specifically interact with quadruplex DNA, have been shown to act as selective inhibitors of telomerase thereby demonstrating some potential as anti-cancer therapeutics (8-11). Cationic porphyrins are known to associate with G-quadruplex DNA. Proposed binding modes include: groove binding (with or without self-stacking along the DNA surface), *end- stacking*, and *intercalation* (12-15). The exact nature of the interactions between cationic porphyrins and G-quadruplex DNA depends on the folding topology and base sequence of the G-quadruplex and on the molecular structure of the porphyrin (16).

Haq *et al.* demonstrated that the saturation stoichiometry for porphyrin binding to G-quadruplex DNA depends on the number, n, of stacked G-tetrads and the formula (n+1) (13). The Lewis group reported that the binding stoichiometry of the cationic porphyrin (5, 10, 15, 20–meso-tetra (N-methyl- 4-pyridyl) porphyrin), TMPyP4, to the c-MYC and Bcl-2 promoter region G-quadruplexes is 4:1 at saturation (17-19). Their microcalorimetric and spectroscopic results were consistent with two TMPyP4 binding modes for both the c-MYC and Bcl-2 promoter quadruplexes; external or *end binding* and *intercalation* (17-19). Wei *et al.* has similarly suggested that TMPyP4 binds to G-quadruplex DNA by a combination of binding modes that include external (end) stacking within the loop region and intercalation between G-tetrads with complex binding ratios of both 2:1 and 4:1 between TMPyP4 and G-quadruplex DNA (20). In addition, Kumar *et al.* has reported that the concentration of porphyrin increases the relative concentration of quadruplex DNA in equilibrium with duplex DNA in dilute solutions (21).

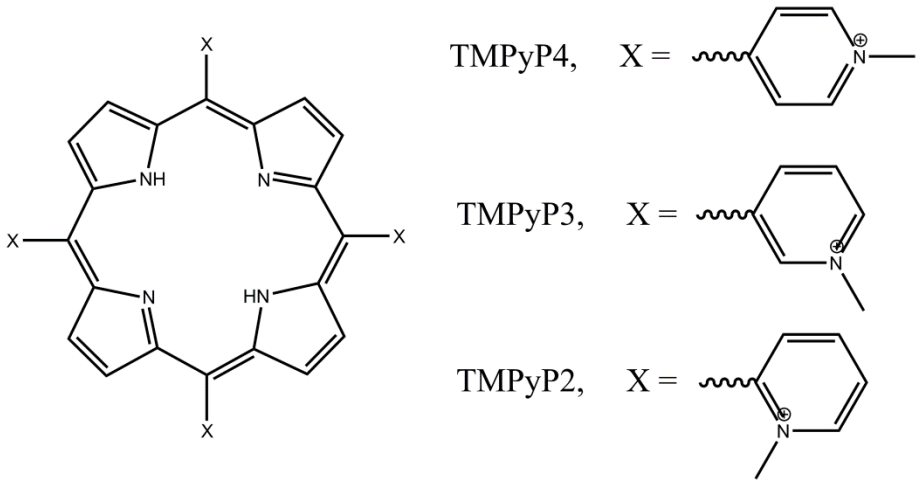
The majority of previous studies have focused on the planar cationic porphyrin TMPyP4, although, Han *et al.* has described the interactions between the related ligands (5, 10, 15, 20–meso-tetra (N-methyl- 2-pyridyl) porphyrin), TMPyP2, and (5, 10, 15, 20–meso-tetra (N-methyl- 3-pyridyl) porphyrin), TMPyP3, with G-quadruplex forming oligonucleotides using gel mobility shift and helicase assays (22). The structures of the three cationic porphyrins differ only in the location of the bulky N^{+}-CH_3 substituent group in the pyridinium rings. Steric hindrances force the four substituent pyridinium rings in the TMPyP2 and TMPyP3 molecules to be out of plane relative to the porphyrin ring. In the present study, we looked at the interaction of three cationic porphyrins (TMPyP2, TMPyP3, and TMPyP4) with a WT 39-mer G-quadruplex forming sequence

from the Bcl-2 promoter region. Two of these ligands are non-planar (TMPyP2 and TMPyP3) and would be too bulky to thread between the stacked G-tetrads of the Bcl-2 promoter sequence G-quadruplex. Absence of a weaker "intercalation" binding mode for these non-planar ligands would serve to support our hypothesis that TMPyP4 binds to the Bcl-2 G-quadruplex (and other G-quadruplexes) by both end stacking and intercalation. The results of this study provide new insight into the origin of porphyrin/G-quadruplex DNA interactions, including the influence of ligand geometry and the influence of the intramolecular DNA folding topology on the thermodynamics for these interactions.

4.3 Material and methods

The WT 39-mer Bcl-2 oligonucleotide sequence used in this study was obtained from Oligos Etc (Wilsonville, OR). The Bcl-2 G-quadruplex forming promoter sequence , 5'-AGGGGCGGGC GCGGGAGGAA GGGGGCGGGA GCGGGGCTG – 3', included six runs of three or more guanines. Bcl-2 stock solutions were prepared by dissolution of weighed amounts of lyophilized oligonucleotide into KBPES {20 mM K_2HPO_4/KH_2PO_4, 2 mM EDTA} buffer with a supporting electrolyte concentration of 130 mM [KCl] and a pH of 7.0 (17). Approximately 1 mL of the oligonucleotide was exhaustively dialyzed (1000 molecular-weight cutoff membrane) with two changes of buffer solution (1 L, 24h each) at 4°C. The concentrations of all DNA solutions were verified using ultraviolet-visible spectrophotometry (UV-Vis). Molar extinction coefficients were determined for each of the four oligonucleotides using a nearest-neighbor method for single stranded DNA (17,23). The extinction coefficient at 260 nm for the 39-mer WT sequence was 3.863×10^5 $M^{-1}cm^{-1}$.

TMPyP2, TMPyP3, and TMPyP4 were obtained from Frontier Scientific (Logan, UT). The chemical structures and naming nomenclature for these porphyrin ligands are given in Figure 4.1.



Abbreviated name	Full chemical name
TMPyP4	meso-tetra-(N-methyl-4-pyridyl) porphine
TMPyP3	meso-tetra-(N-methyl-3-pyridyl) porphine
TMPyP2	meso-tetra-(N-methyl-2-pyridyl) porphine

Figure 4.1 The chemical structures for the planar TMPyP4 and non-planar TMPyP3, and TMPyP2 porphyrin ligands.

All cationic porphyrin solutions were prepared by dissolution of a known amount of porphyrin into a measured volume of the oligonucleotide dialysate solution. This exact matching of the buffer composition in both the titrate (DNA) and titrant (porphyrin) solutions is necessary to minimize any buffer heat of dilution effects in the ITC experiments. The porphyrin solution concentrations were determined from UV-Vis absorbance measurements in the 414-424 nm region and the following molar extinction

coefficients: TMPyP4, ε_{424}=2.26 × 10^5 $M^{-1}cm^{-1}$; TMPyP3, ε_{417}=2.5× 10^5 $M^{-1}cm^{-1}$; and TMPyP2, ε_{414}=1.82 × 10^5 $M^{-1}cm^{-1}$ (13).

CD analysis was performed using an Olis DSM 20 CD Spectrophotometer (Olis, Bogart, GA) equipped with a 1 cm path length quartz cuvette. All CD measurements and titrations were performed at 25ºC in KBPES buffer (pH 7.0) with spectra collected over a wavelength range of 200-500 nm. The Bcl-2 promoter DNA concentration was nominally 5μM in G-quadruplex and a titration involved the addition of the porphyrin ligand to the DNA solution. CD spectra were collected at mole ratios (porphyrin:DNA) of 0:1, 0.5:1, 1:1, 2:1, 3:1, 4:1, 5:1 and 6:1 for each of the three porphyrins and each of the four Bcl-2 oligonucleotides respectively.

ESI-MS experiments were performed using a Bruker MicrOTOFQ mass spectrometer in negative ion mode. Oligonucleotide/porphyrin samples were dissolved in a 50 mM ammonium acetate buffer (pH≈7) containing 20% methanol for these experiments. The WT-39-mer Bcl-2 was prepared at a concentration of approximately 80 μM in the ammonium acetate buffer and was exhaustively dialyzed at 4°C against the 50 mM ammonium acetate {$NH_4^+/CH_3CO_2^-$} buffer. CD spectra for the oligonucleotide in 50 mM ammonium acetate were compared to CD spectra for the same samples in KBPES. The change to an ammonium acetate buffer did not affect the structure of the G-quadruplex, at least as determined by CD.

Stock solutions of the porphyrins were prepared by dissolution of the compound in the final dialysate to obtain a nominal porphyrin concentration of 1.0 mM. The ESI-MS samples were prepared by mixing the DNA and ligand stock solutions to prepare a mixture containing 4 equivalents of the porphyrin per equivalent of DNA. The MS

capillary voltage was set to +4000 V, dry N_2 gas flow was adjusted to 2.0 L/min at 110°C, and the G-quadruplex/porphyrin samples were directly introduced into the MS by using a kD Scientific syringe pump set to a flow rate of 108 μL/hour. Data processing was performed by using Bruker Daltonics Data Analysis program.

Isothermal Titration Calorimetry (ITC) experiments were performed using a Microcal VP-ITC (Microcal, Northampton, MA). All titrations were performed by over filling the ITC cell with ~1.5 mL of stock oligonucleotide solution having a concentration ranging from ~25-50μM. A titration involved the addition of approximately 55x5μL injections of the porphyrin titrant solution. TMPyP2, TMPyP3, and TMPyP4 titrant concentrations were nominally 1×10^{-3} to 3×10^{-3} M to obtain a ligand concentration approximately twenty times greater than oligonucleotide. All titrations were done at 25°C and at a supporting electrolyte concentration of 130 mM [K^+]. Three replicate titration experiments were typically performed. The ITC data were fit using an independent sites model with either one or two independent sites required to fit the data within experimental error. The nonlinear regression fitting was done using CHASM© data analysis program developed in our laboratory (24). The ITC fitting procedures used here have been described previously (17,19,24-26).

Fluorescence emission spectra were collected using a FluoroMax-3 spectrofluorimeter (Horiba Jobin Yvon, Edison, NJ) equipped with a 1 cm path length quartz cuvette. The total concentration of the porphyrin was kept constant throughout the titration (900μM TMPyP2, or TMPyP3, and 1200 μM TMPyP4) while the concentration of the G-quadruplex DNA titrant increased from 0 to 300 μM over the course of the experiment. The TMPyP2, TMPyP3 and TMPyP4 solutions were excited at 424 nm, 417

nm and 433 nm, respectively. Emission spectra were collected from 600 to 800 nm after each injection of DNA into the porphyrin solution. All fluorescence measurements and titrations were performed at 25°C in KBPES buffer (pH 7.0).

4.4 Results

The interactions between the three cationic porphyrins (TMPyP2, TMPyP3, and TMPyP4) and the 39-mer WT Bcl-2 promoter sequence G-quadruplex motifs were probed using CD spectroscopy. The CD spectra for the 39-mer Bcl-2 promoter sequence quadruplex in the presence of saturating amounts TMPyP2, TMPyP3, and TMPyP4 are shown in Figure 4.2.

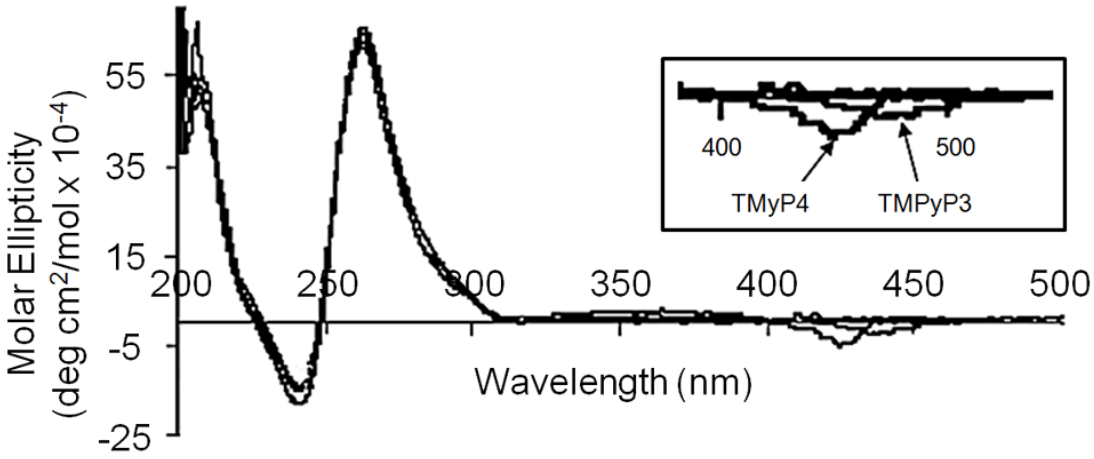


Figure 4.2 CD spectra for complex of the 39-mer Bcl-2 P1promoter G-quadruplex saturated with bound TMPyP4,TMPyP3, or TMPyP2.

The inset shows an expanded view of the CD spectra in the range of the porphyrin absorption spectrum. The very small negative CD signals in the 400-450nm range are only observed in the presence of excess porphyrin for TMPyP4 and TMPyP3.

The Bcl-2 39-mer porphyrin free spectrum is typical for a G-quadruplex structure with a predominantly parallel folding topology. Parallel G-quadruplex motifs typically

exhibit a positive CD band near 264 nm and a negative CD band near 240 nm (1,6,27). The small shoulder at 295 nm is an indication of minor species with anti-parallel folding topology in the 39Bcl-2 quadruplex mixture. The free cationic porphyrins do not possess a CD signal and porphyrin ligands bound to the G-quadruplex do not exhibit a significant induced CD signal. The obvious point to be made is that whether the porphyrin is self-stacking, stacked on the terminal G-tetrads in the quadruplex structure, or intercalated between G-tetrads in the quadruplex, the porphyrin environment in the complex is not asymmetric. The CD spectra for the 39-mer Bcl-2 promoter sequence quadruplex is unchanged by the addition of saturating amounts of TMPyP2, TMPyP3, or TMPyP4 either in CD band intensity or in CD band wavelength. The very low intensity negative bands in the 425 nm region (shown in the inset in Figure 4.2) are the likely result of very weakly bound porphyrin sitting in the quadruplex grooves. Induced CD signals are not seen in any of the TMPyP4 containing solutions, even at porphyrin/DNA ratios in excess of 4:1. The induced CD signals for the non-planar porphyrins only appear at ligand to DNA ratios that greatly exceed the saturation stoichiometry (2:1) for formation of these ligand DNA complexes. Our interpretation of these signals is that the G-quadruplex grooves could provide an asymmetric environment for binding excess porphyrin and that binding to these grooves might result in the observed CD spectrum. In the case of the planar ligand there are no induced CD signals indicating that neither end-stacking nor intercalation yield the required asymmetric environment for the porphyrin to exhibit an induced CD signal.

ESI-mass spectrometry was used to determine the porphyrin/DNA species found in WT 39-mer Bcl-2 G-quadruplex solutions with and without saturating amounts of

porphyrin ligands. The mass spectrum for free DNA is shown in Figure 4.3A, while the mass spectra for $(TMPyP4)_x$/DNA, $(TMPyP3)_x$/DNA and $(TMPyP2)_x$/DNA are shown in Figure 4.3B, 2C, and 2D respectively.

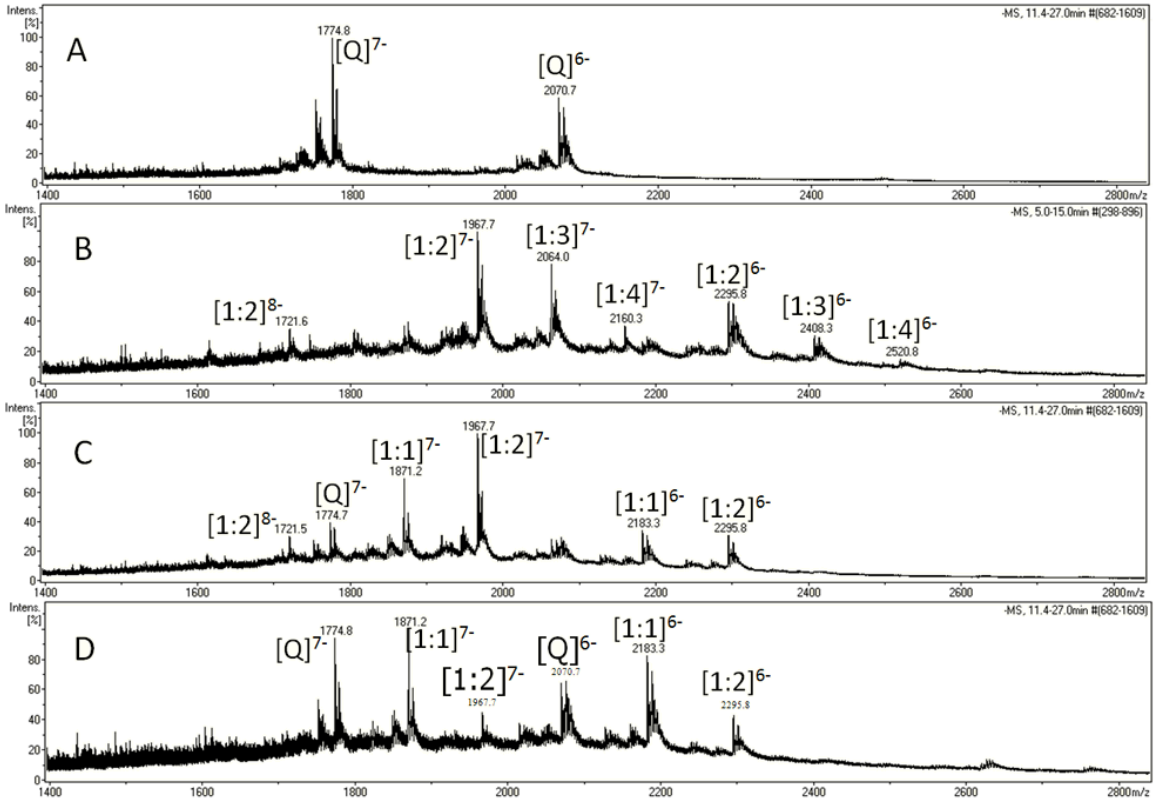


Figure 4.3 Electrospray ionization mass spectra for solutions containing (A) 80 μM WT 39-mer Bcl-2 G-Quadruplex and in complexation with (B) 320 μM TMPyP4or (C) 320 μM TMPyP3 or (D) 320 μM TMPyP2.

The mass spectrum for the TMPyP4/Bcl-2 (4:1) solution shows peaks corresponding to $(TMPyP4)_x$/DNA complex species with x = 1, 2, 3, and 4, and having net charges of -6 to -8. There is no evidence of uncomplexed or free DNA in the TMPyP4 solution. The mass spectrum for the TMPyP3/Bcl-2 (4:1) solution shows peaks

corresponding to uncomplexed DNA, and $(TMPyP3)_x$/DNA complexes with x = 1 or 2 and having net charges of -6 to -8. In addition, there is an m/z peak at 1774.7 corresponding to some free DNA with a net charge of -7. The mass spectrum for the TMPyP2/Bcl-2 (4:1) solution shows peaks corresponding to uncomplexed DNA, and $(TMPyP2)_x$/DNA complexes with x = 1 or 2 and having net charges of -6 and -7. The m/z peaks at 1774.7 and 2070.7 in the mass spectrum for the TMPyP2/Bcl-2 (4:1) solution correspond to free DNA with a net charge of -7 and -6 respectively.

Typical ITC data obtained for the titration of the 39Bcl-2 G-quadruplex DNA with TMPyP2, TMPyP3, or TMPyP4 at 25^{O}C in 100 mM KBPES (pH 7.0) buffer are shown in Figure 4.4.

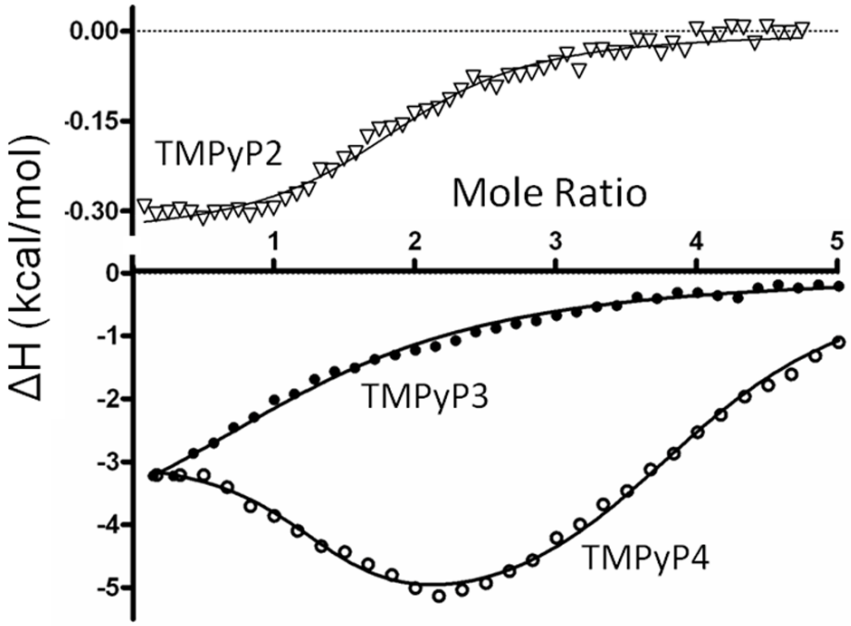


Figure 4.4 Typical ITC titration data (points) and nonlinear regression fits (solid lines) are shown for three titration experiments in which a dilute solution of porphyrin was added to a dilute solution of the annealed oligonucleotide in the calorimeter cell.

The upper panel represents data for the titration of 39-mer Bcl-2 with TMPyP2 (Note the Y-axis is scaled from 0 to .3 kcals/mol). The lower panel represents data for the titration of 39-mer Bcl-2 with TMPyP3 and TMPyP4. The best fit parameters for the nonlinear regression analysis of these ITC data are given in Table 4.1.

Table 4.1 ITC Derived Thermodynamic Parameters for the binding of porphyrin ligands (TMPyP2, TMPyP3, and TMPyP4) to the 39Bcl-2 promoter sequence G-Quadruplex.

Ligand	$K_1 \times 10^{-5}$	ΔG_1 (kcal/mol)	ΔH_1 (kcal/mol)	$-T\Delta S_1$ (kcal/mol)	$K_2 \times 10^{-5}$	ΔG_2 (kcal/mol)	ΔH_2 (kcal/mol)	$-T\Delta S_2$ (kcal/mol)
TMPyP2	0.7 ± 0.5	-6.6	-0.6 ± 0.2	-6.0	-	-	-	-
TMPyP3	2.6 ± 0.5	-7.4	-4.1 ± 0.1	-3.3	-	-	-	-
TMPyP4	37.0 ± 4.0	-9.0	-4.2 ± 0.1	-4.7	0.9 ± 0.5	-6.7	-9.8 ± 0.1	3.1

The data points shown in each of the three representative thermograms are the corrected integrated heat produced per injection, plotted with respect to the mole ratio of porphyrin/DNA at each point in the titration. It needs to be noted that since the TMPyP2

affinity and enthalpy change are much smaller than for the other porphyrins, the TMPyP2 data in Figure 4.4 are shown on a magnified (10X) scale. After making corrections for blank heat effects and titrant dilution, the corrected heat data were fit to a thermodynamic model using a non-linear regression algorithm to obtain *best-fit* values for the equilibrium constant(s), K_i, enthalpy change(s), ΔH_i, and reaction stoichiometry, n, for complex formation. The TMPyP2 and TMPyP3 thermograms were fit with a model having a single binding mode, with two equivalent binding sites per DNA, and a saturation stoichiometry of 2:1. The TMPyP4 thermograms required a more complicated model having two binding modes, with each binding mode having two equivalent sites, and a saturation stoichiometry of 4:1. The *best-fit* values of the thermodynamic parameters estimated for the cationic porphyrin interactions with the WT 39-mer Bcl2 G-quadruplex DNA are listed in Table 4.1. Parameter values listed for the interactions of TMPyP4 with the WT 39-mer Bcl-2 promoter sequence G-quadruplex are very similar to the K_i, ΔG_i, ΔH_i, and $-T\Delta S_i$ values reported previously for the interaction of TMPyP4 with the shorter c-MYC and Bcl-2 promoter G-quadruplexes (17,28).

The interactions between the three cationic porphyrins and the WT 39-mer Bcl-2 promoter sequence G-quadruplex motif were also probed using fluorescence techniques. Emission spectra for the porphyrin ligands were recorded in DNA free solutions and in the presence of varying amounts of Bcl-2 promoter sequence quadruplex DNA. The fluorescence emission spectra observed for TMPyP4 in a G-quadruplex DNA free solution and in solutions containing either 2 moles of TMPyP4/mole of DNA or 4 moles of TMPyP4/mole of DNA are shown in Figure 4.5.

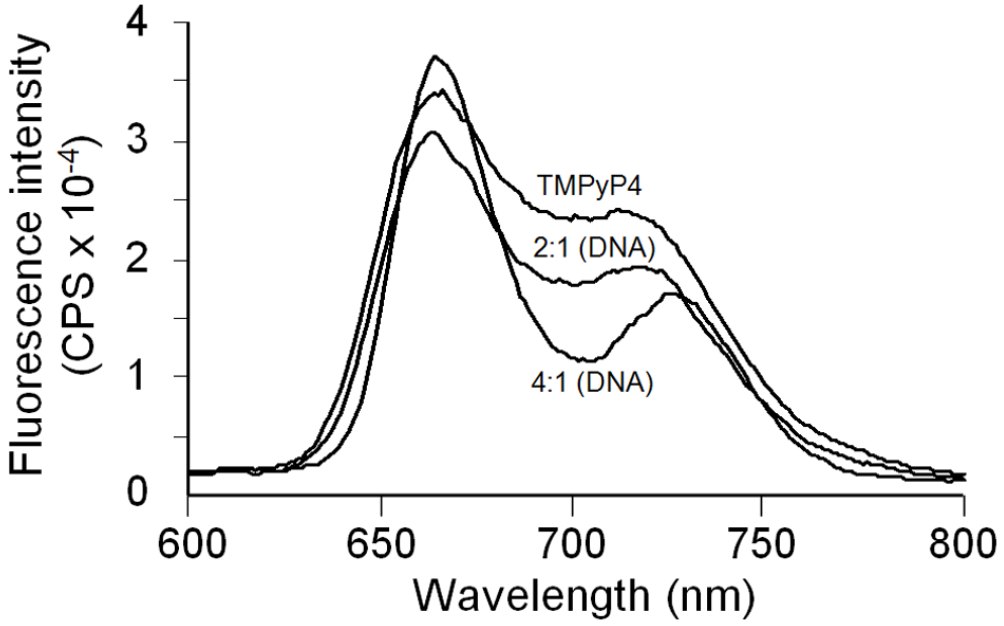


Figure 4.5 Fluorescence emission spectra obtained for theTMPyP4 in the presence or absence of 39-mer Bcl-2 G-quadruplex DNA.

Spectrum 1 labeled TMPyP4 is for the free porphyrin (in the absence of DNA). Spectrum 2 is for TMPyP4in the (TMPyP4)2•39-mer Bcl-2 complex. Spectrum 3 is forTMPyP4 in the (TMPyP4)4•39-mer Bcl-2 complex.

In dilute solution, each of the three porphyrins exhibit complex emission spectra. TMPyP2 exhibits an emission spectrum with emission maxima at 640, 660, and 705 nm, while TMPyP3 exhibits emission maxima at 660 and 705 nm, and TMPyP4 exhibits emission maxima at 660 and in the range of 720 to 730 nm. In all instances, the total fluorescence emission is quenched in the presence of quadruplex DNA. This is a clear indication of porphyrin binding to G-quadruplex DNA. It is obvious, from the fluorescence spectra shown in Figure 4.5, that TMPyP4 fluorescence is changed on formation of the TMPyP4/39Bcl-2 complex. At low mole ratios of TMPyP4 to DNA (e.g. ratio ≤ 2.0), the TMPyP4 fluorescence is attenuated across the entire wavelength range of 625 to 775 nm. At higher mole ratios (e.g. TMPyP4/DNA $\approx$ 4:1), the TMPyP4

fluorescence emission spectrum changes dramatically, exhibiting a significant increase in fluorescence in the 660 nm region, a nonlinear attenuation near 700 nm, and a red shift in the λ_{max} for the 730 nm peak in the emission spectrum. Although in essence we are only showing two points in a DNA titration of TMPyP4 (Figure 4.5), the spectra shown represent the end points for formation of the 2:1 complex (with only end binding) and for the formation of the 4:1 complex with two different binding modes evident.

Changes in the emission spectra for TMPyP2 and TMPyP3 (data not shown) are linear with increasing DNA up to the point of saturation (2:1) porphyrin to DNA. The TMPyP2 and TMPyP3 fluorescence is quenched proportionally over the emission wavelength range (600 to 800 nm) by the DNA interaction. The results are consistent with bound TMPyP4 occupying two different environments in the saturated (4:1) TMPyP4/Bcl-2 promoter sequence quadruplexes while bound TMPyP2 and TMPyP3 appear to occupy a single environment in the saturated (2:1) complexes. The fluorescence experiments reported here for the interactions of all three porphyrin ligands with the WT 39-mer Bcl-2 promoter sequence agree in general with the CD, UV-Vis and ITC results reported previously for interaction of TMPyP4 with shorter c-MYC and Bcl-2 promoter sequence quadruplexes (17,19,28).

4.5 Discussion

The WT 39-mer Bcl-2 oligonucleotide used in this study contains six contiguous runs having three or more consecutive guanines. The WT 39nt sequence has one run with five consecutive guanines, two runs containing four consecutive guanines, and three runs having three consecutive guanines. The Hurley group has reported that the WT 39-mer Bcl-2 P1 promoter sequence forms a mixture of three distinct intramolecular G-

quadruplex structures in the presence of K^+ ions (1,6). Numerous others have shown that cationic porphyrins can bind to G-quadruplex DNA and stabilize the DNA structure through π-π stacking interactions between the aromatic porphyrin core and a G-tetrad or through electrostatic interactions between positively charged nitrogen atoms of the pyridyl rings and negatively charged phosphate oxygen atoms of DNA (1,13,14,22,29-31).

The CD results for the WT 39-mer Bcl-2 sequence used in this study agree with previously published data obtained for shorter Bcl-2 and c-MYC promoter sequence quadruplexes (17,19,28). Specifically, each of the Bcl-2 promoter sequence G-quadruplexes (i.e. 23-mer, 27-mer, 30-mer and 39-mer) are best described as predominantly parallel structures exhibiting a large positive molar ellipticity in the 260 nm region with a smaller percentage of anti-parallel character (indicated by a small shoulder in the 295 nm region). In addition, the CD spectra for the Bcl-2 promoter quadruplexes is unchanged by the addition of the three porphyrin ligands up to saturation (2:1 for TMPyP2 or TMPyP3, and 4:1 for TMPyP4) and there is little or no evidence of an induced CD signal for the bound porphyrin (See Figure 4.2). It appears that bound porphyrin, either *end-stacked* or *intercalated*, has almost no effect on the CD spectra for the G-quadruplex structure or on the equilibrium between the ensemble of Bcl-2 G-quadruplex species.

The ESI-MS results provide spectroscopic evidence for the saturation stoichiometry of the ligand DNA binding. There is no detectable peak for free DNA in the TMPyP4/Bcl-2 mass spectrum. This is the obvious result of the large equilibrium constant for formation of the $(TMPyP4)_x$/DNA complex species ($K_1 \approx 4\times10^6$ and $K_2 \approx$

7×10^4). There are however, detectable peaks for free DNA in the mass spectra for both the $(TMPyP3)_x$/Bcl-2 and $(TMPyP2)_x$/Bcl-2 solutions. The TMPyP4/Bcl-2 spectrum (Figure 4.3B) clearly exhibits m/z peaks corresponding to TMPyP4/DNA species having 2:1, 3:1, and 4:1 complex stoichiometry. To the best of our knowledge, this is the first ESI-MS evidence for a G-quadruplex species having four porphyrin ligands bound to an intramolecular G-quadruplex. To demonstrate that observation of G-Quadruplex species and formation of porphyrin/G4 complexes are independent of the G4 length, we provide ESI mass spectra obtained in similar experiments performed on a shorter 27-mer Bcl-2 G-Quadruplex construct (see Figure 4.6).

