APPLICATION OF DYNAMIC CONBINTORIAL CHEMISTRY TO IDENTIFY NEW COMPOUNDS THAT BIND G-QUADRUPLEX DNA
&
PROBING THE ROLE OF THE CATION-PI INTERACTION BETWEEN THE HP1 CHROMODOMAIN AND METHYLATED LYSINE USING UNNATURAL AMINO ACIDS

Elizabeth Jane Cline

A dissertation submitted to the Faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2013

Approved by:

Marcey Waters

Jeffrey Johnson

Eric Brustad

Michel Gagné

Joseph Templeton

©2013
Elizabeth Jane Cline
ALL RIGHTS RESERVED

ABSTRACT

Elizabeth Jane Cline: Application of Dynamic Combinatorial Chemistry to Identify New Compounds that Bind G-Quadruplex DNA & Probing the Role of the Cation-π Interaction Between the HP1 Chromodomain and Methylated Lysine Using Unnatural Amino Acids
(Under the direction of Marcey L. Waters)

This dissertation discusses two different projects. The first project involves the development of cyclic-peptide acridine conjugates in the effort to identify molecules that can selectively bind to and stabilize G-quadruplex DNA. The second project seeks to probe the role of the cation-π interaction in the binding of the HP1 chromodomain to trimethylated lysine 9 of histone 3 (H3).

In recent years, interest in the G-quadruplex DNA structure has increased enormously due to the unique physical properties of this secondary DNA structure as well as the presence of guanine-rich sequences in biologically functional regions of many genomes. Given the propensity of G-quadruplex structures to control many biological functions, it has become desirable to identify small molecules that can bind to and stabilize G-quadruplexes. This work aims to develop quadruplex ligands that exhibit selectively over not only duplex DNA but also over various quadruplex sequences. It was believed that cyclic peptides could deliver selectivity for the quadruplex structure over duplex DNA while also providing the added advantages of mimicking native protein structure, displaying enhanced metabolic stability and possessing structural preorganization. Using a strategy that has been developed in our lab, we propose screening libraries of cyclic peptides generated *in situ* using thiol-thioester exchange for dynamic combinatorial chemistry (DCC). These libraries can efficiently be screen against different

quadruplex sequences as well as duplex DNA in order to determine the selectivity of each species.

In the second project, we sought to characterize the noncovalent interactions responsible for the recognition of trimethylated lysine 9 of histone 3 by the aromatic pocket of the HP1 chromodomain. Lysine can exist in three distinct methylation states under the control of highly specific methyl transferases or demethylases. These methylation states serve to turn on specific protein-protein interactions with partners that specifically recognize the methylated side chain. Recognition and affinity is mainly derived from cation-π interactions between the positively charged cationic side chain and the electron rich π surfaces of nearby aromatic rings. This interaction can be quantified by incorporation of fluorinated derivatives of the aromatic amino acids responsible for the binding of H3K9Me3 to the HP1 chromodomain. The cation-π interaction between the aromatic pocket of the HP1 chromodomain and H3K9Me3 can be revealed by incorporation of a series of fluorinated amino acid analogues. A linear correlation between binding affinity and the calculated magnitude of the cation-pi interaction of those groups indicates a cation-π interaction. Because many reader proteins for methylated lysine have an aromatic cage in their binding pockets, findings from the investigation of the HP1 chromodomain will provide broad insight into this class of proteins.

ACKNOWLEDGEMENTS

First, I would like to thank Marcey Waters for giving me the opportunity to join her lab. You have been an incredible mentor and friend over the past three years, providing a welcoming and supportive learning environment. Because of this I have learned an incredible amount and have grown considerably as a chemist. I am so grateful for the opportunities that have been given to me since joining your lab.

Additionally, I would like to thank all members of the Waters lab, past and present. It has been a rewarding experience working next to such a diverse group of scientists and as a result, I have learned a little from each of you. I would specifically like to thank Dr. Dale Wilger, Dr. Jes Park and especially Dr. Soumyadip Ghosh for teaching and mentoring me when I first joined the lab.

I am extremely grateful for the presence of Kaiualani, Amber and Effie in the lab. Working next to you has been an incredible experience and has made the lab a friendly, positive and supportive environment to be in. You have been my biggest cheerleaders, sharing my successes and providing much needed motivation in spite of my failures. I will always be grateful for that. In addition to playing a positive role in the lab, you have also been amazing friends outside of the lab as well. I have many experiences with each of you that I will never forget and I am so lucky to call you my friends.

I have made wonderful friends during my time here in Chapel Hill, providing me with many rewarding experiences outside of lab over the past five years. There are so

many people that have made Chapel Hill an enjoyable place to live that it would take too long to list all of you. But, I would specifically like to thank Bryan and Danielle for being remarkable friends to me. I know I never would have made it through graduate school without your friendship and support. Our lives have taken us to different places but I hope our friendships will continue and grow even stronger.

And, finally I need to thank my family. I would not be the person that I am today without the support of my family. To James, you have become one of my best friends over the past five years and I am incredibly proud to call you my brother. I would be absolutely nowhere without my mom and dad. They have been amazing parents to me throughout my entire life. Mom, you have always pushed me to be my best even when I didn't want to and I have become a stronger and more independent woman because of it. Dad, you have always had my back and supported me no matter what. You have always been a voice of reason in my life and I will never forget everything you have done for me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	Amino acid
Ac_2O	Acetic anhydride
Ach	Acetylcholine
AcOH	Acetic acid
AdoMet	S-adenosylmethionine
Arg, R	Arginine
Asp, D	Aspartic acid
C	Carbon
CD	Circular Dichroism
CH_2Cl_2	Dichloromethane
CN	Cyano
CrO_3	Chromium(VI) oxide
Cys, C	Cysteine
DBU	Diazabicycloundecene
DCC	Dynamic combinatorial chemistry
DIC	Diisopropylcarbodiimide
DIPEA	Diisopropylethylamine
DMF	Dimethylformamide
DNA	Deoxyribose nucleic acid
DSB	Double stranded break
dsDNA	Double stranded DNA

EDT	Ethanedithiol
ESI-TOF	Electrospray Ionization-Time of flight
Et_2O	Diethyl ether
EtOAc	Ethyl acetate
EtOH	Ethanol
F	Fluorine
Fe(II)	Iron(II)
FID	Fluorescence indicator displacement
Fmoc	9-Fluorenylmethyl chloroformate
FRET	Fluorescence resonance energy transfer
G	Guanine
G4	G-quadruplex
Gln, Q	Glutamine
Glu, E	Glutamic acid
H_2SO_4	Sulfuric acid
HBTU	O-Benzotriazole-N,N,N',N'-tetramethyluronium hexafluorophosphate
HCl	Hydrochloric acid
HNO_3	Nitric acid
HOBt	1-Hydroxybenzotriazole
HP1	Heterochromatin protein 1
HPLC	High performance liquid chromatography
Htelo	Human telomeric
Hyp	Hydroxyproline

K	Potassium
KCl	Potassium chloride
KDM	Lysine demethylases
KMT	Lysine methyltransferases
LC/MS	Liquid chromatography/mass spectrometry
Lys, K	Lysine
Me	Methyl
MeCN	Acetonitrile
Met, M	Methionine
N	Nitrogen
N_2	Nitrogen (g)
Na	Sodium
nAChRs	Nicotinic acetylchloine receptors
NaHCO3	Sodium bicarbonate
NaOH	Sodium Hydroxide
NH_4OAc	Ammonium acetate
NHPPE	Nuclease hypersensitive polypurine-polypyrimidine element
NMR	Nuclear magnetic resonance
NO_2	Nitro
Phe, F	Phenylalanine
Pro, P	Proline
PTM	Post-translational modification
PyBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate

S	Sulfur
$SnCl_2$	Tin(II) chloride
SPPS	Solid phase peptide synthesis
SPR	Surface plasmon resonance
tBu	*tert*-Butyl
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Trp, W	Tryptophan
TSS	Transcriptional start site
Val, V	Valine

CHAPTER 1: INTRODUCTION
STRUCTURE AND FUNCTION OF G-QUADRUPLEX DNA AND METHODS FOR TARGETING THE G-QUADRUPLEX STRUCTURE

A. Significance

The right-handed double helical structure of DNA is well known. However, growing evidence has suggested that DNA can adopt a variety of alternative conformations based on sequence motifs and interactions with various proteins.[1] One example of these alternative DNA structures that has gained great interest in recent years is the G-quadruplex (G4) structure. Interest has increased enormously due to the unique physical properties of the G-quadruplex structure as well as the presence of guanine (G)-rich sequences in biologically functional regions of many genomes.[2] For example, G-rich regions with the potential to form G-quadruplex structures are found in telomeres, double stranded break (DSB) sites, and transcriptional start sites (TSS). In addition to the important regulatory and structural functions of the G4 structure, it could also be a source of genomic instability leading to cancer, aging and human genetic diseases. Given the propensity of G-quadruplex structures to control many biological functions, it has become desirable to identify small molecules that can bind to and stabilize G-quadruplexes.[3] Moreover, despite the fact that G4 DNA has been studied rigorously *in vitro*, further studies are necessary to determine if this structure forms *in vivo* and what exactly its cellular roles might be. Small molecules that target the quadruplex structure could be useful in further probing the biological function and significance of quadruplex structures *in vivo*.

B. G-Quadruplex Structure

G-rich sequences of nucleic acids can form four-stranded, helical DNA or RNA structures comprised of 4 guanine bases brought together in a square planar arrangement connected by Hoogsten hydrogen bonds (Fig. 1.1a).[5,6] The G-rich core typically consists of two or more G-tetrads stacked via π-π interactions with intervening sequences extruded as single-stranded loops (Fig. 1.1b). The sequence and size of the loop regions can vary, with the loops usually containing 1-7 nucleotides. The stacks are joined together by the normal sugar-phosphate backbone with the structure further stabilized by monovalent cations, typically potassium (K^+) or sodium (Na^+), that occupy the central cavity neutralizing the electrostatic repulsion of the guanine oxygens (Fig. 1.1a).[7,8]

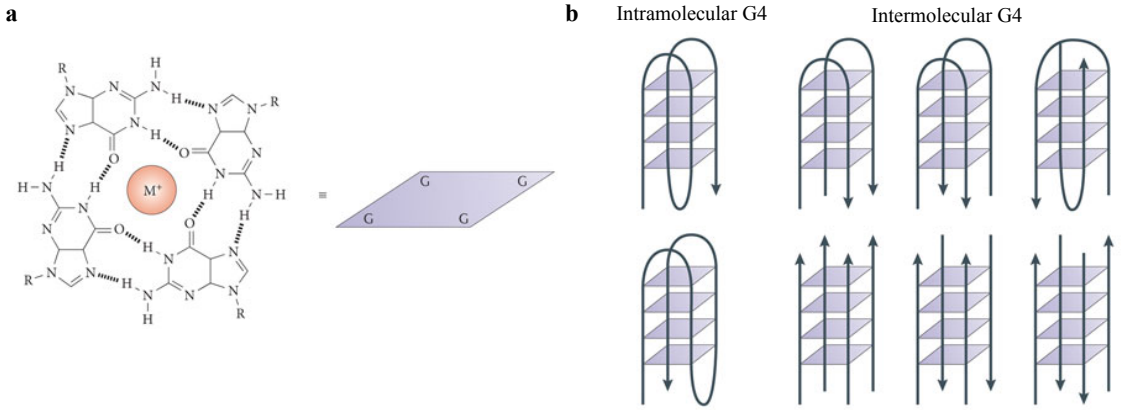

Figure 1.1 The G-quadruplex structure (a) An illustration of the G-quadruplex G-quartet. The quartet is pictured as a square in the other panels of the figure. M^+ denotes a monovalent cation. (b) Schematic of intramolecular (left) and intermolecular (right) G-quadruplex DNA structures. The arrows indicate the direction of the DNA strands.[4]

The quadruplex structure can be extremely stable. However, the stability can vary due to differences in length and sequence of the loop, strand orientation and alignment and the nature of the binding cations. G-quadruplex structures can be parallel, antiparallel or hybrids of the two and can be made up of one (intramolecular), two or four separate strands (intermolecular) (Fig. 1.1b).[6] Loops may link positions on the top (or bottom) of the stacks, forming diagonal (Fig.

1.2c) or lateral loops (Fig. 1.2b) depending on which guanines are linked. Additionally, loops can link a guanine on the top G-tetrad to a guanine on the bottom providing a double-chain reversal loop (Fig. 1.2a).[9-12] Consequently, the G4 motif can adopt a great variety of different topologies with different combinations of strand polarities and loop types. In addition to differences in the length and sequences of the loop, the structure of the quadruplex can vary greatly due to differences in groove size based on relative strand orientation. All quadruplexes have four grooves, defined as the cavities bounded by the phosphodiester backbone. Parallel quadruplexes have all guanine glycosidic angles in an *anti* conformation while anti-parallel quadruplexes contain both *syn* and *anti* guanines arranged in a way that is particular for a given topology and set of strand orientations.[6] Furthermore, the formation and stability of the G-quadruplex is monovalent cation-dependent. A strong negative electrostatic potential is created by the guanine O6 oxygen atoms which form a central channel of the G-tetrad stack where the cations are located (Fig. 1.1a). The precise location of the cation between the tetrads is dependent on the nature of the cation. Potassium (K^+) and sodium (Na^+) are the most prevalent cations, with the smaller Na^+ cation orienting itself in the plane of the G-tetrad while the larger K^+ ion lies equidistant between each tetrad plane. Studies have shown that the change from Na^+ to K^+ can induce profound structural changes.[13,14]

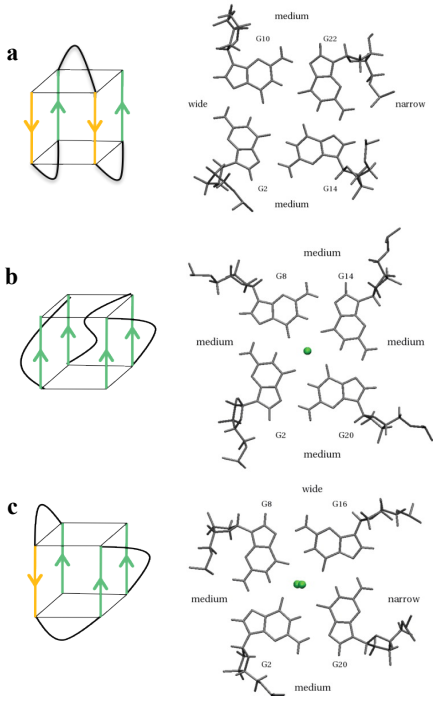

Figure 1.2 Structures of the human unimolecular telomeric quadruplex formed by the sequence d[AGGG(TTAGGG)$_3$].[6] In each case two views are shown with arrows depicting strand polarities. (a) The Na^+ form, determined by NMR containing two anti-parallel and two parallel strands with a diagonal and two lateral loops. (b) K^+ form A, determined by crystallography containing all parallel strands with three strand reversal loops resulting in four equal grooves (c) K^+ form B, showing the topology determined by NMR, containing three parallel strands with one oriented anti-parallel.