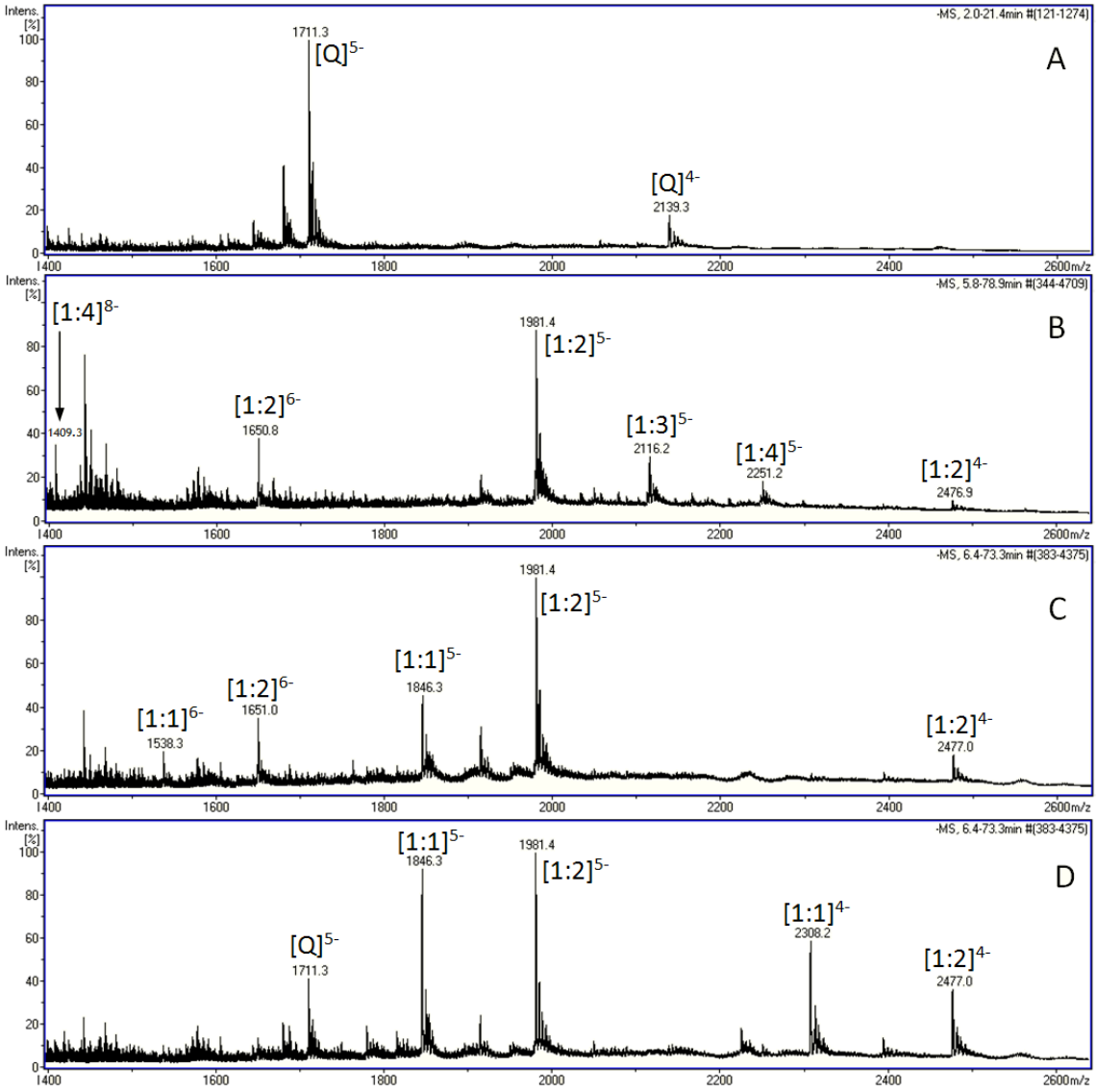


Figure 4.6 Electrospray ionization mass spectra for solutions containing (A) 80 µM WT 27-mer Bcl-2 G-Quadruplex and in complexation with (B) 320 µM TMPyP4 or (C) 320 µM TMPyP3 or (D) 320 µM TMPyP2.

The 39-mer Bcl-2 and 27-mer Bcl-2 ESI-MS results taken together suggest that the saturation stoichiometry of the G-quadruplex DNA/porphyrin ligand complexes is independent of the oligonucleotide length, eliminating the possibility of non-specific binding of porphyrin molecules to either the 5' or the 3' single stranded ends.

The ITC results present a consistent picture regarding the binding of the three porphyrin ligands to the WT 39-mer Bcl-2 promoter sequence quadruplex. We had previously reported on the site and mode of TMPyP4 interactions with the WT 27-mer Bcl-2 quadruplex and two mutant 27nt sequences designed to fold into 3:7:1 and 3:5:3 loop isomer quadruplexes (19). In this previous work, we found that both mutant sequence quadruplexes exhibited lower affinity for TMPyP4 than the WT sequence and that the isomer with the largest (7 Base) unstructured loop showed an approximate 40-fold decrease in TMPyP4 binding affinity. The energetic profile reported here for the interaction of TMPyP4 with the WT 39-mer Bcl-2 sequence is similar to the energetic profiles reported previously for TMPyP4 binding to the WT Bcl-2 27mer, Bcl-2 (3:7:1) 23-mer loop isomer, and the Bcl-2 3:5:3 23mer loop isomer G-quadruplexes (19). All of these quadruplex motifs exhibited two binding modes for TMPyP4 and a saturation stoichiometry of 4:1 in the ITC titration experiments. The higher affinity binding mode combines a favorable enthalpy contribution (-1.8 to -4.6 kcal/mol) with a more significant favorable entropy contribution of -4 to -9 kcal/mole in terms of (-TΔS). The lower affinity mode combines a more significant enthalpy contribution (-6 to -12 kcal/mol) with a smaller entropy term ranging from -3 to +5 kcal/mol. The energetic profiles reported here for the interactions of TMPyP3 and TMPyP2 exhibit only one binding mode that is most similar to the higher affinity TMPyP4 binding mode, at least in terms of the relative entropy and enthalpy contributions to the overall binding free energy change. Mode 1 binding affinity of the non-planar porphyrin molecules is reduced by a factor of 10 for TMPyP3 and more than 60 for TMPyP2 in comparison to TMPyP4 external binding or

end stacking. Neither TMPyP3 nor TMPyP2 show Mode 2 binding, presumably the result of the non-planar molecules inability to *intercalate*.

We speculate here that the entropy-driven energetics for Mode 1 binding, typical for the binding of a hydrophobic ligand to the exterior of duplex DNA, can be used as a model for the exterior or end binding interactions of the TMPyP4, TMPyP3, and TMPyP2 ligands to the top or bottom of the G-quadruplex structure. The relative mode 1 affinities, TMPyP4 > TMPyP3 >> TMPyP2 are the result of the non-planar TMPyP3 and TMPyP2 molecules being unable to stack effectively on the flat or planar terminal G-tetrads. The energetic profile for the second TMPyP4 binding mode is best described as an enthalpy driven process. In the case of the insertion of planar aromatic ligands between the stacked bases in duplex DNA, the typical exothermic enthalpy change is largely the result of increased π-π stacking interactions between the DNA bases and the intercalated aromatic ligand (32-36). We speculate here that the enthalpy driven energetics for Mode 2 binding, typical for the intercalation of a planar aromatic ligand into duplex DNA, can be used as a model for the stacking of the TMPyP4 ligand between two G-tetrads in the G-quadruplex structure. One subtle difference would be that in duplex intercalation there are additional consequences of intercalation, i.e. some unwinding of the DNA duplex.

Although the intercalation binding mode remains hypothetical, there continue to be reports that intercalation of planar ligands, like TMPyP4, can occur into G-quadruplex DNA structures (37-41). Wei *et al*. reported a change in the G-quadruplex DNA conformation on intercalation of TMPyP4 using spectroscopic methods (20). Scheidt *et al*. has shown by X-ray crystallography that the four 4-N-methyl pyridinium rings in a

nitrosyl-Co II- TMPyP4 complex are nearly perpendicular to the plane of the core porphyrin ring system (42). The interaction of TMPyP4 with G-quadruplex DNA in the $(TMPyP4)_4$/DNA complex are maximized by rotation of the pyridinium rings to attain a planar TMPyP4 geometry favoring both end stacking on the terminal G-tetrads and intercalation between stacked G-tetrads in the G-quadruplex structure. The small change in G-quadruplex structure observed by Wei (43) can be attributed to the separation of two stacked G-tetrads required for intercalation to occur.

Gavathiotis *et al.* reported an NMR structure (PDB accession code: 1NZM) having a spacing of 6.9 Å between a G- tetrad and a plane of four Adenines in a parallel stranded DNA quadruplex $d(TTAGGGT)_4$ containing the human telomere repeat sequence and intercalated quinoacridinium ligands (37). Keating and Szalai reported EPR data consistent with the intercalation of CuTMPyP4 into a $d(T_4G_8T_4)_4$ G-quadruplex (38). Hounsou *et al.* reported an NMR structure (PDB accession code: 2JWQ) having a spacing of 5.8 Å between a G- tetrad and a plane of four Adenines, with quinacridine based ligands (MMQs) intercalated between two G and A planes (39). Cavallari *et al.* have reported on an MD simulation study in which they found that TMPyP4 can stack with G-tetrads in the absence of interplane cations (41). They report TMPyP4 G-tetrad stacking distances of 4.3-4.7 Å and that TMPyP4 intercalation depends on the length of the quadruplex, the stoichiometric ratio, and the edge termination motif. Although none of these previous reports directly demonstrates TMPyP4 intercalation between G-tetrads in an intramolecular G-quadruplex, we believe that intercalation of the more weakly bound porphyrin ligands remains a viable model for the $(TMPyP4)_4$/DNA complex at this time.

The TMPyP2, TMPyP3, and TMPyP4 total fluorescence is attenuated on binding to the WT 39-mer Bcl-2 promoter sequence quadruplex. The attenuation of the porphyrin fluorescence on binding exhibits an irregular trend in that the percent fluorescence hypochromicity is largest for TMPyP3, followed by TMPyP4, and then TMPyP2. Porphyrin bound to the WT 39-mer Bcl-2 exhibits hypochromicity in total fluorescence of 53%, 29% and 13% for TMPyP3, TMPyP4, and TMPyP2 respectively at saturation (either 2:1 for TMPyP2 and TMPyP3 or 4:1 for TMPyP4). Whereas TMPyP2 and TMPyP3 exhibit attenuation of their emission spectra over the wavelength range of 600 to 800 nm when bound to quadruplex DNA, TMPyP4 fluorescence changes in a more complex manner on binding to the Bcl-2 quadruplexes. The decrease in fluorescence intensity for either TMPyP2 or TMPyP3 and for TMPyP4, at mole ratios of less than 2:1, is due to binding of the porphyrin externally or to the end of the G-tetrad core structure. The externally bound porphyrin remains exposed to solvent and its fluorescence quenched by water surrounding the externally bound ligand. The increase in the TMPyP4 fluorescence emission intensity, at mole ratios between 2:1 and 4:1 ligand/DNA, results from TMPyP4 intercalating between two G-tetrads resulting in energy transfer between the intercalated porphyrin and the eight guanines located in the two G-tetrads stacked on the TMPyP4. The same argument can be used to explain the apparent lower hypochromicity of bound TMPyP4 in comparison to bound TMPyP3 at low mole ratios (< 2:1). In effect, the increase in fluorescence emission for even a small amount of intercalated TMPyP4 (at low mole ratios) partially compensates for the attenuation of the externally bound TMPyP4 resulting in TMPyP4 exhibiting decreased hypochromicity in comparison to TMPyP3.

The results of the spectroscopic and microcalorimetric experiments described here clearly demonstrate that the planar cationic porphyrin, TMPyP4, binds with high affinity to the WT 39-mer Bcl-2 promoter sequence G-quadruplex by two distinctly different binding modes. We speculate that the two binding modes are an external or *end stacking* mode and an *intercalation* mode. The *intercalation* mode is eliminated when the bulky N^{+}-CH_3 pyridinium substituent groups (in TMPyP3 and TMPyP2) force the ligand to become non-planar, requiring a greater G-tetrad spacing for intercalation to occur. The fluorescence emission spectra can only be explained by the placement of the four bound TMPyP4 ligands in two completely different chemical environments in the saturated complex. The non-planar structure of the TMPyP2 and TMPyP3 ligands results not only in the loss of mode 2 binding but also in decreased affinity for the *end binding* interaction. The TMPyP2 ligand can only weakly stack on the terminal G-tetrads, while the TMPyP3 molecule can apparently partially rotate the pyridinium rings out of the way to gain a better π-π overlap between the aromatic porphyrin core and the terminal G-tetrads. The favorable binding interaction, that is inevitably lost in both the TMPyP2 and TMPyP3 quadruplex interactions, is any loop or terminal base sequence overlap with the end-stacked ligand.

The binding of the two non-planar porphyrins (TMPyP3 and TMPyP2) to the 39-mer Bcl-2 promoter sequence G-quadruplex is clearly different from the binding of the planar TMPyP4 ligand to the same DNA construct. The Structure Activity Relationships in these porphyrin binding interactions are the result of differences in the ligands solvent accessible surface area (SASA), charge location, and planarity. The weaker interactions for the non-planar ligands were anticipated in that all of these factors are working against

the binding of the TMPyP2 and TMPyP3 molecules. Both TMPyP2 and TMPyP3 have somewhat smaller SASA values, both have reduced favorable charge/charge interactions as the positively charged pyridinium groups are moved further away from the negative charges in the DNA backbone in these complexes, and both have weaker π-π stacking interactions with the planar receptor G-tetrad. The non-planar ligands simply cannot stack as closely to an end tetrad or even fit between two stacked tetrads to intercalate. Even though these structure activity effects could have been predicted, their measurement removes any ambiguity relative to the placement of the bound porphyrins and is consistent with the two binding modes observed for TMPyP4, end-stacking and intercalation. The significance of this study is that it may be possible to design drug molecules that specifically target either the *end-stacking* site or the *intercalation* site in G-quadruplex motifs. If our binding model is correct, there are two unique opportunities for the future development of G-quadruplex structure-selective drugs.

4.6 References

1. Dexheimer, T.S., Sun, D. and Hurley, L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J Am Chem Soc*, **128**, 5404-5415.

2. Chao, D.T. and Korsmeyer, S.J. (1998) BCL-2 family: regulators of cell death. *Annu Rev Immunol*, **16**, 395-419.

3. Adams, J.M. and Cory, S. (1998) The Bcl-2 protein family: arbiters of cell survival. *Science*, **281**, 1322-1326.

4. Baretton, G.B., Diebold, J., Christoforis, G., Vogt, M., Muller, C., Dopfer, K., Schneiderbanger, K., Schmidt, M. and Lohrs, U. (1996) Apoptosis and immunohistochemical bcl-2 expression in colorectal adenomas and carcinomas. Aspects of carcinogenesis and prognostic significance. *Cancer*, **77**, 255-264.

5. Heckman, C., Mochon, E., Arcinas, M. and Boxer, L.M. (1997) The WT1 protein is a negative regulator of the normal bcl-2 allele in t(14;18) lymphomas. *J Biol Chem*, **272**, 19609-19614.

6. Dai, J., Dexheimer, T.S., Chen, D., Carver, M., Ambrus, A., Jones, R.A. and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J Am Chem Soc*, **128**, 1096-1098.

7. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, **35**, 406-413.

8. Dixon, I.M., Lopez, F., Esteve, J.P., Tejera, A.M., Blasco, M.A., Pratviel, G. and Meunier, B. (2005) Porphyrin derivatives for telomere binding and telomerase inhibition. *Chembiochem*, **6**, 123-132.

9. Izbicka, E., Wheelhouse, R.T., Raymond, E., Davidson, K.K., Lawrence, R.A., Sun, D., Windle, B.E., Hurley, L.H. and Von Hoff, D.D. (1999) Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res*, **59**, 639-644.

10. Clark, G.R., Pytel, P.D., Squire, C.J. and Neidle, S. (2003) Structure of the first parallel DNA quadruplex-drug complex. *J Am Chem Soc*, **125**, 4066-4067.

11. Haider, S.M., Parkinson, G.N. and Neidle, S. (2003) Structure of a G-quadruplex-ligand complex. *J Mol Biol*, **326**, 117-125.

12. Vialas, C., Pratviel, G. and Meunier, B. (2000) Oxidative damage generated by an oxo-metalloporphyrin onto the human telomeric sequence. *Biochemistry*, **39**, 9514-9522.

13. Haq. I., J.O.T., B. Z. Chowdhry, and T. C. Jenkins. . (1999) Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin. *J Am Chem Soc*, **121**, 1768-1779.

14. Pasternack, R.F., Gibbs, E.J. and Villafranca, J.J. (1983) Interactions of porphyrins with nucleic acids. *Biochemistry*, **22**, 2406-2414.

15. Anantha, N.V., Azam, M. and Sheardy, R.D. (1998) Porphyrin binding to quadrupled T4G4. *Biochemistry*, **37**, 2709-2714.

16. Carvlin, M.J., Mark, E., Fiel, R. and Howard, J.C. (1983) Intercalative and nonintercalative binding of large cationic porphyrin ligands to polynucleotides. *Nucleic Acids Res*, **11**, 6141-6154.

17. Freyer, M.W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L.H. and Lewis, E.A. (2007) Biophysical studies of the c-MYC NHE III1 promoter: model quadruplex interactions with a cationic porphyrin. *Biophys J*, **92**, 2007-2015.

18. Cashman, D.J., Buscaglia, R., Freyer, M.W., Dettler, J., Hurley, L.H. and Lewis, E.A. (2008) Molecular modeling and biophysical analysis of the c-MYC NHE-III1 silencer element. *J Mol Model*, **14**, 93-101.

19. Nagesh, N., Buscaglia, R., Dettler, J.M. and Lewis, E.A. (2010) Studies on the site and mode of TMPyP4 interactions with Bcl-2 promoter sequence G-Quadruplexes. *Biophys J*, **98**, 2628-2633.

20. Wei, C., Jia, G., Yuan, J., Feng, Z. and Li, C. (2006) A spectroscopic study on the interactions of porphyrin with G-quadruplex DNAs. *Biochemistry*, **45**, 6681-6691.

21. Kumar, N. and Maiti, S. (2005) The effect of osmolytes and small molecule on Quadruplex-WC duplex equilibrium: a fluorescence resonance energy transfer study. *Nucleic Acids Res*, **33**, 6723-6732.

22. Han, H., Langley, D.R., Rangan, A. and Hurley, L.H. (2001) Selective interactions of cationic porphyrins with G-quadruplex structures. *J Am Chem Soc*, **123**, 8902-8913.

23. Oligonucleotide MW and Extinction Coefficient Calculator. 2011 [cited03/01/2011]; Available from: http://www.ambion.com/techlib/misc/oligo_calculator.html.

24. Freyer, M.W., Buscaglia, R., Nguyen, B., Wilson, W.D. and Lewis, E.A. (2006) Binding of netropsin and 4,6-diamidino-2-phenylindole to an A2T2 DNA hairpin: a comparison of biophysical techniques. *Anal Biochem*, **355**, 259-266.

25. Freyer, M.W. and Lewis, E.A. (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol*, **84**, 79-113.

26. Dettler, J.M., Buscaglia, R., Cui, J., Cashman, D., Blynn, M. and Lewis, E.A. (2010) Biophysical characterization of an ensemble of intramolecular i-motifs formed by the human c-MYC NHE III1 P1 promoter mutant sequence. *Biophys J*, **99**, 561-567.

27. Miyoshi, D., Nakao, A. and Sugimoto, N. (2003) Structural transition from antiparallel to parallel G-quadruplex of d(G4T4G4) induced by Ca2+. *Nucleic Acids Res*, **31**, 1156-1163.

28. Dettler, J.M., Buscaglia, R., Le, V.H. and Lewis, E.A. (2011) DSC Deconvolution of the Structural Complexity of c-MYC P1 Promoter G-Quadruplexes. *Biophys J*, **100**.

29. Han, F.X.W., Richard T.; Hurley, Laurence H. (1999) Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. . *J. Am. Chem. Soc.*, **121**, 3561-3570.

30. Rezler, E.M., Seenisamy, J., Bashyam, S., Kim, M.Y., White, E., Wilson, W.D. and Hurley, L.H. (2005) Telomestatin and diseleno sapphyrin bind selectively to two different forms of the human telomeric G-quadruplex structure. *J Am Chem Soc*, **127**, 9439-9447.

31. Weisman-Shomer, P., Cohen, E., Hershco, I., Khateb, S., Wolfovitz-Barchad, O., Hurley, L.H. and Fry, M. (2003) The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res*, **31**, 3963-3970.

32. Chaires, J.B. (2001) Analysis and interpretation of ligand-DNA binding isotherms. *Methods Enzymol*, **340**, 3-22.

33. Suh, D. and Chaires, J.B. (1995) Criteria for the mode of binding of DNA binding agents. *Bioorg Med Chem*, **3**, 723-728.

34. Ren, J. and Chaires, J.B. (1999) Sequence and structural selectivity of nucleic acid binding ligands. *Biochemistry*, **38**, 16067-16075.

35. Hyun, K.M., Choi, S.D., Lee, S. and Kim, S.K. (1997) Can energy transfer be an indicator for DNA intercalation? *Biochim Biophys Acta*, **1334**, 312-316.

36. Chaires, J.B. (2006) A thermodynamic signature for drug-DNA binding mode. *Arch Biochem Biophys*, **453**, 26-31.

37. Gavathiotis, E., Heald, R.A., Stevens, M.F. and Searle, M.S. (2003) Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)4 containing the human telomeric repeat. *J Mol Biol*, **334**, 25-36.

38. Keating, L.R. and Szalai, V.A. (2004) Parallel-stranded guanine quadruplex interactions with a copper cationic porphyrin. *Biochemistry*, **43**, 15891-15900.

39. Hounsou, C., Guittat, L., Monchaud, D., Jourdan, M., Saettel, N., Mergny, J.L. and Teulade-Fichou, M.P. (2007) G-quadruplex recognition by quinacridines: a SAR, NMR, and biological study. *ChemMedChem*, **2**, 655-666.

40. Zhang, Z., Dai, J., Veliath, E., Jones, R.A. and Yang, D. (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K+ solution: insights into the interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res*, **38**, 1009-1021.

41. Cavallari, M., Garbesi, A. and Di Felice, R. (2009) Porphyrin intercalation in G4-DNA quadruplexes by molecular dynamics simulations. *J Phys Chem B*, **113**, 13152-13160.

42. Scheidt, W.R. (1974) Stereochemistry of low-spin cobalt porphyrins. 3. The crystal structure and molecular stereochemistry of bis(piperidine)-alpha, beta, gamma, delta-tetraphenylporphinatocobalt(II). *J Am Chem Soc*, **96**, 84-89.

43. Lipscomb, L.A., Zhou, F.X., Presnell, S.R., Woo, R.J., Peek, M.E., Plaskon, R.R. and Williams, L.D. (1996) Structure of DNA-porphyrin complex. *Biochemistry*, **35**, 2818-2823.

CHAPTER V

THE EFFECT OF PYRIDYL SUBSTITUENTS ON THE THERMODYNAMICS OF PORPHYRIN BINDING TO G-QUADRUPLEX DNA

As previously published in Bioorganic & Medicinal Chemistry, 2013, 21,7515-7522.

5.1 Abstract

Most of the G-quadruplex interactive molecules reported to date contain extended aromatic flat ring systems and are believed to bind principally by π-π stacking on the end G-tetrads of the quadruplex structure. One such molecule, **TMPyP4**, (5,10,15,20-tetra(*N*-methyl-4-pyridyl)porphyrin), exhibits high affinity and some selectivity for G-quadruplex DNA over duplex DNA. Although not a realistic drug candidate, TMPyP4 is used in many nucleic acid research laboratories as a model ligand for the study of small molecule G-quadruplex interactions. Here we report on the synthesis and G-quadruplex interactions of four new cationic porphyrin ligands having only1, 2, or 3 (N-methyl-4-pyridyl) substituents. The four new ligands are: **P(5)** (5-(*N*-methyl-4-pyridyl)porphyrin), **P(5,10)** (5,10-di(*N*-methyl-4-pyridyl)porphyrin), **P(5,15)** (5,15-di(*N*-methyl-4-pyridyl)porphyrin), and **P(5,10,15)** (5,10,15-tri(*N*-methyl-4-pyridyl)porphyrin). Even though these compounds have been previously synthesized, we report alternative synthetic routes that are more efficient and that result in higher yields. We have used ITC, CD, and ESI-MS to explore the effects of the number of *N*-methyl-4-pyridyl

substituents and the substituent position on the porphyrin on the G-quadruplex binding energetics. The relative affinities for binding these ligands to the WT Bcl-2 promoter sequence G-quadruplex are: $K_{TMPyP4} \approx K_{P(5,15)} > K_{P(5,10,15)} >>> K_{P(5,10)}, K_{P(5)}$. The saturation stoichiometry is 2:1 for both **P(5,15)** and **P(5,10,15),** while neither **P(5)** nor **P(5,10)** exhibit significant complex formation with the WT Bcl-2 promoter sequence G-quadruplex. Additionally, binding of **P(5,15)** appears to interact by an *"intercalation mode"* while **P(5,10,15)** appears to interact by an *"end-stacking mode"*.

5.2 Introduction

Over expression of Bcl-2 has been observed in a number of cancers including: breast, prostate, cervical, colorectal, and lung carcinoma (1). The human Bcl-2 P1 promoter region is GC-rich and the G-rich strand is known to form a stable G-quadruplex structure (2). The stabilization of oncogene promoter sequence G-quadruplexes and telomere overhang repeat sequence G-quadruplexes by binding small molecules (e.g. ligands or drugs) is a subject of interest in cancer drug discovery. Ligands exhibiting structural recognition (selectivity) and high affinity for G-quadruplex DNA including several porphyrin based compounds have been explored for their potential as anti-cancer drugs (3-6). The porphyrin-based ligands have shown the ability to inhibit telomerase and they typically exhibit low cyto-toxicity (5,6). **TMPyP4**, (5,10,15,20- (N-methyl-4-pyridyl) porphyrin), while not a serious drug candidate, has been used extensively as a model ligand in the characterization of the structure, stability, and ligand binding properties of oncogene promoter sequence G-quadruplexes (5-14).

We have published several studies on the interactions of TMPyP4 with several oncogene promoter sequence G-quadruplexes including c-MYC, Bcl-2, and K-ras (7-11).

We have shown that that TMPyP4 binding to these G-quadruplexes results in a complex having 4 moles of TMPyP4 per mole of quadruplex and that the complex is stabilized significantly over the naked G-quadruplex. TMPyP4 binds to these G-quadruplexes by two distinctly different binding modes, a higher affinity mode which is driven by a favorable entropy change and a weaker mode which is accompanied by an unfavorable entropy change. We believe that these two modes represent first the end-binding of two moles of TMPyP4 followed by the intercalation of another two moles of TMPyP4 so that every G-tetrad is sandwiched between two planar porphyrin molecules. We recently reported on a study of the interactions of three related porphyrins, (TMPyP2, TMPyP3, and TMPyP4), with a WT 39-mer Bcl-2 promoter sequence G-quadruplex (7). Using CD, Fluorescence, ITC, and ESI-MS, we were able to shown that as the substituted porphyrin becomes non-planar, the intercalation binding mode is eliminated so that the saturation stoichiometry for binding TMPyP2 or TMPyP3 becomes 2:1 (15). In addition, the end-stacking mode is also weakened as the porphyrin ligand becomes increasingly less planar with the binding of TMPyP2 << TMPyP3. TMPyP4 like many other G-quadruplex interactive ligands reported to date are planar heterocyclic aromatic ring systems(16-25). The porphyrin ligand charge (e.g. +1, +2, +3, or +4), size (e.g. SASA), shape (e.g. planarity) and symmetry (e.g. substitution pattern), may all play a role in establishing the selectivity and affinity for G-quadruplex complex formation. In this study, we have begun to explore each of these effects by synthesizing new ligands that are systematically reduced in size, charge, and symmetry in comparison to TMPyP4. The new ligands differ from the model TMPyP4 quadruplex interactive compound by having only one, two, or three (*N*-methyl-4-pyridyl) substituents.

We used ITC, CD, and ESI-MS to study the interactions of these new porphyrin ligands with a 27-mer WT Bcl-2 P1 promoter sequence G-quadruplex. Although Goncalves *et al.* has already reported the synthesis for the **P(5)** (5-(*N*-methyl-4-pyridyl)porphyrin), **P(5,10)** (5,10-di(*N*-methyl-4-pyridyl)porphyrin), **P(5,15)** (5,15-di(*N*-methyl-4-pyridyl)porphyrin), and **P(5,10,15)** (5,10,15-tri(*N*-methyl-4-pyridyl)porphyrin) compounds (26), we present here alternative synthetic routes that are more efficient and that result in higher yields. Goncalves *et al.* also reported SPR experiments from which they estimated the binding affinities for these same four ligands with a human telomere repeat sequence G-quadruplex (26). In our experiments, the G-quadruplex is different, but even more importantly we have determined the complete set of thermodynamic parameters for the formation of these four porphyrin ligand complexes with a typical oncogene promoter sequence G-quadruplex. The ITC, CD, and ESI-MS data reported here have been analyzed to yield values for the free energy changes (ΔG), the enthalpy changes (ΔH), the entropy changes ($-T\Delta S$), and the saturation stoichiometry (n) for the formation of the WT Bcl-2 27-mer P1promoter sequence G-quadruplex complexes with **P(5), P(5,10), P(5, 15)** and **P(5,10,15)**.

5.3 Material and methods

5.3.1 Binding studies

The WT 27-mer Bcl-2 oligonucleotide used in this study was obtained from Midland oligos (Midland, TX). The oligonucleotide has the following sequence 5'-CGG GCG CGG GAG GAA GGG GGC GGG AGC -3'. DNA stock solution was prepared by reconstituting the lyophilized oligonucleotide into KBPES buffer with a salt concentration of 130 mM [KCl] and a pH of 7.0. Approximately 2 mL of the

oligonucleotide was dialyzed (1000 Mw cutoff membrane) against three changes of buffer solution (1 L, 24h each) at 4°C. The concentrations of stock DNA solution were verified using UV-Vis. DNA was annealed by quickly heating the sample to 100°C, holding at 100°C for ten minutes then slowly cooling to 5°C over a three hour period. The molar extinction coefficient of the DNA was determined using a nearest-neighbor method for single stranded DNA (27). The extinction coefficient at 260 nm for the WT 27-mer Bcl-2 sequence was: 2.26×10^5 $M^{-1}cm^{-1}$.

CD experiments were performed with an Olis DSM-20 spectropolarimeter (Bogart, GA). All measurements were done at 25°C using a 1 cm quartz cuvette and covering a spectral range of 225 - 325 nm. All CD samples were prepared such that they had a nominal absorbance of less than 1.0 at 260 nm.

ITC experiments were performed using a VP-ITC calorimeter (GE-Heathcare). A typical ITC experiment involved the addition of 28 (10μL) injections of a dilute (115 μM) WT 27-mer Bcl-2 DNA solution into ~ 1.5 mL of a dilute porphyrin solution (5 μM). Reverse titrations were employed because of the low solubility of the substituted porphyrins. Three replicate experiments were done for each of the four porphyrins. Blank titrations were done by injecting the titrant (115 μM) G-quadruplex) into buffer solution. Corrected titration curves were achieved by subtracting the blank titration data from the ITC-data for the porphyrin-DNA titrations. Corrected ITC titrations were fit with a nonlinear regression algorithm again using the CHASM© ITC data analysis program developed in our laboratory (28).

ESI-MS experiments on Bcl-2 and the Bcl-2 porphyrin complexes were carried out on a Bruker MicrOTOFQ mass spectrometer. Data acquisition was set to operate in

negative ion mode. All experiments were performed in 50 mM ammonium acetate buffer (pH=7.0) containing 20% HPLC grade methanol. The WT Bcl-2 27-mer G-quadruplex sample was prepared at a concentration of approximately 25 μM in the ammonium acetate buffer and was dialyzed against three changes of buffer 12 hours each at 4°C. Stock solutions of the porphyrin ligands were prepared in the final dialysate buffer at concentration as high as 50 μM. The ESI-MS samples were prepared by mixing the DNA and ligand stock solutions to yield a mixture containing 4 equivalents of each porphyrin ligand per equivalent of DNA. The MS capillary voltage was set to +3500 V, dry N_2 gas flow was adjusted to 0.5 L/min at 110°C, and the G-quadruplex/porphyrin samples were directly infused into the MS by using a kD Scientific syringe pump set to a flow rate of 200 μL/hour. Data processing was performed by using Bruker Daltonics Data Analysis program.

5.3.2 Porphyrin ligand synthesis

The overall synthesis for the four porphyrin ligands studied here is shown in Figure 5.1. We began our synthesis of the substituted porphyrins by synthesizing the key dipyrromethane building blocks **1** and **4** by condensation of the respective aldehydes with pyrrole (29,30). Formylation of **1** utilizing a Vilsmeier reaction gave compound **2** in 45% yield (26). Functionalized dipyrromethane **5** was synthesized by the diformylation of pyridyl dipyrromethane **4** also utilizing a Vilsmeier reaction in 28% yield (26). The reduction of formylated dipyrromethane **5** with sodium borohydride provided compound **6,** which was used in subsequent steps without purification (26).

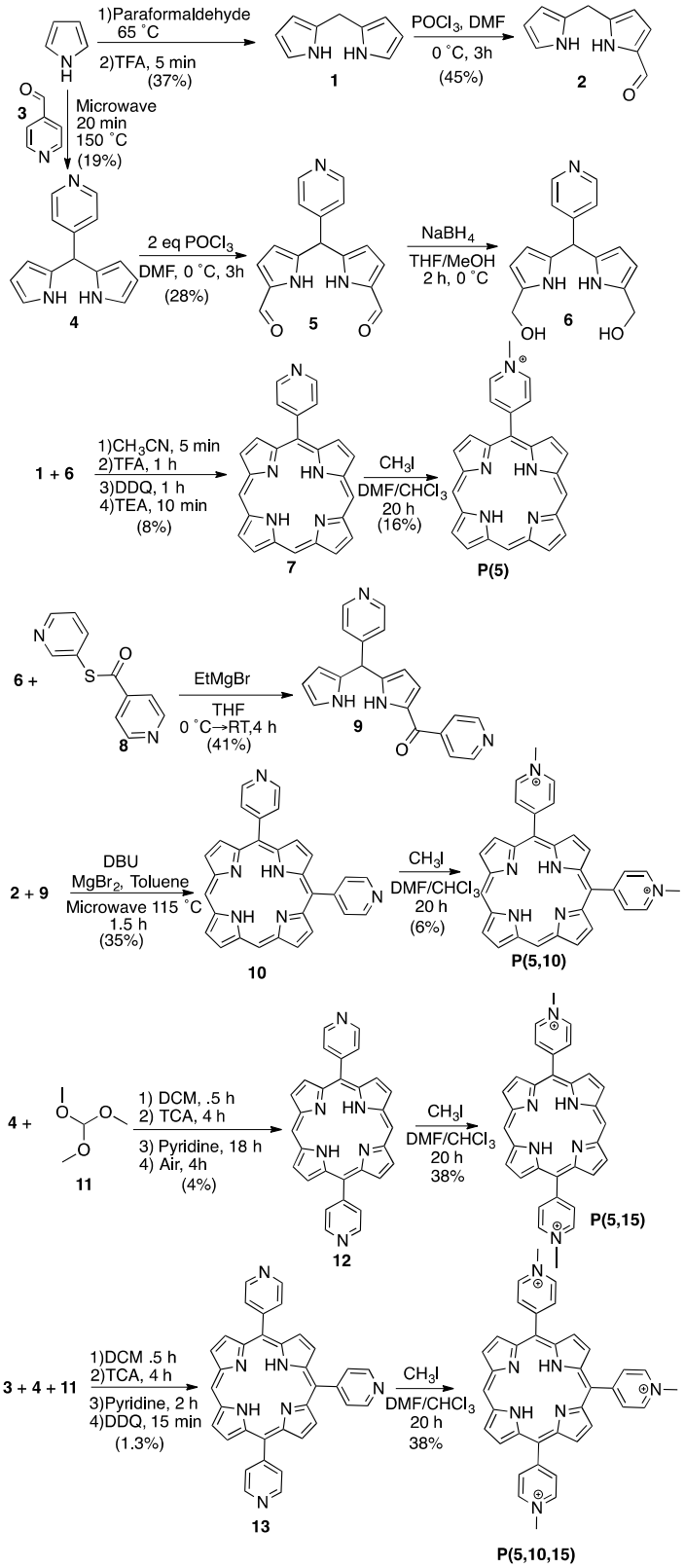


Figure 5.1 Synthetic routes for the four cationic porphyrin ligands used in this study.