It has been predicted that intramolecular quadruplex structures can form at specific G-rich regions *in vivo*. These G-rich regions share a common sequence motif of at least four guanines, with each G-tract containing most often at least three guanines. Computational analyses have revealed that there are >375,000 G4 motifs in the human genome including those in ribosomal and telomeric DNA. However, it is still unclear how many of these motifs can form stable G-quadruplex structures *in vivo* and, if they do, when they form.[15-18]

C. Biological Relevance

Computational studies have revealed that G4 motifs are not randomly located within the genome, but tend to cluster in particular genome regions such as promoters. Due to the nonrandom location and evolutionary conservation of the position of G4 motifs within the genome, it is suggested that the G-quadruplex structure plays one or more positive functions within the cell.[19,20] Telomeres are regions of the chromosome that contain a high concentration of G4 motifs due to their high GC content and the single-stranded nature of the telomeric overhang. G4 DNA motifs are also common G-rich regions up- and downstream of TSSs and near transcription factor binding sites.

i. G4 Structures at telomeres

Telomeres make up a nucleoprotein complex at the ends of linear chromosomes that protect chromosomes from degradation, end-to-end fusions, and being recognized as DSBs.[21] They contain a double-stranded region and a single-stranded G-rich 3' overhang (Fig. 1.3). In most telomeric DNAs, guanines and cytosines are distributed asymmetrically between two DNA strands with the G-rich strand being longer than its complement, resulting in single-stranded 'G-tails' at the termini of the chromosome. Despite differences in the exact sequence, the G-rich strand of various telomeric sequences can usually form stable G4 structures *in vitro*.[22-24]

Due to the biochemical properties of DNA polymerases, they cannot replicate the very ends of linear chromosomes. Therefore, as a healthy cell undergoes replication, there is a progressive shortening of the telomeric region following each cell division. Normal cells have a limited replicative lifespan and once a critical shortening of the telomere is reached, the cell is signaled for growth arrest or apoptosis. However, a reverse transcriptase enzyme, telomerase, can be activated to lengthen the G-strand of the telomeric region (Fig. 1.3). Human telomerase is

inactive in most somatic cells but is found to be upregulated in many cancers where it is thought to promote the lifespan of malignant cells.[25] The G-quadruplex structure can inhibit the activity of telomerase by preventing the enzyme from annealing to G-strand overhangs.[26,27] Because telomerase is active in a large number of human cancers and can be influenced by the formation of the G-quadruplex structure, there has been great effort to design a variety of small molecules that can bind and stabilize G4 structures (Fig. 1.3).[28]

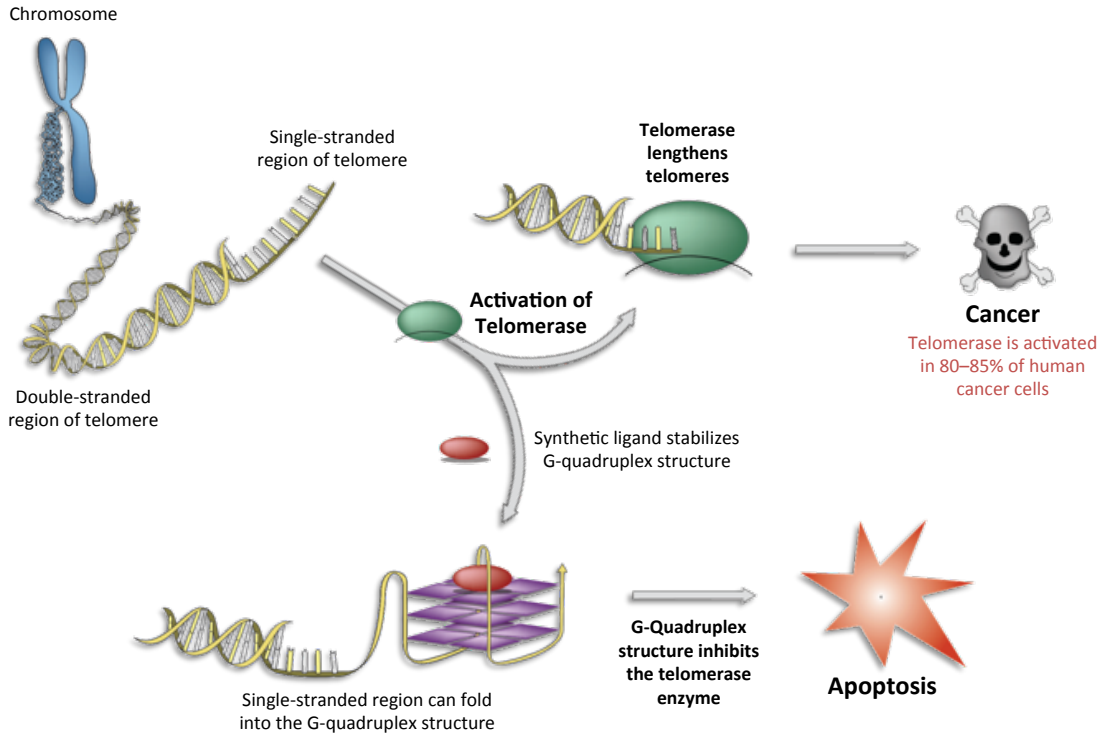

Figure 1.3 Schematic outlining the role of human telomerase in cancer. Telomeric sequences can fold into G-quadruplex structure inhibiting the activity of the reverse transcriptase telomerase. Due to the ability of the G4 motif to inhibit the enzyme telomerase, it has become desirable to design small molecules that can bind to and stabilize the quadruplex structure for use as novel anticancer therapies.

ii. Effects of G4 structures on DNA transcriptional regulation

In addition to their presence in telomeres, a high concentration of G4 motifs have been found in the promoter regions of genes suggesting that G4 structures play a potential role in the

regulation of gene transcription. Bioinformatics has shown that the promoter region of human oncogenes and regulatory genes (i.e. transcription factors) are more likely than the average gene to contain G4 motifs.[29] There is an equilibrium between two forms of the DNA during transcription. One side of the equilibrium is the double helix DNA, while on the other side one strand is separated and has folded up into a G-quadruplex.[30] As a result, studies have suggested that the presence of some G4 structures during DNA replication could influence transcription in positive and negative ways. Depending on which DNA strand contains the G4 motif, the structure could inhibit transcription (if the motif is on the template strand) or enhance transcription (if the motif is on the non-template strand) (Fig. 1.4a and 1.4b). Also, proteins that bind to the G4 structures may also affect transcription (Fig. 1.4c and 1.4d).[31-33] A number of oncogenes have been found to contain the G4 motifs in their promoter region leading to the possibility of novel therapeutics that target the quadruplex structure. The addition of a G-quadruplex ligand will energetically favor the G4 structure, thus providing a way to manipulate or control transcription of a particular gene.

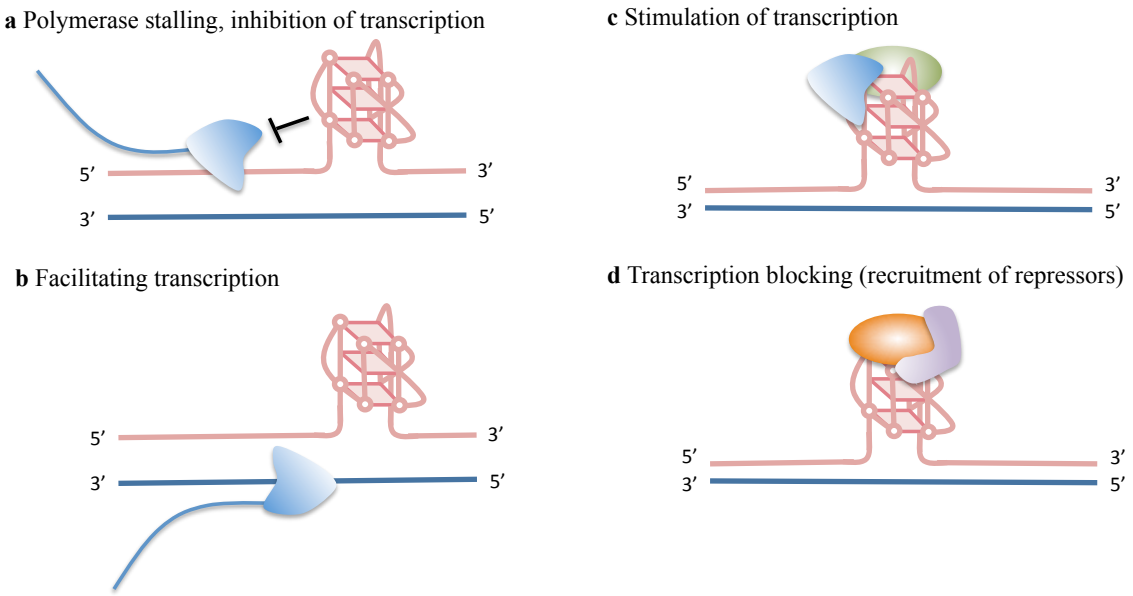

Figure 1.4 Proposed functional roles of G-quadruplex structures during transcription.[4] (a) G4 structures are thought to block transcription by inhibiting polymerase (blue). (b) G4 structures are postulated to facilitate transcription by keeping the transcribed strand in the single-stranded conformation. (c) G4 structure may stimulate transcription by recruiting proteins (green) that recruit or stimulate polymerase (d) G4 structures are suggested to block transcription by recruiting G4 binding proteins (orange) that can directly or indirectly repress transcription (purple).

D. Methods for Determination of Quadruplex Topology and Structure

i. CD Spectroscopy

A number of quadruplex studies have used biophysical chemistry, mainly circular dichroism (CD) to assign topology. CD is sensitive to stereochemical variations, allowing it to be an important technique for studying subtle conformational changes and supramolecular interactions.[34] Using CD spectroscopy it is possible to discriminate between quadruplex topologies having differences in parallel and anti-parallel strand orientation due to different arrangements of *anti/syn* glycosidic angles. Therefore, CD is a useful and rapid method for establishing an overall fold pattern as it requires little sample and is suited to examining a wide range of solution conditions and their influence on quadruplex formation. Classic parallel and

anti-parallel quadruplexes provide similar traces but with maxima at distinct wavelengths. For quadruplexes assigned to be parallel-stranded, a maximum is present at ~260 nm and a minimum at ~240 nm while the maximum and minimum for an anti-parallel quadruplex are typically at around 290 and 260 nm, respectively.[35-37] CD has predominantly been used to study telomeric sequences and those with regular repeating loop regions. More complex quadruplex-forming show decreased reliability when assigning topology since they may not conform to observed telomeric quadruplexes, multiple species cannot be identified by CD and non-telomeric loop sequences may perturb the CD spectra in unforeseen ways.[6,38]

ii. X-ray Crystallography and NMR Spectroscopy

X-ray crystallography and high-field NMR spectroscopy are other ways to determine topological assignment as well as a more detailed atomic-level structure determination.[39] However, there can be limitations to each of these methods. In order to successfully determine a structure by NMR, the sequence needs to form a kinetically stable species in solution while the presence of multiple species limits the structural information that can be obtained. It is common to set up a screen of mutated sequences and other variants until one is found that produces a well-resolved NMR spectrum showing a single species amenable to analysis.[40] Similarly, crystallography uses site mutations and/or sequence screening to find sequences that will crystallize.[6] The various structures formed in solution as variants of the human telomeric two-repeat sequence show that such mutations and changes cannot always reliably preserve a particular topology and will inevitably alter the equilibrium between different structures.[41] Therefore, generalizations from any one NMR or crystal structure needs to be done carefully and need to acknowledge the contribution of modified/additional nucleotides.

E. G-Quadruplex Ligands

The G-quadruplex structure has been implicated in several biological dysfunctions that selectively alter the integrity of cancer cells. As a result, G-quadruplex binders that stabilize the G4 structure promise to lead to the discovery of novel anticancer agents. A common binding motif for a ligand is a flat aromatic molecule that binds to the G-quartet of the external face of the quadruplex resulting in stabilization of the structure through π-π stacking and electrostatic interactions (Figure 1.5b). The G-quartet has a large flat surface therefore an efficient quadruplex ligand should feature a large aromatic surface, much larger than that of a duplex binder, improving the aromatic-aromatic overlap and providing selectivity over duplex DNA.[6,41,42] Electrostatic interactions between positively charged sidechains of the ligands and the negatively charged phosphate backbone of the G-quadruplex DNA scaffold also contribute to stabilization of the G4 structure (Figure 1.5b). Generally, the ligands that have been synthesized to date can be classified into four different categories based on their cationic nature, *i.e.* cationic (1) upon *in situ* protonation of an amine appendage, (2) *via* N-methylation of an aza-aromatic moiety, (3) presence of a metal center, or (4) non-cationic ligands (Figure 1.6).[43]

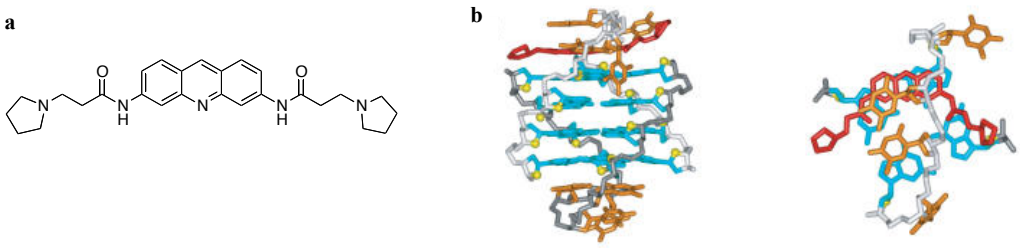

Figure 1.5 Despite differences in the many ligands developed that successfully target G-quadruplex DNA over duplex DNA, G4 ligands tend to have two common features including a large flat aromatic surface for π stacking with the G-tetrad and protonable side chains for electrostatic interaction with the quadruplex grooves. (a) Shown is an acridine ligand that targets G-quadruplex DNA over duplex DNA and displays the two common characteristics of G4 ligands. The acridine core provides a large flat aromatic surface for binding while the protonable pyrrolidine side chains provide positive electrostatic interactions. (b) X-ray structure of the acridine ligand (red) complexed with a G4 structure (quartet=blue, backbone=grey, loops=orange) illustrating how the ligands can end stack with the G-tetrads of the quadruplex structure.[43]

The usual way to target G-quadruplex DNA is the introduction of protonable sidearms around an aromatic core. Neidle, Hurley and co-workers followed this method with the development of a bisamidoanthraquinone G-quadruplex ligand and telomerase inhibitor.[44] While this class of molecule was shown to have telomerase inhibition activity, these molecules had insufficient selectivity for G4 DNA over duplex DNA for further biological studies. To improve the selectivity for quadruplex DNA, Neidle and coworkers progressively modified the core and sidearms from anthraquinone to fluorenone, then acridone and acridine.[45,46] A member of the 3,6-disubstituted acridine series showed particular promise against G-quadruplex DNA (Figure 1.5a).[47] This structure displayed hydrophobic-π-stacking interactions between the flat aromatic core of the acridine and the square planar guanine residues of the G-tetrad in addition to electrostatic interactions between the two protonable sidechains of the ligand and the quadruplex grooves. Neidle and coworkers further optimized this class of molecules by designing BRACO-19, a molecule able to interact simultaneously with three G-quadruplex grooves due to the

addition of a third side arm.[48] The optimized BRACO-19 displayed high levels of quadruplex stabilization as determined by FRET (fluorescence resonance energy transfer)-melting assay and by SPR (surface plasmon resonance) method that revealed 31-fold selectivity for the quadruplex structure. Additionally, BRACO-19 shows a strong potency for telomerase inhibition as well as inhibition of cancer cell proliferation however, pharmacological use of this molecule has been limited by poor membrane permeability or cellular uptake.[49] Despite these limitations, the simple acridine motif appears to be very valuable for G-quadruplex recognition and one that we have incorporated into our system design.

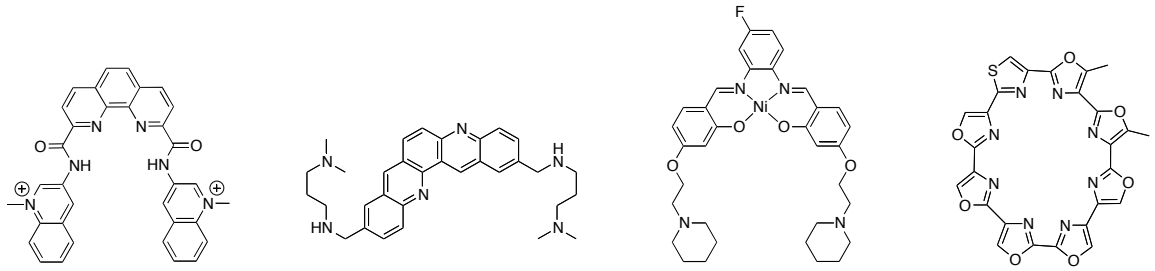

Figure 1.6 Molecules exemplifying the four broad categories of G-quadruplex binding ligands.

Despite the number of ligands designed that successfully target G-quadruplex DNA over duplex by targeting the common G-tetrad of the G4 motif, a major challenge still remains to design ligands with specificity for individual quadruplex structures given the wide variety of topologies specific quadruplex sequences can adopt. It may be possible to obtain selectivity by targeting the loops that vary in length and sequence as well as the grooves, which can vary in size due to difference in relative strand orientation. A common strategy that has been employed to achieve greater high selectivity and specificity is conjugation, using a ligand that can target the planar surface of the G-tetrad and appending it to a variable substituent that can contact the loops and grooves of the quadruplex structure.

An initial example comes from Pedroso et al., where a library of acridine-oligonucleotide conjugates were synthesized and studies were carried out to examine their interaction with DNA quadruplexes.[50] The acridine ligand was expected to exhibit selectivity for quadruplex DNA over duplex while the oligonucleotide attached was complementary to the G-rich telomere strand. These studies showed that the acridine-oligonucleotides did stabilize the telomeric quadruplex structure as determined by thermal denaturation experiments; however, no other sequences were screened against the library to determine selectivity.

Arya et al. published a more recent example of a novel perylene-neomycin conjugate that was shown to bind preferentially to telomeric G-quadruplex DNA in the presence of other nucleic acids, including DNA, RNA, DNA-RNA hybrids, and other higher order structures (single strands, duplexes, triplexes, other G-quadruplexes, and the i-motif).[51] The perylene moiety is known to end stack with the G-tetrads of the quadruplex structure, while recent studies have shown that neomycin can bind in the wide groove of G4 DNA. Arya et al. showed through fluorescence intercalator displacement (FID) assays that the conjugated species displayed tighter binding than each of the parent constituents.

In a final example, Neidle et al. published studies performed with a common G-tetrad interacting acridine core conjugated to various tetrapeptide substituents that have the ability to discriminate between different quadruplex types.[52] Based on the substitution patterns selectivity could be seen for the parallel quadruplex derived from the human *N-ras* gene over the human telomeric quadruplex while the peptide substituent KRSR proved to show marginal superior selectivity and affinity over FRHR While linear peptides provided modest binding selectivity, we felt we could further develop this method by using cyclic peptides to deliver selectivity.

F. Methods to Investigate G-Quadruplex/Ligand Interactions

A variety of techniques have been employed to study the interaction of natural or synthetic compounds with G-quadruplex structures. These methods can vary from evaluating properties like ligand affinity, to more sophisticated methods used for determining kinetic, thermodynamic, stoichiometric and conformational data for structure-activity relationship studies. Methods being used to investigate these interactions should be able to detect and measure the ligand selectivity for quadruplexes over duplexes or other secondary structures. Each method has its own advantages and disadvantages therefore more than one method is usually necessary to obtain complete information about quadruplex DNA/ligand interactions.

i. Melting Temperature Measurements

Melting temperatures can be measured, providing information about the stabilization (or destabilization) of the quadruplex structure by the ligand under investigation. Quadruplex nucleic acids show a strong absorbance around 295 nm, which is greatly reduced in unstructured DNA due to the π-stacking of nucleobases in folded quadruplex structures (Figure 1.7).[53] There is a decrease in absorbance at 295 nm upon denaturation by heating. This simple thermal melting experiment is performed to demonstrate the stabilization or destabilization of a quadruplex structure by a ligand. Preparation of the sample is straightforward and reasonably low amounts are required.

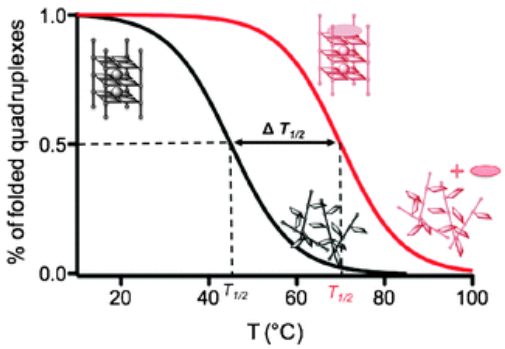

Figure 1.7 Schematic representation of melting temperature experiments using CD at 295 nm.[53]

ii. CD Spectroscopy

Circular dichroism (CD) spectroscopy is extremely useful in studying the conformation of nucleic acids. As discussed previously, CD is a widely used technique to study the G-quadruplex conformation in which the CD signal is influenced by stacking interactions between adjacent quartets.[54] The titration of G4-ligands to quadruplexes can induce changes in the CD spectrum, correlating to the properties of the ligand under investigation.

iii. Surface Plasmon Resonance

Surface Plasmon Resonance (SPR) has become a popular method to study the kinetic and thermodynamic parameters of intermolecular interactions. SPR is an optical technique that utilizes the decrease in intensity of a circular polarized light reflected from the surface of a glass prism coated with a thing metal film. The minima in reflected light density occur at a specific angle at which surface plasmons of the metal are excited. The angle at which the resonance occurs becomes the function of refractive index of the medium in the vicinity of the surface. The variation of value at this angle is recorded, detecting any interactions involving free species in solution (ligand) and one that is covalently bound to the surface (DNA). A variety of quadruplex forming oligonucleotides have been immobilized to sensor chips to study binding interactions with small organic molecules and proteins.[53]

G. Purpose of this work

Given the important regulatory and structural functions of the G4 motif, we aim to develop quadruplex ligands that exhibit selectively over not only duplex DNA but also over various quadruplex sequences. In the effort to discover molecules that can discriminate against assorted quadruplex structures via groove recognition, we have screened libraries of cyclic peptide-acridine using dynamic combinatorial chemistry (DCC). We chose to focus on cyclic peptides due to their prevalence in nature and their promise as biologically relevant molecules. Additionally, the use of cyclic peptides to target the quadruplex structure seemed promising due to the ability of the macrocyclic, natural product telomestatin to bind G-quadruplex DNA. Furthermore, in order to screen libraries of cyclic peptides in a high-throughput format, a method has been developed in our lab using thiol-thioester exchange in DCC. Using this method, we can generate libraries of cyclic peptides from simple linear peptide building blocks in solution, within hours, and at a neutral pH. These libraries can efficiently be screened against different quadruplex sequences as well as duplex DNA in the effort to determine the selectivity of each species. Once a hit has been discovered, the molecule can easily be resynthesized as a more stable analogue for use in biological studies.

References

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

2. Wu, Y.; Brosh, R. B. G-Quadruplex nucleic acids and human disease. *FEBS Journal* **2010,** *277,* 3470-3488.

3. Balasubramanian, S.; Neidle, S. G-Quadruplex nucleic acids as therapeutic targets *Curr. Opin. Chem. Biol.* **2009,** *13*, 345−353.

4. Bochman, M. L.; Paeschke K.; Zakian V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Rev. Genetics* **2012,** *13,* 770-780.

5. Huppert, J. L. Structure, location and interactions of G-quadruplexes *FEBS Journal* **2010,** *277,* 3452-4358.

6. Burge, S.; Parkinson, G. N.; Hazel P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006,** *34,* 5402-5414.

7. Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. Monovalent cation induced structure of telomeric DNA: the G-quartet model. *Cell* **1989,** *59,* 871-880.

8. Wong, H. M.; Payet, L.; Huppert, J. L. Function and targeting of G-quadruplexes *Curr. Opin. Mol. Ther.* **2009,** *11,* 146-155.

9. Hazel, P.; Parkinson, G. N.; Neidle, S. Predictive modeling of topology and loop variations in dimeric DNA quadruplex structures. *Nucleic Acid Res.* **2006,** *34,* 2117-2127

10. Hardin, C. C.; Perry A. G.; White, K. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **2000,** *56,* 147-194.

11. Guedin, A.; Gros, J.; Alberti, P.; Mergny, J. L. How long is too long? Effects of loops size on quadruplex stability. *Nucleic Acids Res.* **2010,** *38,* 7858-7868.

12. Bagaut, A.; Balasubramanian, S. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **2008,** *47,* 689-697.

13. van Mourik, T.; Dingley, A. J. Characterization of the monovalent ion position and hydrogen-bond network in guanine tetrads by DFT calculations of NMR parameters. *Chemistry* **2005,** *7,* 6064-6079.

14. Dingley A. J.; Peterson R. D.; Grzesiek, S.; Feigon, J.; Characterization of the cation and temperature dependence of DNA quadruplex hydrogen bond properties using high-resolution NMR. *J. Amer. Chem. Soc.* **2005,** *127,* 1446-14472.

15. Capra, J. A.; Paeschke, K.; Singh, M.; Zakian, V. A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae. Plos Comput. Biol.* **2010,** *6,* e1000861

16. Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005,** *36,* 144-156.

17. Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex forming sequences in *Saccharomyces cerevisiae. Nucleic Acids Res.* **2008,** *33,* 144-156.

18. Huppert, J. L.; Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33,* 2908-2916.

19. Rawal, P. *et al.* Genome-wide prediction of G4 DNA as regulatory motifs; role in *Escherichia coli* global regulation. *Genome Res.* **2006,** *16,* 644-655.

20. Eddy, J.; Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **2006,** *34,* 3887-3896.

21. Zakian, V. A. Telomeres: the beginning and ends of eukaryotic chromosomes. *Exp. Cell Res.* **2012,** *318,* 1456-1460.

22. Henderson, E. *et al.* Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **1987,** *51,* 899-908.

23. Sundquist, W. I.; Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989,** *342,* 825-829.

24. Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by gunine-rich motifs in DNA and its implications for meiosis. *Nature* **1988,** *334,* 364-366.

25. Shay, J. W.; Wright, W. E. Role of telomeres and telomerase in cancer. *Seminars Cancer Biol.* **2011,** *21,* 349-353.

26. Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **1991,** *350,* 718-720.

27. Oganesian, L; Graham, M. E.; Robinson, P. J.; Bryan, T. M. Telomerase recognizes G-quadruplex and linear DNA as distinct substrates. *Biochemistry* **2007,** *46,* 11279-11290.

28. Neidle, S. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **2010,** *277,* 1118-1125.

29. Huppert, J. L.; Balasubramanian, S. Prevalance of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33,* 2908-2916.

30. Shirude, P. S.; Okumus, B.; Ying, L. M.; Ha, T.; Balasubramanian, S. Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene *c-kit. J. Am. Chem. Soc.* **2007,** *129,* 7484-7485.

31. Sun, D.; Hurley, L. H.; The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.* **2009,** *52,* 2863-2874.

32. Brooks, T. A.; Kendrick, S.; Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **2010,** *277,* 3459-3469.

33. Qin, Y.; Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochemie* **2008,** *90,* 1149-1171.

34. van Dijk, L.; Bobbert, P. A.; Spano, F. C. *J. Phys. Chem. B,* **2010**, *114*, 817.

35. Balagurumoorthy, P; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic Acids Res.,* **1992**, *20*, 4061.

36. Jin, R.; Gaffney, B. L.; Wang, R. A.; Jones, A.; Breslauer, K. J.; *Proc. Natl. Acad. Sci.,* 1992, 89, 8832.

37. Lu, M.; Guo, Q.; Kallenbach, N. R. *Biochemistry*, **1993**, *32*, 598.

38. Smith, F.W.; Feigon, J. *Nature*, **1992**, *356,* 164.

39. Lane, A. N.; Chaires, B. J.; Gray, R. D.; Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36,* 5482-5515.

40. da Silva, M. W. NMR methods for studying quadruplex nucleic acids. *Methods,* **2007**, *43,* 264-277.

41. Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomerase sequences in K+ solution. *Nucleic Acids Res.,* **2007**, *35,* 6517-6525.

42. Phan, A.T.; Kuryavyi, V.; Patel, D. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 288.

43. Monchaud, D.; Teulade-Fichou, M. P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* **2008**, *6*, 627-636.

44. Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. *J. Med. Chem.* **1997,** *40*, 2113.

45. Perry, P. J.; Read, M. A.; Davies, R. T.; Gowan, S. M.; Reszka, A. P.; Wood, A. A.; Kelland, L. R.; Neidle, S. *J. Med. Chem.* **1999***, 42,* 2679.

46. Harrison, R. J.; Reszka, A. P.; Haider, S. M.; Romagnoli, B.; Morrell, J.; Read, M. A.; Gowan, S. M.; Incles, C. M.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2004**, *14,* 5845.

47. Haider, S. M.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2003**, *326,* 117.

48. Schultes, C. M.; Guyn, B.; Cuesta, J.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2004***, 14,* 4347.

49. Lehr, C. –M. *et al. Pharm. Res.* **2006,** *23,* 1031.

50. Casals, J.; Debethune, L.; Alvarez, K.; Risitano, A.; Fox, K. R.; Grandas, A.; Pedroso, E. Directing quadruplex-stabilizing drugs to the telomere: synthesis and properties of acridine-oligonucleotide conjugates. *Bioconjugate Chem.* **2006,** *17,* 1351-1359.