With the key building in hand, we focused on the synthesis of the desired porphyrin analogues. The synthesis of **P(5)** was accomplished by the condensation of dipyrromethanes **1** and **6** in the presence of trifluoracetic acid. The resulting mixture was allowed to stir for one hour, oxidized with DDQ, and treated with triethylamine to provide porphyrin **7** in 8% yield. Methylation of **7** with iodomethane followed by washing with ether provided **P(5)** in 16% yield (26).

The synthesis of **P(5,10)** was accomplished by deprotonation of compound **6** with ethyl magnesium bromide followed by the addition of Mukiyama reagent **8** resulted in the formation of compound **9** in 41% yield (31). The condensation of compounds **2** and **9** resulted in the formation of porphyrin **10** in 35% yield (32). Methylation of **14** provided the **P(5,10)** in 6% yield (26).

Porphyrin **P(5,15)** was synthesized by the condensation of compound **4** with trimethylorthoformate **11** followed by methylation provided **P(5,15)** in 2% overall yield. Condensation of compounds **3**, **4**, and **11** followed by methylation resulted in the formation of porphyrin **P(5,10,15)** (26).

5.3.2.1 Dipyrromethane 1 (30):

Paraformaldehyde (1.73g, 57.7mmol) and pyrrole (100mL, 1.44 mol) were added to a 250mL two-necked round-bottom flask equipped with a condenser and an internal thermometer. The mixture was brought to an internal temperature of 65°C. The flask was removed from the heat and TFA (444 μL, 5.77 mmol) was added. The mixture was purified via flash chromatography eluting at 8-10% EtOAc in hexanes affording 3.17g of **1** as a white solid in 37% yield. ^{1}H NMR (300MHz, DMSO-d_6): δ=3.807 (s, 2H), 5.744 (s, 2H), 5.884-5.893 (d, 2H), 6.583 (s, 2H), 10.502 (br s, 2H).

5.3.2.2 1-formyl dipyrromethane **2** (26):

DMF (24.2 mmol, 1.9 mL) was added to a dry RBF and cooled to 0 ^{0}C. $POCl_3$ (0.3 mL, 3.01 mmol) was added drop wise and stirred for 10 min to yield the Vilsmeier reagent. In another dry RBF, dipyrromethane **1** (0.500 g, 3.42 mmol) was dissolved in DMF (7.5 mL) under argon and cooled to 0 ^{0}C. To the reaction mixture, the freshly prepared Vilsmeier reagent (1.9 mL, 3.36 mmol,) was added drop wise and stirred at 0 ^{0}C for 2 h. After 2 h, a biphasic solution of Ethyl acetate (50 mL) and saturated aqueous sodium acetate (50 mL) was added and stirred at room temperature for 4 h. Ethyl acetate was separated and the aqueous layer was extracted three times with ethyl acetate and the combined organic layer was washed with brine and water, dried over anhydrous Na_2SO_4 and evaporated *in vacuo* to get crude product. The desired compound was isolated using automated flash chromatography, eluting at 20% EtOAc in DCM and evaporate *in vacuo* to yield **2** as a pale yellow solid (270 mg, 45.2%). ^{1}HNMR (300 MHz, DMSO-d_6): δ =11.93 (br s, 1H), 10.606 (br s, 1H), 9.328 (s, 1H), 6.867-6.887 (m, 1H), 6.602-6.607 (m, 1H), 5.98-5.999 (m, 1H), 5.866-5.892 (m, 1H), 5.765 (br s, 1H), 3.868 (s, 2H) ppm.

5.3.2.3 5-(4-pyridyl)dipyrromethane 4 (29):

4-pyridylcarboxaldehyde **3** (1.5mL, 16.4 mmol) and pyrrole (16mL, 230 mmol) were added to a microwave vial. The vial was sealed under argon and stirred for 20 min at 150°C. The reaction mixture was cooled and concentrated under reduced pressure to yield a dark red liquid which was left on high vacuum overnight to yield a dark oil. The oil was dissolved in minimal DCM and wet-loaded onto a Biotage 50G column, packed with alumina oxide and purified via flash chromatography with DCM:EtOAc. Fractions containing product were collected and concentrated under vacuum. Minimal DCM was

added to the resulting solid to create a slurry. Excess hexanes were added and the mixture was filtered, collecting **4** as a pale grey solid (645.1mg, 18.9% yield). ^{1}H NMR (300MHz, $CDCl_3$): δ= 5.541 (s, 1H), 5.874 (s, 2H), 6.168-6.197 (q, 2H), 6.738-6.761 (q, 2H), 7.313-7.340 (d, 2H), 7.599-7.627 (d, 2H), 7.973 (br s, 2H).

5.3.2.4 1,9-Diformyl-5-(4-pyridyl)dipyrromethane **5** (26):

The procedure of 1-formyldipyrromethane **2** was followed with DMF (6 mL, 58 mmol), $POCl_3$ (0.9 mL, 9.66 mmol) and replacing the dipyrromethane with 5-(4-pyridyl)dipyrromethane **11** (0.300 g, 1.34 mmol). The desired compound was isolated using automated flash chromatography, eluting at 9% MeOH in DCM and concentrated *in vacuo* to yield **5** as a brown oil (0.103 g, 28%). ^{1}HNMR (300 MHz, $CDCl_3$), δ =10.96 (br s, 2H), 9.24 (s, 2H), 8.58 (d, J = 5.1 Hz, 2H), 7.28 (d, J = 5.1 Hz, 2H), 6.91 (s, 2H), 6.09 (s, 2H) and 5.60 (s, 1H) ppm.

5.3.2.5 (5,5'-(pyridine-4-ylmethylene)bis(1H-Pyrrole-5,2-diyl)dimethanol **6** (26):

In a dry RBF under argon, 1,9-Diformyl-5-(4-pyridyl)dipyrromethane **5** (0.103 g, 0.365 mmol) was dissolved in THF-Methanol (10:1 , 20 mL) and cooled to 0 ^{0}C. $NaBH_4$ (0.29 g, 7.66 mmol) was added in roughly 50 mg portions every 2 min under a stream of argon. The reaction mixture was stirred at room temperature for 1 h and poured into a biphasic mixture of saturated aqueous ammonium chloride and DCM (1:1, 50 mL). The organic layer was washed with water twice and dried over anhydrous Na_2SO_4 and evaporated *in vacuo* to get crude **6** (100 mg) which was used directly in the next step.

5.3.2.6 4-Pyridylporphyrin **7** (26)**:**

The above **11** stirred at room temperature for 5 min, then TFA (0.33 mL, 4.39 mmol) was added and the mixture was stirred for 1 h. DDQ (0.247 g, 1.09 mmol) in toluene (6 mL) was added and stirred for1 h. Triethylamine (0.6 mL, 4.39 mmol) was added and stirred for 10 min. The reaction mixture was evaporated *in vacuo*. Crude **7** was isolated at 2-10% MeOH in DCM as a purple solid (150 mg) using flash chromatography and purified by preparative TLC (eluted by 5% MeOH in DCM). The purple band was collected and extracted from the silica with MeOH. Evaporation of the solvent *in vacuo* gave **12** as a purple-red solid (6 mg, 8%). ^{1}HNMR (600 MHz, DMSO-d_6): δ=10.70 (s, 2H), 10.64 (s, 1H), 9.74-9.72 (m, *J* = 4.4 Hz, 6H), 9.33 (d, *J* = 5.9 Hz, 2H), 9.27 (d, *J* = 4.4 Hz, 2H), 8.92 (d, *J* = 5.9 Hz, 2H) and -3.57 (br s, 2H, NH) ppm.

5.3.2.7 S-2-Pyridyl isonicotinothioate **8** (31,32):

An oven dried flask was charged with 2-mercaptopyridine (1.000 g, 8.99 mmol) and purged with argon. The solid was dissolved in anhydrous THF (20 mL) with stirring. The solution was treated with isonicotinoyl chloride hydrochloride (1.600 g, 8.99 mmol). The resulting slurry was stirred overnight. The reaction mixture was filtered. The filtrate was washed with hexanes. The resulting orange solid was added to a biphasic solution of diethyl ether and saturated sodium bicarbonate and stirred until it no longer bubbled. The organic layer was collected and the aqueous layer was washed with diethyl ether (3 X 50 mL). The organic layers were combined and dried with Na_2SO_4 and concentrated *in vacuo* to yield a yellow solid. Minimal THF was added to the solid to form a slurry which was filtered and washed with hexanes to yield **8** (0.8 g, 45%). ^{1}HNMR (300 MHz,

DMSO-d_6): δ= 8.65 (m, 2H), 8.719 (m, 1H), 7.820-7.854 (m, 3H), 7.749 (d, J = 7.8 Hz, 1H), 7.40 (m, 1H).

5.3.2.8 1-Isonicotinoyl-5-(4-pyridyl)dipyrromethane **9** (31,32):

To a solution of **9** (0.844 g, 3.77822 mmol) in THF (15 mL) under argon was added dropwise 1.0 M EtMgBr (9.6 mL, 9.6 mmol) at room temperature. The flask was stirred for 10 min and cooled to -78 °C. S-2-pyridyl isonicotinothioate in THF (15 mL) was added dropwise over 10 min and stirred for an additional 10 min before warming to RT and stirring for 4 h. The reaction was quenched at RT with 25 mL saturated aqueous NH_4Cl. The slurry was extracted with EtOAc (3x25 mL). The organic layer was washed with DI water (3x20 mL), brine (3x20mL), dried with Na_2SO_4 and evaporated *in vacuo* to yield a yellow-brown solid. Flash chromatography gave **9** (0.506, 1.54 mmols) as a light brown solid in 41% yield eluting at 12% MeOH in DCM. ^{1}HNMR (300 MHz, DMSO-d_6): δ= 12.290 (br s, 1H), 10.809 (br s, 1H), 8.738 (d, J = 4.8, 2H), 8.493 (d, J = 4.8), 7.650 (d, J = 4.8), 7.159 (d, J = 4.8, 2H), 6.774 (m, 1H), 6.693 (m, 1H), 6.067 (m, 1H), 5.942 (m, 1H), 5.761 (m, 1H), 5.601 (s, 1H).

5.3.2.9 5,10-Di(4-pyridyl)porphyrin **10** (32):

To a stirred solution of 1-formyldiprromethane **7**(0.018 g, 0.103 mmol) and 1-isonicotinoyl-5-(4-pyridyl)dipyrromethane **13** (0.035 g, 0.106 mmol) in 2.4 mL toluene in a microwave reactor vial was added DBU (0.35 ml, 2.3 mmol) and stirred for 10 min. $MgBr_2$ (0.125 g, 0.67mmol) was added to the vial, which was then capped in atmosphere and heated to 115 °C for 90 min in the microwave reactor with magnetic stirring. The vial was emptied and rinsed with THF into a pear-shaped flask and evaporated. The

crude solid was dissolved in DCM (15 mL) and washed with DI water (2x30 mL) and brine (2x30mL) dried with Na_2SO_4 and evaporated. The solid was dissolved in 4 mL DCM and demetallated with trifluoroacetic acid (0.050 mL) and neutralized with triethylamine (0.040 ml). After 5 min the mixture was washed with DI water (1x10 mL) and brine (2x10 mL), dried with Na_2SO_4 and evaporated to dryness. The crude product was then purified by column chromatography eluting at 45-50% MeOH in EtOAc. Evaporation of the solvent gave **10** (0.017g) in 35% yield. ^{1}H NMR (300 MHz, $CDCl_3$): δ=10.341 (s, 2H), 9.519 (s, 2H), 9.439 (m, 2H), 9.079 (m, 4H), 9.012 (m, 2H), 8.936 (s, 2H), 8.206 (m, 4H), -3.422 (s, 2H).

5.3.2.10 5, 15-Di-4-pyridylporphyrin 12 (26):

5-(4-pyridyl)dipyrromethane **9** (1.000 g, 4.51 mmol), trimethyl orthoformate **11** (36 mL, 329 mmol) and DCM (700 mL) were added to a flask covered with aluminum foil, stirred and degassed by bubbling with Argon for 30 min. A solution of TCA (17.6 g, 108 mmol) in DCM (300 mL) was added drop wise to the flask over 15 min. The reaction stirred under argon for 4 h. The mixture was quenched with pyridine (31.2 mL) and stirred for18 h. The reaction mixture was bubbled with air for 10 min. The aluminum foil was removed and the reaction was stirred under ambient light for 4 h. The solvent was removed *in vacuo* to yield a black solid. The solid was filtered through a silica pad, and washed with DCM and 0.5% MeOH in DCM. The fraction collected was concentrated *in vacuo* and dissolved in DCM and filtered through a Buchner funnel to eliminate some black solid. The filtrate was concentrated to 3 mL and wet loaded onto a flash chromatography column. Automated flash chromatography gave pure fractions at 1% MeOH in DCM. Fractions were collected and concentrated *in vacuo* to afford **12** as a

purple solid (80 mg, 4%). ^{1}H NMR (300 MHz, $CDCl_3$): δ=10.416 (s, 2H), 9.485 (d, J=4.65, 4H), 9.11 (m, 4H), 9.075 (d, J=4.67, 4H), 8.260 (m, 4H).

5.3.2.11 5, 10, 15-Tri-4-pyridylporphyrin 13 (26):

5-(4-pyridyl)dipyrromethane **4** (500 mg, 2.25 mmol), trimethyl orthoformate **1** (9 mL, 82.5 mmol), 4-pyridylcarboxaldehyde **3** (0.4 mL, 4.1 mmol) and anhydrous DCM (400 mL) were added to a flask and degassed by bubbling with argon for 30 min. In a separate flask, TCA (8.83 g, 54 mmol) was dissolved in anhydrous DCM (100 mL). The solution was added to the reaction mixture slowly over 15 min and stirred for 4 h. The reaction mixture was quenched with pyridine (15.6 mL) and stirred for 2 h. DDQ (0.3 g, 1.3 mmol) was added and stirred for 15 min. The aluminum foil was removed and the solvent was evaporated *in vacuo* to yield a black solid, which was left on high vacuum overnight. The black solid was preadsorbed onto silica and purified twice via automated flash chromatography using 5% MeOH in DCM. Preparative TLC, run in 8% MeOH in DCM gave **12** as a purple band, isolated as a purple solid (15 mg, 1.3%). ^{1}H NMR (300 MHz, $CDCl_3$): δ=10.355 (s, 1H), 9.436 (d, J=4.7 Hz, 2H), 9.081 (m, 6H), 9.032 (d, J=4.7 Hz, 2H), 8.923 (m, 4H), 8.218 (m, 6H), -3.118 (s, 2H). Exact mass: calculated: 542.21; found: 542.2071 ($M+H^+$)

5.3.2.12 General procedure for the quaternization of porphyrin pyridyl groups with methyl iodide (26):

To an oven dried flask cooled under vacuum porphyrin (0.044 mmol) was added and purged with 3x with argon. 32 mL dry $CHCl_3$ followed by 8 mL dry DMF was added via syringe. 3.5 mL (excess) CH_3I was added and the solution was stirred for 12 hours at RT. The solution was concentrated under vacuum to 1/4 the original volume.

To the solution was then added 60 mL Et_2O several times with subsequent decanting. The solid was then triturated with CH_3CN, decanted and dried *in vacuo* to yield product.

5.3.2.12.1 5-(N-methylpyridynium-4-yl)porphyrin P(5):

Trituration yielded 0.002 g **P(5)**. Exact mass: calculated: 402.1713 ; found: 402.1771

5.3.2.12.2 5,10-Di(N-methylpyridinium-4-yl)porphyrin P(5,10):

Note that **P(5,10)** was soluble in CH_3CN, so the last wash was filtered and evaporated to dryness to yield 0.001 g **P(5,10)**. Exact mass: calculated: 494.22 ; found: 494.2140 (M)

5.3.2.12.3 5,15-Di(N-methylpyridinium-4-yl)porphyrin P(5,15):

Trituration yielded 0.012 g **P(5,15)**. ^1H NMR (300 MHz, DMSO-d_6): δ = 10.849 (s, 2H), 9.859 (d, J=4.7, 4H), 9.516 (d, J=5.91, 4H), 9.245 (d, J=4.7, 4H), 9.083 (d, J=5.92, 4H), 4.745 (s, 6H), -3.333 (br s, 2H). Exact mass: calculated 494.22; found: 494.2228.

5.3.2.12.4 5,10,15-tri(N-methyl pyridinium-4-yl)porphyrin P(5,10,15):

Trituration yielded 0.008 g **P(5,10,15)** ^1H NMR (300 MHz, $CDCl_3$): δ = 10.866 (s, 1H), 9.860 (br s, 2H), 9.518 (m, 6H), 9.210 (m, 4H), 9.183 (m, 2H), 9.030 (m, 4H), 8.970 (m, 2H), 4.981 (s, 9H), -3.660 (s, 2H). Exact mass: calculated: 195.42; found: 195.4242

5.4 Results

The CD spectra shown in Figure 5.2 indicate that the WT Bcl-2 27-mer P1 promoter sequence exists as a mixed parallel/anti-parallel G-quadruplex in the KBPES buffer in the absence of any added porphyrin ligand. The naked Bcl-2 27-mer CD spectrum shown in the Figure 5.2, (P/DNA = 0), exhibits typical G-quadruplex characteristic peaks and troughs at 245 nm, 265 nm and a shoulder at 290 nm (8). The dashed line shown along the x-axis in Figure 5.2 is the measured ellipticity of the porphyrin in a DNA free solution. The CD spectra for the **P(5,10,15)/Bcl-2** complexes having porphyrin to DNA ratios of 1, 2, 3, or 4 show that the CD signal for the G-quadruplex is largely the same for the complexes and the naked DNA although there is some attenuation of the characteristic peaks and troughs in the complex spectra. It is obvious that binding **P(5,10,15)** to the WT Bcl-2 27-mer P1 promoter sequence G-quadruplex does not perturb the folding of the DNA in the G-quadruplex. The CD spectra shown in Figure 5.3 are for the 2:1 complexes of the Bcl-2 G-quadruplex with P(5,15), P(5,10,15) and TMPyP4 (P(5,10,15,20)). The CD spectra shown in Figure 5.3 indicate that there is no change in the G-quadruplex structure upon complex formation with theses ligands.

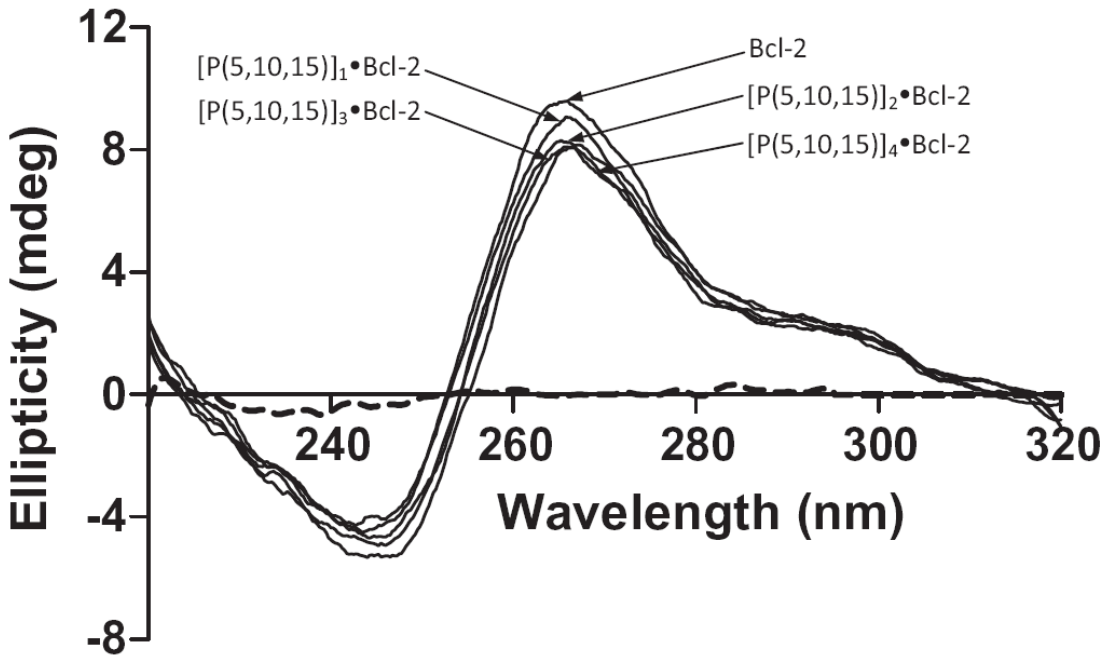


Figure 5.2 CD spectra of WT 27-mer Bcl-2 G-quadruplex in the absence and presence of the **P(5,10,15)** ligand. The ellipticity at 265 nm is attenuated progressively as the porphyrin to DNA ratio is increased from 0 to 4. The dashed line near y = 0 is the ellipticity for the **P(5,10,15)** in a DNA free solution.

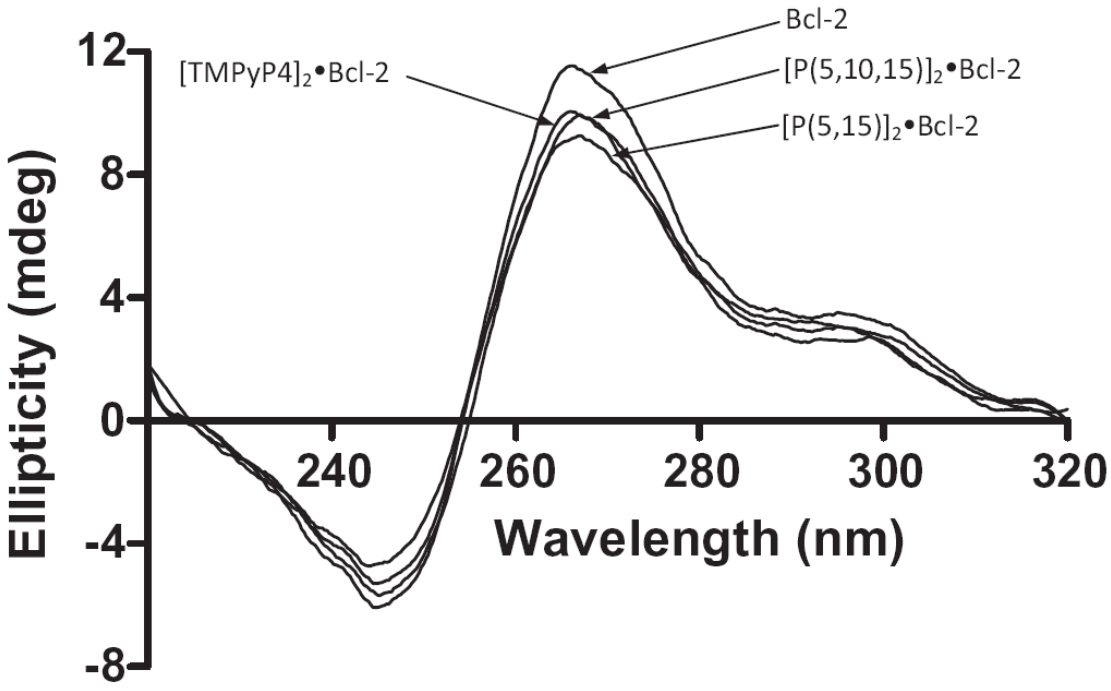


Figure 5.3 CD spectra are shown for the Bcl-2 G-quadruplex and its 2:1 complexes with **P(5,15)**, **P(5,10,15)** and **TMPyP4**. The ellipticities at 265 nm and 295 nm are essentially the same for all three ligands and the free Bcl-2 G-quadruplex.

The ESI mass spectrum for the WT Bcl-2 27-mer P1 promoter sequence G-quadruplex, and the spectra for the saturated complexes of the WT Bcl-2 27-mer P1 promoter sequence G-quadruplex with **P(5)**, **P(5,10)**, **P(5, 15)** and **P(5,10,15)** are shown in Figure 5.4.

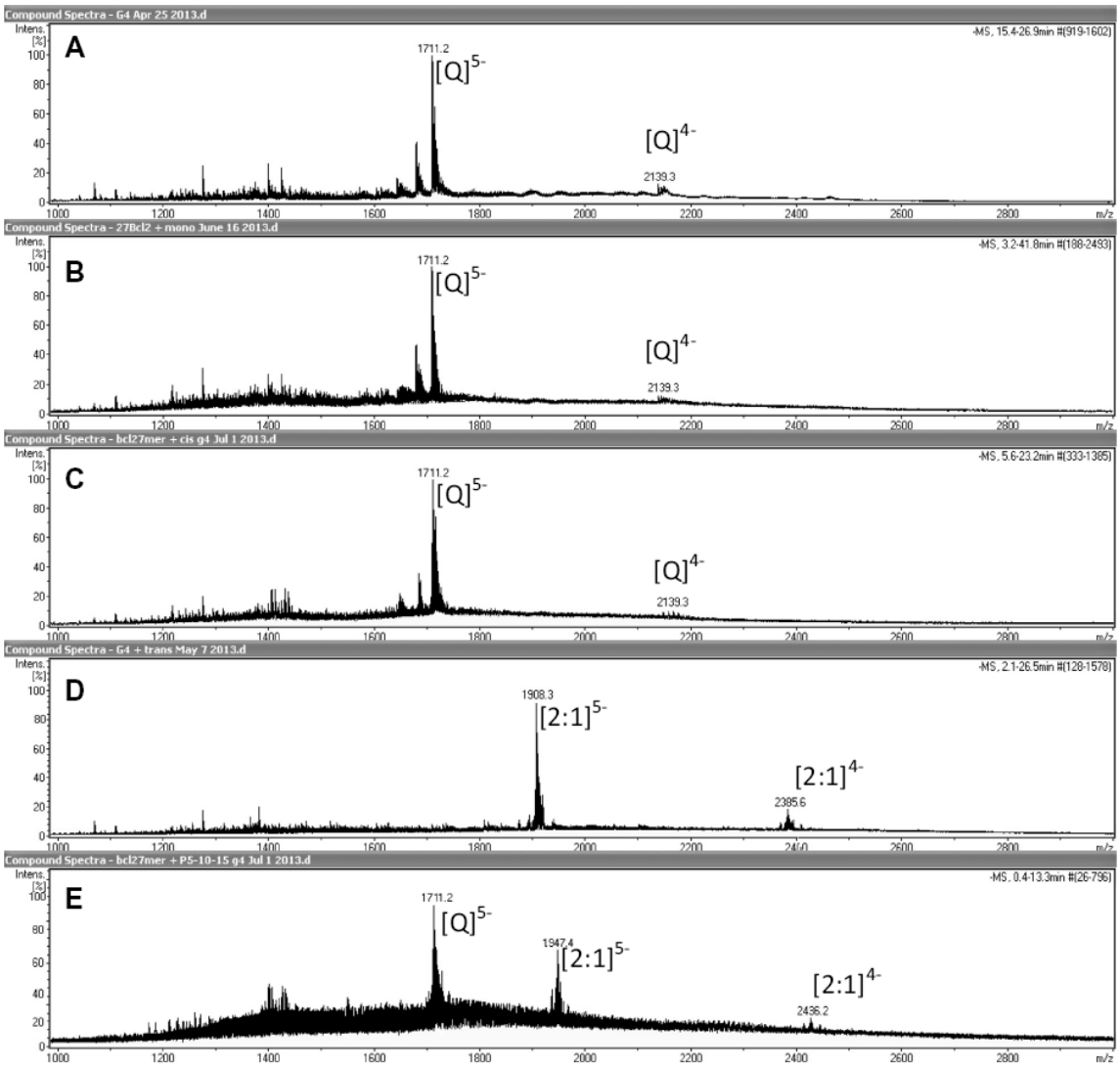


Figure 5.4 Electrospray ionization mass spectra for WT 27-mer Bcl-2 G-quadruplex and its complexes with **P(5)**, **P(5,10)**, **P(5,15)**, and **P(5,15,20)**.

Panel A shows the mass spectrum obtained for a solution containing only the Bcl-2 oligonucleotide at a concentration of 10 μM. Panels B–E show the mass spectra for the Bcl-2 oligonucleotide and its porphyrin complexes obtained in solutions containing 40 μM **P(5)**, 40 μM **P(5,10)**, 40 μM **P(5,15)** and 40 μM **P(5,15,20)**, respectively.

The ESI mass spectra for the solutions containing **P(5)** or **P(5,10)** along with the Bcl-2 G-quadruplex, shown in Figure 5.4B and C respectively, provide no evidence for Bcl-2 complex formation with these ligands. The only peaks in the mass spectra shown

in Panel B and Panel C of Figure 5.4 are for the parent DNA with a charge of -5 or -4, exactly the same as the spectrum shown in Panel A for the naked DNA. The ESI mass spectrum for the solution containing **P(5,15)** along with the Bcl-2 G-quadruplex is shown in Panel D of Figure 5.4. The spectrum for this solution which contains four moles of **P(5,15)** per mole of Bcl-2 G-quadruplex is consistent with significant complex formation. The new peaks with m/z values of 1908.3 and 2385.6 are attributed to a **P(5,15)/Bcl-2** complex having a molar ratio of 2:1 and a charge of -5 or -4 respectively. It should also be noted that no free or uncomplexed Bcl-2 is present in this solution. This result is consistent with a very large K_a for formation of the 2:1 **P(5,15)/Bcl-2** complex. The ESI mass spectrum for the solution containing **P(5,10,15)** and the Bcl-2 G-quadruplex is shown in Panel E of Figure 5.4. The spectrum for this solution, which contains four moles of **P(5,10,15)** per mole of Bcl-2 G-quadruplex, is consistent with weaker complex formation. The new peaks with m/z values of 1947.4 and 2436.2 are attributed to a **P(5,10,15)/Bcl-2** complex having a molar ratio of 2:1 and a charge of -5 or -4. In this spectrum we also see a peak with an m/z value of 1711.2 which corresponds to uncomplexed G-quadruplex with a charge of -5. The same peak (m/z = 1711.2) is also seen as the predominant peak in the spectrum obtained for DNA in the ligand free solutions (Figure 5.4 Panel A). The ESI-MS spectra were obtained using ammonium acetate as the buffer and supporting electrolyte (in place of either Na^+ or K^+). The greater lability of the ammonium ion in comparison to either sodium or potassium ions reduces DNA counterion adduction yields a cleaner mass spectrum. Under these conditions, the observation of the loss of the two NH_4^+ ions from between the stacked G-tetrads in the quadruplex is problematic.

Due to the limited solubility of the P(5), P(5,10), P(5,15) and P(5,10,15) porphyrin ligands, the ITC experiments were all done as reverse titrations in which a dilute solution of the DNA was added to a dilute solution of the porphyrin in the calorimeter cell (see Figure 5.5).

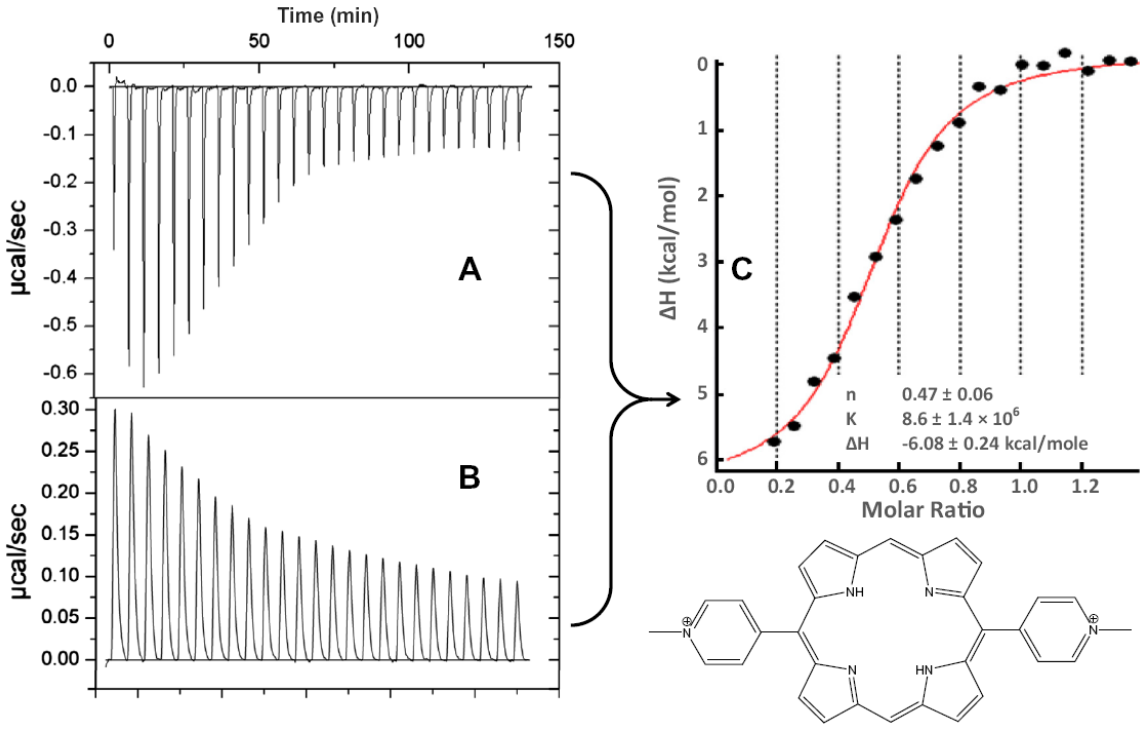


Figure 5.5 Panel A shows the thermogram for the addition of the 115 µM WT-27mer Bcl-2 titrant into 5 µM **P(5,15)**. Panel B shows the thermogram for heat of dilution of the porphyrin solution, that is, the addition of buffer into 5 µM **P(5,15)**. Panel C shows the corrected enthalpogram for titration DNA into porphyrin ligand along with the model fit (—) and best fit parameters for formation of the 2:1 complex.

Reverse titration is a common practice for systems with low solubility and yields thermograms that are easily fit to an independent site model with our CHASM© program (28) (see ITC data in Figure 5.5 panel C). Typical ITC data for the titration of the

P(5,15) porphyrin ligand with the Bcl-2 G-quadruplex are shown in Figure 5.5 panel A. The data shown in Figure 5.5 panel B are for the dilution of **P(5,15)**, i.e. the curve shown is for the addition of buffer to the **P(5,15)** porphyrin titrate. The heats of dilution for these porphyrins were significant, a sign that the porphyrins in solution are undergoing some self- association. The data shown in Figure 5.5 panel C are the integrated heat data corrected for the heat of dilution. The best fit parameters and associated uncertainties for the non-linear regression fit of the ITC heat data are shown in the box in Figure 4C. The expected value of *n* would be 0.5 for a complex having two ligands per mole of DNA since these data are for the reverse titration. The ITC data for the **P(5,15)** and **P(5,10,15)** titrations were similar in that these two ligands bind with similar stoichiometry (2:1) and similar affinity ($K_a \approx 1 \times 10^7$). The ITC data for the **P(5)** and **P(5,10)** titrations were similar in that these two ligands were not observed to bind and the titrations were almost identical to the blank titrations. The corrected data for the **P(5,15)** and **P(5,10,15)** titrations were fit with CHASM© to yield best fit values for *K*, ΔH, and *n*. Values of ΔG and -TΔS were calculated from the best fit parameters and are reported in Table 5.1. The relative affinities for binding the four new porphyrin ligands and TMPyP4 to the WT Bcl-2 promoter sequence G-quadruplex are: $K_{TMPyP4} \approx K_{P(5,15)} > K_{P(5,10,15)} >>> K_{P(5,10)}, K_{P(5)}$. The ITC and ESI-MS data are in agreement with respect to the relative affinities.

Table 5.1 A Comparison of the thermodynamic parameters for binding the porphyrin ligands **P(5)**, **P(5,10)**, **P(5,15)**, **P(5,10,15)** and **TMPyP4** to G-quadruplex DNA

Ligand	SPR[a]	ITC[b]					
	ΔG (kcal/mol)	*'End-stacking'*			*'Intercalation'*		
		ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
P(5)	na	*No interaction*			*No interaction*		
P(5,10)	−6.1 ± 0.1	*No interaction*			*No interaction*		
P(5,15)	−8.3 ± 0.1	–			-9.4 ± 0.1	-6.1 ± 0.2	−3.3
P(5,10,15)	−7.1 ± 0.1	−6.7 ± 0.1	−2.5 ± 0.4	−4.2	–		
TMPyP4	−8.4 ± 0.1	−10.4 ± 0.1	−1.8 ± 0.2	−8.6	−7.6 ± 0.1	−6.1 ± 0.1	−1.5

[a] The SPR data are for the interaction of the porphyrin ligands with a 26-mer human telomere G-quadruplex having the sequence: 5'-GGA TTG GGA TTG GGA TTG GGA TTG GG-3' (26)
[b] The ITC data reported here are for the interaction of the porphyrin ligands with a 27-mer Bcl-2 G-Quadruplex having the sequence: 5'-CGG GCG CGG GAG GAA GGG GGC GGG AGC-30. The data for the binding of TMPyP4 to the same G-quadruplex are from Nagesh *et al.* (9)

5.5 Discussion

The synthetic routes for all four model porphyrin ligands are shown in Figure 5.1 and all previous synthesis and synthetic methods are referenced. The synthesis of porphyrins **P(5)**, **P(5,15)**, and **P(5,10,15)** was accomplished utilizing the route reported by Goncalves *et al.* (26). This route provided these porphyrins in good yield. The synthesis of **P(5,10)** was accomplished by the modification of a synthetic route reported by Dogutan *et al.* (32). This synthetic route provided us with a more efficient means for the synthesis of the **P(5,10)** analogue as compared to the Goncalves route.

The binding of the new porphyrin ligands is compared to TMPyP4 for the purpose of determining the effects of the ligand's solvent accessible surface area (SASA), charge, symmetry, and geometry on the thermodynamics of their interactions with G-quadruplex DNA. The thermodynamic data given in Table 5.1 yield a couple of surprises. The binding of the **P(5,15)** and to a lesser degree the **P(5,10,15)** compounds seems to be tighter than expected, while the binding of the **P(5,10)** compound appears to be weaker

than expected. We have previously suggested that the two binding modes observed for the interaction of TMPyP4 with G-quadruplexes in general and more specifically with c-MYC, Bcl-2, and K-ras promoter sequence G-quadruplexes are 1) and "end binding" mode, and 2) an "intercalation mode" (7,8,10). The end-binding mode is generally accepted in the literature and is not controversial in that stacking of the planar porphyrin ligand, TMPyP4, on a terminal G-tetrad is consistent with simple π-π stacking and similar to guanine self-association. On the other hand, the proposed intercalation mode is neither proven nor does it resemble simple self-association or the π-π stacking of the end binding interaction. However, we as well as others have published numerous studies that suggest that intercalation is plausible, not only from the saturation stoichiometry which varies as the number of stacked tetrads plus one (12), but also from the reduction in bound ligands from 4 to 2 as the ligand becomes non-planar (15). Modeling studies show that TMPyP4 can intercalate between the G-tetrads and that the structure of the G-quadruplex is not affected (33). All of these results are also supported by CD data that indicate no change in the G-quadruplex CD spectrum on formation of the 4:1 TMPyP4/DNA complex. There is also no induced CD signal for bound TMPyP4 until excess ligand is present, with excess ligand more weakly bound to the exterior of the G-quadruplex (15), possibly in the grooves. The thermodynamic signatures for these two modes are characteristic in that the "end binding" mode is accompanied by a small favorable ΔH and a large favorable -TΔS term which drives the complex formation, while the "intercalation" mode is driven by a large favorable ΔH and a negligible change in entropy (7). Using these same arguments, we describe the binding of the **P(5,15)** ligand as "intercalation" and the binding of the **P(5,10,15)** ligand as "end binding". In

comparison to TMPyP4 which forms G-quadruplex complexes having a saturation stoichiometry of 4:1, the **P(5,15)** and **P(5,10,15)** only form 2:1 complexes with the Bcl-2 promoter sequence G-quadruplex.

The affinities we report in Table 5.1 for binding **P(5,15)** and **P(5,10,15)** to the Bcl-2 G-quadruplex are in general agreement with the affinities reported by Goncalves *et al.* (26) for binding these same ligands to a human telomere G-quadruplex. In comparison to our ITC results for G-quadruplex complex formation with these ligands, the SPR experiments of Goncalves *et al.* (26) yield a smaller ΔG° value for the interaction of **P(5,15)** (-8.3 kcal/mol vs. -9.4 kcal/mol by ITC), a similar ΔG value for the interaction of **P(5,10,15)** (-7.1 kcal/mol vs. -6.7 kcal/mol by ITC), and a smaller ΔG° value for the interaction of TMPyP4 (-8.4 kcal/mol vs. -9.0 kcal/mol, the average ΔG° for the mode 1 and mode 2 reactions as determined by ITC). All of these values are in good agreement particularly since the Goncalves results are for binding to a human telomere G-quadruplex instead of the Bcl-2 G-quadruplex used in our study which exists as an ensemble of at least three folded isomers. The biggest differences are that Goncalves *et al.* (26) reports a significant affinity for binding the **P(5,10)** ligand to a human telomere G-quadruplex while we did not observe any complex formation between the **P(5,10)** ligand and the Bcl-2 promoter sequence G-quadruplex in either our ITC or ESI-MS experiments. Additionally Goncalves *et al.* (26) failed to find the expected 4:1 stoichiometry for the TMPyP4 complex with the human telomere (34). The SPR study of Federici *et al.* (35) reports two binding modes and a saturation stoichiometry of 4:1 for the binding of TMPyP4 to the c-MYC promoter sequence G-quadruplex. Their two reported affinities are: $K_1 \approx 1.3\times10^8$ (ΔG = -11.0 kcal/mol) and $K_2 \approx 7.7\times10^5$ (ΔG° = -8.0

kcal/mol). These are in excellent agreement with our previously reported values for the interaction of TMPyP4 with the c-MYC, Bcl-2, and K-ras promoter sequence G-quadruplexes (7,8,10).

Whether the binding of **P(5,15)** is through end-binding, intercalation, or binding in a G-quadruplex groove, its apparent high affinity points us in a new direction with respect to designing G-quadruplex interactive ligands. In comparison to TMPyP4, the **P(5,15)** ligand has a smaller SASA, and has a charge of only +2. However, **P(5,15)** binds to the Bcl-2 G-quadruplex with almost the same affinity as TMPyP4, ΔG° = -9.4 kcal/mol vs. -10.4 kcal/mol for TMPyP4 (for binding the first two moles of TMPyP4). It also appears to bind by what we are calling the "intercalation mode" vs. the mode 1 interaction for TMPyP4 which has been assigned to "end binding". Perhaps even better G-quadruplex ligands could be synthesized to take advantage of the *trans* (5,15) substitution pattern on the core porphyrin, but with changes in the substituent charge or shape to produce stronger interactions with either the G-tetrad(s), loops, or grooves of the G-quadruplex (36-38). The weaker affinity of the **P(5,10,15)** ligand in comparison to TMPyP4 and its apparent end binding are more difficult to explain. For example, why does the removal of a single pyridinium substituent from TMPyP4 result in the loss of almost 3.5 kcal in the binding energy? Perhaps the explanation is that the substituent pyridinium groups in TMPyP4 are interacting with the loops, grooves and or tail residues in the G-quadruplex and the loss of one interaction is destabilizing. The Arya group has begun to exploit dual recognition with conjugated ligands incorporating both a planar moiety to base stack with an aminoglycoside to bind in the quadruplex groove (36,37). A better understanding of the G-quadruplex interactions with model porphyrin ligands used

in this study as well as other new ligands will most certainly lead to better G-quadruplex stabilizing drugs.

5.6 References

1. Baretton, G.B., Diebold, J., Christoforis, G., Vogt, M., Müller, C., Dopfer, K., Schneiderbanger, K., Schmidt, M. and Löhrs, U. (1996) Apoptosis and immunohistochemical bcl-2 expression in colorectal adenomas and carcinomas: Aspects of carcinogenesis and prognostic significance. *Cancer*, **77**, 255-264.

2. Heckman, C., Mochon, E., Arcinas, M. and Boxer, L.M. (1997) The WT1 protein is a negative regulator of the normal bcl-2 allele in t(14;18) lymphomas. *Journal of Biological Chemistry*, **272**, 19609-19614.

3. Gavathiotis, E., Heald, R.A., Stevens, M.F.G. and Searle, M.S. (2003) Drug Recognition and Stabilisation of the Parallel-stranded DNA Quadruplex d(TTAGGGT)4 Containing the Human Telomeric Repeat. *Journal of Molecular Biology*, **334**, 25-36.

4. Haider, S.M., Parkinson, G.N. and Neidle, S. (2003) Structure of a G-quadruplex–Ligand Complex. *Journal of Molecular Biology*, **326**, 117-125.

5. Han, H., Langley, D.R., Rangan, A. and Hurley, L.H. (2001) Selective interactions of cationic porphyrins with G-quadruplex structures. *Journal of the American Chemical Society*, **123**, 8902-8913.

6. Dixon, I.M., Lopez, F., Estève, J.P., Tejera, A.M., Blasco, M.A., Pratviel, G. and Meunier, B. (2005) Porphyrin derivatives for telomere binding and telomerase inhibition. *ChemBioChem*, **6**, 123-132.

7. Freyer, M.W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L.H. and Lewis, E.A. (2007) Biophysical Studies of the c-MYC NHE III1 Promoter: Model Quadruplex Interactions with a Cationic Porphyrin. *Biophysical Journal*, **92**, 2007-2015.

8. Nagesh, N., Buscaglia, R., Dettler, J.M. and Lewis, E.A. (2010) Studies on the Site and Mode of TMPyP4 Interactions with Bcl-2 Promoter Sequence G-Quadruplexes. *Biophysical Journal*, **98**, 2628-2633.

9. Nagesh, N., Sharma, V.K., Ganesh Kumar, A. and Lewis, E.A. (2010) Effect of Ionic Strength on Porphyrin Drug Interactions with Quadruplex DNA Formed by the Promoter Region of c-MYC and Bcl-2 Oncogenes. *J. Nucleic Acids*, 2009.

10. Dettler, J.M. and Lewis, E.A. (2011) Biophysical Studies of the Structure, Stability and Ligand Properties of G-Quadruplex DNA: Thoughts and Comparisons of the K-ras, c-MYC, and Bcl-2 Oncogene promoter Sequence Quadruplexes. *Frontiers in Nucleic Acids*, **1082 VOL.**, 33-50.

11. Dettler, Jamie M., Buscaglia, R., Le, Vu H. and Lewis, Edwin A. (2011) DSC Deconvolution of the Structural Complexity of c-MYC P1 Promoter G-Quadruplexes. *Biophysical Journal*, **100**, 1517-1525.

12. Haq, I., Trent, J.O., Chowdhry, B.Z. and Jenkins, T.C. (1999) Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin. *Journal of the American Chemical Society*, **121**, 1768-1779.

13. Arora, A. and Maiti, S. (2008) Effect of loop orientation on quadruplex - TMPyP4 interaction. *Journal of Physical Chemistry B*, **112**, 8151-8159.

14. Kimura, T., Kawai, K., Fujitsuka, M. and Majima, T. (2006) Detection of the G-quadruplex-TMPyP4 complex by 2-aminopurine modified human telomeric DNA. *Chemical Communications*, 401-402.

15. Le, V.H., Nagesh, N. and Lewis, E.A. (2013) Bcl-2 promoter sequence G-quadruplex interactions with three planar and non-planar cationic porphyrins: TMPyP4, TMPyP3, and TMPyP2. *PloS one*, **8**, e72462.

16. Han, H., Bennett, R.J. and Hurley, L.H. (2000) Inhibition of unwinding of G-quadruplex structures by Sgs1 helicase in the presence of N,N'-Bis[2-(1-piperidino)ethyl]-3,4,9,10-perylenetetracarboxylic diimide, a G-quadruplex-interactive ligand. *Biochemistry*, **39**, 9311-9316.

17. Neidle, S. (2010) Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS Journal*, **277**, 1118-1125.

18. Collie, G.W., Promontorio, R., Hampel, S.M., Micco, M., Neidle, S. and Parkinson, G.N. (2012) Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. *Journal of the American Chemical Society*, **134**, 2723-2731.

19. Monchaud, D., Granzhan, A., Saettel, N., Guedin, A., Mergny, J.L. and Teulade-Fichou, M.P. (2010) "One ring to bind them all"-part I: the efficiency of the macrocyclic scaffold for g-quadruplex DNA recognition. *Journal of nucleic acids*, **2010**.

20. Li, Q., Xiang, J.F., Yang, Q.F., Sun, H.X., Guan, A.J. and Tang, Y.L. (2013) G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic acids research*, **41**, D1115-1123.

21. Yaku, H., Murashima, T., Miyoshi, D. and Sugimoto, N. (2013) In Vitro Assays Predictive of Telomerase Inhibitory Effect of G-Quadruplex Ligands in Cell Nuclei. *The journal of physical chemistry. B*.

22. Iida, K., Majima, S., Nakamura, T., Seimiya, H. and Nagasawa, K. (2013) Evaluation of the interaction between long telomeric DNA and macrocyclic hexaoxazole (6OTD) dimer of a G-quadruplex ligand. *Molecules*, **18**, 4328-4341.

23. Tan, W. and Yuan, G. (2013) Electrospray ionization mass spectrometric exploration of the high-affinity binding of three natural alkaloids with the mRNA G-quadruplex in the BCL2 5′-untranslated region. *Rapid Communications in Mass Spectrometry*, **27**, 560-564.

24. Zhu, L.N., Wu, B. and Kong, D.M. (2013) Specific recognition and stabilization of monomeric and multimeric G-quadruplexes by cationic porphyrin TMPipEOPP under molecular crowding conditions. *Nucleic acids research*, **41**, 4324-4335.

25. Blankson, G.A., Pilch, D.S., Liu, A.A., Liu, L.F., Rice, J.E. and Lavoie, E.J. (2013) Macrocyclic biphenyl tetraoxazoles: Synthesis, evaluation as G-quadruplex stabilizers and cytotoxic activity. *Bioorganic and Medicinal Chemistry*, **21**, 4511-4520.

26. Gonçalves, D.P.N., Ladame, S., Balasubramanian, S. and Sanders, J.K.M. (2006) Synthesis and G-quadruplex binding studies of new 4-N-methylpyridinium porphyrins. *Organic and Biomolecular Chemistry*, **4**, 3337-3342.

27. Plum, G.E. (2000) Determination of oligonucleotide molar extinction coefficients. *Current Protocols in Nucleic Acid Chemistry*, **73**, 1-17.

28. Le, V.H., Buscaglia, R., Chaires, J.B. and Lewis, E.A. (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Analytical biochemistry*, **434**, 233-241.

29. Gryko, D. and Lindsey, J.S. (2000) Rational synthesis of meso-substituted porphyrins bearing one nitrogen heterocyclic group. *Journal of Organic Chemistry*, **65**, 2249-2252.

30. Littler, B.J., Miller, M.A., Hung, C.H., Wagner, R.W., O'Shea, D.F., Boyle, P.D. and Lindsey, J.S. (1999) Refined synthesis of 5-substituted dipyrromethanes. *Journal of Organic Chemistry*, **64**, 1391-1396.

31. LINDSEY, J.S., CALLINAN, J.B., DHARMA RAO, P. and GEIER, G.R. (2001) A survey of acid catalysts in dipyrromethanecarbinol condensations leading to meso-substituted porphyrins. *Journal of Porphyrins and Phthalocyanines*, **05**, 810-823.

32. Dogutan, D.K., Ptaszek, M. and Lindsey, J.S. (2008) Rational or statistical routes from 1-acyldipyrromethanes to meso-substituted porphyrins. Distinct patterns, multiple pyridyl substituents, and amphipathic architectures. *Journal of Organic Chemistry*, **73**, 6187-6201.

33. Cashman, D.J., Buscaglia, R., Freyer, M.W., Dettler, J., Hurley, L.H. and Lewis, E.A. (2008) Molecular modeling and biophysical analysis of the c-MYC NHE-III1 silencer element. *Journal of Molecular Modeling*, **14**, 93-101.

34. Martino, L., Pagano, B., Fotticchia, I., Neidle, S. and Giancola, C. (2009) Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *Journal of Physical Chemistry B*, **113**, 14779-14786.

35. Federici, L., Arcovito, A., Scaglione, G.L., Scaloni, F., Lo Sterzo, C., Di Matteo, A., Falini, B., Giardina, B. and Brunori, M. (2010) Nucleophosmin C-terminal leukemia-associated domain interacts with G-rich quadruplex forming DNA. *Journal of Biological Chemistry*, **285**, 37138-37149.

36. Xue, L., Ranjan, N. and Arya, D.P. (2011) Synthesis and spectroscopic studies of the aminoglycoside (Neomycin)-perylene conjugate binding to human telomeric DNA. *Biochemistry*, **50**, 2838-2849.

37. Ranjan, N., Davis, E., Xue, L. and Arya, D.P. (2013) Dual recognition of the human telomeric G-quadruplex by a neomycin-anthraquinone conjugate. *Chemical Communications*, **49**, 5796-5798.

38. Neidle, S. (2012). Therapeutic Applications of Quadruplex Nucleic Acids; Academic Press, Boston, pp. 151.

CHAPTER VI

CAPPED G-QUADRUPLEXES FLANKED BY DUPLEX DNA: A MODEL FOR HUMAN C-MYC NHE-III$_1$

6.1 Abstract

In vitro studies of intramolecular G-quadruplex structure and function have typically been restricted to studies of single strand G-rich DNA constructs. We report here on the biophysical characterization of double stranded intramolecular G-quadruplexes that model the human c-MYC NHE-III$_1$. The model NHE-III$_1$ was assembled by annealing a 17-mer oligonucleotide (short polypyrimidine strand) having a mid-sequence run of 5T's (or C's) and tail sequences of 3'-AAATTT and TTTAAA-5', to a 32-mer oligonucleotide (polypurine strand) having a 20-mer c-MYC G-quadruplex forming core and complimentary tail sequences of 5'-TTTAAA and AAATTT-3'. CD spectra for the model NHE-III$_1$ motifs exhibit a shoulder at 290 nm indicative of increased anti-parallel character in these double stranded G-quadruplex structures. The single stranded c-MYC and model NHE-III$_1$ contain at least two folded G-quadruplex conformations (e.g. 1:2:1 and 1:6:1 isomers). In the model NHE-III$_1$, this conformational equilibrium is shifted, with the 1:6:1 loop isomer becoming the predominant G-quadruplex species. The model NHE-III$_1$ exhibits a significant increase in TMPyP4 binding affinity over the single stranded construct ($\delta\Delta G_i \approx 2$ kcal/mol). Although this NHE-III$_1$ model may still be too simple, it appears to provide a new perspective for drug

targeting G-quadruplexes formed in the c-MYC NHE-III_1 or G-Quadruplexes formed in other oncogene promoters..

6.2 Introduction

Putative G-quadruplex forming sequences have been found in the regulatory regions of approximately 40% of all human genes and in the promoter regions of several oncogenes that are associated with a broad spectrum of cancers (1-7). Our research group has a strong interest in the c-MYC oncogene, which is overexpressed in 60–80% of all cancers (8-12). The c-MYC oncogene has been known to associate with a number of human cancers, including lung, breast, colon, colorectal, and hematopoietic (4,5,13,14). Located at -142 to -115 bp upstream of the P1 promoter of the c-MYC oncogene is a region called the nuclease hypersensitivity element, or NHE-III_1. As the name suggests, this region of DNA is particularly susceptible to DNA degradation by DNase I (15) due to the slow equilibrium between the duplex B-DNA form to higher order DNA structures known as G-quadruplex and i-motif (16-19). The NHE-III_1 controls 80-90% of the transcription in the c-MYC gene which makes it a promising anti-cancer drug target for down regulating c-MYC expression. Binding small molecules to these DNA secondary structures can stabilize the G-quadruplex or i-motif, thereby inhibiting the formation of the transcriptionally active DNA/ protein complex (11,12,18,20-22). Understanding the structure and stability of the G-quadruplex is key to designing small molecules that can specifically target the G-quadruplex with high affinity.

Previous studies have shown that the typical G-quadruplex forming DNA sequence results in an ensemble of folded structures exhibiting a number of folding topologies. The ensemble of folded structures result from changes in the specific bases

which reside in the end or lateral loops and/or the direction of the backbone at each of the four corners of the stacked G-tetrads. The c-MYC WT polypurine P1 promoter sequence has been shown to fold into 1:2:1 and/or 1:6:1 loop isomers with a predominantly parallel backbone orientation (10,11,20,22). We have shown that a wild type 24-mer c-MYC sequence spontaneously folds in solution to yield an equilibrium mixture of two independently melting folded conformers having either the 1:2:1 or the 1:6:1 structures (20,23) as previously described by Phan *et al.* (10). These two foldamers are named for the number of bases present in the 1st, 2nd, and 3rd lateral loops, hence the two names being 1:2:1 or 1:6:1 and having either 1, 2, and 1 bases in the three lateral loops, or 1, 6, and 1 bases in the three lateral loops (23). The 1:6:1 loop isomer is less stable and has the larger number of unstructured loop bases in the core region of the G-quadruplex forming sequence. All previous *in vitro* studies of G-quadruplex structure, stability and/or ligand binding have employed relatively short single stranded guanine rich oligonucleotides to form the various G-quadruplexes of interest (7,10-12,16,18,20-30). These previous studies have been somewhat limited in their physiological relevance since these higher order DNA structures are found in a constant equilibrium with their duplex form. In order to fully understand the physical interactions between the G-quadruplex and small molecules, a better DNA model must be developed. Currently the procedure to create complimentary non-duplex forming purine and pyrimidine rich constructs has not been developed.

We have assembled model NHE-III_1 constructs by pairing a c-MYC WT G-rich sequence in the longer strand (to form the looped out G-quadruplex) with a shorter strand and a partially complimentary "capping" sequence consisting of three distinct regions 1)

a complimentary 3' duplex forming region, 2) a non-bonding/ non complimentary loop/bridging region and 3) a second complimentary 5' duplex forming region. Essentially we "capped" the G-quadruplex, allowing the higher order structure to form freely while the tail regions (not used in G-quadruplex formation) form flanking duplex regions. The length of the loop/bridge has also been investigated to determine its effects on G-quadruplex formation.

Preliminary data were collected in our lab on the original c-MYC 36-mer WT with "capping" sequences. Capping constructs consisting of 4 thymines (4T), 5 thymines (5T), or 6 thymines (6T) were used to bridge the flanking duplex tail regions of the 36-mer WT. All constructs showed the same results with the addition of a "capping" construct; a shift in the equilibrium between the 1:2:1 and the 1:6:1 conformers. The relative percentages of the two conformers present without the addition of the capping construct were 70% and 30% for the 1:2:1 and 1:6:1 respectively. With the addition of the "capping" construct the relative percentages shifted to 30% and 70% for the 1:2:1 and 1:6:1, respectively. The change in relative percentages or shift in equilibrium observed with the "capping" sequence demonstrated that the 1:6:1 became the predominant structure observed within the WT rather than the 1:2:1 that is observed when no capping sequence is present.

Current studies are being performed on shorter 32-mer c-MYC sequences which consist of sticky, mutant ends. The non-complimentary "capping" sequence consisted of the same type of three distinct regions as used in the preliminary studies but simply consisting of five thymines within the non-bonding/ non-complimentary bridging region. The complimentary duplex forming ends were mutated to ensure proper association

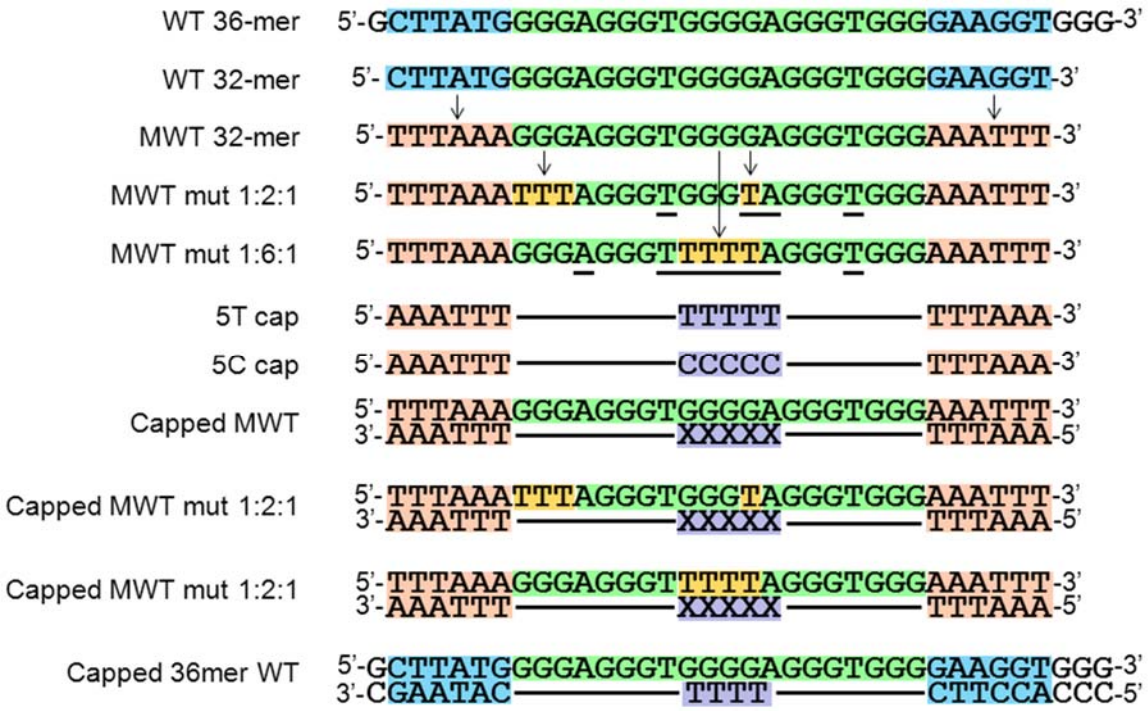
between the G-quadruplex forming sequence and the “capping” construct and to prevent self-association of the G-quadruplex sequence ends. A shift in the equilibrium of the 1:2:1 and the 1:6:1 conformers was similarly observed when the “capping” sequence was added. The 1:6:1 was observed to be the most predominant conformation when a capping sequence was present. There was no significant deviation observed in G-quadruplex structure or thermal stability with the addition of the “capping” construct.

The binding of small molecules to G-quadruplexes and i-Motifs, have become a focus for new anticancer drug research (31). Previous studies of meso-tetra (N-methyl-4-pyridyl) porphine (TMPyP4) has shown a high binding affinity for G-quadruplex DNA (11,20,32,33). TMPyP4 increases the stability of the quadruplex by inhibiting the transcriptional abilities of the c-MYC gene (11,12,18,20,33). With the addition of the capping sequences, the CD data support the presence of a mixture of parallel and anti-parallel folded G-quadruplex structures (23). At physiological pH and conditions (130 mM K^+) the 1:6:1 binds preferentially to TMPyP4 over the WT and 1:2:1. The shift towards the lower-melting 1:6:1 conformer was observed. This may lead to a significant change in the targeting of anti-cancer drugs to folded G-quadruplex structures.

6.3 Material and methods

In this study five different oligonucleotides were used to model for the double stranded c-MYC P1 promoter G-quadruplexes. All oligos were obtained from Oligos Etc. (Wilsonville, OR, USA). The 32-mer MWT sequence was designed to have the 20-mer wild type c-MYC P1 G-rich promoter sequence flanked by six nucleotides on both the 5’ end and the 3’ end making up the 32-mer construct. The 32-mer MWT **(**modified wild type) oligonucleotide serves as the wild-type scaffold having five runs of three or more

guanines. Subsequent **G→T** and **G→A** mutations were made to create the 1:6:1 and 1:2:1 mutants having specific folding topology. The sequences for the three different 32-mer oligonucleotides and the two capping constructs are shown in Scheme 6.1. The flanking duplex sequence regions of the three c-MYC 32-mer sequences were selected to provide sticky ends for capping and to prevent unwanted self-association reactions. The capping sequences were prepared by inserting five thymines or five cytosines between the complimentary flanking regions.



Scheme 6.1 Oligonucleotides sequences used for the assemble of the model NHE-III$_1$. Shown at the top is the WT 36-mer c-MYC promoter sequence G-quadruplex. Base modifications are shown with color boxes and arrows. Bases in the bridging regions are either T's or C's and represented as X's.

All chemicals used in this study were purchased from Sigma-Aldrich (St. Louis, MO, USA) and used without further purification. All oligonucleotide solutions were

prepared by dissolution of the oligonucleotide in BPES buffer (30 mM (KH_2PO_4/ K_2HPO_4) 1 mM EDTA, 100 mM KCl and adjusted to pH 7.0 by addition of KOH). Small volumes of the oligonucleotide stock solutions were dialyzed (1000 molecular-weight cutoff membrane) against two changes of buffer at 4 ºC (1 L, 24 h each). 1:1 mixtures of G-quadruplex strand and capping strand were annealed by heating to 100°C, holding at 100°C for ten minutes then slowly cooling to 5 °C over a three hour period. DNA concentrations were verified using extinction coefficients determined using a nearest-neighbor calculation for single-stranded DNA (34). The extinction coefficients determined at 260 nm for the 32-mer MWT, 32-mer 1:2:1, 32-mer 1:6:1, 17-mer 5T cap, 17-mer 5C cap are 3.298 x 10^5 $M^{-1}cm^{-1}$, 3.240 x 10^5 $M^{-1}cm^{-1}$, 3.226 x 10^5 $M^{-1}cm^{-1}$, 1.665 x 10^5 $M^{-1}cm^{-1}$, and 1.620 x 10^5 $M^{-1}cm^{-1}$ respectively.

Meso-tetra (N-methyl-4-pyridyl) porphine (TMPyP4) was obtained from Frontier Scientific (Logan, UT, USA). Ligand solutions were prepared by dissolving weighed amounts of TMPyP4 in 5 to 10 mL of the oligonucleotide final dialysate solution. The concentrations of the TMPyP4 solutions were verified with absorbance measurements using the published molar extinction coefficient of ε_{424} = 2.26 x 10^5 $M^{-1}cm^{-1}$ (35).