51. Xue, L.; Ranjan, N.; Arya, D. P. Synthesis and Spectroscopic Studies of the Aminoglycoside (Neomycin)-Perylene Conjugate Binding to Human Telomeric DNA *Biochemistry* **2011,** *50,* 2838-2849.

52. Redman, J. E.; Granadino-Roldan, J. M.; Schouten, J. A.; Ladame, S.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. Recognition and discrimination of DNA quadruplexes by acridine-peptide conjugates. *Org. Biomol. Chem.* **2009***, 7,* 76-84.

53. Murat, P.; Singh, Y.; Defrancq, E. Methods for investigating G-quadruplex DNA/ligand interactions *Chem. Soc. Rev.* **2011,** *40,* 5293-52307.

54. Masiero, S.; Trotta, R.; Pieraccini, S.; De Tito, s.; Perone, R.; Randazzo, A.; Spada, G. P.; *Org. Biol. Chem.* **2010**, *8,* 2683-2692.

CHAPTER 2: DESIGN OF DYNAMIC COMBINATORIAL LIBRARIES AND EXPERIMENTS

A. Background

i. Cyclic Peptides

Peptides are well suited to act as modulators of biological function given that all enzyme sites and protein-protein interactions are comprised of amino acids.[1] Additionally, cyclic peptides can be considered naturally occurring privileged structures due to their ability to mimic biologically relevant regions of protein diversity, such as β-turns, which are important recognition elements of peptides and proteins.[2,3] The constrained nature of cyclic peptides has been shown to enhance binding affinity by restricting conformational freedom, causing cyclic peptides to bind with a higher affinity to targets than their linear counterparts.[5] This increase in binding affinity is due to the entropic advantage gained by the constrained scaffold of cyclic peptides. Furthermore, cyclic peptides display a higher metabolic stability because without exposed termini, they are inaccessible to cellular proteases.[5] The advantages offered by cyclic peptides are seen in nature considering many natural product antibiotics, such as tyrocidine A and gamidin S, are based on cyclic scaffolds.[6,7] In addition, the commonly marketed drugs octreotide and the immunosuppressant cyclosporin A also contain cyclic peptide structures.[8] Therefore, given their high affinity, target specificity and metabolic stability, cyclic peptides continue to hold promise as powerful biological tools.

Despite the promise of cyclic peptides as useful molecular tools, current methods for

generating libraries of structurally diverse cyclic peptides in a high-throughput format are limiting. Methods for synthesizing libraries of cyclic peptides are often difficult and labor intensive due to the need for orthogonal deprotection strategies as well as time-consuming purifications at different stages of the syntheses. Syntheses of cyclic peptides often suffer from low yields due to side reactions and mixtures of cyclic and linear products.[5] Due to these complications, our lab has developed a method using dynamic combinatorial chemistry to generate libraries of cyclic peptides *in situ* in a high-throughput format.

ii. Dynamic Combinatorial Chemistry

Dynamic combinatorial chemistry is a method for the in situ generation of a complex mixture of macrocycles from smaller building blocks and is combinatorial chemistry that is under thermodynamic control with all library constituents under equilibrium.[9] This requires that all library members be able to interconvert through a reversible chemical process, which can involve either covalent or noncovalent interactions. The composition of the library is determined via the thermodynamic stability of each of the library members. Thermodynamic control of the library implies that introduction of changes in conditions of the experiment will induce changes in the library composition. When a template is introduced and binds to a library member, this species is stabilized. The bound species is thus amplified, as the remaining building blocks will establish a new equilibrium increasing the concentration of the selected library member at the expense of the remaining library members.[10-12] Therefore, in theory the amplification should be selective for the species that binds most tightly with the template.

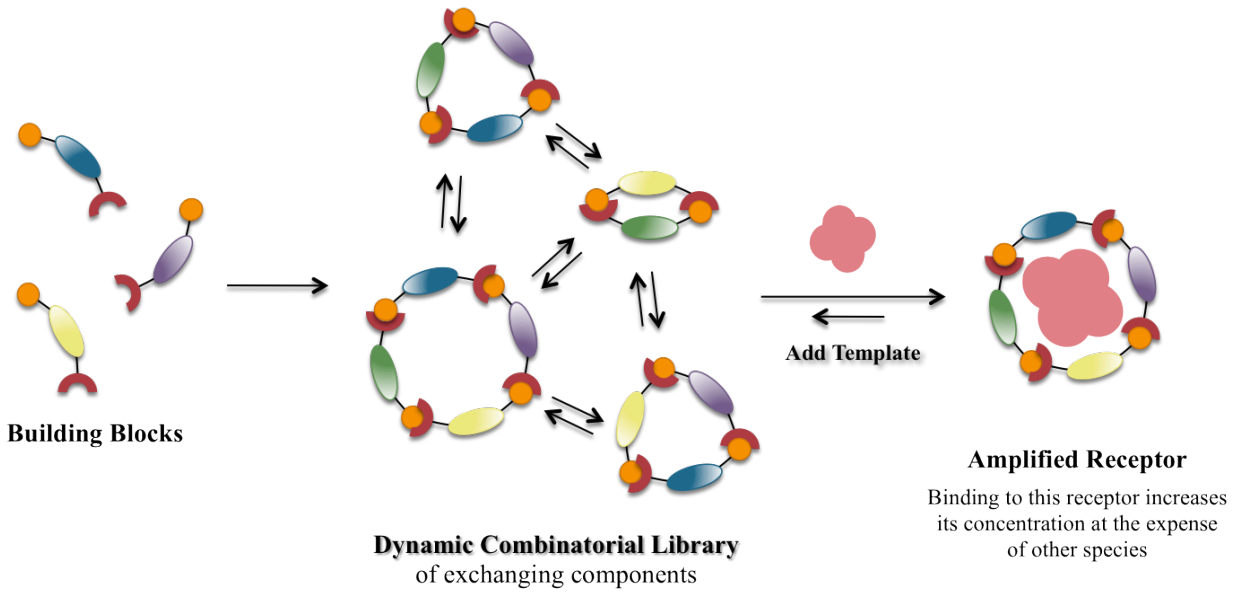

Figure 2.1 Scheme representing dynamic combinatorial chemistry as a way to select specific members of a dynamic combinatorial library through noncovalent interactions by the introduction of a separate guest.

An important feature of dynamic combinatorial chemistry is the reversible exchange reaction that is necessary in order for the building blocks to be capable of interconverting between library members. A number of requirements are necessary for this reaction: (i) the reaction needs to be reversible on a reasonable timescale (ii) the reaction needs to be compatible with the experimental conditions including the functional groups on the building blocks and template, the solvent, and the pH (iii) the reaction conditions need to be mild so as not to interfere with the delicate noncovalent interactions involved in molecular recognition (iv) it needs to guarantee the solubility of all library members so as not to interfere or alter the equilibrium (v) it should be possible to turn off the reaction and thus kinetically "freeze" the selected library members enabling their isolation and characterization and (vi) ideally all library members should be isoenergetic to prevent production of biased libraries.[5]

Three main different types of reversible reactions have been used including: noncovalent (hydrogen bonds), coordinative, and covalent bonds. Our lab has focused only on covalent

reversible reactions therefore only reactions most relevant to the Waters lab will be discussed. The first example is the acid-catalyzed condensation of an amine with a ketone or aldehyde to form an imine. While imine formation is reversible, imines condense and hydrolyze quickly at neutral to acidic pH making it an impractical exchange reaction in for biologically relevant templates.[13] On the other hand, hydrazones can be thermodynamically stable in the presence of water even at lower pHs.[14] The hydrazine exchange reaction is reversible and compatible with a wide range of solvents and functional groups. Most applications of hydrazones in dynamic combinatorial chemistry involves an acyl hydrazine where the acyl group moderates the stabilizing influence of the amine subunit in C=N-NRR'.[5] The drawback to using hydrazones as the exchange reaction is that exchange occurs at a pH of 4, which is not compatible with most biomolecules. Another example of a covalent exchange reaction that is commonly employed in dynamic combinatorial chemistry is disulfide exchange, which plays an important role in biology due to its role in the folding of proteins and the redox state of cells.[15] The mechanism includes the nucleophilic displacement of a thiolate anion from the disulfide through attack by another thiolate anion.[16] This process requires deprotonation of the thiol therefore the exchange is highly pH dependent and allows the process to be halted by lowering the pH. Disulfide exchange can be carried out at near neutral conditions (pH 7-9) making it compatible with most biomolecules. However, when a hit has been identified, the disulfide bond is not stable within the cellular environment and there is no straightforward synthetic replacement. Additionally, this reaction is exceedingly slow, taking days to weeks to reach equilibrium.

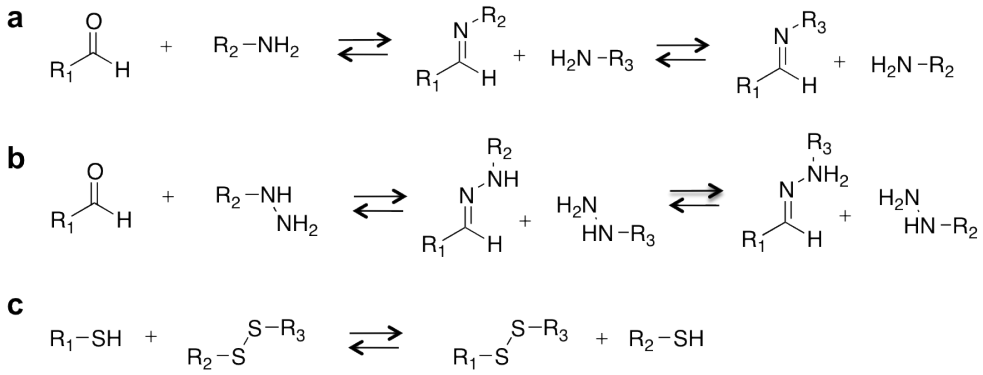

Figure 2.2 Reversible covalent reactions used for dynamic combinatorial chemistry. (a) Acid-catalyzed exchange reaction between an amine and an aldehyde to form an imine. (b) Reaction of a hydrazine and aldehyde to form a hydrazone. (c) Disulfide exchange, which occurs between a thiolate and a disulfide bond to release a free thiolate anion.

iii. Dynamic Cyclic Peptide Libraries from Thiol-Thioester Exchange

Our lab has recently published a method utilizing thiol-thioester exchange as a reversible reaction for DCC to form cyclic peptides.[17] The advantage of using thiol-thioester exchange over other exchange reactions is that it is rapid in aqueous solution at neutral pH and provides a native-like linkage that is successively replaceable by a more robust amide or ester functionality.[18,19] Equilibrium is reached within 1 to 2 hours depending on the sequence, with no significant hydrolysis of monomers during that time. Previous reports of the use of thioester exchange in DCC experiments has involved thioesters that were unsubstituted at the α-position, limiting the diversity of structures possible in a library of cyclic peptides.[19,20] Our lab performed systematic studies to determine the reactivity of peptide thiol-thioester exchange to its scope and limitation in its application in DCC.

Each monomer was a four-residue peptide containing a thioester at the C-terminus and a thiol at AA_2 (Cys). The structures of each monomer were varied at AA_1, AA_3 and AA_4. The monomers were designed so that AA_2-AA_3-AA_4 form the macrocycle while AA_1 remained exocyclic. Charged amino acids were incorporated to improve water solubility and prevent

aggregation. Proline was initially included in AA_3 as a turn residue to favor macrocycles. Various amino acids were incorporated into AA_4 to study their effect on the rate of macrocyclization, including positively charged amino acids (Lys, Arg), a negatively charged amino acid (Glu), and a hydrogen-bonding and neutral amino acid (Gln), and hydrophobic and sterically bulky amino acids (Phe, Val).

Upon dissolution of the monomers in potassium phosphate butter (pH 7), the monomers underwent facile thiol-thioester exchange to form a mixture of macrocycles, with the dimeric 20-atom macrocyclic hexapeptides as the major product. The reaction mixture was monitored over time by HPLC, allowing a simple two-step mechanism to be proposed. First, two monomers react to form an oligodimer followed by a ring closure intramolecular trans thioesterification reaction. The oligodimer intermediate was not observed to accumulate, indicating that the intermolecular step is rate limiting.

Thiol-thioester exchange can tolerate a variety of amino acids and structure-function studies indicated that the rate of macrocycle formation is dependent on the amino acid sequence. The observed reactivity of monomers with variation at the C-terminus corresponded to differences in sterics, with the incorporation of β-branched amino acids such as Val resulting in measurable hydrolysis before cyclization was complete. When Glu was placed at the C-terminus, it reacted more slowly than its Gln counterpart and can isomerize through anhydride intermediates. Additionally, the effect of positively charged amino acids (Lys and Arg) at positions AA_1 and AA_4 were investigated were found to react significantly faster than their negatively charged analogues suggesting that positively charged amino acids could stabilize the buildup of negative charge in the transition state. Finally, Trp was placed at the turn residue to determine if Pro is necessary for rapid cyclization and it was found that the Trp substituted

monomer actually reacted faster than the hydroxyproline (Hyp) analogue. Stereochemistry also played a role in the rate of macrocyclization as D-Pro reacted considerably faster than Hyp suggesting that chirality has an influence on the accessibility of the thioester or thiol.

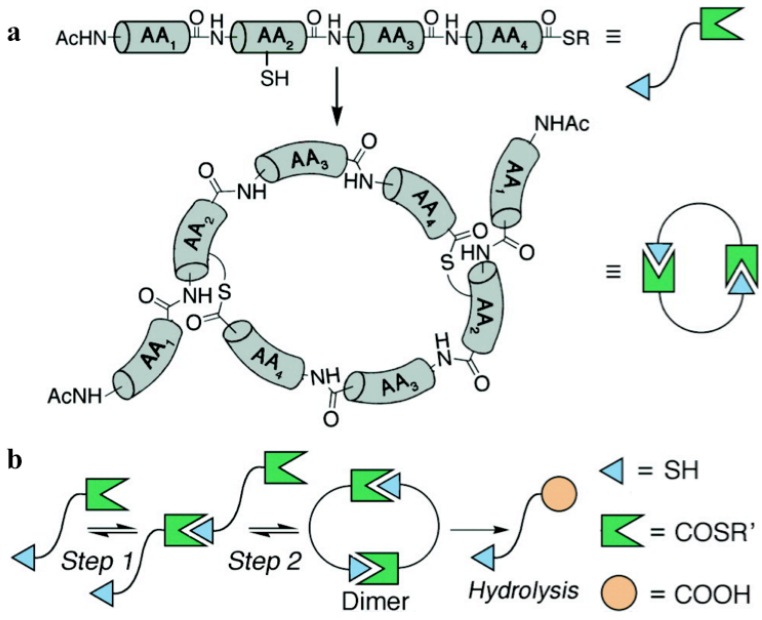

Figure 2.3 Scheme depicting macrocyclization of peptides via thiol-thioester exchange. (a) General design of monomers and major products (b) Sequence of steps toward dimeric cyclic thiodepsipeptide.

In summary, our lab has developed an efficient method to rapidly generate libraries of macrocyclic thiodepsipeptides at netrual pH for high-throughput screening. Upon mixing two or more monomers, we can form a complex library of cyclic thiodepsipeptides generated in situ under thermodynamic control. These libraries can then be screened against a particular target efficiently within hours. Given the transient stability of thioesters in vivo, a "hit" in the high-throughput screening can be resynthesized as a more stable analogue by replacing the thioester group with an amide or ester. Our lab then looked to apply this method of high-throughput screening to find cyclic peptides binders for G-quadruplex DNA.

B. System Design

i. Dynamic Combinatorial Library

Our strategy for the discovery of high-affinity ligands with selectivity between quadruplexes uses a common acridine core to target the planar surface of the G-tetrad, appended with variable peptide substituents to contact the loops and grooves that distinguish each quadruplex structure. We chose the acridine ligand due to the high selectivity it provides for quadruplex DNA over duplex DNA. Previously, Neidle and coworkers screened a number of linear peptide-acridine ligands in the effort to gain selectivity amongst individual quadruplex structures. While linear peptides provided modest binding selectivity in Neidle's work, we felt that cyclic peptides could deliver greater selectivity and potentially fit nicely into the grooves of the quadruplex structure. Using methods previously developed in our lab, we wanted to create libraries of cyclic peptides using thiol-thioester exchange in dynamic combinatorial chemistry. The dynamic combinatorial libraries consisted of tetrapeptides appended to an acridine core to target the quadruplex structure with the tetrapeptides designed to undergo thiol-thioester exchange by incorporation of a Cys at AA_2 and a thioester at the C-terminus. Since the loops and peripheral grooves of the quadruplex structure provide a pattern of hydrogen bonding, hydrophobic π-surfaces, and negative charges unique to a particular quadruplex sequence, we chose to focus on the incorporation of amino acids with hydrogen bonding groups, positive charges or aromaticity to facilitate favorable interaction.

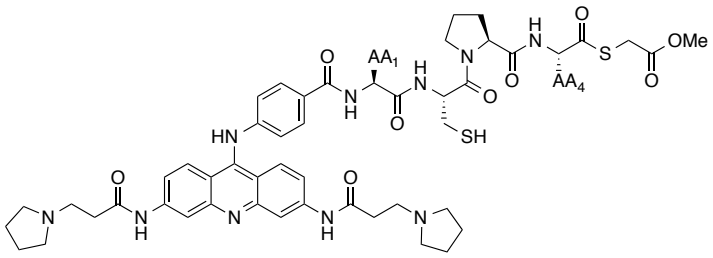

Figure 2.4 General design of monomer library with acridine ligand incorporated at the N-terminus to target the G-tetrad of the quadruplex structure. Cysteine is incorporated at AA_2 along with a thioester installed at the C-terminus to facilitate thiol-thioester exchange, while a Pro was included at the AA_3 as a turn residue to aid in cyclization.

ii. Biological Targets

The goal of this project is to find molecules that can bind to, and differentiate between various G-quadruplex structures. With this in mind, it was necessary to choose different DNA sequences to screen our libraries against. The first sequence we chose was the human telomeric (*Htelo*) sequence given the extensive studies that have been carried out surrounding the role of quadruplex formation in the single-stranded telomeric region and its role in the inhibition of the reverse-transcriptase enzyme telomerase. Secondly, we chose three oncogenes whose promoter regions are known to contain G-quadruplex folding patterns and structures, *cKit21, cMyc22* and *NRas*. Finally, we chose a dsDNA sequence containing guanines as a control to determine the selectivity of our library for quadruplex DNA versus duplex.

The first oncogene we chose to screen is the *c-myc* oncogene, which is one of the most commonly malfunctioning genes in human cancers.[22] The *myc* family of oncogenes encodes phospho-proteins that activate genes and encourage forward cell growth. Normally, the human *c-myc* gene is tightly regulated and it is the overexpression of this gene that leads to the progression of many cancers. The *c-myc* gene utilizes four promoters with the nuclease-hypersentive element III_1 accounting for 75-85% of total *c-myc* transcription.[23] It is this region of

the *c-myc* promoter region that can form G-quadruplexes under physiological conditions thereby playing an important role in the transcriptional regulation of this oncogene.

The promoter region of *c-kit* is the second sequence we chose to screen against. *c-kit* is an oncogene which codes for a tyrosine kinase receptor and is pivotal for relaying extracellular signals. When activated, KIT stimulates proliferation, differentiation, migration and survival.[24] Overexpression of the *c-kit* gene results in uncontrolled cell proliferation and is considered the primary pathogenic event in gastrointestinal stromal tumors. A quadruplex forming 21-nucleotide sequence upstream of the transcription initiation site has been identified on the G-rich strand, which occupies a site required for core promoter activity.[25] Until recently, all cancer therapies targeting kinases inhibit the kinase protein after expression rather than controlling its expression. The position of the quadruplex sequence in the promoter region of *c-kit* makes it an attractive target for the regulation of *c-kit* at the transcriptional level.

A final oncogene that we chose to focus on is the *NRAS* gene that is another of the most frequently mutated oncogenes detected in human cancer. The *NRAS* gene encodes for a guanine nucleotide (GTP/GDP)-binding protein that acts as a signal transducer. Mutated forms of the gene can result in a protein locked activated state, transmitting constitutively signals for cell proliferation to the nucleus. The promoter of the *NRAS* gene contains a nuclease hypersensitive polypurine-polypyrimidine element (NHPPE) that is essential for transcription and has the ability to form the G-quadruplex structure.[26] Studies have again shown that formation of G4 DNA in the control region of the gene may contribute to the regulation of expression of the *NRAS* gene. Due to its role in the pathogenesis of cancer, the quadruplex structure of the *NRAS* gene is an ideal target for anticancer drugs.

C. Conclusions

The goal of this project is to identify small molecules that can bind to and stabilize the G-quadruplex structure of DNA. Previous studies have shown modest success in differentiating the various possible quadruplex structures and topologies through the use of acridine-linear peptides conjugates. We felt strongly that the use of cyclic peptides could deliver greatly improved selectivity while also providing the added advantages of mimicking native protein structure, displaying enhanced metabolic stability and possessing structural preorganization thus reducing the entropic cost of binding. Despite the therapeutic potential of cyclic peptides, the options for synthesizing structurally diverse libraries in a high-throughput format are limiting. Using a strategy that has been developed in our lab, we propose screening libraries of solution-phase cyclic peptides generated using thiol-thioester exchange for DCC. Our system has been designed as tetrapeptides containing a thioester installed at the C-terminus with the acridine ligand attached at the N-terminus. We selected the four quadruplex forming sequences of *Htelo, c-kit, c-myc* and *NRAS* as our biological targets due to their biological importance and implications in a number of different cancers. The next step was to synthesis the monomers of our library followed by optimization of the dynamic combinatorial experiments.

References

1. Troitskaya, L. A.; Kodadek, T. Peptides as modulators of enzymes and regulatory proteins. *Methods,* **2004,** *32,* 406-415.

2. Andrianov, A. M. *Mol. Biol.* **1999**, *33,* 534-538.

3. Freidinger, R. M.; Veber, D. F.; Perlow, D. S.; Brooks, J. R. *Science,* **1980**, *210,* 656-658.

4. Katz, B; Liu, B.; Cass, R.; Structure based design tools: structural and thermodynamic comparison with biotin of a small molecule that binds to streptavidin with micromolar affinity. *J. Am. Chem. Soc.* **1996,** *118,* 7914-7920.

5. Horton, D. A.; Bourne, T. T. Smythe, M. L. Exploring privileged structures: The combinatorial synthesis of cyclic peptides. *J. Comput. Aided Mol. Design,* **2002**, *16,* 415-430.

6. Mootz, H. D.; Marahiel, M. A. *J. Bacteriol* **1997**, *197,* 6843-6850.

7. Kratzschmar, J.; Krause, M.; Marahiel, M. A.; *J. Bacteriol.* **1989**, *171,* 5422-529.

8. Emmel, E. A.; Verweij, C. L.; Durand, D. B.; Higgins, K. M.; Lacy, E.; Crabtree, G. R. *Science,* **1989***, 246,* 1617-1620.

9. Corbett, P.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. –L.; Sanders, J. K. M.; Otto, S. Dynamic combinatorial chemistry. *Chem. Rev.* **2006,** *106,* 3652-3711.

10. Hoss, R.; Vogtle, F. *Angew. Chem. Int. Ed. Engl.* **1994**, *33,* 375.

11. Anderson, S.; Anderson, H. L.; Sanders, J. K. M. *Acc. Chem. Res.* **1993,** *26,* 469.

12. Busch, D. H. *Incl. Phenom. Mol. Recognit. Chem.* **1992**, *12,* 389.

13. Giuseppone, N.; Lehn, J. –M. *J. Am. Chem. Soc.* **2004,** *126,* 11448.

14. Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry Part A,* Plenum Press: New York, 1990, p. 451.

15. Gilbert, H. F. *J. Biol. Chem.* **1997,** *272,* 29399.

16. Fernandes, P. A.; Ramos, M. J. *J. Chem.-Eur. J.* **2004**, *10,* 257.

17. Ghosh, S.; Ingerman, L. A.; Frye, A. G.; Lee, S. J.; Gagne, M. R.; Waters, M. L. Dynamic cyclic thiodepsipeptide libraries from thiol-thioester exchange. *Org. Lett.* **2010***, 12,* 1860-1863.

18. Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *Science,* **2009**, *325,* 73-77.

19. Woll, M. G.; Gellman, S. H. *J. Am. Chem. Soc.* **2004**, *126,* 11172-11174.

20. Larsson, R.; Pei, Z. C.; Ramstrom, O. *Angew. Chem. Int. Ed.* **2004**, *43,* 3716.

21. Redman, J. E.; Granadino-Roldan, J. M.; Schouten, J. A.; Ladame, S.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. Recognition and discrimination of DNA quadruplexes by acridine-peptide conjugates. *Org. Biomol. Chem.* **2009***, 7,* 76-84.

22. Brooks, T. A.; Hurley, L. H. The role of supercoiling in transcriptional control of MYC and its importanct in molecular therapeutics. *Nat. Rev. Cancer* **2009**, *9,* 849-861.

23. Gonzalez, V.; Hurley, L. H. The c-MYC NHE III_1: function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50,* 111-129.

24. Shirude, P. S.; Okumus, B.; Ying, L.; Ha, T.; Balasubramanian, S. Single-molecular conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit. *J. Am. Chem. Soc.* **2007,** *129,* 7484-7485.

25. Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M. Parkinson, G. N.; Todd, A. K.; Ladame, S. Balasubramanian, S.; Neidle, S. Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **2005**, *127,* 10584-10589.

26. Xodo, L. E.; Cogoi, S. G-Quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nuc. Acids. Res.* **2006**, *34,* 2536-2549.

CHAPTER 3: SYNTHESIS OF MONOMERS FOR DYNAMIC COMBINATORIAL LIBRARIES

A. Monomer Synthesis

i. Synthesis of acridine ligand

The monomer system design consisted of a tetrapeptide containing a cysteine residue and thioester installed at the C-terminus to facilitate thiol-thioester exchange. Additionally, an acridine ligand was installed at the N-terminus to target the quadruplex structure (Figure 3.1). The tetrapeptide is synthesized using solid phase peptide synthesis followed by coupling of the free N-terminus to the acridine ligand. Given the excellent selectivity of the acridine motif for the quadruplex structure over the duplex DNA, the synthesis of BRACO-19 (**1**) is well precedented.[1,2] However, the structure synthesized was slightly modified as shown in structure **2** to include a carboxylic acid group off of the phenyl side chain allowing for coupling to the N-terminus of the peptide substituents.

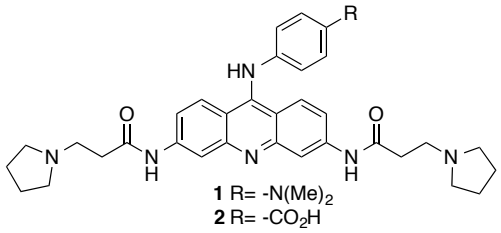

Figure 3.1 Structure of the G-quadruplex ligand BRACO-19 (**1**) and the modifications made to the acridine ligand (**2**) to allow incorporation via solid phase synthesis at the N-terminus of the tetrapeptide monomers.

The synthesis of the acridine ligand began with the nitration of diphenyl methane followed by oxidation to the ketone (**2**) using chromium oxide. The nitro groups were reduced in

the presence of tin (II) dichloride in refluxing hydrochloric acid to provide the cyclized 3,6-diaminoacridone product (**3**). This step was problematic, as it was difficult to remove the tin from the final product. This problem was circumvented by first dissolving the product in water after filtration from hydrochloric acid, followed by basification using 2N NaOH, converting the tin byproduct to a water soluble compound. After addition of the base, the mixture was heated to reflux for 15 minutes to ensure removal of the tin byproduct. Compound **4** was then acylated by treatment with 3-chloropropionyl chloride followed by subsequent amination with pyrrolidine providing compound **8**.

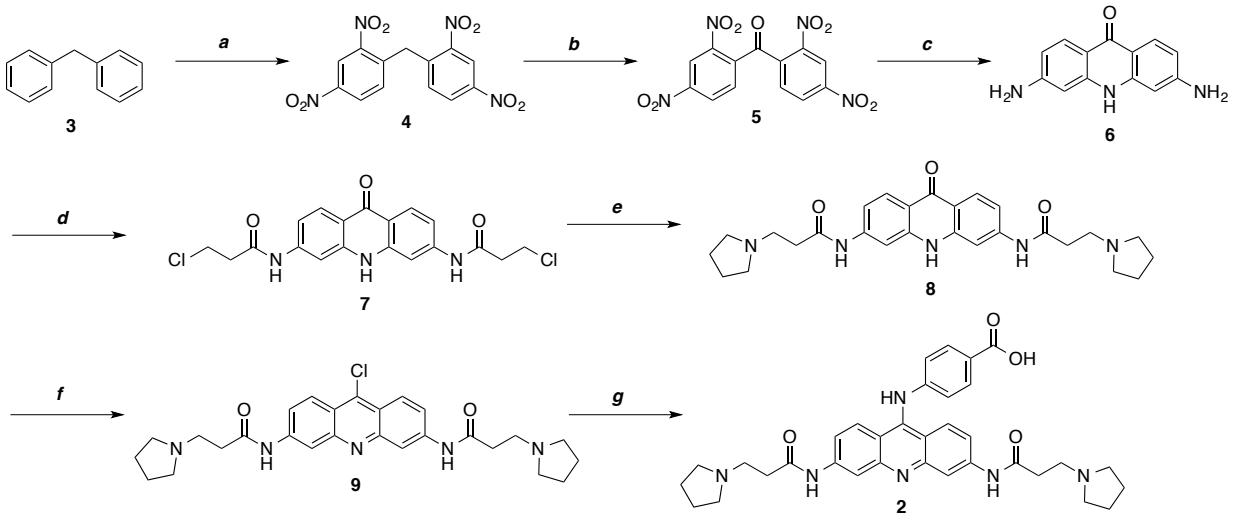

Figure 3.2 Synthesis of the acridine ligand (**2**). Conditions: (a) HNO_3 (10 equiv) at 0°C in H_2SO_4, warm to rt over 4 hours (88%) (b) CrO_3 (4 equiv) in AcOH, reflux, 10 hours (quant.) (c) $SnCl_2$ in conc. HCl (0.3M) and EtOH (0.67M), reflux, 8 hours (49%) (d) 3-chloropropionyl chloride (20 equiv), 80°C, 10 hours (77%) (e) Pyrrolidine (20 equiv), reflux, 10 hours (87%) (f) $POCl_3$, 100°C, 3 hours (g) 4-aminobenzoic acid, EtOH, reflux, 12 hours (40% over two steps).