ESI-MS experiments were performed using a Bruker (Bellirica, MA, USA) MicrOTOFQ mass spectrometer in negative ion mode. DNA samples were dissolved in a 50 mM ammonium acetate buffer (pH ≈ 7) containing 20% methanol for these experiments. All DNA samples were prepared at a concentration of approximately 80 μM in the ammonium acetate buffer and were exhaustively dialyzed at 4°C against the 50 mM ammonium acetate NH_4^+/CH_3COO^- buffer. Samples were directly infused into the MS by using a KD Scientific (Holliston, MA, USA) syringe pump set to a flow rate of

200 μL/hour. The MS capillary voltage was set to +3500 V, and dry N_2 gas flow was adjusted to 0.5 L/min at 110°C. Data processing was performed by using Bruker Daltonics Data Analysis program.

Circular Dichroism (CD) experiments were performed using an Olis DSM 20 CD Spectrophotometer (Bogart, GA, USA) equipped with a 1 cm path length quartz cuvette. Spectra were collected at 0.25 nm resolution over the wavelength range of 220 to 320 nm. All the CD spectra were collected on annealed oligonucleotide solutions having nominal a concentration of approximately 3 μM.

Differential scanning calorimetry (DSC) was performed using a Microcal VP-DSC (Northampton, MA, USA). The concentration for all of the oligonucleotide samples used in the DSC experiments was nominally 250 μM. Experiments were performed over a temperature range from 10 to 95 ºC with a scan rate of 90 ºC/hour. At least five identical scans were obtained to ensure reproducibility. All of the DSC data were fit with two overlapping "non-two-state" processes using Origin 7.1 (Microcal, Inc., Northampton, MA).

Isothermal titration calorimetry (ITC) experiments were performed using a Microcal VP-ITC (Northampton, MA, USA). All of the ITC titrations were done by overfilling the calorimeter cell with ~1.5 mL of the dilute oligonucleotide solution and then injecting up to 50 (5 μL) injections of the dilute ligand solution. The dilute oligonucleotide solution concentration was nominally 50 μM and the TMPyP4 titrant concentration was nominally 1.5 mM. All titrations were done at 25 ºC and all experiments were done in triplicate. The ITC data were analyzed using our CHASM© analysis program (36).

Molecular modeling and MD simulations were performed using Accelrys Discovery Studio v.3.1 (San Diego, CA, USA). An ensemble of the c-MYC parallel G-quadruplex structures was generated from the solution NMR structure adapted from the Protein Date Bank (PDB accession code 1xav) (22). Using the nucleic acid building module, two separate flanking ds B-DNA constructs were built with the sequences in the polypurine strand being 5'-TTTAAA and AAATTT-3'. The WT c-MYC 20 base G-quadruplex core sequence was fused between these two B-DNA constructs to model the polypurine strand in the c-MYC NHE-III_1 and a short single strand sequence of 4, 5, or 6 T's (or C's) was fused between the two B-DNA constructs to complete the polypyrimidine strand in the model NHE-III_1. Similar constructs were built using the 1:2:1 and 1:6:1 mutant 20 base G-quadruplex core sequences. The three resulting constructs, each having two duplex tails with a looped out G-quadruplex in the polypurine strand and a short unstructured poly-T (or poly-C) loop in the polypyrimidine strand, were typed with the CHARMm27 forcefield (37,38) using the Momany-Rone partial charge method.

All NHE-III_1 model systems were solvated using an Explicit Periodic Boundary to create a solvation shell extending 10 Å away from the boundary of the DNA construct. Counterions were added to a concentration of 0.15 M. Each system was subjected to a minimization routine involving as many as 8,000 steps using a RMS gradient of less than 0.1. The Smart Minimizer algorithm was implemented with atom constraints for all nucleobase atoms in the G-tetrads. The SHAKE constraint (39) was applied in order to preserve the double stranded nature of the flanking duplex regions by retaining the hydrogen bonds throughout the simulation. TMPyP4 was manually docked into each

binding mode ('end binding' and/or 'intercalation' modes) and additional minimization routines were performed. Minimization was achieved by placing a harmonic restraint with a force constant of 10 kcal/mol $Å^2$ on all atoms in the porphyrin ligand and relaxing the capped DNA construct. After minimization, the interaction energy between each docked ligand and the DNA was calculated. For the 32-mer 1:6:1 capped structure, the six-bases-lateral loop was folded-back manually into a position allowing maximum overlap with an end bound TMPyP4.

Molecular dynamics (MD) simulations were initiated with heating dynamics to raise the temperature of the system from 50 to 300 K, followed by an equilibration period and lastly a production step both executed at 300 K. The heating steps were executed with gradual removal of atom constraints on the DNA(40). The total equilibration time was simulated for 20 ps at 1fs step followed by an additional of 100 ps of production step. NVT production type and the Leapfrog Verlet algorithm was used to integrate the dynamics. Structure and Trajectory analysis was performed using the Analyze Trajectory protocol in Accelrys DS 3.1.

6.4 Results

ESI mass spectrometry was used to provide direct evidence for the formation of a 1:1 complex between the 32-mer MWT DNA construct and the 17-mer 5-T capping construct. Figure 6.1A shows the ESI mass spectrum for the uncapped 32-mer MWT oligonucleotide. The three significant peaks that are observed in the mass spectrum correspond to m/z values of 1263.2, 1689.6, and 2027.7, and all three are attributed to the MWT oligonucleotide having a charge of -8, -6 or -5 respectively. Figure 6.1B shows the ESI mass spectrum for the 17-mer 5-T cap. Two significant m/z peaks are observed at

m/z values of 1289.6 and 1719.9. These peaks correspond to the 17-mer 5-T cap DNA species having a charge of -4 or -3 respectively. Figure 6.1C provides mass spectral evidence for the formation of the 32-mer MWT•5-T capped complex. Multiple m/z peaks are observed at m/z values of 1176.1, 1273.1, 1392.1, 1702.6, and 1910.1. These peaks are correlated with the 32-mer MWT•5-T capped complex having a net charge from -8 to -13. There is no evidence of either the free 32-mer MWT oligonucleotide or the free 17-mer 5-T oligonucleotide in these solutions.

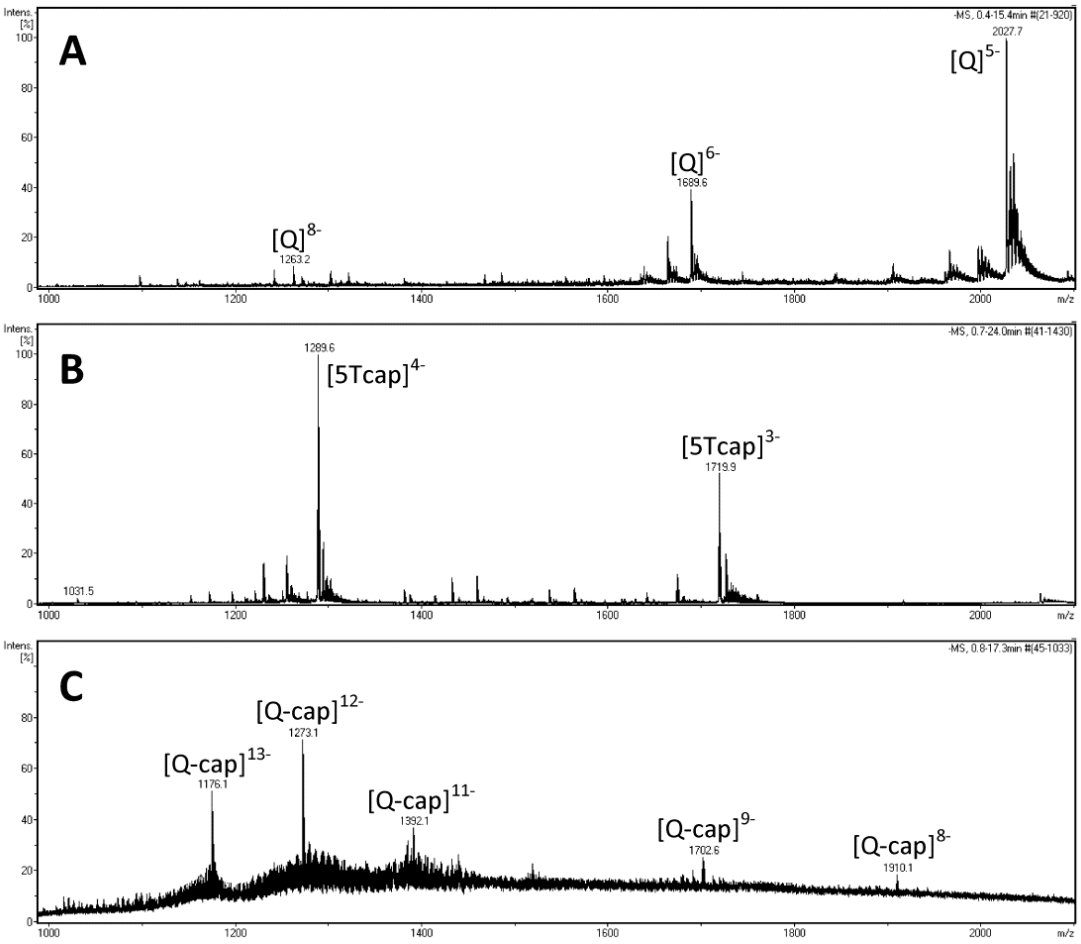


Figure 6.1 Electrospray ionization mass spectra for solutions containing A) 80 μM 32-mer MWT construct B) 80 μM 17-mer 5T capping construct and C) 1:1 mixture solution of 32-mer MWT•5T capped complex.

CD spectra at 20 °C for the uncapped 32-mer MWT, the 32-mer MWT capped with 5 T's, and the 12bp dsDNA are shown in Figure 6.2 Panel A. For comparison, CD spectra at 20 °C for the uncapped 32-mer MWT and the 5T-capped 32-mer MWT after removal of 12bp dsDNA CD signal are shown in Figure 6.2 Panel B.

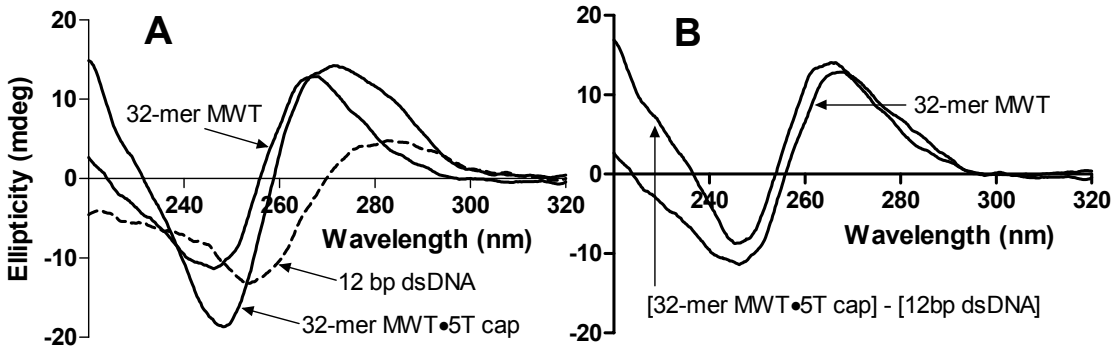


Figure 6.2 CD spectra for the 32-mer MWT construct with and without 5 T's cap and a 12bp dsDNA (panel A). CD spectra for the 32-mer MWT and the 5T capped 32-mer MWT after removal of 12bp dsDNA CD signal (panel B).

For comparison, the CD spectra for the uncapped 32-mer MWT and the uncapped 32-mer 1:2:1 and 32-mer 1:6:1 mutant sequences at 20 °C are shown in Figure A.1. Each of the independent spectra shows distinct positive maximums in the ellipticity at 264 and minimums in the ellipticity at 245 nm which are consistent with previous reports on parallel intramolecular G-quadruplexes (11). The 20 °C CD spectrum for the uncapped MWT (32-mer with sticky ends) shown in Figure 6.2A is essentially unchanged from the published CD spectra for the c-MYC WT 24-mer and WT 27-mer G-quadruplexes (11,12,20-22). In comparison, the 20 °C CD spectrum for the capped MWT construct

shown in Figure 6.2A, exhibits a pronounced shoulder in the region from approximately 270 nm to 300 nm. The longer wavelength shoulder in the CD spectrum for the capped MWT G-quadruplex must be either result of additional anti-parallel folding topology or contribution from 6bp dsDNA from either side of the G-quadruplex core. The CD spectrum for a 12bp dsDNA is shown in Figure 6.2 Panel A. It can be easily seen that the CD spectrum for dsDNA is very distinguished from the CD spectrum for G-Quadruplex DNA with almost zero CD signal at characteristic wavelength (264 nm) for parallel quadruplex and pronounced positive peak in the 280 to 290 nm region. Our 12bp dsDNA CD spectrum is very similar to the CD spectrum observed for B-DNA reported by Miyahara *et al.* (41). The concentration of the dsDNA for this CD experiment was calculated to be approximately equal the amount of duplex presence in the 5T capped MWT construct. Figure 6.2 Panel B shows comparison between CD spectra for the uncapped 32-mer MWT and the 5T-capped 32-mer MWT after removal of 12bp dsDNA. The two spectra are essentially very similar to one another with the exception that the 5T capped MWT construct exhibit slight increase in the amount of quadruplex formed.

A typical DSC thermogram for melting the uncapped 32-mer MWT G-quadruplex ensemble is shown in Figure 6.3A. This thermogram cannot be fit for a single melting transition, and the two independent overlapping melting transitions obtained by fitting the uncapped MWT thermogram for multiple independent "two-state" melting processes having T_m values of 61.1 °C and 66.7 °C are shown as subordinate peaks in panel A of Figure 6.3.

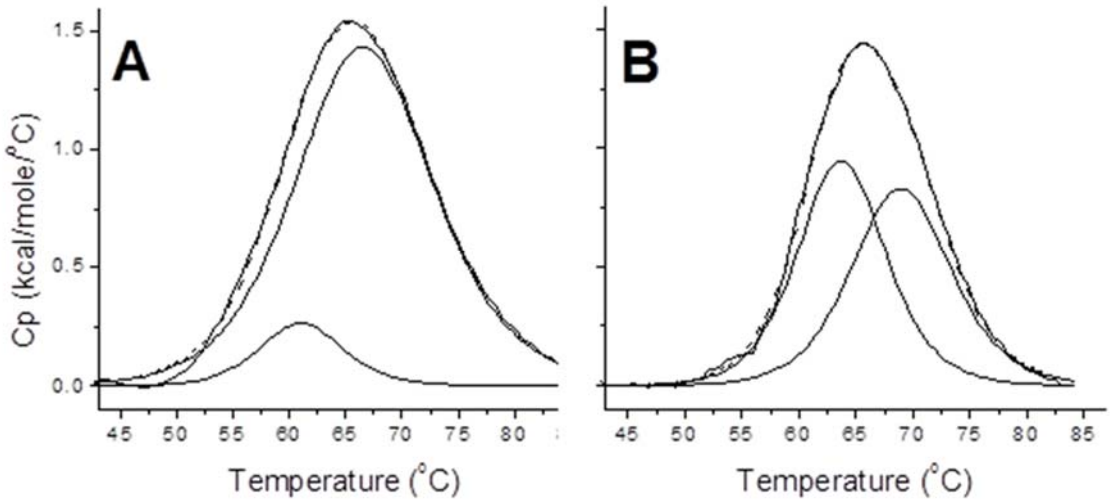


Figure 6.3 A) Thermal denaturation of the 32-mer MWT G-quadruplex. B) Thermal melting profile of the c-MYC 32-mer MWT•5T capped complex.

The two independent melting transitions imply that there are at least two different G-quadruplex conformations existing simultaneously in these solutions with the lower melting transition attributed to a minor species as judged from the relative areas of the lower temperature and high temperature subordinate peaks. The thermogram for thermal denaturation of the 5-T capped MWT is shown in Figure 6.3 Panel B. Again using the same fitting procedures, this thermogram implies that there is an equilibrium mixture of two folded G-quadruplex species for the capped construct. Both folded isomers in this solution exhibit T_m values (63.8 °C, and 69.1 °C) that are increased by approximately 2 °C over the T_m values exhibited for the uncapped MWT and a dramatic increase in the fraction of the lower melting species which is now the predominant species in solution. Although it is tempting to attribute the lower melting temperature to the 1:6:1 loop conformation and the higher melting temperature to the 1:2:1 loop conformation, the T_m

values do not exactly correspond to the expected values, and in all cases the T_m values are much lower ($\delta T_{m's} \approx$ -15 to -25 °C) than previously observed for the uncapped 24-mer and 27-mer ss c-MYC WT and mutant sequences (20,23). The DSC melting curves for the 32-mer 1:2:1 and 1:6:1 loop isomers are shown in Figure A.2 along with the melting curves for their 5-T (and 5-C) capped constructs. The 1:2:1 and 1:6:1 DSC data have been fit for a single two-state process. The melting transitions for the uncapped and 5-T capped 1:2:1 constructs showed melting temperatures of 66.8°C and 67.1 °C respectively, while the thermal transitions for the uncapped and 5-T capped 1:6:1 constructs exhibited T_m values of 55.2 °C and 56.8 °C respectively. The T_m values exhibited by the 5-C capped 1:2:1 and 1:6:1 isomers were similar to the T_m values for 5-T capped constructs. The lower values for the T_m's observed for both ss and capped MWT and mutant MWT c-MYC sequences must be attributed to the presence of the sticky ends which place AAA sequences immediately before and after the core G-quadruplex forming WT, 1:2:1 and 1:6:1 sequences.

Figure 6.4 shows typical ITC titration data for TMPyP4 titrations of the uncapped 32-mer MWT and the uncapped 1:2:1 and 1:6:1 mutant c-MYC constructs. All three oligonucleotides exhibit a saturation stoichiometry of 4 moles of TMPyP4 bound per mole of G-quadruplex DNA. The uncapped MWT and mutant quadruplex titrations were corrected for a heat effect attributed to ligand induced folding (the shaded areas in Figure 6.4) and then fit for two independent binding processes as described in an earlier study of the TMPyP4 interactions with ss G-quadruplexes formed by shorter WT or mutant c-MYC promoter sequences (20).

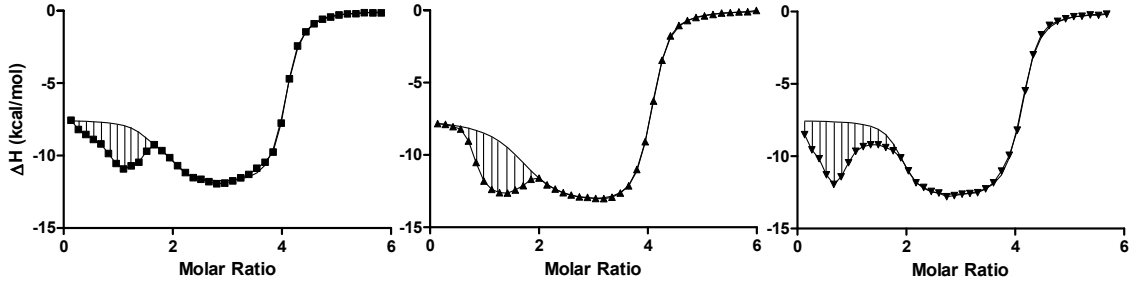


Figure 6.4 Titrations of TMPyP4 into 32-mer MWT (■), 1:2:1 (▲), and 1:6:1 (▼) G-quadruplex DNAs.

Hatched areas bounded by the best fit lines and the raw data indicate the heats attributed to ligand induced restructuring of G-quadruplex DNAs. These integrated areas were calculated between 0 and 2.0 Molar Ratio and averaged to be 2.1 ± 0.3 kcal/mol.

The two binding processes exhibit characteristic thermodynamic signatures in that the higher affinity binding process includes a very favorable entropic contribution to the free energy while the lower affinity process is almost completely the result of a favorable change in enthalpy. The thermodynamic parameters (ΔG_i, ΔH_i, and $-T\Delta S_i$, where i=1 or 2) obtained by fitting the data in Figure 6.4 to the two-process thermodynamic model are listed in Table 6.1. Typical corrected ITC data for TMPyP4 titrations of the 5-T capped MWT and 5-T capped mutant MWT G-quadruplexes are shown in Figure 6.5 along with typical raw ITC data for the TMPyP4 titration of the 5-T capped MWT (MWT•5T).

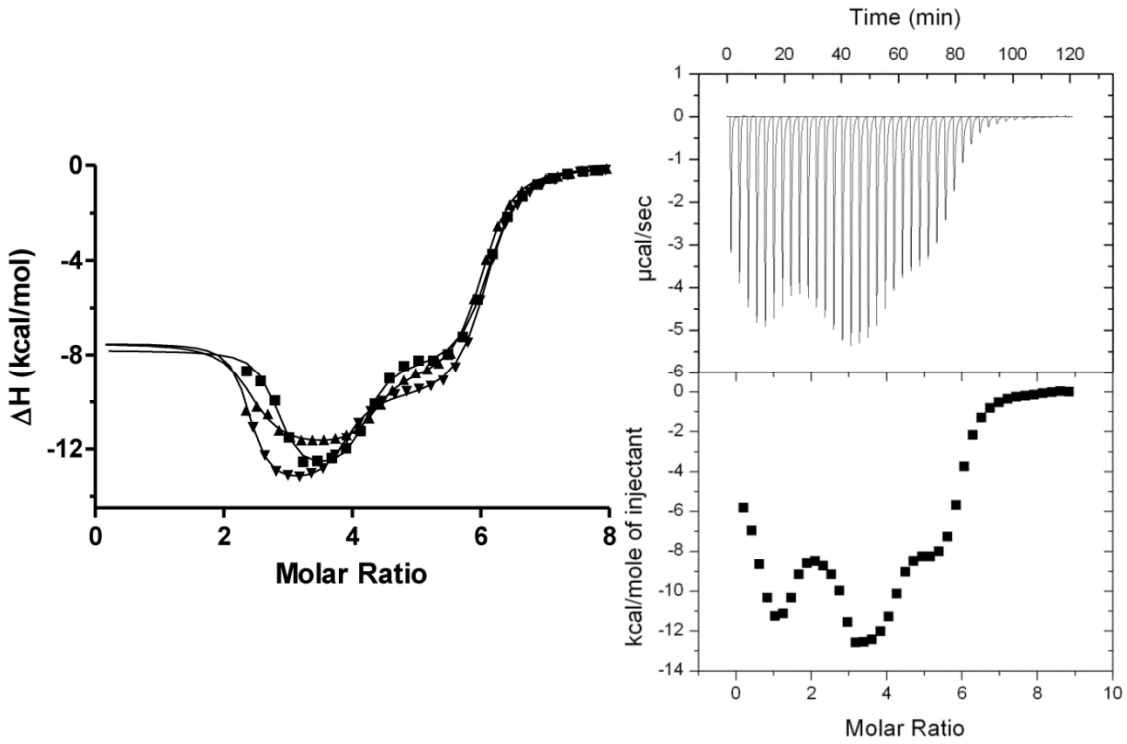


Figure 6.5 Titrations of TMPyP4 into 32-mer MWT•5T capped complex (-■-), 1:2:1•5T capped complex (▲), and 1:6:1•5T capped complex (▼).

Heat data between 0 and 2.0 Molar Ratio are attributed to ligand induced refolding of the capped DNA construct and therefore excluded from the fit. The solid lines represent the non-linear best fits. The right panel shows raw and integrated heats for a typical titration of TMPyP4 into 32-mer MWT•5T capped complex.

The derived thermodynamic parameters for TMPyP4 Mode I and Mode II binding to the 5-T capped constructs were obtained by fitting the corrected ITC data in Figure 6.5 to a three-process thermodynamic model. The ΔG_i, ΔH_i, and $-T\Delta S_i$ values for these systems are again given in Table 6.1.

Table 6.1 Thermodynamic parameters (in kcal/mol) for the titrations of TMPyP4 into 32-mer MWT, 1:2:1, and 1:6:1 G-quadruplexes in the absence and presence of the 5Tcap.

DNA Construct	ΔG_1	ΔH_1	$-T\Delta S_1$	ΔG_2	ΔH_2	$-T\Delta S_2$
32-mer MWT	-11.5±0.2	-7.8±0.4	-3.7	-8.4±0.1	-11.9±0.2	3.6
32MWT•5T cap	-13.2±0.5	-7.8±0.2	-5.4	-10.5±0.3	-13.3±0.6	2.8
1:2:1	-10.4±0.2	-7.5±0.3	-2.9	-8.4±0.0	-13.7±0.2	5.3
1:2:1•5T cap	-12.9±0.4	-7.5±0.1	-5.4	-10.4±0.2	-12.0±0.3	1.6
1:6:1	-11.2±0.2	-7.6±0.3	-3.6	-8.3±0.1	-13.0±0.1	4.7
1:6:1•5T cap	-13.5±0.4	-7.6±0.2	-5.9	-10.5±0.2	-13.6±0.3	3.1

Three significant differences become apparent when comparing the ITC data and derived thermodynamic parameters for the MWT constructs and data reported earlier for TMPyP4 titrations of ss c-MYC 27-mer WT and mutant promoter sequence G-quadruplexes (20). The first difference is that in the MWT titrations (for both uncapped and capped constructs), an additional heat producing process is observed in the early part of the titration that was not observed in titrations of the unmodified c-MYC sequences. This additional heat is attributed to ligand induced G-quadruplex stabilization or refolding. The implication is that the presence of modified the tail sequences, TTTAAA and AAATTT, located immediately before and immediately after the core G-quadruplex forming sequence results in some G-quadruplex unfolding. The excess heat observed in the MWT titrations, corresponds to renaturation of less than 10% of the total G-quadruplex in these solutions. The second difference is noted when comparing the Mode I and Mode II thermodynamic parameters for TMPyP4 binding to the uncapped MWT construct vs. TMPyP4 binding to the ss WT 27-mer G-quadruplex. The uncapped MWT constructs exhibit a more favorable free energy change for Mode I TMPyP4 binding

($\delta\Delta G_1 \approx$ -1.5 kcal/mol). In contrast, the free energy change for Mode II TMPyP4 binding is unchanged when comparing the capped MWT constructs to the ss WT 27-mer G-quadruplex. However, in titrations of the capped constructs, the free energy changes for both Mode I and Mode II TMPyP4 binding become significantly more favorable ($\delta\Delta G_1$ and $\delta\Delta G_2 \approx$ -2.0 kcal/mol). The last difference is that all three capped constructs exhibit a saturation stoichiometry of 6 moles of TMPyP4 bound per mole of the capped G-quadruplex DNA. The two additional binding sites, attributed to TMPyP4 interactions with the flanking duplex regions of the capped construct are 10^2 weaker than the lowest affinity G-quadruplex binding sites. Typical ITC data for TMPyP4 titrations of the 5-C capped MWT and mutant MWT constructs are shown in Figure A.3, with the derived thermodynamic parameters obtained for these systems give in Table A.1. Overall, the 5-C cap data are very similar to the 5-T cap data. One difference is that the 5-C cap appears to be more destabilizing to the G-quadruplex as evident from a larger heat of refolding in the early part of these ITC titrations (see Figure A.3). The excess heat in the 5-C titrations is consistent with the ligand induced refolding of as much as 40% of the G-quadruplex in these solutions.

Results of the docking studies, i.e. the non-bonded interaction energy scores for binding TMPyP4 to both the uncapped and capped MWT 1:2:1 and 1:6:1 G-quadruplex constructs are listed in Table 6.2. The calculated interaction energies (given in kcal/mol) are the estimated energy of binding (ΔH) for binding one molecule of TMPyP4 per one molecule of the specific G-quadruplex.

Table 6.2 Calculated interaction energies for binding TMPyP4 to the G-quadruplex DNA motif formed in c-MYC MWT mutant 1:2:1 and 1:6:1 sequences.

Loop isomer	Cap	Interaction mode	Electrostatic	VDW	Total
1:2:1	na	end-stacking	-17.0	-16.3	-33.3
		intercalation	-13.6	-60.3	-73.9
1:6:1	na	end-stacking	-10.8	-20.5	-31.3
		Intercalation	-9.6	-65.2	-74.8
1:2:1	5-T	end-stacking	-15.5	-15.9	-31.4
		intercalation	-15.5	-67.2	-82.7
	5-C	end-stacking	-18.1	-14.9	-33.0
		intercalation	-13.3	-70.6	-83.9
1:6:1	5-T	end-stacking	-11.9	-22.3	-34.2
		end-stacking with fold-back	-12.7	-78.7	-91.4
		intercalation	-10.9	-81.9	-92.8
	5-C	end-stacking	-11.6	-14.1	-25.7
		end-stacking with fold-back	-9.2	-61.3	-70.5
		intercalation	-11.9	-80.4	-92.3

These calculations employed the CHARM-27 force field and the TMPyP4 docking was done for both the uncapped and capped MWT mutant constructs in an explicit solvent environment.

Several observations can be made from the docking studies. First, docking free energies indicate that TMPyP4 interactions with the uncapped 1:2:1 and uncapped 1:6:1 mutant MWT are very similar to one another (regardless of interaction mode: end-stacking or intercalation). Second, the total non-bonded interaction energies for TMPyP4 end-stacking appear to be insensitive to G-quadruplex folding (e.g. 1:2:1 vs. 1:6:1) and to the presence of flanking duplex regions (e.g. uncapped vs. capped). The one exception appears to be TMPyP4 binding by the end-stacking mode to the capped 1:6:1 mutant MWT construct. In this case, the six base lateral loop can fold-over the end-stacked TMPyP4 increasing the binding affinity by from 45 to 57 kcal/mol. Third, docking scores for TMPyP4 binding via the end-stacking mode with lateral loop folded-back yield interaction energies that are similar to TMPyP4 binding by the intercalation mode.

Fourth, docking scores predict that TMPyP4 intercalation is more favorable than end-stacking in all cases and that TMPyP4 intercalation into the capped MWT mutants is more favorable than into the uncapped MWT mutants. Fifth, the docking energies are similar for binding TMPyP4 to either the 5-T or 5-C capped MWT mutants. Lastly, docking scores for binding TMPyP4 to the flanking duplex regions on either end of the capped G-quadruplex are much less favorable than binding TMPyP4 to the G-quadruplex DNA itself.

Molecular dynamics simulations were run to estimate the relative stabilities of the NHE-III_1 model constructs with and without bound ligand. Representative stable NHE-III_1 structures for the 5-T capped MWT 1:6:1 mutant construct are shown in Figure 6.6. (The structures shown in Figure 6.6 were observed at the end (last 5 ps) of a 100 ps production period in a typical MD simulation run.) Again, several observations can be made from these MD simulations. First, trajectory analysis of the entire 100 ps production period reveals that base paring in the flanking B-DNA regions is loosened up during the course of the MD simulation. The flanking duplex regions in the capped 1:2:1 MWT mutant construct end up having only 4 bases (out of 6 bases) paired in the 5'-tail region and 5 bases (out of 6 bases) paired in the 3'-tail region. In contrast, the flanking duplex regions in the capped 1:6:1 MWT mutant construct end up having 5 bases paired in both the 5'and 3'-tail regions. Second, the polypurine G-quadruplex forming strand is less dynamic than the polypyrimidine strand in all of the constructs. Third, the large six-base-lateral loop in the capped 1:6:1 mutant MWT is the most dynamic part of the NHE-III_1 model structure. Lastly, the motion of the six-base lateral loop becomes very

restricted only when TMPyP4 is bound to either end of the G-tetrad stack and the six-base loop is folded back over the end-bound ligand.

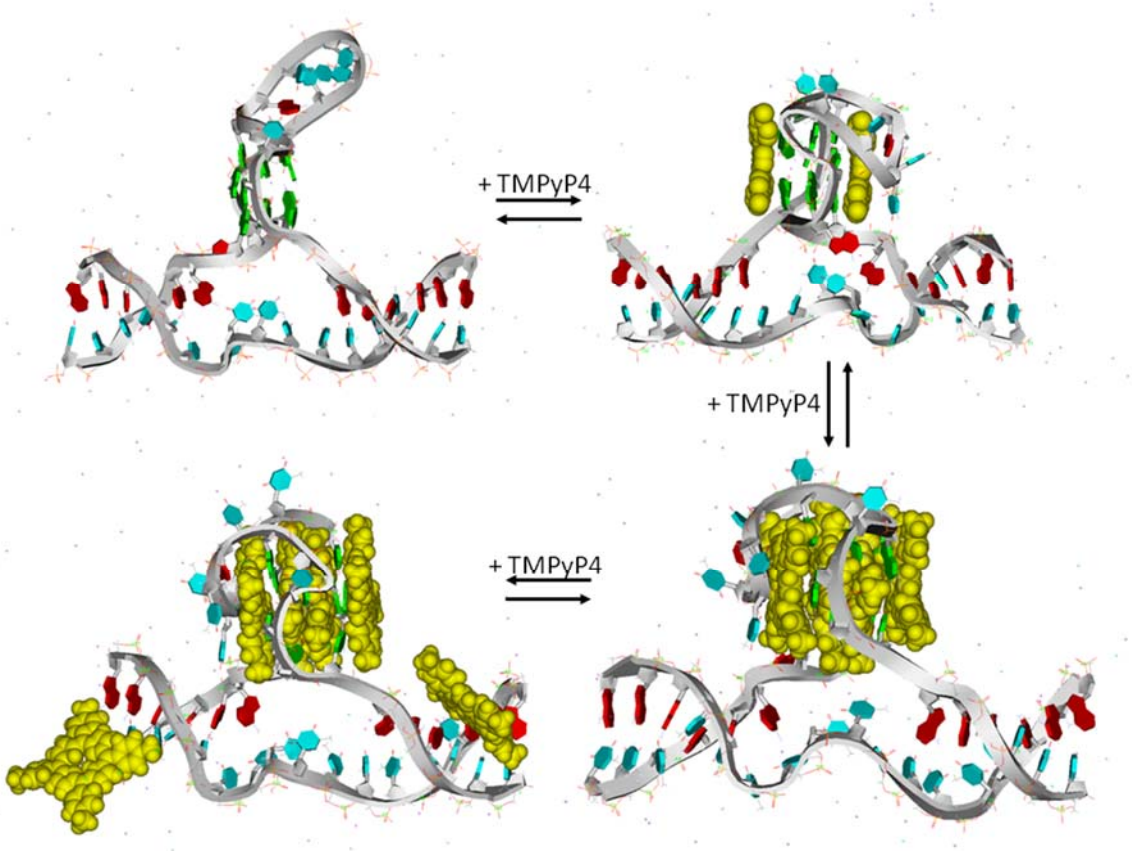


Figure 6.6 Proposed binding mechanism for TMPyP4 interactions with 32-mer MWT•5T.