The next steps of the synthesis introduced some complications. Following literature procedure, the acridone compound **8** was heated to 100°C in phosphorous oxychloride. After one hour, the reaction was cooled and diethyl ether was added. The precipitated product was filtered and then added to a mixture of ice-aqueous ammonia-chloroform in an ice bath. Further aqueous ammonia was added to ensure that the mixture remained alkaline and the chloroform layer was

separated. When following this literature, it was found that upon filtration the product would form an oil rather than the reported solid, most likely due to hydrolysis of the 9-chloroacridine product. In order to avoid hydrolysis of the product, upon completion of the reaction diethyl ether was added and the reaction mixture poured into falcon tubes. The precipitated product was then centrifuged and the supernatant decanted off. The solid product was rinsed thoroughly with ether and centrifuged two more times. The solid was dried *in vacuo* and taken forward as the salt to the next step. By using centrifugation rather than filtration, the exposure of the product to the atmosphere was minimized thus preventing hydrolysis of the product.

The final product was formed by dissolving 9-chloroacridine compound **9** in ethanol with the addition of 4-aminobenzoic acid. Given the deviations in the previous step from literature procedure, it was necessary to also optimize the final reaction of the synthesis. Since compound **9** was carried forward as the salt, it had poor solubility in chloroform. The compound showed better solubility in methanol, but the reaction was messy with a number of side products being formed. Ethanol proved to be the best solvent providing the final product cleanly in 40% yield over two steps. The final acridine compound **10** used without further purification in the synthesis of the monomers.

ii. Synthesis of monomers

The peptide portion of the monomers was synthesized via traditional solid phase peptide synthesis (SPPS). The synthesis began by pre-loading the first amino acid onto 2-chlorotrityl chloride resin, which provides a carboxylic acid at the C-terminus upon cleavage (Figure 3.3). The synthesis of the peptide then proceeded as usual with Fmoc deprotection in 20% piperidine and coupling of subsequent amino acids using the activators HBTU/HOBt. After synthesis of the peptide portion of the monomer, the N-terminus was either acetyl capped (Figure 3.4) or the

acridine ligand was coupled (Figure 3.5) in the presence of PyBOP, DIPEA and HOBt in DMF. Following capping of the peptide, the protected peptide fragment was orthogonally cleaved from the resin using 1% TFA in dichloromethane.

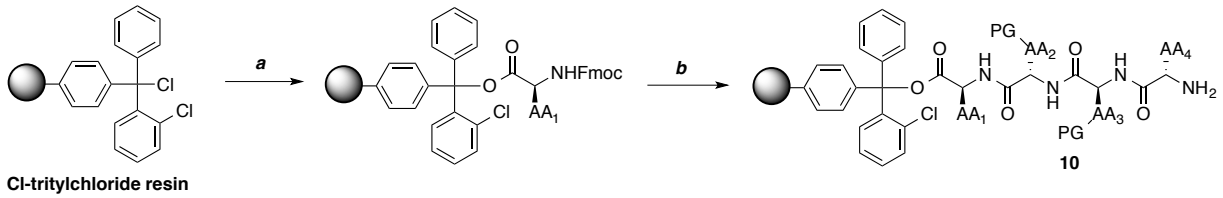

Figure 3.3 Fmoc solid phase synthesis of peptides on Cl-tritylchloride resin. Conditions (a) Fmoc-amino acid (2 equiv), DIPEA (8 equiv), CH_2Cl_2, 2 x 1hour (b) Standard SPPS with 3 amino acid residues; Coupling steps: Fmoc-amino acid (4 equiv), HBTU (4 equiv), HOBt (4 equiv), DIPEA (4 equiv), DMF, 1.5 hour; Deprotection steps: 20% piperidine, DMF, 15 minutes.

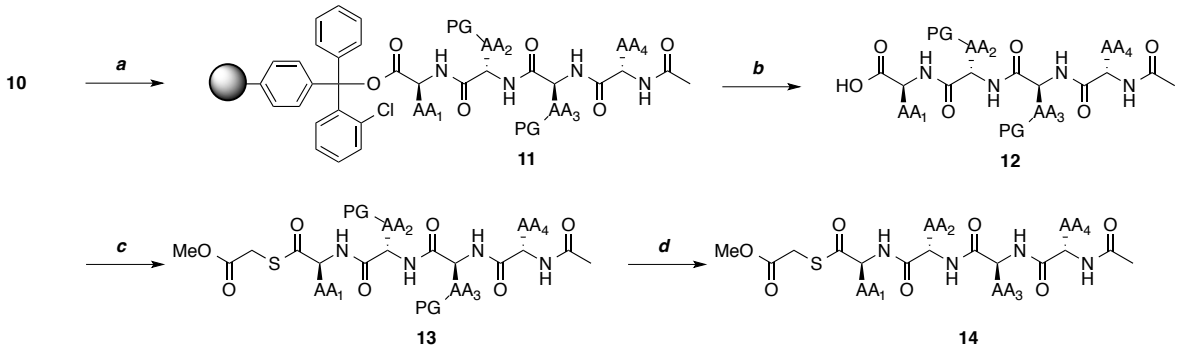

Figure 3.4 Synthesis of the acetyl capped peptide monomers. Conditions: (a) Ac_2O (5%), 2,6-lutidine (6%), CH_2Cl_2 (b) 1% TFA, CH_2Cl_2, 3 x 3 minutes (c) Methylthioglycolate (4 equiv), PyBOP (4 equiv), DIPEA (4 equiv), 1 hour (d) TFA: water: TIPS: EDT (94:2.5:1:2.5), 4 hours.

The next step in the synthesis is formation of the thioester. As previously published in our lab, the thioester was synthesized by activation of the C-terminus with PyBOP and DIPEA in DMF in the presence of methylthioglycolate (Figure 3.4). In the case of the acetyl-capped monomers, this reaction was carried out smoothly producing the desired thioester product. However, when the peptide was capped at the N-terminus with the acridine ligand, it was found

after workup that there remained a 50:50 mixture of thioester and carboxylic acid. It was hypothesized that incomplete reaction of the carboxylic acid was due to poor solubility. Therefore, alternative coupling reagents and solvents were used, finding that DIC in CH_2Cl_2 led to complete conversion of the carboxylic acid to the thioester (Figure 3.5). Moreover, it was found that the addition of the reagents was very important to prevent epimerization of the amino acid at the C-terminus. Initially, activator (DIC or PyBOP) was added to the peptide in the presence of DIPEA to activate the carboxylic acid before addition of the methylthioglycolate. However, this led to two peaks of identical mass when purified by HPLC, which was believed to be the result of epimerization. To correct for this, methylthioglycolate was added first to the reaction mixture, followed by addition of the base and finally, addition of the activator. Upon formation of the thioester, the peptide was deprotected in the presence of 94:2.5:1:2.5 trifluoroacetic acid (TFA): water: triisopropylsilane (TIPS): ethanedithiol (EDT) and purified via RP-HPLC.

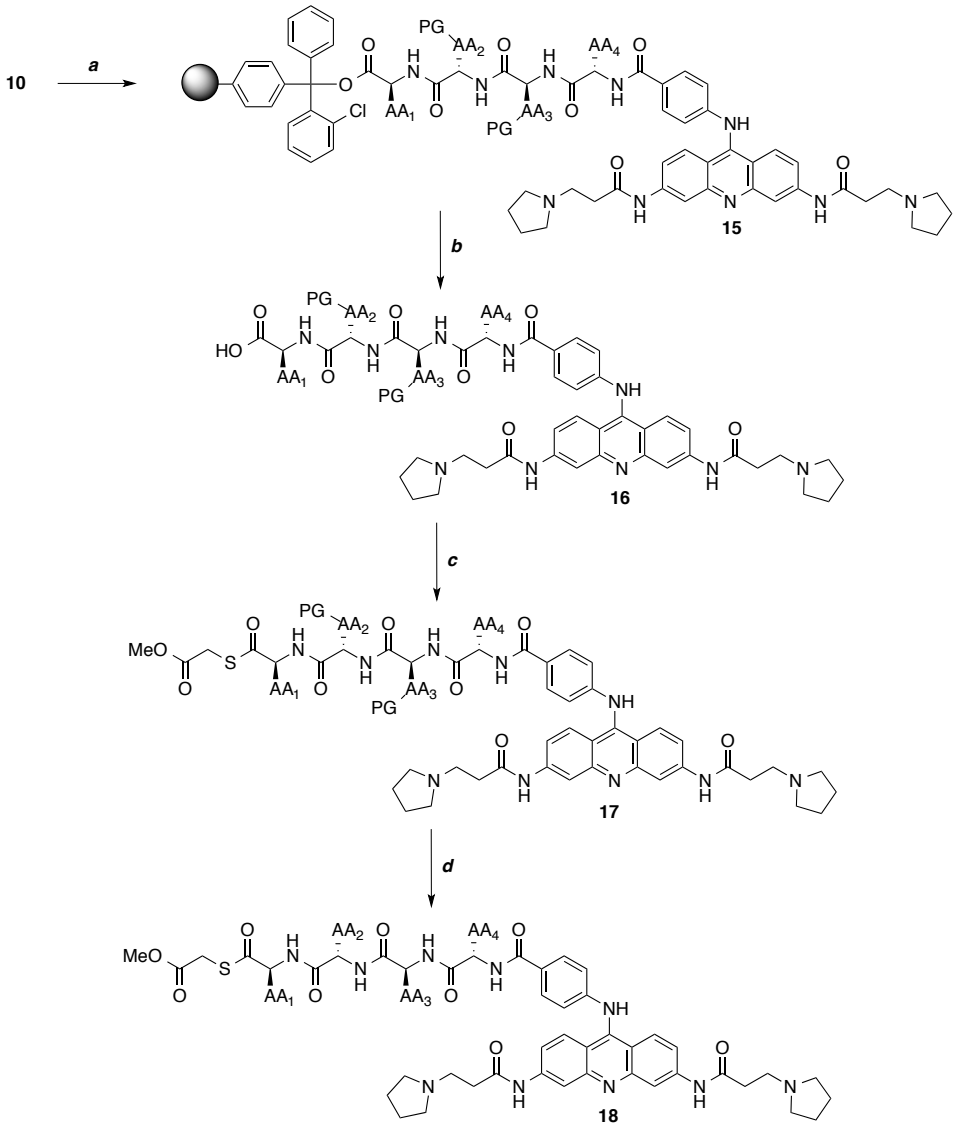

Figure 3.5 Synthesis of the acridine-peptide monomers. Conditions: (a) **2** (1.2 equiv), PyBOP (4 equiv), HOBt (4 equiv), DMF, 10 hours (b) 1% TFA, CH_2Cl_2, 3 x 3 minutes (c) Methylthioglycolate (30 equiv), DIC (20 equiv), DIPEA (20 equiv), 10 hours (d) TFA: water: TIPS: EDT (94:2.5:1:2.5), 4 hours.

B. Experimental Section

i. Materials and General Methods

Solvents were of HPLC or reagent grade quality and purchased commercially from Sigma-Aldrich. Starting materials were purchased commercially from Fisher Scientific unless otherwise noted and used without further purification.

ii. Synthesis of acridine ligand

Bis(2,3-dinitrophenyl)methane (4) A suspension of diphenylmethane (5 g, 29.7 mmol) was stirred in 150 mL of H_2SO_4. The mixture was cooled to -10° C in a salt/ice bath. 12.4 mL of HNO_3 was added slowly and the reaction mixture allowed to warm slowly to room temperature over a period of 4 hours. The reaction was quenched by pouring the reaction mixture over ice in a 500 mL Erlenmeyer flask. The resulting precipitate was filtered and washed thoroughly with water followed by multiple washes with MeOH. (9.1 g, 26.1 mmol, 88%).

Bis(2,3-dinitrophenyl)methanone (5) Compound 4 (9.1 g, 26.1 mmol) was dissolved in 250 mL of acetic acid. Chromium(VI) oxide (10.4 g, 104.6 mmol, 4 equiv) was added carefully and portion wise to the reaction mixture. The reaction was then heated to reflux overnight. The reacted was stopped by cooling to room temperature and pouring onto ice in a 500 mL Erlenmeyer flask. The resulting precipitate was collected by filtration and washed thoroughly with water. (9.5 g, 26.1 mmol, quant.)

3,6-diaminoacridin-9(10H)-one (6) Tin(II) chloride (26.2 g, 138.1 mmol, 10 equiv) was heated to reflux in concentrated hydrochloric acid (40 mL) under nitrogen for 15 minutes in a three neck roundbottom flask. After 15 minutes, the mixture was allowed to cool slightly and compound 5 (5 g, 13.8 mmol) was added in one of the side necks, keeping the reaction under an inert atmosphere. 20 mL of EtOH was added to the reaction and heated to reflux overnight. The reaction was then cooled to room temperature and filtered without washing. The resulting solid was dissolved in water and basified using 2 N NaOH. The mixture was heated to reflux for 10 minutes to ensure removal of the tin byproduct. After heating, the reaction was cooled to room temperature and the product isolated by filtration and washed thoroughly with water to provide

acridone compound 6 as a tan solid. (1.5 g, 6.7 mmol, 49%).

N,N'-(9-oxo-9,10-dihydroacridine-3,6-diyl)bis(3-chloropanamide) (7) Compound 6 (1.5 g, 6.7 mmol) was suspended in 12.5 mL of 3-chloropropionyl chloride (133.2 mmol, 20 equiv). The reaction was heated to 80°C and left to stir overnight under nitrogen. The reaction was cooled to room temperature and 12.5 mL of Et_2O was added and allow to stir for 15 minutes. The resulting precipitate was filtered and added to a solution of 1:1 Et_2O and aqueous $NaHCO_3$. The suspension was stirred for another 15 minutes and filtered. The product was washed thoroughly with Et_2O to provide a brown solid. (2.1 g, 5.2 mmol, 77%).

N,N'-(9-oxo-9,10-dihydroacridine-3,6-diyl)bis(3-(pyrrolidine-1-yl)propanamide) (8) Compound 7 (2.1 g, 5.2 mmol) was heated to reflux in pyrrolidine (10 mL, 104.0 mmol, 20 equiv) and left to stir overnight under nitrogen. The reaction was cooled to room temperature and aqueous $NaHCO_3$ (10 mL) was added and the mixture was allowed to stir for 1 hour. The resulting precipitate was filtered and washed thoroughly with Et_2O. (1.8 g, 3.8 mmol, 73%).