6.5 Discussion

The first objective of the work described here was to assemble a model NHE-III$_1$ construct. The model NHE-III$_1$ was assembled by annealing a 17-mer oligonucleotide (short polypyrimidine strand) having a mid-sequence run of 5 T's (or C's) and tail sequences of 3'-AAATTT and TTTAAA-5', to a 32-mer oligonucleotide (polypurine

strand) having a 20-mer c-MYC G-quadruplex forming core and complimentary tail sequences of 5'-TTTAAA and AAATTT-3'. ESI mass spectra for the model NHE-III$_1$ motif demonstrated the presence of the typical c-MYC promoter sequence G-quadruplex in the new double stranded structure. The addition of sticky ends on both the 5' and 3' ends of the 20-mer c-MYC WT G-quadruplex forming core was necessary for the assembly of our model NHE-III$_1$. Several factors were taken into consideration when designing the sticky end sequences. We elected to eliminate guanines in both the 5' and 3' tails to avoid introducing additional complexity into the ensemble of G-quadruplex folded species. The sticky end sequences, 5'-TTTAAA and AAATTT-3', were selected in part to prevent hairpin formation. Both sticky end sequences included three point mutations in comparison to the c-MYC WT 32-mer. The three mutations in the 5'-tail were: **C→T**, **T→A**, and **G→A**, while the three mutations in the 3'-tail were: **G→A**, and two **G→T** mutations. The minimum number of bases in each of the sticky ends was chosen to be at least equal to six bases, the number of unstructured bases in the largest bridging loop (e.g. 6 T's or 6C's) of the 17-mer capping sequence. The ESI mass spectrum showed that while the un-annealed 17-mer capping and un-annealed 32-mer MWT and mutant MWT sequences exhibited some end to end polymerization (see unassigned peaks in Figure 6.1A and 6.1B), the annealed NHE-III$_1$ construct exhibited a single species with a mass consistent for a 1:1 base paired complex of the 17-mer and 32-mer constructs, consistent with the correct pairing of the sticky ends. The ESI-MS experiments (taken together with results from the CD, DSC, and ITC experiments discussed below) provide direct evidence that a model NHE-III$_1$, containing an ensemble

of G-quadruplex conformers and flanked on both the 5' and 3' sides by duplex B-DNA had been assembled.

The CD experiments shown in both Figure 6.2 Panel A and B clearly indicate that the capping construct has little influence on the G-quadruplex structure of the model NHE-III$_1$. However, the DSC results indicates a shift in equilibrium favoring the 1:6:1 loop isomer over 1:2:1 loop isomer in the model NHE-III$_1$. The challenge is that both 1:6:1 loop isomer and 1:2:1 loop isomer have the same strand orientation (parallel) resulting in a indistinguishable CD spectra. In fact, for the human telomere sequence G-quadruplex, both the basket form and the chair form share the same CD spectrum for anti-parallel G-quadruplex (29). In other word, because CD spectroscopy is not sensitive to changes in loop isomerization, the shift in equilibrium favoring 1:6:1 loop isomer is not detected during CD experiment.

A shift in the predominant folded structure of the c-MYC P1 promoter G-quadruplex was observed in the 32-mer c-MYC G-quadruplex capped with either the 5-T or 5-C capping construct. Previous studies concur that the c-MYC promoter sequence is capable of forming multiple conformational isomers (11,22). The conformational isomers were reported by Patel to have the folding patterns of either 1:2:1 bases or 1:6:1 bases within the connecting loops of the G-quadruplex (11). These structures were studied in depth through CD and DSC experiments. The 1:6:1 folding pattern correlates with the lower melting transition and the 1:2:1 folding pattern coordinates with the higher melting transition (12,20). Although the T_m values for the uncapped 32-mer 1:6:1 and 1:2:1 mutants MWT (55.2 and 66.8 °C respectively) are much lower than T_m values reported previously for the 1:6:1 and 1:2:1 mutants for ss WT c-MYC G-quadruplex

(76.4 and 91.2 °C respectively), there is clearly a systematic decrease in melting temperature ($\Delta T_{m,average} \approx 23°C$) for both the uncapped 32-mer 1:6:1 and 1:2:1 mutants MWT. As previously mentioned, this systematic destabilization must be attributed to the presence of the sticky ends which place AAA sequences immediately before and after the core G-quadruplex. To further support the deleterious effect of placing AAA adjunct to the G4 core, we melted another mutant oligonucleotide namely MWT[T_6] 32-mer. This mutant has the same 20-mer c-MYC G-quadruplex core flaked by 6T's on both the 5' end and 3' ends. As shown in Table 6.3, MWT[T_6] 32-mer melts at 69 and 83 °C, exactly 10°C less than the melting temperature for both folded species (1:2:1 and 1:6:1) in the WT 24-mer ensemble.

Table 6.3 DSC determined melting temperatures, T_ms, for all c-MYC P1 DNA constructs in the presence and absence of capping strands.

c-MYC P1 Construct	T_m1	T_m2
WT 24-mer *	79	93
mut 1:2:1 24-mer *	-	90
mut 1:6:1 24-mer *	78	-
WT 27-mer #	76	91
WT 32-mer	75	89
MWT[T_6] 32-mer	69	83
MWT[A_3T_3] 32-mer	61	67
mut 1:2:1[A_3T_3] 32-mer (w/ wo cap)	-	67
mut 1:6:1[A_3T_3] 32-mer (w/ wo cap)	55-57	-
MWT[A_3T_3]32-mer • 5C cap	61	67
MWT[A_3T_3] 32-mer • 5T cap	63	69
WT 36-mer ‡	75	89
WT 36-mer/4T cap ‡	75	89

* These T_m values have been reported by Freyer *et al.* (20)
# These T_m values have been reported by Dettler *et al.* (23)
‡ These T_m values were obtained from preliminary data which are shown in Figure S4

With the addition of a "capping" construct, an obvious change in the relative equilibrium of the folded structures in solution is observed. Deconvolution of the 32-mer MWT melting curve suggests that the 1:2:1 folded species contributes 90% of the total population while the 1:6:1 only contributed 10%. However, upon complexation with the capping strand, the new relative population distribution is 40% for 1:2:1 and 60% for 1:6:1. This shift in equilibrium is most likely due to structural rearrangement to release strain imposed on the DNA when the capping construct is added.

There is an important feature that needs to be emphasized from DSC thermograms as shown in Figure 6.3. The thermal unfolding of the "capped" vs. the "uncapped" MWT are visually distinguishable by the shapes of the melting curves: the melting curve for "capped" complex is more narrow than the "uncapped" . The melting of the flanking duplex A_3T_3 regions are not shown for clarity purpose. An important question needs to be raised is *"Upon melting the flanking duplex regions, why does the equilibrium favoring 1:6:1 over 1:2:1 not return back to [90%-1:2:1 and 10%-1:6:1] distribution such as observed in the 'uncapped' MWT?"*.

To answer this question, we look at physical changes of the DNA species as the temperature changes during the DSC experiments from 10 to 100 °C. At 30 °C, both duplex regions on either end of the "capped" MWT are melted, resulting in a "nascent" single stranded form of the MWT. However, unlike the "uncapped" MWT-32mer, this "nascent" ss MWT predominantly exists in 1:6:1 conformation. One plausible explanation for why the 1:6:1 G4 folded population is still accessible well after flanking duplex removal is that the unfolding rate of this folded species is slower than the scanning rate of the DSC experiment. Based on a scan rate of 90 °C/hour and the melting

curve for "capped" MWT as shown in Figure 6.3B, the duration between completely melting the duplex ends and thermally unfolding the 1:6:1 foldamer is less than 30 minutes. The real question now is whether there is any evidence showing the interconversion rate of 1:6:1 foldamer to 1:2:1 foldamer on the order of 30 minutes or 1800 seconds time scale.

Single molecule FRET is employed by many research groups to measure the number of folded and unfolded species of G-quaduplex in a particular solution condition. Lee *et al.* used smFRET and observed interconversion between three G-quadruplex conformations coexist in the human telomere repeat $(GGGTTA)_3GGG$ under low $[K^+]$ solution condition (42). Each observed conformation is further divided into having a long-lived and short-lived species. The reported lifetime for some of these long-lived intermediate species are in excess of 250 seconds (42). In 2010, Chaires reported a thermodynamic and kinetic study on the same telomeric repeat, and reported three distinct intermediate species linking two folded structures of the Na^+ form and the K^+ form telomeric quadruplex (43). The third intermediate species has a dwelling time of over 600 seconds before rearranging to fold into the K^+ form quadruplex structure (43). Two years later, Gray *et al.* used CD and FRET in combination with SVD analysis to perform a mechanistic study of the same human telomere. They reported a sequential unfolding model with two intermediate species, one of which involves a triple-helical structure (44). In the most recent study, the same group demonstrated that intermediate species (e.g. I_2, I_3) can take anywhere between 10^2 to 10^4 seconds for the rearranging process (e.g. $I_2 \rightarrow I_3$ takes 3700 s, $I_3 \rightarrow F$ takes 760 s) (45).

Even though there hasn't been any similar study performed on the c-MYC sequence G-quadruplex, it would not be a stretch to imagine c-MYC would behave in some way similarly to human telomere quadruplex in terms of their polymorphism along the unfolding pathway.

The ultimate goal of this study was to develop a double stranded model for c-MYC P1 promoter G-quadruplexes and the investigation of its structural complexity, stability and ligand interactions. This has been accomplished via our model NHE-III_1 by introducing shorter partial complementary strand to the G-quadruplex forming strand. So far, evidences spectroscopic (CD) and the calorimetric (DSC) studies have shown an slight increase in the stability of the G-quaduplex as a well as a structural selection in favor of a "previously-thought" less stable conformation: the 1:6:1 loop isomer. We expected that ligand binding experiment would exhibit some small increases in binding affinities (dictated by unanimously increase in entropy contribution due to enhancement in stability of the G-quadruplex), yet we found the results of the binding study for the capped constructs as a model for NHE-III_1 rather interesting and more complex than anticipated.

The first notable difference between the capped constructs ITC thermogram in comparison to the uncapped constructs is the presence of some additional heats in the early part of the titration which we attributed to ligand induced refolding (LIR). LIR process observed in the ITC experiments caused by the 5-T versus the 5-C caps to the quadruplex is different depending on how the bridging bases (5-T's vs. 5-C's) in the capping sequence affect the structural stability of the G-quadruplex. It is very likely that the 5-C cap destabilizes the G-quadruplex structure by pulling out structured guanines

from the quadruplex core and forming G-C base pairing. A decrease in the concentration of the folded G-quadruplex species was observed with the addition of the 5-C capping construct in comparison to the 5-T capping construct (see Supplementary Figure A.5). The overall concentration of the oligo remains the same; however, the concentration of the folded species in solution is decreased in comparison. The decreased concentration of folded G-quadruplex species may correspond to the degree of additional heats generated by LIR process as seen in the early parts of these ITC thermograms. It is speculated that 5-C capped oligos without TMPyP4 present show only a small fraction of folded G-quadruplex which can then be increased with the addition of a G-quaduplex stabilizing ligand such as TMPyP4. The presence of TMPyP4 is thought to increase the structural order of the G-quadruplex, therefore possibly increasing the binding affinities for those capped sequences. This increase in G-quadruplex formation can be seen in the integrated heat data as the first peak that is not fit to be a binding site. Therefore the integrated heat data for the capped oligos can be described overall as a total of four reactions comprising of an initial increased formation of G-quadruplex followed by three binding reactions (end binding, intercalation, and duplex binding) as shown in Figure 6.6. The proposed binding mechanism as shown in Figure 6.6 provides a plausible explanation for the titration data obtained in this study.

ITC data indicate that again the capped ds G-quadruplex structures exhibits four TMPyP4 binding sites for the core G-quaduplex. The four sites are predicted to be two end binding sites and two intercalation binding. The affinity for binding TMPyP4 to the capped G-quadruplex sites is significantly greater than for binding TMPyP4 to the uncapped G-quadruplex sites (i.e. $K_{i,\ capped} \approx 10^2 \times K_{i,\ uncapped}$, where i = 1 or 2). This may

be due, at least in part, to an increase in the stability of the G-quadruplex motif in the capped constructs as pointed out in previous CD and DSC discussion. The increase in TMPyP4 binding affinity for 'end-binding' mode, is solely the result of a more favorable entropy change for the Mode I interaction with the capped G-quadruplex. And since the enthalpy change for Mode I binding is the same for both the capped and uncapped G-quadruplexes, we speculate that this may be the result of the entropy lost upon formation of the capped construct being returned as a larger entropy change for the release of additional waters or some flexibility gained in the DNA upon complex formation. On the contrary, the increase in TMPyP4 binding affinity for intercalation is more complicated. Mode II affinity is increased for the capped G-quadruplex due to a more favorable entropy change in all cases (with $-T\Delta S$ being more favorable by from 0.8 to 3.7 kcal/mol). For capped MWT and capped 1:6:1 TMPyP4 binding experiments, the ΔH values for TMPyP4 binding via intercalations are also more favorable by from 0.6 to 1.4 kcal/mol for binding TMPyP4 to the capped MWT or to the capped 1:6:1 construct in comparison to their uncapped counterparts. However, the enthalpy change is less favorable by 1.7 kcal/mol for binding TMPyP4 to the capped 1:2:1 construct. Is this the result of a poorer fit between the ligand and a strained 1:2:1 conformation in the capped construct ?

Results from the docking study and MD simulation are generally in agreement with the biophysical results performed in this study. Docking studies showing very similar results with the uncapped 1:2:1 and uncapped 1:6:1 are in excellent agreement with our previously published ITC binding study (23). In this study, two mutants 1:2:1 and 1:6:1 were used to represent the two folded species in the 24-mer WT c-MYC

sequence G-quadruplex. ITC results performed at 25 °C and in 130 mM [K^+] showing that binding of TMPyP4 to either 1:2:1 or 1:6:1 loop isomer are very similar binding free energy and enthalpy changes for both end-stacked and intercalative sites (23). The increases in the interaction energies for TMPyP4 intercalating into "capped" MWT over "uncapped" is observed experimentally in the current ITC study. Perhaps, the presence of the flanking duplex regions "tighten up" the core quadruplex structure and leads to increases in the intercalative interactions. Furthermore, docking energies for binding TMPyP4 to the flanking duplex regions also corroborate with result from ITC study. One caveat when using docking scores to predict ligand interactions with G-quaduplex is that the interaction energy does not include entropy calculation rendering these docking scores having limited implications when compared to thermodynamic values determined experimentally.

In 2009, Onyshchenko *et al.* used a short peptide nucleic acids (PNAs) that bind to the complementary C-rich strand leaving the G-rich strand of the Bcl-2 capable of forming G-quaduplex structure (46). One caveat for this technique is that due to the complementarity between the G-rich and C-rich strand as well as C-rich and PNAs, this competition effect that will cause some destabilization of the G-quadruplex. In fact, we have performed biophysical analysis on a similar Bcl-2 sequence and reported T_m values for at least three folded species for the Bcl-2 sequence ensemble. These T_m values are approximately 68, 75, and 90 °C (47). However, using UV melt experiments, Onyshchenko *et al.* only observed a single thermal transition at ~ 60°C (46). Two years later, Phan and colleagues reported an NMR structure for RGG peptide bound to a RNA duplex-quadruplex junction (48). Although this model is also useful, it has some

limitations due to presence of only one duplex-quadruplex junctions as supposed to two duplex-quadruplex in our model NHE-III_1. Furthermore, binding of G-quadruplex interactive ligand is sterically hindered at this duplex-quadruplex junction. To the best of our knowledge, this is the first report for successful assembly of a G-quaduplex structure in double stranded environment but more importantly the first in-depth biophysical characterization of this model NHE-III_1. This model NHE-III_1 will prove to be useful to for designing anti-cancer drugs to folded G-quadruplex structures in its native environment.

6.6 References

1. Hurley, L.H. (2001) Secondary DNA structures as molecular targets for cancer therapeutics. *Biochemical Society transactions*, **29**, 692-696.

2. Mergny, J.L., Riou, J.F., Mailliet, P., Teulade-Fichou, M.P. and Gilson, E. (2002) Natural and pharmacological regulation of telomerase. *Nucleic Acids Res*, **30**, 839-865.

3. Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res*, **33**, 2901-2907.

4. Marcu, K.B., Bossone, S.A. and Patel, A.J. (1992) myc function and regulation. *Annual review of biochemistry*, **61**, 809-860.

5. Pelengaris, S., Rudolph, B. and Littlewood, T. (2000) Action of Myc in vivo - proliferation and apoptosis. *Current opinion in genetics & development*, **10**, 100-105.

6. Slamon, D.J. and Cline, M.J. (1984) Expression of cellular oncogenes during embryonic and fetal development of the mouse. *Proc Natl Acad Sci U S A*, **81**, 7141-7145.

7. Hazel, P., Huppert, J., Balasubramanian, S. and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J Am Chem Soc*, **126**, 16405-16415.

8. Dexheimer, T.S., Sun, D. and Hurley, L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J Am Chem Soc*, **128**, 5404-5415.

9. Kumar, P., Verma, A., Maiti, S., Gargallo, R. and Chowdhury, S. (2005) Tetraplex DNA transitions within the human c-myc promoter detected by multivariate curve resolution of fluorescence resonance energy transfer. *Biochemistry*, **44**, 16426-16434.

10. Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J Am Chem Soc*, **126**, 8710-8716.

11. Seenisamy, J., Rezler, E.M., Powell, T.J., Tye, D., Gokhale, V., Joshi, C.S., Siddiqui-Jain, A. and Hurley, L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J Am Chem Soc*, **126**, 8702-8709.

12. Grand, C.L., Powell, T.J., Nagle, R.B., Bearss, D.J., Tye, D., Gleason-Guzman, M. and Hurley, L.H. (2004) Mutations in the G-quadruplex silencer element and their relationship to c-MYC overexpression, NM23 repression, and therapeutic rescue. *Proc Natl Acad Sci U S A*, **101**, 6140-6145.

13. Facchini, L.M. and Penn, L.Z. (1998) The molecular role of Myc in growth and transformation: recent discoveries lead to new insights. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **12**, 633-651.

14. Spencer, C.A. and Groudine, M. (1991) Control of c-myc regulation in normal and neoplastic cells. *Advances in cancer research*, **56**, 1-48.

15. Siebenlist, U., Hennighausen, L., Battey, J. and Leder, P. (1984) Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in Burkitt lymphoma. *Cell*, **37**, 381-391.

16. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, **35**, 406-413.

17. Simonsson, T., Pribylova, M. and Vorlickova, M. (2000) A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochemical and biophysical research communications*, **278**, 158-166.

18. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*, **99**, 11593-11598.

19. Simonsson, T., Pecinka, P. and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res*, **26**, 1167-1172.

20. Freyer, M.W., Buscaglia, R., Kaplan, K., Cashman, D., Hurley, L.H. and Lewis, E.A. (2007) Biophysical studies of the c-MYC NHE III1 promoter: model quadruplex interactions with a cationic porphyrin. *Biophys J*, **92**, 2007-2015.

21. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*, **34**, 5402-5415.

22. Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048-2058.

23. Dettler, J.M., Buscaglia, R., Le, V.H. and Lewis, E.A. (2011) DSC deconvolution of the structural complexity of c-MYC P1 promoter G-quadruplexes. *Biophys J*, **100**, 1517-1525.

24. Parkinson, G.N. (2006) *Fundamentals of Quadruplex Structures. Quadruplex Nucleic Acids*. The Royal Society of Chemistry, Cambridge.

25. Lu, M., Guo, Q. and Kallenbach, N.R. (1993) Thermodynamics of G-tetraplex formation by telomeric DNAs. *Biochemistry*, **32**, 598-601.

26. Balagurumoorthy, P. and Brahmachari, S.K. (1994) Structure and stability of human telomeric sequence. *J Biol Chem*, **269**, 21858-21869.

27. Balagurumoorthy, P., Brahmachari, S.K., Mohanty, D., Bansal, M. and Sasisekharan, V. (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res*, **20**, 4061-4067.

28. Lu, M., Guo, Q. and Kallenbach, N.R. (1992) Structure and stability of sodium and potassium complexes of dT4G4 and dT4G4T. *Biochemistry*, **31**, 2455-2459.

29. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res*, **34**, 2723-2735.

30. Kaushik, M., Bansal, A., Saxena, S. and Kukreti, S. (2007) Possibility of an antiparallel (tetramer) quadruplex exhibited by the double repeat of the human telomere. *Biochemistry*, **46**, 7119-7131.

31. Risitano, A. and Fox, K.R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry*, **42**, 6507-6513.

32. Grand, C.L., Han, H., Munoz, R.M., Weitman, S., Von Hoff, D.D., Hurley, L.H. and Bearss, D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol Cancer Ther*, **1**, 565-573.

33. Han, H. and Hurley, L.H. (2000) G-quadruplex DNA: a potential target for anti-cancer drug design. *Trends Pharmacol Sci*, **21**, 136-142.

34. Plum, G.E. (2001) Optical methods. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]*, **Chapter 7**, Unit 7 3.

35. Le, V.H., Nagesh, N. and Lewis, E.A. (2013) Bcl-2 Promoter Sequence G-Quadruplex Interactions with Three Planar and Non-Planar Cationic Porphyrins: TMPyP4, TMPyP3, and TMPyP2. *PloS one*, **8**, e72462.

36. Le, V.H., Buscaglia, R., Chaires, J.B. and Lewis, E.A. (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Anal Biochem*, **434**, 233-241.

37. Foloppe, N. and MacKerell, J., A.D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem*, **21**, 86-104.

38. MacKerell, J., A.D. and Banavali, N.K. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comp. Chem*, **21**, 105-120.

39. Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys*, **23**, 327-341.

40. Cashman, D.J., Buscaglia, R., Freyer, M.W., Dettler, J., Hurley, L.H. and Lewis, E.A. (2008) Molecular modeling and biophysical analysis of the c-MYC NHE-III1 silencer element. *Journal of molecular modeling*, **14**, 93-101.

41. Miyahara, T., Nakatsuji, H. and Sugiyama, H. (2013) Helical structure and circular dichroism spectra of DNA: a theoretical study. *The journal of physical chemistry. A*, **117**, 42-55.

42. Lee, J.Y., Okumus, B., Kim, D.S. and Ha, T. (2005) Extreme conformational diversity in human telomeric DNA. *Proc Natl Acad Sci U S A*, **102**, 18938-18943.

43. Chaires, J.B. (2010) Human telomeric G-quadruplex: thermodynamic and kinetic studies of telomeric quadruplex stability. *The FEBS journal*, **277**, 1098-1106.

44. Gray, R.D., Buscaglia, R. and Chaires, J.B. (2012) Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *J Am Chem Soc*, **134**, 16834-16844.

45. Gray, R.D., Trent, J.O. and Chaires, J.B. (2014) Folding and Unfolding Pathways of the Human Telomeric G-Quadruplex. *Journal of molecular biology*.

46. Onyshchenko, M.I., Gaynutdinov, T.I., Englund, E.A., Appella, D.H., Neumann, R.D. and Panyutin, I.G. (2009) Stabilization of G-quadruplex in the BCL2 promoter region in double-stranded DNA by invading short PNAs. *Nucleic Acids Res*, **37**, 7570-7580.

47. Dettler, J.M. and Lewis, E.A. (2011) In Winkle, R. W. S. a. S. A. (ed.), pp. 33-50.

48. Phan, A.T., Kuryavyi, V., Darnell, J.C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D. *et al.* (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nature structural & molecular biology*, **18**, 796-804.

CHAPTER VII

A CALORIMETRIC INVESTIGATION OF THE HOST-GUEST INTERACTIONS IN FULLERENE BUCKYCATCHER COMPLEXES

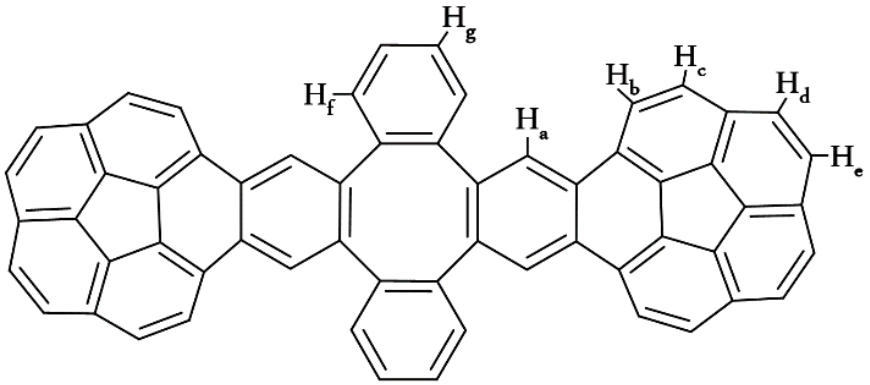
7.1 Abstract

Reliable thermodynamic data are needed to develop better models for computing the energetics of supramolecular assemblies. In this study, we used NMR and isothermal titration calorimetry (ITC) to determine the association constants, K_a, for the formation of the 1:1 complexes of C_{60} or C_{70} buckyballs with $C_{60}H_{28}$ buckycatcher. NMR measurements were done in toluene-d_8 and chlorobenzene-d_5 at 288, 298, and 308 K. ITC measurements were done in toluene, chlorobenzene, *o*-dichlorobenzene, anisole, and 1,1,2,2-tetrachloroethane from 278 to 323 K. K_a values obtained with both techniques are in excellent agreement. Van't Hoff enthalpy changes were determined from the temperature dependence of the NMR K_a values. K_a as well as ΔH, and $-T\Delta S$ values were determined in the ITC experiments. The NMR and ITC ΔH values are also in good agreement at the mid-temperature (298 K). ΔC_p values were calculated from the temperature dependence of the calorimetric ΔH values. The temperature dependence of the thermodynamic data exhibit classical enthalpy/entropy compensation, a phenomena that has been widely reported for complexation in aqueous solutions but has only rarely been reported in organic solutions. The small favorable ΔG values for

buckyball•buckycatcher complexation, result from small differences in the solvation energies for the buckyball and buckycatcher.

7.2 Introduction

The discovery of fullerenes and other forms of elemental carbon with curved surfaces introduced a novel aspect of supramolecular assembly based on the relatively weak dispersion forces between the convex surfaces of the conjugated carbon networks and the appropriate molecular receptors. Buckybowls, curved-surface polycyclic aromatic hydrocarbons (PAH) structurally related to fullerenes, appear to be good candidates for receptors due to the complementarity of their accessible concave surfaces with the convex surfaces of the fullerenes. While supramolecular assemblies of fullerenes with the smallest buckybowl corannulene have not been detected in solution, we have shown that molecular receptors for both C_{60} and C_{70} can be constructed if at least two corannulene pincers are pre-organized on a proper tether(1). The buckycatcher $C_{60}H_{28}$ (referred from hereon as Clip) consisting of two corannulene subunits on a tetrabenzocyclooctatetraene tether was shown to form inclusion complexes with fullerenes in both solid state and in solution. (1,2) The chemical structure of the buckycatcher is shown in Scheme 7.1, which also labels the seven different types of protons in the Clip.



Scheme 7.1 Chemical structure for the $C_{60}H_{28}$ Clip. One of the four equivalent protons for each of the seven different Clip protons are labeled (a-g). The chemical shift of proton H_b was monitored in the NMR titration experiments.

In 2007, Sygula *et al.* reported the crystallographic structure for the inclusion complex C_{60}•Clip. They also reported an NMR determined association constant for formation of the C_{60}•Clip complex at 298 K.(1) Since that time, the C_{60}•Clip inclusion complex has become a prototypical system for large dispersion-driven supramolecular systems and, as such, has been the subject of several computational studies performed at various levels of theory.(3,4)

Theoretical gas-phase binding energies for formation of the C_{60}•Clip complex vary dramatically thus emphasizing the difficulty of accurately computing the energetics of dispersion forces. Hartree-Fock based calculations as well as several commonly employed DFT functionals predict either repulsion or negligible binding energies for the assembly(5) while the dispersion-sensitive DFT functionals predict strong gas-phase binding energies in the range -20 to -44 kcal/mol.(2,3,6) Several recent studies predict binding energies in the range of -30 to -36 kcal/mol.(7) In 2007, Sygula *et al.* reported a crystal structure for the inclusion complex C_{60}•Clip.[1] They also reported both theoretical

and experimental ΔG values for a series of inclusion complexes (including C_{60}•Clip and C_{70}•Clip).(1,2) The calculated ΔG values for formation of these complexes in solution differed on average by only ± 2 kcal/mol from the experimental data.(1,2) Considering the simplicity of the computational model used, the agreement between the computational and experimental results was quite impressive. However, upon closer inspection serious limitations of the model were revealed, e.g. Grimme's model, over-estimated the binding energy for both C_{60} and C_{70} in toluene by -3.3 and -3.9 kcal/mol respectively.(1,2) These differences are significantly larger than the average error for the series of inclusion complexes. Obviously, a larger set of experimental thermodynamic data is needed to assess the accuracy of the computational methods and to improve the theoretical models used to describe solvation.

In this work, we have used isothermal titration calorimetry (ITC) to further explore the energetics for complex formation between C_{60} or C_{70} with the $C_{60}H_{28}$ Clip. The use of the ITC method allowed us to obtain a complete set of thermodynamic parameters (K_a (or ΔG), ΔH, and -TΔS) for formation of the C_{60}•Clip and C_{70}•Clip complexes in a number of different solvents and at a number of different temperatures. We also repeated some of the earlier NMR titrations at lower concentration and at three temperatures. This allowed us to calculate van't Hoff enthalpy changes from the NMR data for comparison with the calorimetric enthalpy changes. The heat capacity changes, ΔC_p, for formation of the two complexes (C_{60}•Clip and C_{70}•Clip) were obtained from the temperature dependence of the calorimetric enthalpy changes.

In all cases the observed interactions were weak, $K_a<10^4$, as expected for systems where the assembly is driven entirely through dispersion forces. Surprisingly for most

systems, the enthalpy change, ΔH, appears to be the major contributor to the free energy change, ΔG, for complex formation. While the entropic contribution to the free energy change is typically smaller than the contribution from the enthalpy change, the –TΔS values are typically negative (favorable). This result is in sharp contrast to the large entropic penalty predicted for the formation of these complexes.[12] As expected, the binding equilibrium is perturbed by changes in temperature as well as the use of different organic solvents. It turned out that the enthalpy term was perfectly compensated by entropy term producing a small net change in binding free energy; a well-known phenomenon known as enthalpy-entropy compensation.

7.3 Material and methods

7.3.1 Sample Preparation

C_{60} and C_{70} were obtained from SES Research (Houston, TX). The buckycatcher (i.e. Clip) was synthesized in our laboratory according to the procedure we previously reported.(1) Anhydrous toluene, chlorobenzene, and *o*-dichlorobenzene were obtained from Sigma Aldrich (St. Louis, MO). Anhydrous 1,1,2,2-tetrachloroethane and anisole were obtained from VWR (Radnor, PA). Toluene-d_8 and chlorobenzene-d_5 were obtained from Cambridge Isotope Laboratories (Tewksbury, MA). Solutions of fullerenes and Clip were prepared at varying concentrations ranging from 0.07 to 1.0 mM [fullerene] titrate concentration, and from 0.7 to 10 mM for the [Clip] titrant solution.

7.3.2 NMR titrations

NMR titrations were performed according to the procedure reported previously(2) but at lower concentrations of both fullerenes and Clip. Several proton peaks on the Clip

exhibited measureable changes in chemical shift upon complexation with either C_{60} or C_{70} fullerenes. The association constant K_a was determined from solving the following equation (Eq. 7.1):

$$\Delta\delta = \frac{L(1+K_aX+K_aY) - \sqrt{(L^2(1+K_aX+K_aY)^2 - 4K_a^2XYL^2}}{2K_aY} \quad (7.1)$$

where: $X = [C_{60}]_{total}$; $Y = [Clip]_{total}$; and $L = \Delta\delta_{max}$ (i.e. $\Delta\delta$ at 100% complexation). Values of K_a and L were obtained from the non-linear regression using the curve-fitting tools of the Origin v.8.5 (Northampton, MA).

7.3.3 Isothermal titration calorimetry

ITC experiments were performed using a Microcal-GE (Northampton, MA) VP-ITC. Titrations were typically done at temperatures ranging from 278 K to 323 K and involved overfilling the ITC cell with ~1.5 ml of fullerene solution (C_{60} or C_{70}) and adding as many as 20 injections (14 μL each) of the Clip titrant solution. Typically three replicate measurements were performed. The raw calorimetric data were corrected for the heat of dilution of the Clip and Buckyball by subtracting the heats from the appropriate blank titrations. These heats were negligible in comparison to the binding interaction heats. Corrected ITC titrations were fit with a nonlinear regression algorithm using the CHASM© ITC data analysis program developed in our laboratory (8).

7.3.4 Atmospheric Pressure Photoionization Mass Spectrometry

APPI-MS experiments on buckyball and Clip complexes were carried out on a Bruker (Billerica, MA) MicrOTOFQ mass spectrometer. Data acquisition was set to operate in positive ion mode. All experiments were performed in 99.99% anhydrous

toluene. The buckyball solutions were prepared at a concentration of approximately 100 μM in anhydrous toluene. Solutions of the Clip were prepared at concentration as high as 300 μM. The APPI-MS samples were prepared by mixing the clip and buckyball solutions to yield a mixture containing an excess of Clip per mole of buckyball. The MS capillary voltage was set to +4500 V, dry N_2 gas flow was adjusted to 12 L/min at 453 K, and the Buckyball/Clip samples were directly infused into the MS by using a kD Scientific syringe pump set to a flow rate of 200 μL/h. Data processing was performed by using Bruker Daltonics Data Analysis program.

7.4 Results

Upon the addition of C_{60} to a solution of the Clip, several Clip proton NMR peaks exhibit measurable changes in their chemical shifts. Although the NMR signals for several of the Clip protons are changed upon fullerene complex formation, we have focused on the changes in the NMR signal for Clip proton ^1H_b (see Scheme 7.1) to construct the Job plot shown in Figure 7.1. Following the changes in chemical shift for proton ^1H_b and using the method of continuous variation, the maximum change in chemical shifts is observed at or near a mole fraction, [Clip]/([Clip]+[C_{60}]), of 0.5. This analysis is consistent with a saturation stoichiometry of 1:1 for formation of the C_{60}•Clip complex.

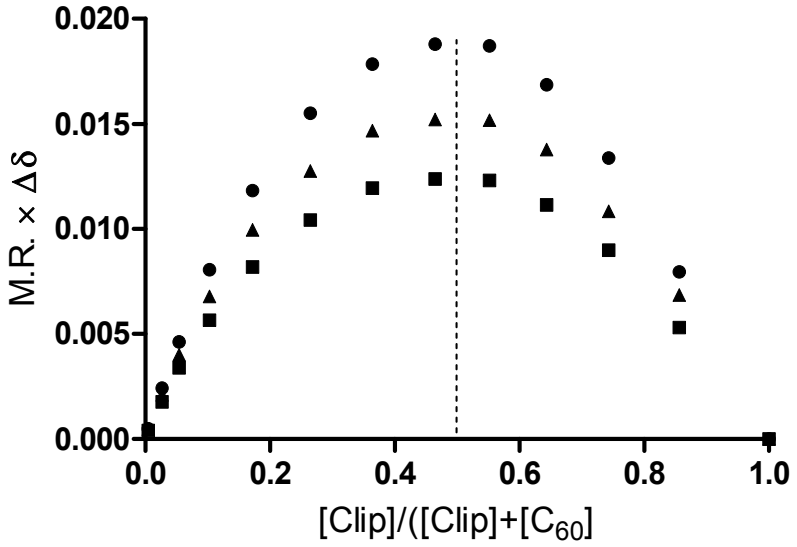


Figure 7.1 Job plot constructed from NMR titrations of the Clip with C_{60} in toluene. Values of the [molar ratio × Δδ] are plotted vs. mol fraction of Clip at three temperatures, 288 K (●), 298 K (▲), and 308 K (■). The dashed line indicates that the maximum change in the chemical shift for ^1H_b occurs at a molar ratio of 0.5. This is consistent with a saturation stoichiometry of 1:1.

Similar results were obtained when signals from the other Clip protons were used (data not shown). Using the same analysis method (Job plot), a similar result was obtained for the formation of the C_{70}•Clip complex. To further confirm the saturation stoichiometry for the host guest inclusion complexes between the C_{60} (or C_{70}) and the Clip, APPI mass spectrometry experiments were performed. Figure 7.2 shows the APPI mass spectra obtained for each of the following chemical species in toluene: free C_{60}, free C_{70}, free Clip (and $Clip_2$), and the C_{60}•Clip and C_{70}•Clip complexes.

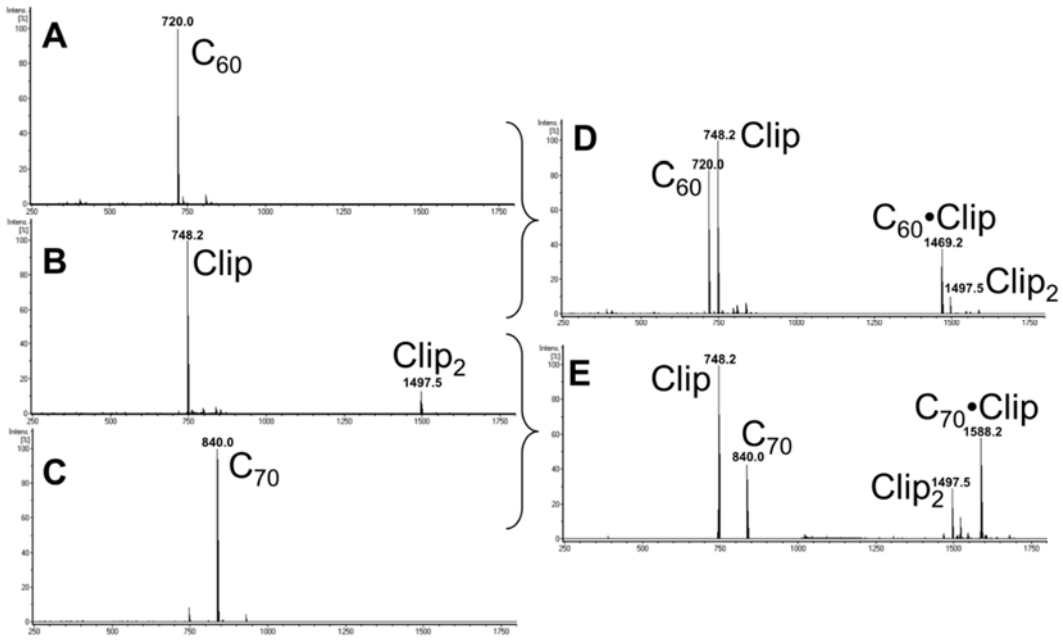


Figure 7.2 APPI mass spectra for toluene solutions containing C_{60} (A), Clip (B), C_{70} (C), and the two inclusion complexes, C_{60}•Clip (D) and C_{70}•Clip (E).

Panels A, B, and C show the APPI mass spectra for solutions containing a single chemical species, C_{60}, Clip, or C_{70} respectively. Panels A and C show only a single peak, e.g. the C_{60}^{-1} and C_{70}^{-1} parent fullerene monovalent anion peaks. Panel B exhibits a spectrum with two notable m/z peaks with masses of 748.2 and 1497.5 g/mol. The m/z peak at 1497.5 indicates that the Clip undergoes self-association to form a small amount of dimer in dilute toluene solutions. The APPI mass spectrum for a mixture containing an excess of Clip (more than 2:1) in addition to C_{60} is shown Panel D. The four peaks observed have m/z values of 720.0, 748.2, 1469.2, and 1497.5 g/mol. The peaks in order are for the free C_{60} parent anion, the free Clip (monomer), the 1:1 C_{60}•Clip complex and the Clip dimer ($Clip_2$). Panel E shows the APPI mass spectrum obtained for a mixture containing an excess of Clip (again more than 2:1) in addition to C_{70}. The four peaks

observed here have m/z values of 748.2, 840.0, 1497.5 and 1588.2 g/mol. The peaks in order are for the free Clip (monomer), the free C_{70} parent anion, the free Clip dimer ($Clip_2$), and the 1:1 C_{70}•Clip complex.

Isothermal titration calorimetry experiments were performed, wherein a dilute solution of the Clip titrant was added to a dilute solution of the fullerene titrate. A typical ITC thermogram for the addition of Clip to C_{60} in toluene at 298 K is shown in Figure 7.3. The solid line through the data points represents a non-linear fit of the data to a thermodynamic model for the formation of a 1:1 inclusion complex. This analysis of the ITC data yields a complete set of thermodynamic parameters (K_a (or ΔG), ΔH, and $-T\Delta S$) for the formation of the fullerene•Clip complexes.

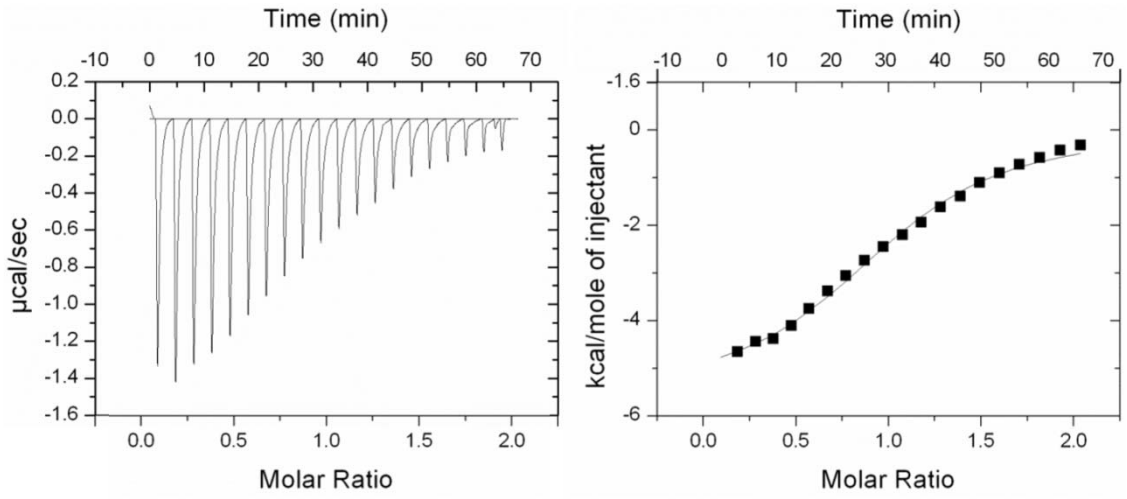


Figure 7.3 ITC data for the titration of Clip into C_{60}.

The left panel shows the baseline-corrected raw ITC signal for a typical titration experiment in which 20 separate injections of dilute Clip titrant solution ([Clip] = 0.7mM in toluene, injection volume = 14 µL) were made into the ITC cell filled with the a dilute C_{60} solution ([C_{60}] = 70 µM in toluene). The right panel shows the ΔH for each injection (■) along with the best-fit non-linear regression line (—) for a 1:1 inclusion complex model (with n having a fixed value of 1.0).

The thermodynamic data for the formation of the C_{60}•Clip and C_{70}•Clip complexes in toluene, chlorobenzene, and *o*-dichlorobenzene at 298 K are listed in Table 7.1.

Table 7.1 Summary of the thermodynamic parameters obtained from fitting ITC titration experiments to a 1:1 inclusion complex model. Values for ΔG, ΔH, and –TΔS have units of kcal/mol.

	toluene		chlorobenzene		*o*-dichlorobenzene	
	C_{60}	C_{70}	C_{60}	C_{70}	C_{60}	C_{70}
K_a	3200 ± 150	4600 ± 170	800 ± 10	900 ± 30	200 ± 20	200 ± 10
ΔG	-4.77 ± 0.03	-4.99 ± 0.02	-3.98 ± 0.01	-4.06 ± 0.02	-3.11 ± 0.09	-3.11 ± 0.03
ΔH	-4.61 ± 0.11	-2.95 ± 0.04	-4.03 ± 0.04	-3.76 ± 0.06	-1.87 ± 0.06	-1.97 ± 0.06
-TΔS	-0.16 ± 0.12	-2.04 ± 0.06	0.04 ± 0.04	-0.29 ± 0.07	-1.25 ± 0.15	-1.15 ± 0.06

Similar thermodynamic data for the formation of the C_{60}•Clip and C_{70}•Clip complexes in other solvents and at other temperatures are given in the supplementary information. The association constants are fairly weak for all of these systems (ranging from K_a = 4600 for formation of the C_{70}•Clip complex in toluene, to K_a = 200 for formation of the C_{70}•Clip complex in *o*-dichlorobenzene. Two general trends are observed in these data. First, the association constant for C_{70} with Clip is typically greater than for C_{60} with Clip. Second, complex formation becomes less favorable as the solvent becomes a better solvent for either the fullerene or the Clip. In effect, the K_a for complexation formation is weaker in solvents where the solubility of the fullerenes and Clip is greater (e.g. the K_a in *o*-dichlorobenzene is weaker than the K_a in toluene). As seen in the data in Table 7.1, at 298 K the favorable free energy change, ΔG, for complex formation is principally the result of a favorable change in enthalpy, ΔH. The entropy term, -TΔS, for formation of the fullerene Clip complexes is smaller than the enthalpy

change in every instance. With the exception of the entropy changes for the formation of the C_{70}•Clip complex in toluene and the C_{60} and C_{70} Clip complexes in *o*-dichlorobenzene, the value of the $-T\Delta S$ term is approximately zero.

The association constants, K_a, for formation of the C_{60}•Clip complex in deuterated toluene and deuterated chlorobenzene at 288, 298, and 308 K, were determined by NMR using the Clip proton chemical shift data (e.g. ^1H_b), $\Delta\delta$, and Eq. 7.1. The NMR derived K_a values are listed in Table 7.2 along with the ITC derived K_a values determined in the same solvents and at the same temperatures. The association constants , K_as, for the formation of the C_{60}•Clip complex in either toluene and chlorobenzene, and as determined in either ITC or NMR titration experiments at 288, 298, and 308 K are listed in Table 7.2.

Table 7.2 Comparison of measured K_a values for binding of C_{60} and CLIP as obtained from ITC and NMR titration experiments.

	288 K		298 K		308 K	
	toluene	chlorobenzene	toluene	chlorobenzene	toluene	chlorobenzene
ITC	4200 ± 340	1060 ± 30	3200 ± 150	800 ± 10	1900 ± 50	720 ± 40
NMR	4370 ± 230	770 ± 50	2820 ± 75	450 ± 40	2150 ± 50	350 ± 30

In general, the ITC and NMR derived K_a values for association of C_{60} and Clip are in excellent agreement with one another in toluene, almost overlapping within the experimental uncertainties. The differences in the NMR and ITC reported ΔG values for complex formation in toluene are -0.02, +0.07, and -0.07 kcal/mol at 288, 298, and 308 K respectively. The ITC and NMR derived K_a values for association of C_{60} and Clip are in good agreement with one another in chlorobenzene with the NMR derived K_a values lower in all instances. The differences in the NMR and ITC reported ΔG values for

complex formation in chlorobenzene are +0.18, +0.34, and +0.44 kcal/mol at 288, 298, and 308 K respectively.

The NMR K_a data in Table 7.2 were used to estimate the enthalpy change for formation of the C_{60}•Clip complex by constructing a van't Hoff plot shown in Figure 7.4.

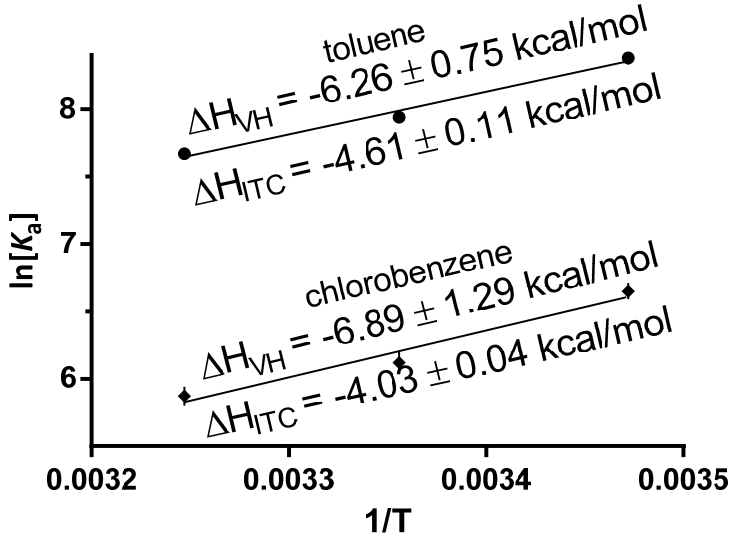


Figure 7.4 The natural log of the K_a values determined in NMR titrations are plotted vs. the reciprocal of the measurement temperature in K. The slopes of the two linear regression lines are proportional to the van't Hoff enthalpy changes, ΔH_{VH}, for formation of the C_{60}•Clip complex in toluene (●) and chlorobenzene (◆).

Values for the van't Hoff enthalpy changes, ΔH_{VH}, for formation of the C_{60}•Clip complex in toluene-d_8 and in chlorobenzene-d_5 are -6.26 ± 0.75 kcal/mol and -6.89 ± 1.29 kcal/mol respectively. These data indicate that there is a significant difference between the calorimetric enthalpy changes, ΔH_{cal}, from the ITC experiments at 298 K and the van't Hoff enthalpy changes, ΔH_{VH}, calculated from the temperature dependence of the NMR derived K_as at the midpoint in temperature (298 K). The NMR van't Hoff

enthalpy changes are more exothermic by -1.7 and -2.8 kcal/mol in toluene-d_8 and chlorobenzene-d_5 respectively. The differences are outside any expected experimental error in either the ΔH_{VH} or ΔH_{cal} values and must reflect real differences in the solvation enthalpies for the fullerene and Clip in the deuterated solvents versus the non-deuterated solvents.

ITC experiments, similar to the titration shown in Figure 7.3, were done by adding Clip to either C_{60} or C_{70} in toluene, anisole, chlorobenzene, 1,1,2,2-tetrachloroethane, and *o*-dichlorobenzene at temperatures from 278 to 323 K. The calorimetric enthalpy changes, ΔH_{cal}, for formation of the C_{60}•Clip and C_{70}•Clip complexes in toluene at 278, 288, 298, and 308 K were plotted versus temperature to yield an estimate of the heat capacity change, ΔC_p, for the formation of the two fullerene Clip complexes and is shown in Figure 7.5.

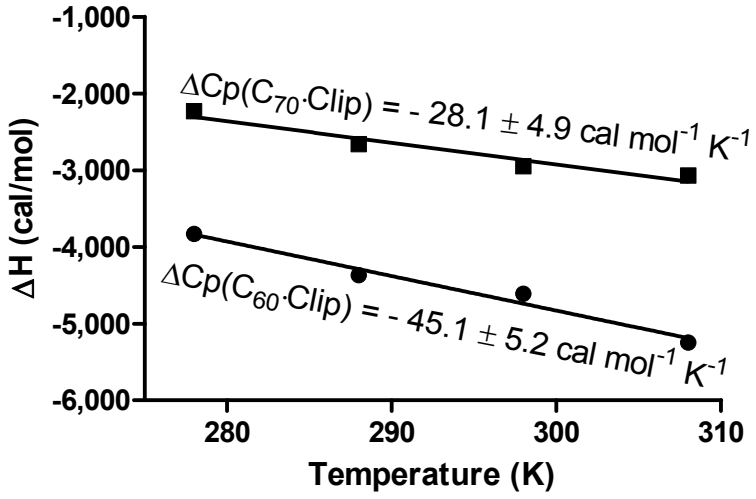


Figure 7.5 The enthalpy change, ΔH, for formation of the C_{60}•Clip (●) and the C_{70}•Clip (■) complexes are plotted as a function of temperature. The slope of the linear regression lines correspond to values for the change in the heat capacity for formation of the two complexes.

The enthalpy changes for formation of both the C_{60} and C_{70} buckyball•buckycatcher complexes are linearly dependent on T over the experimental temperature range. The estimated heat capacity changes for formation of the two host-guest complexes are -45.1 ± 5.2 cal mol^{-1} K^{-1} and -28.1 ± 4.9 cal mol^{-1} K^{-1} for the C_{60}•Clip and C_{70}•Clip complexes respectively. It is clear that these two ΔC_p values are significantly different.

7.5 Discussion

In this study we have used ITC methods to develop a complete thermodynamic description for the formation of C_{60} and C_{70} fullerene buckycatcher complexes. Where possible we have compared our thermodynamic data (K_a, ΔG, ΔH, and $-T\Delta S$) with both existing experimental and computational data. In general the ITC values for K_a were in agreement with the NMR K_a data and some simple computational results. However, a more careful inspection of the ITC results indicated systematic differences in the ITC versus NMR K_a values and even larger differences between the ITC enthalpy change data, ΔH_{cal}, and the NMR van't Hoff enthalpy change, ΔH_{VH}, estimates. In addition to providing values for K_a, ΔG, ΔH, and $-T\Delta S$ for formation of these dispersion complexes, the ITC experiments provided estimates for the ΔC_p values for complex formation, and unexpected evidence for a general enthalpy entropy compensation in the temperature dependence of the free energy change and in the change in free energy with changes in the solvent dielectric constant. It is important to note that the ITC experiments reported here were only possible for these weak complexes because the complex stoichiometry was determined in complimentary mass spectrometry experiments and because we were able to work at reasonably high and accurately known concentrations for both the

fullerenes and the Clip. In other words, we were able to design ITC experiments wherein we were able to measure the heats for the formation of the complex and able to determine the K_a values from the observed curvature in the titration curve. These conditions were met even for systems exhibiting K_a values as low as 200.

Stoichiometric information obtained from both Job plot analysis of the NMR titrations and from the APPI mass spectrometry experiment clearly show a saturation stoichiometry of 1:1 for both the C_{60} and C_{70} buckyball•buckycatcher complexes. These results are in agreement with the previously reported crystallographic structure for the 1:1 inclusion complex of C_{60}•Clip.(1) In addition to showing the expected 1:1 species for the fullerene and buckycatcher complex, the APPI mass spectra indicate the presence of free fullerenes and Clip, and Clip dimers a clear indication of weak complex formation.

In previous studies, we reported NMR derived K_a values for the association of both C_{60} and C_{70} with the Clip in toluene at 298 K. The earlier K_a values were 8600 ± 500 M^{-1} for formation of C_{60}•Clip complex and 6800 ± 400 M^{-1} for formation of C_{70}•Clip complex.(1,2) In the current study, the NMR titrations were repeated but at lower concentrations for both C_{60} and Clip. The latest NMR derived K_a value for formation of C_{60}•Clip complex at 298 K in toluene is 2820 ± 75 M^{-1} which is in excellent agreement with the ITC determined K_a value of 3200 ± 150 M^{-1}. We concluded that K_a values from both previous studies had been over-estimated due to some unobserved aggregation of either C_{60}, Clip or the C_{60}•Clip complex in solution at the higher concentrations. More importantly, we now find as expected that the Clip binds C_{70} with a higher affinity than exhibited for C_{60} under all of the solution conditions studied here (see Table 7.1 and the Supplemental materials).

The heat capacity change, ΔC_p, for formation of a fullerene Clip complex in toluene was determined from the slope of the linear regression line for a plot of ΔH_{cal} versus temperature from 278 to 308 K. The ΔC_p values for formation of both the C_{60}•Clip and C_{70}•Clip complexes were -45 and -28 cal mol^{-1} K^{-1}. These small negative values for ΔC_p indicate that the fullerene•Clip complexes are slightly less structured than the free fullerene and Clip. The observation of a negative heat capacity change is typically attributed to the release of solvent molecules upon complex formation. In the fullerene buckycatcher system, it appears that solvent molecules must be expelled from the interacting surfaces of the fullerene and the buckycatcher with the net negative change in ΔC_p resulting from desolvation of the complex. The larger negative ΔC_p value for formation of the C_{60}•Clip complex (-45 cal mol^{-1} K^{-1} vs. -28 cal mol^{-1} K^{-1}) suggests that C_{60} fits better into the Clip pocket and that more solvent is released in the formation of the C_{60} complex than for formation of the C_{70}•Clip complex. The larger K_a measured for the formation of the C_{70}•Clip complex would be consistent with a lower desolvation penalty for formation of the C_{70} inclusion complex.

Thermodynamic data obtained from fitting ITC experiments for the addition of Clip into either C_{60} or C_{70} solutions performed at several different temperature ranging from 278to 308 K in toluene are plotted in Figure 7.6.

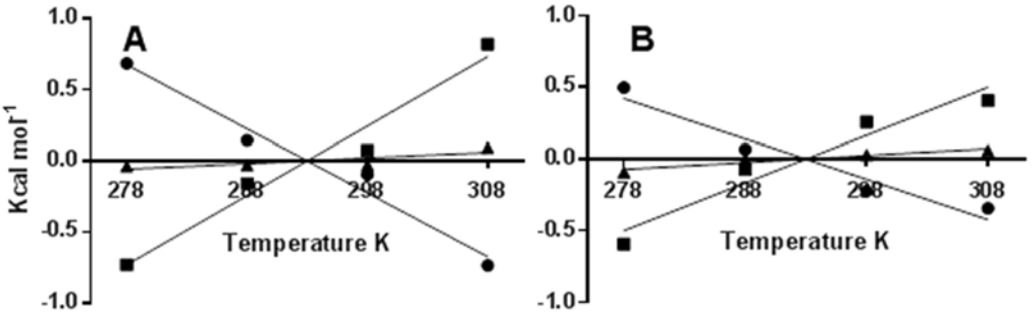


Figure 7.6 Normalized values for the thermodynamic parameters ($\Delta G - \Delta G_{ave}$), ▲, ($\Delta H - \Delta H_{ave}$), ●, and ($-T\Delta S + T\Delta S_{ave}$), ■, for the formation of the fullerene Clip complexes in toluene are plotted at four different temperatures 278, 288, 298, and 308 K.

Panel A shows the thermodynamic data for formation of the C_{60}•Clip complex. Panel B shows the thermodynamic data for the formation of the C_{70}•Clip complex.

The change in the free energy change, $\delta\Delta G$, for fullerene•Clip complex formation over the temperature range, 278 to 308 K, is very small. For example the change in the free energy change, $\delta\Delta G$, for formation of the C_{60}•Clip complex is less than 0.1 kcal/mol while the change in the free energy change, $\delta\Delta G$, for formation of the C_{70}•Clip complex is less than 0.2 kcal/mol. Variations in the enthalpy and entropy changes for formation of the fullerene•Clip complex are much greater over the same temperature range, with the changes in ΔH and in $-T\Delta S$ compensating for each other. Enthalpy-entropy compensation as brought about by changes in temperature has only infrequently been observed or reported for reactions taking place in organic solvents.(9-17) Referring to the extensive enthalpy-entropy compensation literature for reactions taking place in aqueous solution, we are not surprised by this result since its origin is typically attributed to changes in solvation.(18-28) The formation of fullerene•Clip complexes must involve

solvent removal from the interacting surfaces of the associated fullerene guest and Clip host pocket.

It was noted in the Results section that fullerene•Clip complex formation becomes less favorable in solvents where the solubility of the fullerene or the Clip is greater. To further explore the nature of this observation, ITC experiments were done using five different solvents, toluene, anisole, chlorobenzene, 1,1,2,2-tetrachloroethane and *o*-dichlorobenzene, at 298 K. These thermodynamic data which can be found in Table 7.1 and in the supplementary materials section are plotted as a function of solvent dielectric constant in Figure 7.7.

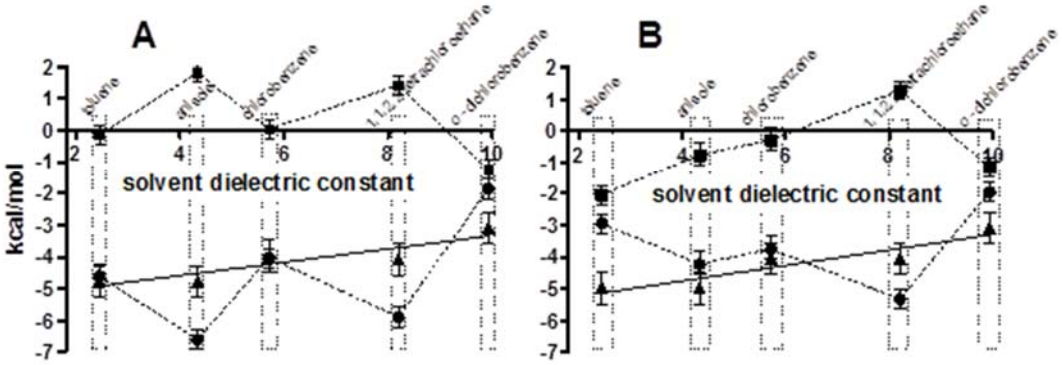


Figure 7.7 Panels A and B demonstrate a new type of enthalpy entropy compensation for the formation of the C_{60}•Clip and C_{70}•Clip complexes respectively. The thermodynamic parameters for fullerene•Clip formation, ΔG (▲), ΔH (●), and $-T\Delta S$ (■) are plotted as a function of solvent dielectric constant for five different organic solvents at 298 K.

Again, we observe a new type of enthalpy-entropy compensation in which the free energy change for complex formation is less dependent on the solvent properties

(e.g. polarity, hydrogen bonding, dielectric constant, etc.) than is either the enthalpy or entropy change. In fact the while the free energy change is observed to vary linearly with solvent dielectric constant, becoming 1.5 to 2 kcal/mol less favorable in *o*-dichlorobenzene than in toluene, both the enthalpy and entropy changes vary unpredictably while exhibiting a high degree of compensation. In effect, every change in ΔH is opposed by a compensating change in $-T\Delta S$. Again, the explanation for this phenomena must reside in the fact that complex formation proceeds with the release of solvent from the fullerene•Clip complex.

7.6 Conclusion

In conclusion, this work has resulted in the most complete set of thermodynamic data obtained to date for the formation of C_{60}•Clip and C_{70}•Clip inclusion complexes. The thermodynamic data, including K_a, ΔG, ΔH, and $-T\Delta S$, for the formation fullerene buckycatcher inclusion complexes are reported over a range of temperatures and in several solvents. The thermodynamic data for these prototypical large dispersion-driven supramolecular systems will be invaluable to the further development of computational methods and models for estimating the energetics of π-π interacting systems in solution. These data will be particularly important in predicting dispersion driven complex formation in aromatic or π bonding solvents.

This work has also yielded fundamental insight into the changes in solvation that must occur for C_{60}•Clip and C_{70}•Clip inclusion complexes to form. For example, the ΔC_p data, as well as the temperature dependent enthalpy-entropy compensation, and the solvent dependent enthalpy entropy compensation phenomena that were observed are new to the dispersion-driven supramolecular assembly literature. Lastly, this work has

shown the power of the ITC method to investigate these supramolecular assembly processes even though the association constants are quite small.

7.7 References

1. Sygula, A., Fronczek, F.R., Sygula, R., Rabideau, P.W. and Olmstead, M.M. (2007) A double concave hydrocarbon buckycatcher. *Journal of the American Chemical Society*, **129**, 3842-3843.

2. Muck-Lichtenfeld, C., Grimme, S., Kobryn, L. and Sygula, A. (2010) Inclusion complexes of buckycatcher with C(60) and C(70). *Physical chemistry chemical physics : PCCP*, **12**, 7091-7097.

3. Zhao, Y. and Truhlar, D.G. (2008) Computational characterization and modeling of buckyball tweezers: density functional study of concave-convex pi...pi interactions. *Physical chemistry chemical physics : PCCP*, **10**, 2813-2818.

4. Henne, S., Bredenkotter, B., Dehghan Baghi, A.A., Schmid, R. and Volkmer, D. (2012) Advanced buckyball joints: synthesis, complex formation and computational simulations of centrohexaindane-extended tribenzotriquinacene receptors for C60 fullerene. *Dalton Transactions*, **41**, 5995-6002.

5. Wong, B.M. (2009) Noncovalent interactions in supramolecular complexes: a study on corannulene and the double concave buckycatcher. *Journal of computational chemistry*, **30**, 51-56.

6. Tkatchenko, A., Alfè, D. and Kim, K.S. (2012) First-principles modeling of non-covalent interactions in supramolecular systems: The role of many-body effects. *Journal of Chemical Theory and Computation*, **8**, 4317-4322.

7. Risthaus, T. and Grimme, S. (2013) Benchmarking of London dispersion-accounting density functional theory methods on very large molecular complexes. *Journal of Chemical Theory and Computation*, **9**, 1580-1591.

8. Le, V.H., Buscaglia, R., Chaires, J.B. and Lewis, E.A. (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Analytical biochemistry*, **434**, 233-241.

9. Adriaenssens, L., Gil-Ramirez, G., Frontera, A., Quinonero, D., Escudero-Adan, E.C. and Ballester, P. (2014) Thermodynamic Characterization of Halide-pi Interactions in Solution Using "Two-Wall" Aryl Extended Calix[4]pyrroles as Model System. *Journal of the American Chemical Society*.

10. Ravi, V., Binz, J.M. and Rioux, R.M. (2013) Thermodynamic profiles at the solvated inorganic-organic interface: the case of gold-thiolate monolayers. *Nano letters*, **13**, 4442-4448.

11. Gonnelli, A., Ortore, M.G., Baldassarri, E.J., Spada, G.P., Pieraccini, S., Perone, R.C., Funari, S.S. and Mariani, P. (2013) Small-angle X-ray scattering study of self-assembling lipophilic guanines in organic solvents: G-quadruplex formation and cation effects in cyclohexane. *The journal of physical chemistry. B*, **117**, 1095-1103.

12. Gebreyohannes, K.G. and McGuffin, V.L. (2010) Thermodynamic and kinetic study of chiral separations of coumarin-based anticoagulants on derivatized amylose stationary phase. *Journal of chromatography. A*, **1217**, 5901-5912.

13. Kabir ud, D. and Koya, P.A. (2010) Effects of solvent media and temperature on the self-aggregation of cationic dimeric surfactant 14-6-14, 2Br- studied by conductometric and fluorescence techniques. *Langmuir : the ACS journal of surfaces and colloids*, **26**, 7905-7914.

14. Fukuhara, G., Mori, T. and Inoue, Y. (2009) Competitive enantiodifferentiating anti-Markovnikov photoaddition of water and methanol to 1,1-diphenylpropene using a sensitizing cyclodextrin host. *The Journal of organic chemistry*, **74**, 6714-6727.

15. Sun, H., Zhao, Y., Huang, Z., Wang, Y. and Li, F. (2008) 1H NMR study on the self-association of quinacridone derivatives in solution. *The journal of physical chemistry. A*, **112**, 11382-11390.

16. Matsumoto, M. and Swaddle, T.W. (2004) The decamethylferrocene(+/0) electrode reaction in organic solvents at variable pressure and temperature. *Inorganic chemistry*, **43**, 2724-2735.

17. Egan, T.J. and Ncokazi, K.K. (2004) Effects of solvent composition and ionic strength on the interaction of quinoline antimalarials with ferriprotoporphyrin IX. *Journal of inorganic biochemistry*, **98**, 144-152.

18. Guthrie, K.M., Parenty, A.D., Smith, L.V., Cronin, L. and Cooper, A. (2007) Microcalorimetry of interaction of dihydro-imidazo-phenanthridinium (DIP)-based compounds with duplex DNA. *Biophysical chemistry*, **126**, 117-123.

19. Muralidhara, B.K., Negi, S.S. and Halpert, J.R. (2007) Dissecting the thermodynamics and cooperativity of ligand binding in cytochrome P450eryF. *Journal of the American Chemical Society*, **129**, 2015-2024.