N,N'-(9-chloroacridine-3,6-diyl)bis(3-pyrrolidin-1-yl)propanamide (9) Compound 8 (1.8 g, 3.8 mmol) was suspended in 40 mL of $POCl_3$ and stirred at 100° C under nitrogen for three hours. Upon completion of the reaction, the mixture was cooled to room temperature and poured into two falcon tubes. Et_2O (15 mL) was added to each tube causing the product to precipitate out of solution. The falcon tubes were centrifuged briefly to form a pellet of the product at the bottom and the Et_2O layer decanted off. This was repeated two more times by adding Et_2O (15 mL) to rinse the product and centrifuged and decanted off. The product was dried under vacuum before immediate use in the next reaction.

4-((3,6-bis(3-(pyrrolidine-1-yl)propanamido)acridin-9yl)benzoic acid (2) Compound 9 was used immediately after the previous step and was suspended in 40 mL of EtOH. 4-

aminobenzoic acid (1.0 g, 7.6 mmol, 2 equiv) was added to the reaction mixture and refluxed overnight. The EtOH was evaporated off and 0.1% TFA (50 mL) was added to the reaction mixture and washed three times with EtOAc to remove excess 4-aminobenzoic acid. The aqueous layer was lyophilized down to provide the product as an orange powder. (1.2 g, 2.0 mmol, 53% over 2 steps).

iii. Synthesis of monomers

9-Fluorenylmethoxycarbonyl (Fmoc) protected amino acids were purchased from Advanced Chemtech. HBTU and HOBt were purchased from Advanced Chemtech. PyBOP was purchased from Novabiochem. 2-Chlorotritylchloride resin was purchased from ChemPep. The first amino acid was preloaded onto 2-chlorotrityl resin (0.1 mmol). Resin was activated with 1mL of thionyl chloride in 9 mL of CH_2Cl_2 for 1 hour in a peptide flask and bubbled with nitrogen. The first amino acid was loaded onto the resin with DIPEA (8 equiv) and Fmoc-[N]-protected amino acid (2 equiv) in CH_2Cl_2 for one hour, two times while being bubbled with nitrogen. The remaining portion of the peptides was synthesized by automated solid-phase synthesis on a Thuramed Tetras peptide synthesizer (version X94) using Fmoc protected amino acids. Activation of amino acids was performed with HBTU and HOBT in the presence of DIEPA in DMF. Peptide deprotection was carried out in 2% DBU (1,8 diazabicyclo[5.4.0]undec-7-ene) and 2% piperidine in DMF for approximately 15 min. Double coupling cycles (60 min each) were used for each amino acid coupling step. Peptides were capped with either Ac_2O or compound 2 using the following procedures. After formation of the thioester the peptide was globally deprotected using 95:2.5:1:2.5 trifluoroacetic acid (TFA): water: triisopropylsilane (TIPS): ethanedithiol (EDT) for 4 h. TFA was evaporated and cleavage products were precipitated with cold ether. The peptide was extracted into water and lyophilized. It was then

purified by reverse-phase HPLC, using a C-18 semipreparative column and a gradient of 0 to 100% B over 50 min, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. After purification, the peptide was lyophilized to powder and identified with ESI-TOF mass spectroscopy.

Acetyl capped monomers Following the above general procedure, monomers were manually acetylated at the N-terminus with 5% acetic anhydride and 6% lutidine in DMF for 30 min. Cleavage of the peptide from the resin was performed in 1% TFA in dry dichloromethane at three times for three minutes. The filtrate was concentrated by rotary evaporation to obtain the cleavage product as a yellow oil. The product was dried under high vacuum and then was dissolved in 3.6 mL of dry DMF. To that solution was added 40 µL N,N-Diisopropylethylamine (DIPEA), 22 µL methyl thioglycolate, 86 mg PyBOP and the reaction was stirred under N_2 at room temperature for 1 h. DMF was removed under high vacuum to obtain yellow oil. The peptide was deprotected using the above general procedure.

Monomers capped with acridine ligand Following synthesis of the peptide on the peptide synthesizer, the peptide was transferred to a peptide flask and reacted with compound 2 (1.5 equiv), PyBOP (4 equiv), HOBt (4 equiv) and DIPEA (4 equiv) in DMF while bubbling with nitrogen overnight. The peptide was cleaved from the resin using 1% TFA in CH_2Cl_2 for three times at three minutes. The thioester was formed by dissolving the peptide in CH_2Cl_2 (0.1 M) and treatment of the fully protected peptide with methylthioglycolate (30 equiv), DIC (20 equiv) and finally DIPEA (30 equiv). The reaction was stirred overnight under nitrogen. The solvent was removed by rotory evaporation. Peptide was deprotected using the above general procedure.

References

1. Redman, J. E.; Granadino-Roldan, J. M.; Schouten, J. A.; Ladame, S.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. Recognition and discrimination of DNA quadruplexes by acridine-peptide conjugates. *Org. Biomol. Chem.* **2009**, *7,* 76-84.

2. Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc.* **2008**, *130,* 6722-6724.

CHAPTER 4: DYNAMIC COMBINATORIAL EXPERIMENTS

A. Background

The molecular recognition of DNA is fundamental to many biochemical processes related to transcription, regulation and gene expression. Therefore, there has been considerable interest in identifying small molecules that bind selectively and with strong affinity to quadruplex DNA, amongst other nucleic acid structures. Dynamic combinatorial chemistry (DCC) has become an increasingly popular tool for understanding molecular recognition.[1] While an enormous number of studies have demonstrated the potential of DCC to identify unexpected and novel receptor molecules, there have been limited studies with nucleic acids. Additionally, studies have been performed to identify small molecules with affinity for duplex DNA, quadruplex DNA and RNA however; almost all examples have utilized the reversible thiol-disulfide oxidation reaction.[2] Herein we report the design of a system using thiol-thioester exchange as a means to effectively screen libraries of cyclic peptides against various quadruplex forming sequences in a high-throughput fashion using dynamic combinatorial chemistry. Thiol-thioester exchange is rapid in aqueous solution at neutral pH and provides a native-like linkage that is successively replaceable by a more robust amide or ester functionality. Furthermore, cyclic peptides are advantageous because they mimic native protein structure, exhibit enhanced metabolic stability and are structurally preorganized thus reducing the entropic cost of binding.

B. Dynamic Combinatorial Experiments

i. Optimization of Experimental Conditions

For DCC to be compatible with nucleic acids, the building blocks as well as the library members need to be soluble in aqueous solution in the pH range of 5-8, which can be difficult to achieve with aromatic chromophores. In the previously reported studies, the exchange reactions were carried out at pH 6.75 in NH_4OAc buffer. However, it was found that monomers containing the acridine ligand were not soluble under these conditions. Therefore, various buffer conditions were screened, and it was found that 25 mM of Tris buffer at a pH of 7.4 was optimal. Additionally, the ideal concentration of potassium chloride (KCl) necessary to facilitate folding of the quadruplex structure needed to be determined. A CD spectrum was taken of the *cKit21* sequence at various concentrations of KCl in 25 mM Tris buffer at pH 7.4. It was desirable to use the least amount of KCl in the dynamic combinatorial experiments while still facilitating full folding of the quadruplex structure. By comparison of the CD spectrums, it was determined that 50 mM of KCl was the optimal concentration to run the dynamic combinatorial experiments as 25 mM of KCl did not facilitate complete folding of the quadruplex structure, which can be seen by the decrease in the CD signal (Figure 4.1). Initially, monomers were mixed in the absence of DNA to confirm that they behaved as expected and cyclized, forming a library of various species under the conditions previously determined. Figure 4.2 shows that upon mixing three monomers, there was a mixture of monomers, homodimers, heterodimers and even trimers.

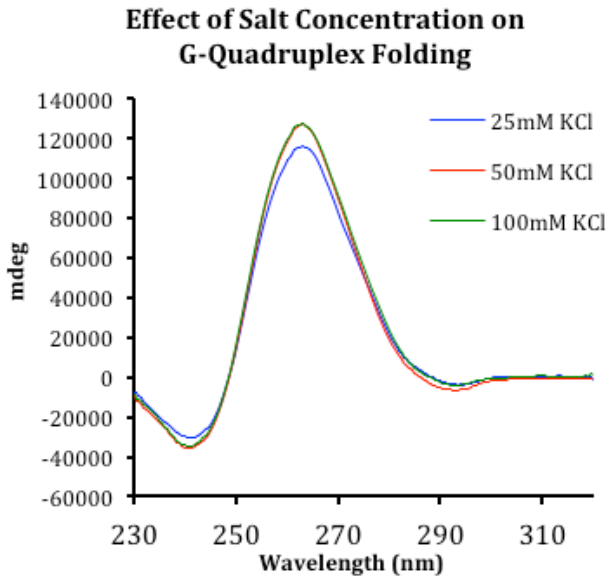

Figure 4.1 Circular dichroism spectra of *cKit-21* (20 μM) at 25°C in 25 mM Tris pH 7.4 with varying concentrations of KCl.

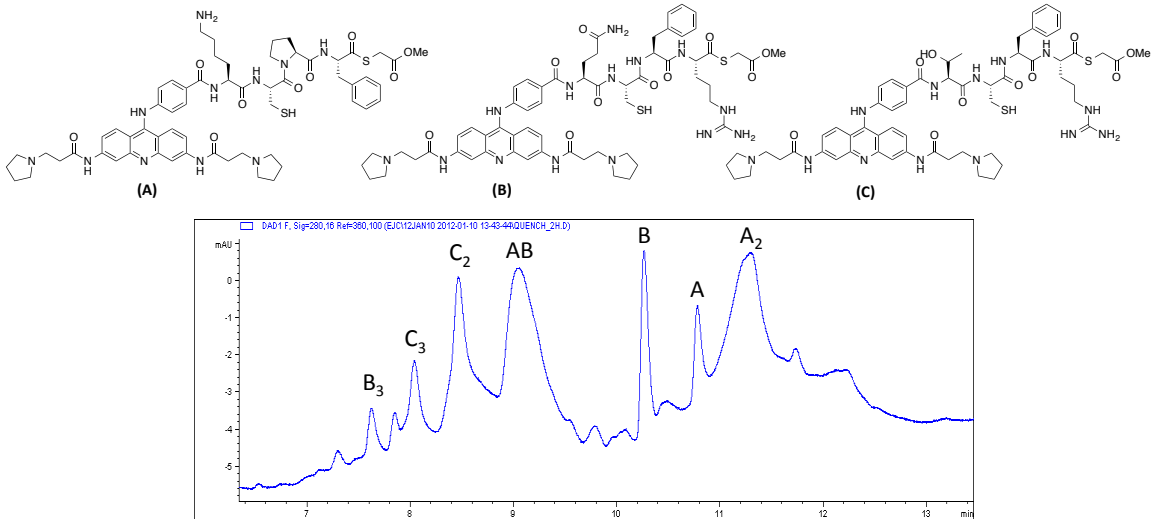

Figure 4.2 Chromatograph of dynamic combinatorial library at 280 nm. Libraries equilibrated for 1 hour in 25 mM Tris buffer with 50 mM KCl at pH 7.4.

ii. Dynamic Combinatorial Experiments

The most common method used to detect and analyze the amplified products of DCC utilizes a biotinylated oligonucleotide target molecule that can be immobilized onto streptavidin-functionalized magnetic beads by the strong biotin-streptavidin interaction.[2] This allows ready

separation of the DNA from the DCL members. Upon removal of the supernatant, washing, denaturing the DNA and removing the beads, the compounds that were bound to the DNA are identified by LC/MS. Similarly, we designed a system containing a DNA sequence with a biotin tag (Figure 4.3), allowing the library mixture to be added to streptavidin-coated beads to isolate the DNA from the library samples. In the initial system we utilized streptavidin-coated agarose or Ultralink resins purchased from Fisher Scientific. This system allowed library washes to be isolated by centrifugation through a fritted column. After thorough rinsing at room temperature, the streptavidin beads could be heated to 80 °C to denature the DNA and thus release any bound library species. Both washes could then be analyzed via LC/MS.

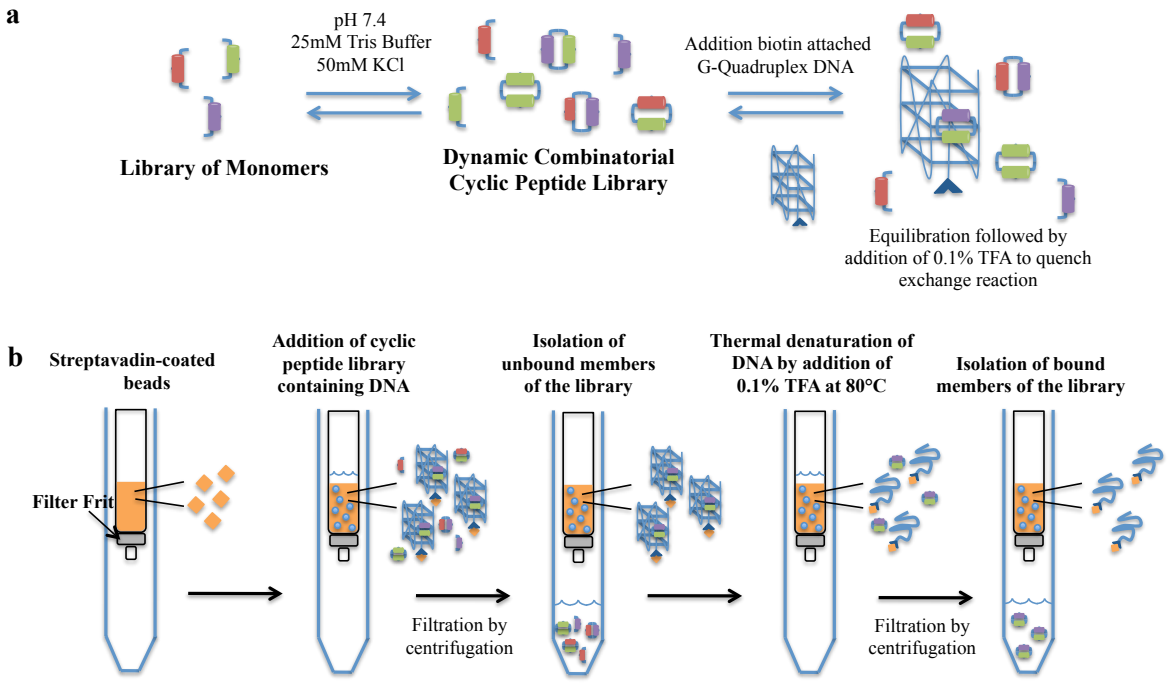

Figure 4.3 Schematic of dynamic combinatorial experiments. A) DCC libraries equilibrated in the presence of quadruplex DNA attached to biotin B) Isolation of DCC libraries from DNA sample to allow for subsequent analysis by LC/MS.