20. Lumry, R. and Rajender, S. (1970) Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers*, **9**, 1125-1227.

21. De, M., You, C.C., Srivastava, S. and Rotello, V.M. (2007) Biomimetic interactions of proteins with functionalized nanoparticles: a thermodynamic study. *Journal of the American Chemical Society*, **129**, 10747-10753.

22. Ortiz-Salmeron, E., Yassin, Z., Clemente-Jimenez, M.J., Las Heras-Vazquez, F.J., Rodriguez-Vico, F., Baron, C. and Garcia-Fuentes, L. (2001) Thermodynamic analysis of the binding of glutathione to glutathione S-transferase over a range of temperatures. *European journal of biochemistry / FEBS*, **268**, 4307-4314.

23. Krueger, S., Gregurick, S., Shi, Y., Wang, S., Wladkowski, B.D. and Schwarz, F.P. (2003) Entropic nature of the interaction between promoter bound CRP mutants and RNA polymerase. *Biochemistry*, **42**, 1958-1968.

24. Soto, A.M., Kankia, B.I., Dande, P., Gold, B. and Marky, L.A. (2002) Thermodynamic and hydration effects for the incorporation of a cationic 3-aminopropyl chain into DNA. *Nucleic acids research*, **30**, 3171-3180.

25. Hegde, S.S., Dam, T.K., Brewer, C.F. and Blanchard, J.S. (2002) Thermodynamics of aminoglycoside and acyl-coenzyme A binding to the Salmonella enterica AAC(6')-Iy aminoglycoside N-acetyltransferase. *Biochemistry*, **41**, 7519-7527.

26. Deinum, J., Gustavsson, L., Gyzander, E., Kullman-Magnusson, M., Edstrom, A. and Karlsson, R. (2002) A thermodynamic characterization of the binding of thrombin inhibitors to human thrombin, combining biosensor technology, stopped-flow spectrophotometry, and microcalorimetry. *Analytical biochemistry*, **300**, 152-162.

27. Wieprecht, T., Apostolov, O. and Seelig, J. (2000) Binding of the antibacterial peptide magainin 2 amide to small and large unilamellar vesicles. *Biophysical chemistry*, **85**, 187-198.

28. Davies, T.G., Hubbard, R.E. and Tame, J.R. (1999) Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA-peptide complexes. *Protein science : a publication of the Protein Society*, **8**, 1432-1444.

CHAPTER VIII

OVERALL DISCUSSION OF RESULTS AND CONCLUSIONS

The power and versatility of ITC techniques are demonstrated by the amount of ITC data that has been added to the literature in recent years. However, most of the data are analyzed using the typical software provided by ITC manufacturers (GE/VP or TA-instruments), having limited models and functionalities. More complex data are rarely encountered in the literature due to two reasons: 1) complex systems are seldom studied by calorimetric techniques 2) the absence of analysis software that is user-friendly publicly available/searchable, and contains the appropriate model.

Nevertheless, in the past decade, there has been an increasing emergence of complex ITC data in the literature. Interestingly, five different ITC analysis software packages (including our recently published program CHASM©) were introduced in the past seven years. SEDPHAT by Schuck, P. (1) was the first software to become available besides the commercial software provided by GE/VP and TA-instruments. SEDPHAT was then followed by IC-ITC, created by Haq, I. (2). Both programs have front-end user interfaces and offer a wide variety of functionalities to the point of being overwhelming to users. Open-ITC, developed by Mohanty, S. (3) and suITCase by Piñeiro, A. (4) were both introduced in 2011. Open-ITC is written in GSL library functions and does not have a front-end interface which requires the users to be familiar with C++ programming and Linux. In the same year, suITCase, introduced by Piñeiro, is a web–based program that

only offers analysis for "sequential binding sites" models. Overall, all four programs have their own unique features and functionalities. While some are more useful than others, they all lack the analysis for multiple overlapping equilibria (>2 sites). This is where our program fits in.

In a little over a year, when CHASM© was launched , we tallied over 370 unique downloads from over 25 different countries and 31 states. Some of these users included Washington University, UC Berkeley, Purdue University, the FDA, the NIH, and Harvard University.

Isothermal titration calorimetry plays a central role throughout this dissertation and the course of my PhD training. All chapters in this dissertation contain ITC data in addition to other biophysical techniques. The following conclusions for each chapter are presented according to the order of the chapter as it appeared in this dissertation.

In Chapter IV, we studied a Bcl-2 Promoter Sequence G-Quadruplex interactions with three planar and non-planar cationic porphyrins: TMPyP4, TMPyP3, and TMPyP2. Spectroscopic results (CD and fluorescence) both suggested that TMPyP4 behaved differently from TMPyP3 and TMPyP2 upon binding to quaduplex DNA. Whereas TMPyP2 and TMPyP3 exhibit gradual attenuation of their emission spectra over the wavelength range of 600 to 800 nm when bound to quadruplex DNA, TMPyP4 fluorescence changes in a more complex manner upon binding to the Bcl-2 quadruplexes: first decreased from 0 to 2 moles and then increased from 2 to 4 moles. On the other hand, the CD spectra for the Bcl-2 promoter quadruplexes is unchanged by the addition of the three porphyrin ligands up to saturation (2:1 for TMPyP2 or TMPyP3, and 4:1 for TMPyP4). Only when saturation ratios are exceeded do we begin to observe induced CD

signal for the bound porphyrin; again (2:1 for TMPyP2 or TMPyP3, and 4:1 for TMPyP4). The ESI-MS results provide spectroscopic evidence for the saturation stoichiometry of the ligand DNA binding. The clear indication of a 4:1 ratio peak for TMPyP4/Bcl-2 the lack of free/unbound DNA are in excellent agreement with ITC binding studies. On the other hand, the absence of the 4:1 peaks and the apparent peaks for free DNA in the mass spectra for complexation of both the TMPyP3 and TMPyP2 to Bcl-2 solutions suggested both TMPyP3 and TMPyP2 exhibited weaker binding affinity to Bcl-2 quadruplex with a maximum stoichiometry of 2:1. To the best of our knowledge, this is the first ESI-MS evidence for a G-quadruplex species having four porphyrin ligands bound to an intramolecular G-quadruplex. The ITC results present a consistent picture regarding the binding of the three porphyrin ligands to the WT 39-mer Bcl-2 promoter sequence quadruplex. For TMPyP4, the higher affinity binding mode combines a favorable enthalpy contribution (-1.8 to -4.6 kcal/mol) with a more significant favorable entropy contribution of -4 to -9 kcal/mole in terms of (-TΔS) while the lower affinity mode combines a more significant enthalpy contribution (-6 to -12 kcal/mol) with a smaller entropy term ranging from -3 to +5 kcal/mol.

The results of the spectroscopic and microcalorimetric experiments described here clearly demonstrate that the planar cationic porphyrin, TMPyP4, binds with high affinity to the WT 39-mer Bcl-2 promoter sequence G-quadruplex by two distinctly different binding modes. Our interpretation of two binding modes being an external or *end stacking* mode and an *intercalation* mode can describe the observed ITC data reasonably well. The *intercalation* mode is eliminated when the bulky N^{+}-CH_3 pyridinium substituent groups (in TMPyP3 and TMPyP2) force the ligand to become non-planar,

requiring a greater G-tetrad spacing for intercalation to occur. Even though the first report for TMPyP4 interaction with G-quadruplex dates back to 16 years ago, there is still controversy on the site and mode of its interactions with quadruplex DNA. We were hoping the data and the careful discussion provided in the study can convince the scientific community that TMPyP4 is capable of end-stacking and intercalation.

In Chapter V, we investigated the effect of Pyridyl substitution on the thermodynamics of porphyrin binding to Bcl-2 27-mer G-quadruplex DNA. The initial goal of this study was to understand where the high binding affinity of TMPyP4 towards quadruplex structure originates from. We took an unconventional approach: we removed the arms (N-methyl Pyridyl substituent) from the parent compound TMPyP4 systematically and reevaluated the binding interactions. The thermodynamic properties for binding these truncated porphyrins to a Bcl-2 G-quaduplex yielded a couple of surprises. It was previously suggested that the two binding modes observed for the interaction of TMPyP4 with G-quadruplexes in general and more specifically with c-MYC, Bcl-2, and K-ras promoter sequence G-quadruplexes are 1) an "end binding" mode, and 2) an "intercalation mode". We described the binding of the **P(5,15)** ligand as "intercalation" and the binding of the **P(5,10,15)** ligand as "end binding". In comparison to TMPyP4, which forms G-quadruplex complexes having a saturation stoichiometry of 4:1, the **P(5,15)** and **P(5,10,15)** only form 2:1 complexes with the Bcl-2 promoter sequence G-quadruplex. The biggest surprise is that Goncalves *et al.* reports a significant affinity for binding the **P(5,10)** ligand to a human telomere G-quadruplex while we did not observe any complex formation between the **P(5,10)** ligand and the Bcl-2 promoter sequence G-quadruplex. We believe that the finding of **P(5,10)** in an SPR study reported

by Goncalves *et al.* is erroneous as they failed to find the expected 4:1 binding stoichiometry for the TMPyP4 complex as pointed out by another SPR study by Federici *et al.* on the same human telomere G-quadruplex construct. The apparent high affinity intercalation of the **P(5,15)** ligand points us in a new direction with respect to designing G-quadruplex interactive ligands. In comparison to TMPyP4, the **P(5,15)** ligand has a smaller SASA and a charge of only +2. However, **P(5,15)** binds to the Bcl-2 G-quadruplex with almost the same affinity as TMPyP4. Perhaps even better G-quadruplex ligands could be synthesized to take advantage of the *trans* (5,15) substitution pattern on the core porphyrin, but with changes in the substituent charge or shape to produce stronger interactions. By no surprise, we have recently encountered examples in the literature reporting new G4 interactive ligands combining the planar heterocyclic aromatic naphthalene diimide (NDI) ring and the trans substitution pattern, creating a new family of ligands exhibiting high binding affinity and better selectivity towards the G-quadruplex than duplex DNA.

In Chapter VI, we reported the assembly of a model NHE-III$_1$ construct and reported its structural complexity, stability and ligand interactions with a classical quadruplex interactive ligand, namely TMPyP4. The model NHE-III$_1$ was assembled by annealing a 17-mer oligonucleotide (short polypyrimidine strand) having a mid-sequence run of 5 T's (or C's) and tail sequences of 3'-AAATTT and TTTAAA-5', to a 32-mer oligonucleotide (polypurine strand) having a 20-mer c-MYC G-quadruplex forming core and complimentary tail sequences of 5'-TTTAAA and AAATTT-3'. ESI-MS and CD experiments for the model NHE-III$_1$ construct confirmed the presence of the typical c-MYC promoter sequence G-quadruplex in the new double stranded structure. DSC

experiments suggested a change in relative equilibrium of the folded structures in solution. Deconvolution of the 32-mer MWT melting curve suggests that the 1:2:1 folded species contributes 90% of the total population while the 1:6:1 only contributed 10%. However, upon complexation with the capping strand, the new relative population distribution is 40% for 1:2:1 and 60% for 1:6:1. We have searched the literature extensively in order to fully understand this behavior and have found that our explanation is viable. But more importantly, it suggested that the predominant 1:6:1 foldamer maybe a better candidate for designing drug targeting c-MYC G-quadruplex. Results of the ITC experiments for binding TMPyP4 to our model NHE-III_1 construct are more complex than we anticipated. The availability of an ITC analysis program (CHASM©) that is capable of discerning and fitting multiple overlapping processes has enabled us to extract the thermodynamic properties for binding of TMPyP4 to the model NHE-III_1. Aside from the duplex binding as a byproduct of TMPyP4 binding to the NHE-III_1 construct, we observed two important features that resulted from the ITC experiments: ligand induced refolding (LIR) and an enhancement in the binding affinity.

Unintentionally, the flanking duplex regions (modeled by runs of Ts and As) placed adenines adjacent to the quaduplex core on both ends, causing significant destabilization to the c-MYC quadruplex. Addition of TMPyP4, a classical and well-studied quadruplex-interactive-ligand, to the model NHE-III_1 construct initiated the ligand-induced-refolding process using TMPyP4 as a scaffold. We did not observe this phenomenon when we used the 4T capped WT 36-mer during preliminary study. We altered the DNA sequence for a less important reason but this serendipitous result

demonstrated that even a well-studied (and on-the-verge of being overlooked) ligand such as TMPyP4 can produce surprising yet exciting results.

We are not the first to contemplate the idea of modelling quadruplex in double stranded form. Two other research groups have presented their dsG4 models using PNAs and RGG peptide bound to a RNA duplex-quadruplex junction. Both previous models have serious limitations as discussed in Chapter VI. Although our model is not perfect, the novel aspect of our discovery was that we provided evidence of a conformational selection in the ensembles of quaduplex and we also provided evidence for enhanced binding capability of ligand to the model NHE-III_1 construct. It is important to emphasize that without the evidence from the ITC binding study, the remaining spectroscopic data seems to be very circumstantial or even conjecture. Our three overlapping binding equilibria, proceeded by a ligand-induced-reorganization, is the simplest yet most useful explanation to describe the observed ITC data. Famous mathematician John von Neumann used to say "… *With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."*

In Chapter VII, we reported for the first time calorimetric evidences of supramolecular PAH complexes brought about by weak dispersion forces. The spectroscopic evidence indicated the interactions of Clip and C_{60} is dominated by Van der Waals interactions through $\pi \bullet \bullet \pi$ stacking. Results combined from crystallographic structure, Job's plot analysis, and APPI-MS demonstrated a 1:1 binding stoichiometry for formation of the buckyball•buckycatcher complex. Both C_{60} and C_{70} fullerenes were used to model the concave surfaces and act as receptors for the doubly-convex-corannulene rings of the buckycatcher (or Clip).

The formation of these complexes is the result of a weak binding event ($\Delta G° = -4.0 \pm 1.0$ kcal/mol). Interestingly, the solubility of the reactants (fullerenes or Clip) and the measured association constants follow Le Chatelier's principle: the more soluble the reactants are in a particular solvent, the lower the values of the association constant measured in that solvent. The K_a values determined from ITC and NMR titration experiments are in excellent agreement with one another. Van't Hoff plot were constructed from variable temperature NMR titrations and yield ΔH_{VH} values that are in good agreement with the ΔH_{cal} values determined over similar temperature range. The ΔC_p accompanying complex formation and K_a values suggested that C_{70} seems to be a better structural fit for this Clip than C_{60} . By varying the temperature and the solvent of the ITC experiments, we observed both a classical phenomenon called enthalpy-entropy compensation that is well recognized in aqueous medium. However, our system seems to be more regulated with temperature changes than chemical changes that brought about variations in solvent dielectrics.

To the best of our knowledge, this is the first report of a model system exhibiting enthalpy-entropy compensation through purely van der Waals' interactions in different organic solvents. The ultimate goal of this project was to provide experimental data to assess the quality of current computational methods and to construct better solvent models for these theoretical studies. We have also demonstrated the richness of data collected from just a few ITC experiments. In this study, the thermodynamic data obtained from ITC experiments enabled us to better understand the assembling process of supramolecular complexes. Understanding these fundamental interactions is of premium

importance because they have many implications in chemistry, physics, biology, medicine, and material sciences.

8.1 References

1. Houtman, J.C., Brown, P.H., Bowden, B., Yamaguchi, H., Appella, E., Samelson, L.E. and Schuck, P. (2007) Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein science : a publication of the Protein Society*, **16**, 30-42.

2. Buurma, N.J. and Haq, I. (2008) Calorimetric and spectroscopic studies of Hoechst 33258: self-association and binding to non-cognate DNA. *Journal of molecular biology*, **381**, 607-621.

3. Krishnamoorthy, J. and Mohanty, S. (2011) Open-ITC: an alternate computational approach to analyze the isothermal titration calorimetry data of complex binding mechanisms. *Journal of molecular recognition : JMR*, **24**, 1056-1066.

4. Brocos, P., Banquy, X., Diaz-Vergara, N., Perez-Casas, S., Pineiro, A. and Costas, M. (2011) A critical approach to the thermodynamic characterization of inclusion complexes: multiple-temperature isothermal titration calorimetric studies of native cyclodextrins with sodium dodecyl sulfate. *The journal of physical chemistry. B*, **115**, 14381-14396.

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER VI

Table A.1 Thermodynamic parameters (given in kcal/mole) for the titrations of TMPyP4 into 32-mer MWT•5C capped construct, 1:2:1•5C capped construct, and 1:2:1•5C capped construct.

Construct	ΔG_1	ΔH_1	$-T\Delta S_1$	ΔG_2	ΔH_2	$-T\Delta S_2$	ΔG_3	ΔH_3	$-T\Delta S_3$
32MWT•5Ccap	-13.8±0.5	-7.8±0.1	-6.0	-10.3±0.6	-12.5±0.4	2.2	-7.6±0.1	-10.0±0.3	2.6
1:2:1•5Ccap	-12.6±0.6	-7.5±0.1	-5.1	-10.2±0.2	-12.3±0.2	2.2	-7.6±0.1	-9.4±0.2	1.8
1:6:1•5Ccap	-13.2±0.6	-7.5±0.2	-5.7	-9.2±0.1	-13.0±0.2	3.8	-6.9±0.1	-9.5±0.3	2.6

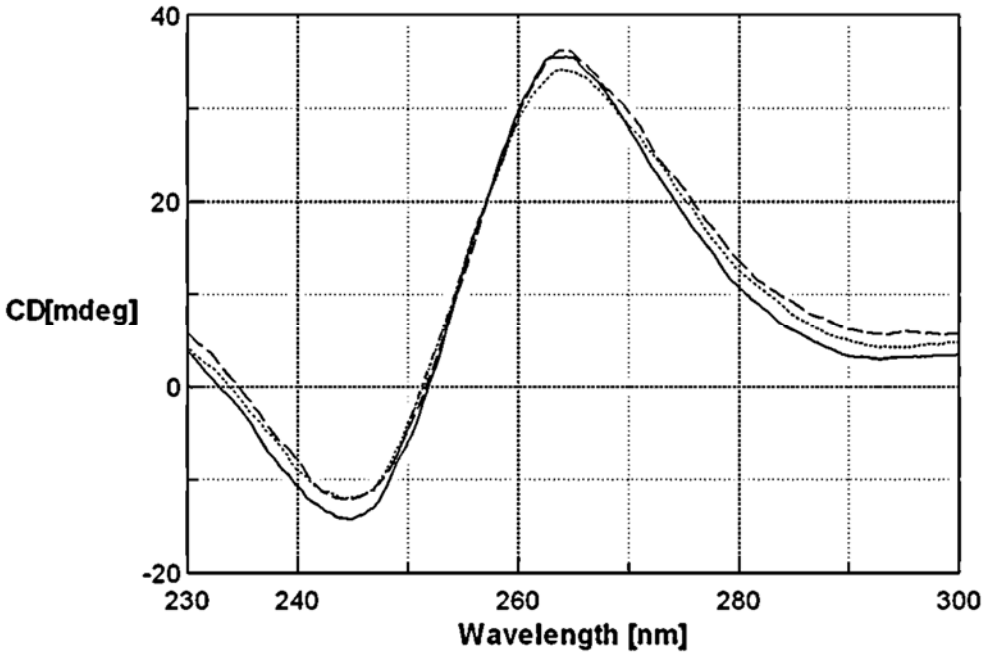


Figure A.1 Overlay CD spectra obtained for the uncapped 32-mer MWT (——), uncapped 1:2:1 mutant (— —), and uncapped 1:6:1 mutant (- - -).

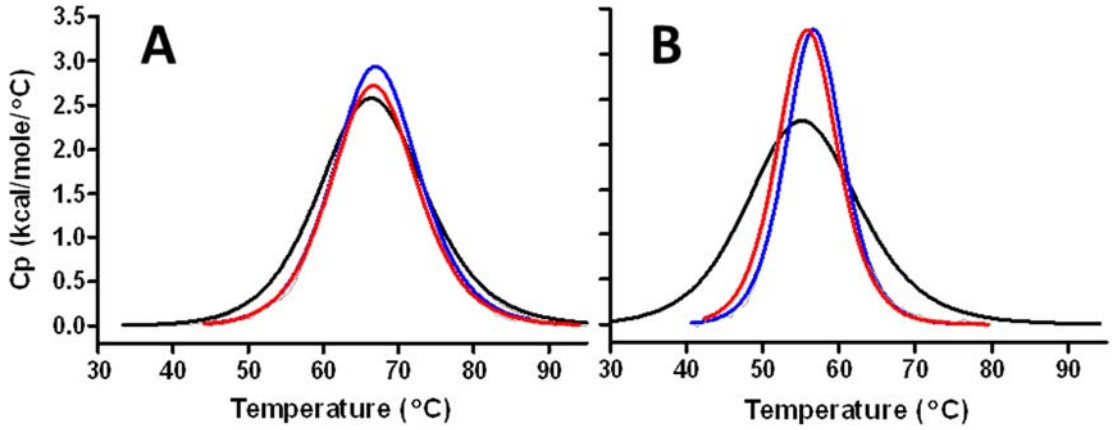


Figure A.2 A) DSC melting profiles for the thermal denaturation of 1:2:1 (—), 1:2:1•5T(—), and 1:2:1•5C (—)capping construct. B) DSC melting profiles for the thermal denatures of 1:6:1(—), 1:6:1•5T(—), and 1:6:1•5C(—) capping construct.

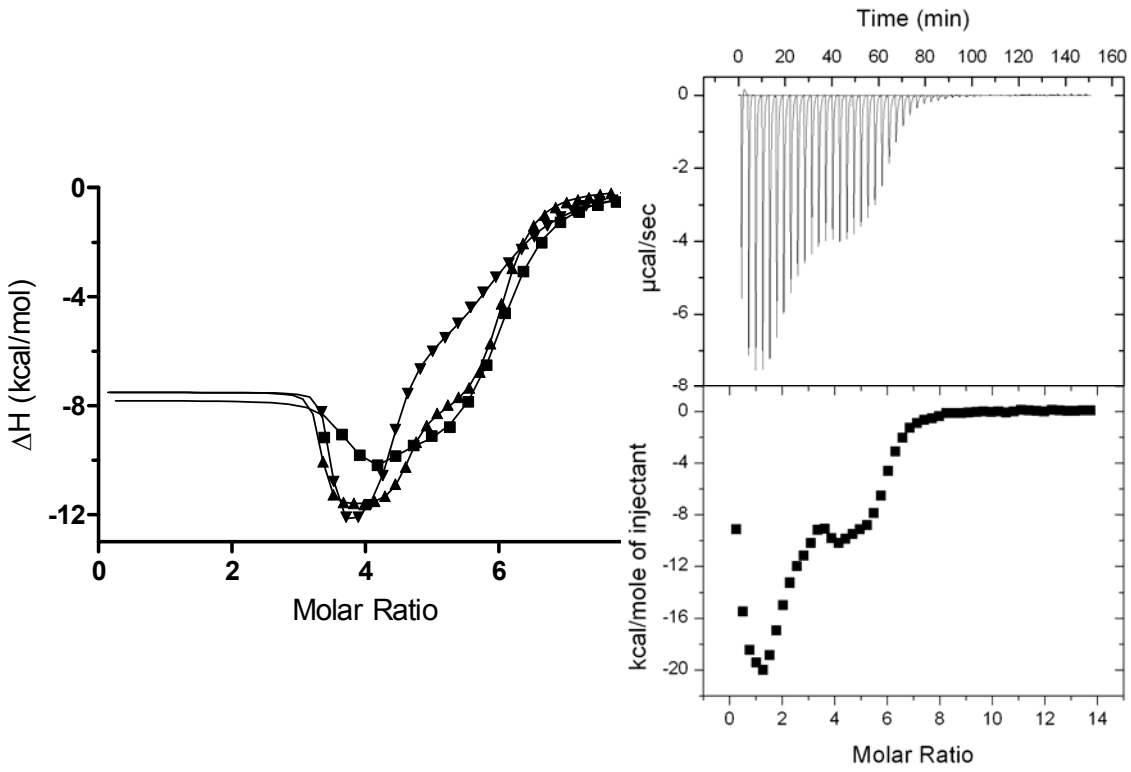


Figure A.3 Titrations of TMPyP4 into 32-mer MWT•5C capped construct (-■-), 1:2:1•5C capped construct (-▲-), and 1:6:1•5C capped construct (-▼-) are shown in the left panel.

Heat signatures between 0 and 3.0 Molar Ratio are attributed to ligand induced restructuring and excluded from the non-linear fit. The solid lines represent the non-linear best fits. Raw and integrated heats for a typical titration of TMPyP4 into 32-mer MWT•5C capped DNA complex are shown in the right panel.

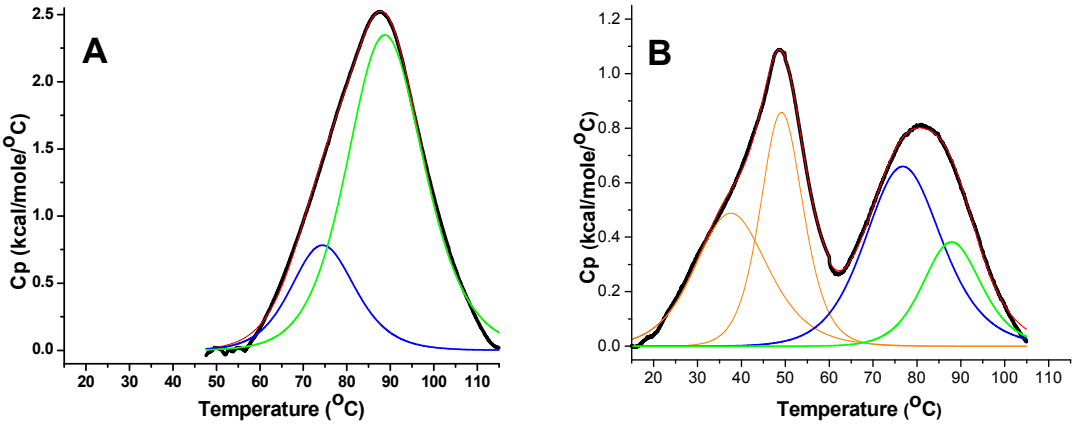


Figure A.4 A) DSC melting and deconvoluted melting curves for the thermal denaturation of 36-mer WT c-MYC. B) DSC melting for the thermal denaturation of the 36-mer WT•4T capping construct.

Orange curves (—) represent the deconvoluted melting curves for double stranded flanking regions. Blue curves (—) represent the deconvoluted melting curves for 1:6:1 quadruplex foldamer. Green curves (—) represent the deconvoluted melting curves for 1:2:1 quadruplex foldamer.

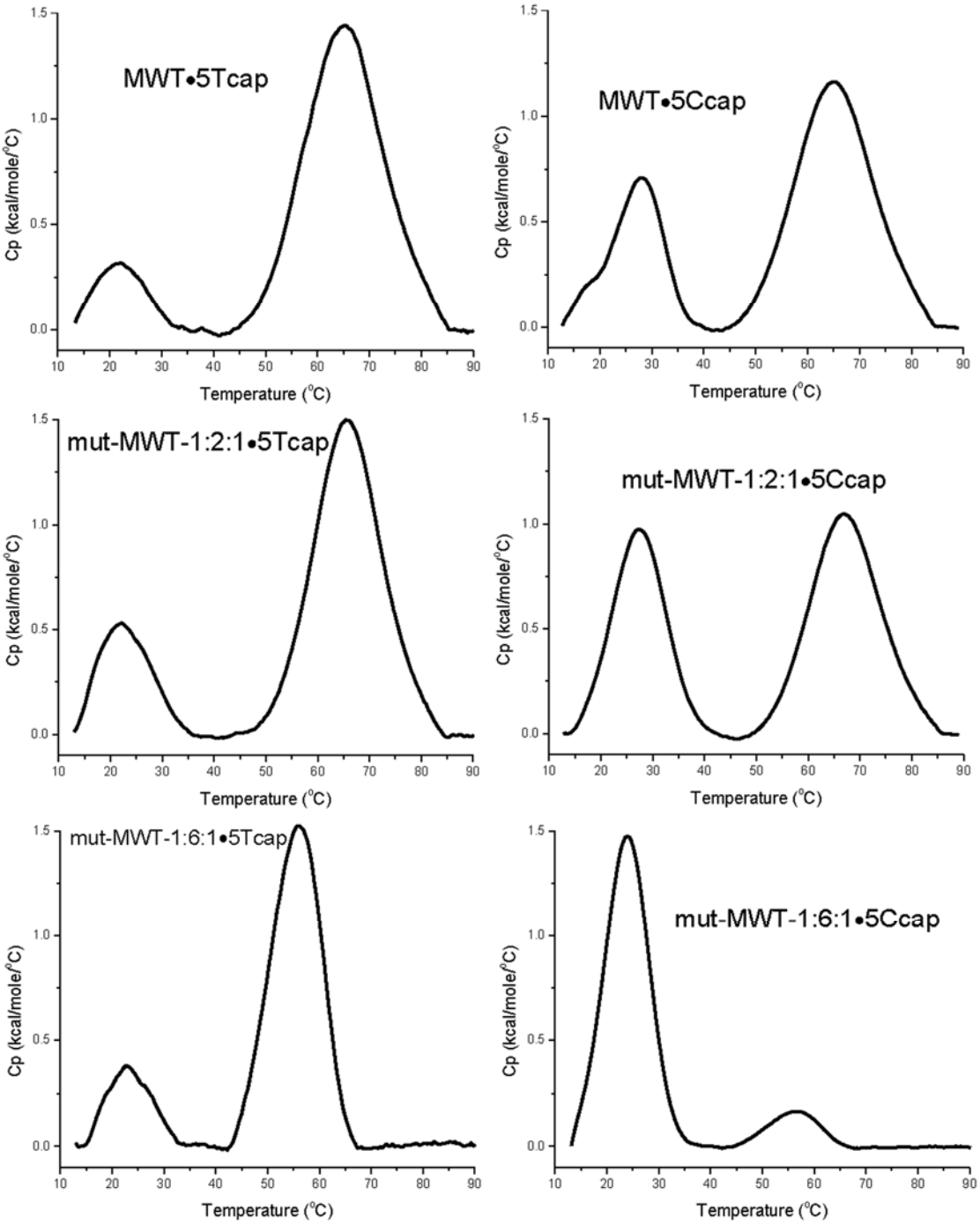


Figure A.5 DSC thermograms showing the deleterious effect of the 5-C cap on the thermal stability of the "capped" MWT and "capped" mutant MWT. DSC thermograms for 5-T "capped" complexes are shown on the left, DSC thermograms for 5-C "capped" complexes are shown on the right. The 5-C "capped" complexes exhibited increased amount of duplex melting evidenced by the melting peaks centered around 27 °C.

APPENDIX B

SUPPLEMENTARY MATERIALS FOR CHAPTER VII

Table B.1 Thermodynamic parameters (given in kcal/mol) for the complexation formation of C_{60} and C_{70} with Clip in toluene

	C_{60}				C_{70}			
T (°C)	K_a	ΔG	ΔH	- TΔS	K_a	ΔG	ΔH	- TΔS
5	5700	-4.78±0.01	-3.83±0.05	-0.96±0.04	10500	-5.11±0.03	-2.23±0.06	-2.89±0.07
15	4200	-4.77±0.02	-4.37±0.06	-0.39±0.08	6500	-5.02±0.01	-2.66±0.05	-2.37±0.06
25	3200	-4.77±0.02	-4.61±0.11	-0.16±0.02	4600	-4.99±0.02	-2.95±0.04	-2.04±0.06
35	1900	-4.65±0.41	-5.25±0.54	0.59±0.07	3400	-4.96±0.47	-3.07±0.04	-1.89±0.11

Table B.2 Thermodynamic parameters (given in kcal/mol) for the complexation formation of C_{60} and C_{70} with Clip in chlorobenzene

	C_{60}				C_{70}			
T (°C)	K_a	ΔG	ΔH	- TΔS	K_a	ΔG	ΔH	- TΔS
5	1300	-3.96±0.01	-3.35±0.01	-0.31±0.01	1300	-3.91±0.01	-3.99±0.03	0.08±0.03
25	800	-3.98±0.01	-4.03±0.04	0.04±0.04	900	-4.06±0.02	-3.76±0.06	-0.29±0.07
50	500	-4.02±0.05	-4.08±0.01	0.06±0.05	500	-4.04±0.03	-4.07±0.02	0.03±0.04

Table B.3 Thermodynamic parameters (given in kcal/mol) for the complexation formation of C_{60} and C_{70} with Clip in *o*-dichlorobenzene

	C_{60}				C_{70}			
T (°C)	K_a	ΔG	ΔH	- TΔS	K_a	ΔG	ΔH	- TΔS
5	300	-3.12±0.05	-1.53±0.06	-1.59±0.11	270	-3.08±0.07	-1.83±0.06	-1.25±0.12
25	200	-3.11±0.09	-1.87±0.06	-1.25±0.15	200	-3.13±0.03	-1.98±0.06	-1.15±0.06
50	100	-2.95±0.02	-3.50±0.06	0.55±0.05	100	-2.95±0.07	-3.15±0.06	0.26±0.03

Table B.4 Thermodynamic parameters (given in kcal/mol) for the complexation formation of C_{60} and C_{70} with Clip in anisole

	C_{60}				C_{70}			
T (K)	K_a	ΔG	ΔH	- TΔS	K_a	ΔG	ΔH	- TΔS
298	3290	-4.78±0.05	-6.59±0.03	1.81±0.08	4705	-4.99±0.05	-4.24±0.06	-0.75±0.11

Table B.5 Thermodynamic parameters (given in kcal/mol) for the complexation formation of C_{60} and C_{70} with Clip in 1,1,2,2-tetrachloroethane

	C_{60}				C_{70}			
T (K)	K_a	ΔG	ΔH	- TΔS	K_a	ΔG	ΔH	- TΔS
1007	200	-4.08±0.07	-5.89±0.05	1.41±0.12	957	-4.05±0.03	-5.33±0.06	1.28±0.09