Initial screens were done with acetyl-capped monomers due to the more time-consuming synthesis of the monomers containing the acridine ligand. We began by using the High-Capacity

Pierce Streptavidin Agarose Resin due to the cost efficiency and high loading of streptavidin. Initial screening resulted in incomplete recovery of the library members. As shown in Figure 4.4, the combined signals for the supernatant and denature washes did not equal the signal for the library in buffer in the absence of a DNA template. In the effort to overcome this problem, the libraries were quenched with acetic acid rather than trifluoroacetic acid (TFA) since too much TFA often resulted in the DNA crashing out of solution. In addition, different methods of heating were attempted in the denature step. Initially denaturation was done by adding a hot solution of 0.1% TFA. However, this seemed insufficient so heat resistant parafilm was used to seal the bottom of the columns so that they could be heated directly in a sand bath. Also, different heating times were investigated from 10 minutes to 30 minutes, but none of these changes seemed to improve recovery of the library species. Further screens with different monomers were completed and displayed the same problem, showing incomplete recovery of the library members. However, despite seeing incomplete recovery of the library species, we did see amplification of the Ac-KCPW-SR dimer species in the presence of *Htelo* (Figure 4.5). This result was encouraging, and we continued to work to optimize the experiment to determine what was happening to the library members.

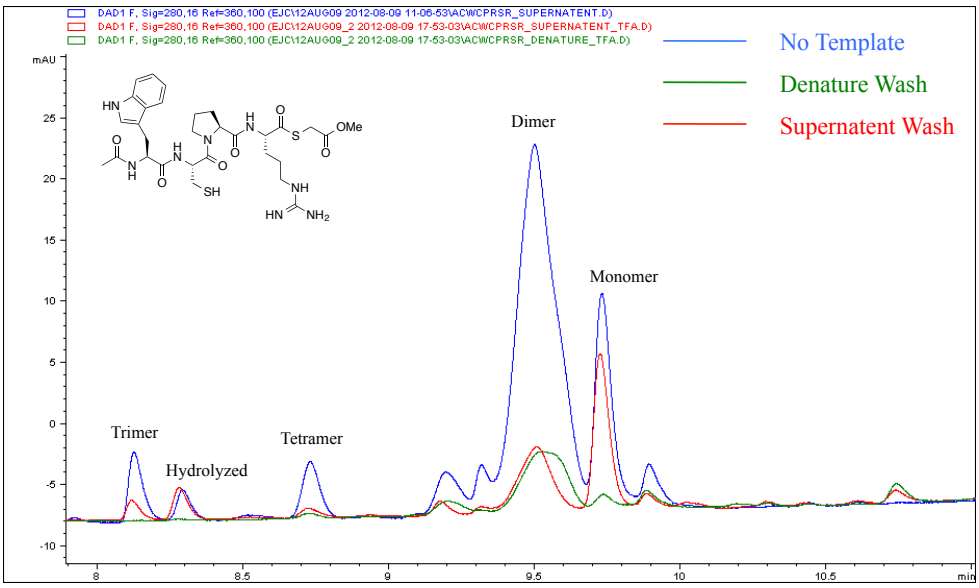

Figure 4.4 Chromatograph of monomer Ac-WCPR-SR screened against quadruplex sequence *Htelo* at 280 nm. Library equilibrated for 1 hour at 500 μM in 25 mM Tris buffer with 50 mM KCl at pH 7.4 at 25 °C.

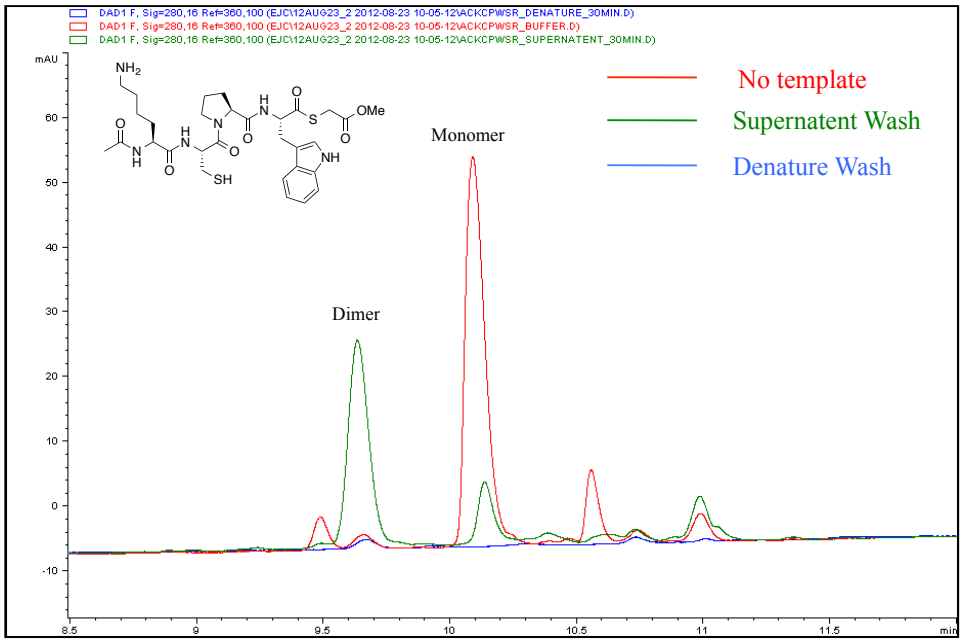

Figure 4.5 Chromatograph of monomer Ac-KCPW-SR screened against quadruplex sequence *Htelo* at 280 nm. Library equilibrated for 1 hour at 500 μM in 25 mM Tris buffer with 50 mM KCl at pH 7.4 at 25 °C.

We had a number of theories for why we were not recovering all of the library members including: 1) The DNA was not being completely denatured, 2) The library members were interacting with the streptavidin beads and 3) the thioester bond was reacting with the DNA. In the effort to circumvent the second problem, we attempted to carry out the experiments using Plus capacity Pierce Streptavidin Ultralink Resin that contains a polyacrylamide resin as opposed to agarose. The dynamic combinatorial experiments were carried out the same way as before with the only difference being the resin used. This seemed to result in improved, if not complete, recovery of the library members (Figure 4.6). Dimeric as well as trimeric species were amplified while the monomeric species decreased. Both of the dimeric and trimeric species were also pulled out in the denature wash. However, some species were amplified that could not be identified.

Pleased with this result, we moved forward to incorporate monomers containing the acridine ligand at the N-terminus. Following the same protocol and using the Plus capacity Pierce Streptavidin Ultralink Resin we screened three monomers containing the acridine ligand against the Htelo sequence. Again however, we encountered issues with recovering all of the library members (Figure 4.7) as no species were recovered in the supernatant wash and less than half were recovered in the denature wash. The concentration of the DCC screens was decreased to 100μM given the high binding affinity the acridine ligand has for the quadruplex structure. We rationalized that maybe having two ligands (once the monomers have dimerized) was too much, leading to non-specific interactions with the quadruplex DNA. To correct for this, libraries containing acetyl-capped monomers were incorporated as well as those containing the acridine ligand at the N-terminus (Figure 4.8). This also did not result in complete recovery of the library species. It was interesting however, that most of the acetyl capped library members

were recovered while little to none of the library members containing the acridine ligand were recovered (Figure 4.8). This led us to believe that the DNA was not being completely denatured, therefore some library species were remaining bound.

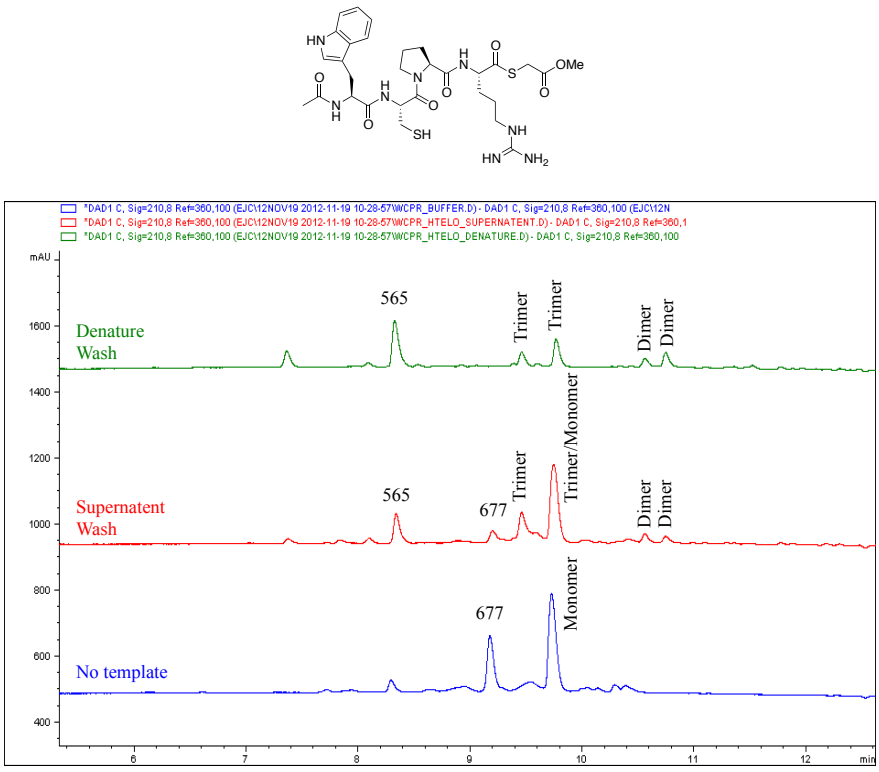

Figure 4.6 Chromatograph of monomer Ac-WCPR-SR screened against quadruplex sequence *Htelo* at 210 nm. Library equilibrated for 1 hour at 500 μM in 25 mM Tris buffer with 50 mM KCl at pH 7.4 at 25 °C.

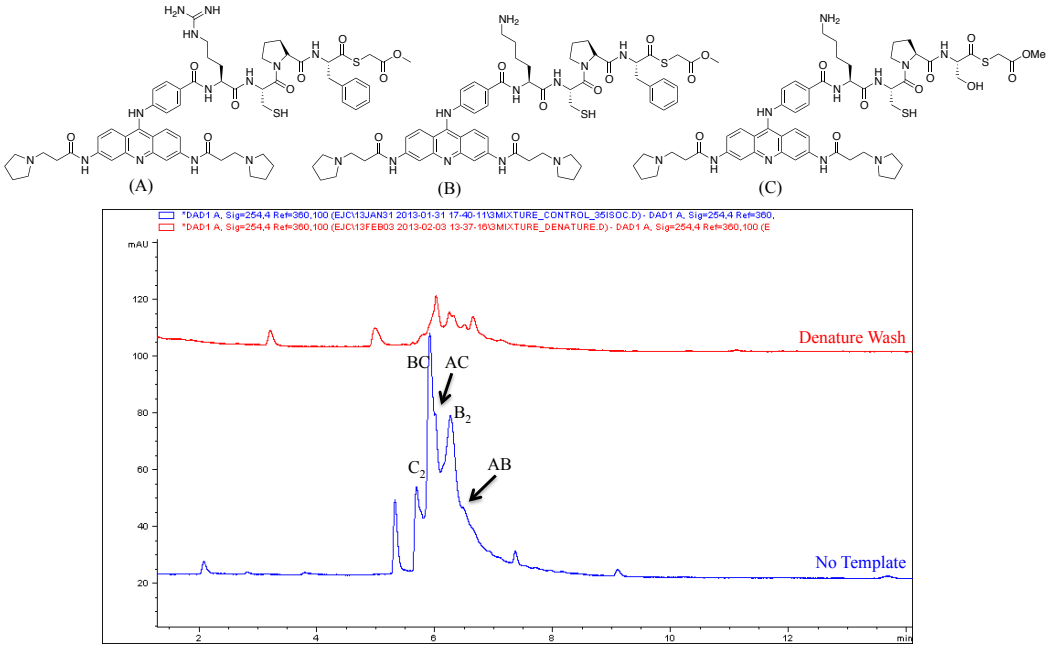

Figure 4.7 Chromatograph of monomers Lig-RCPF-SR, Lig-KCPF-SR and Lig-KCPS-SR screened against quadruplex sequence *Htelo* at 254 nm. Library equilibrated for 1 hour at 100 µM in 25 mM Tris buffer with 50 mM KCl at pH 7.4 at 25 °C.

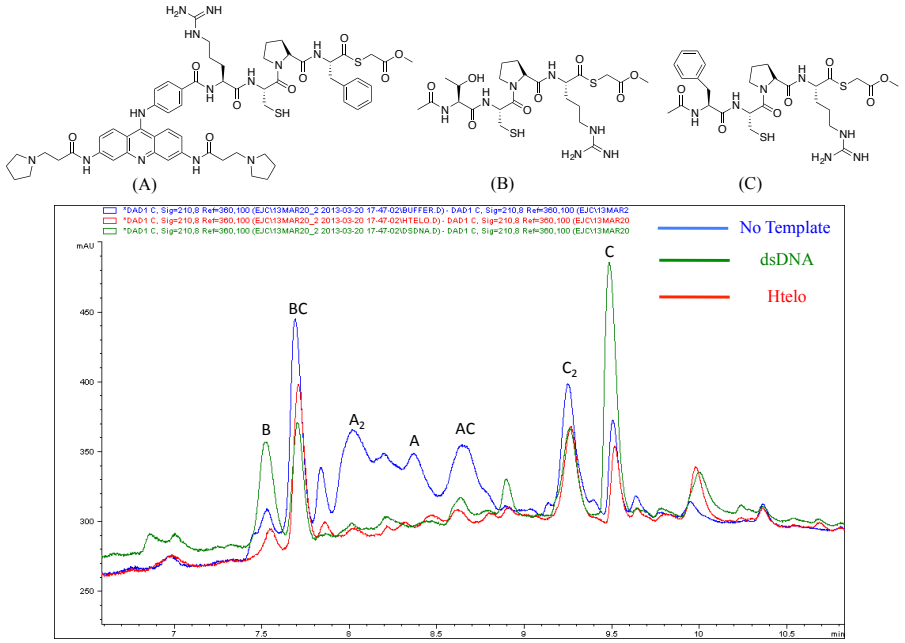

Figure 4.8 Chromatograph of monomers Lig-RCPF-SR, Ac-TCPR-SR and Ac-FCPR-SR screened against quadruplex sequence *Htelo* and dsDNA at 210 nm. Library equilibrated for 1 hour at 100 µM in 25mM Tris buffer with 50 mM KCl at pH 7.4 at 25°C. Spectrums shown are the combination of the supernatant and denature washes.

At this time, the libraries were also screened against dsDNA in addition to *Htelo, cKit21* and *cMyc22* to determine if there is any selectivity. Interestingly, in Figure 4.8, it appears that the dsDNA favors the monomers. However, nothing conclusive can be determined since not all of the library members in the screen were recovered. Figure 4.9 shows screens of Lig-RCPF-SR, Ac-TCPR-SR, Ac-FCPR-SR against *Htelo, cKit21* and *cMyc22*. Despite not recovering all of the library members, there is little difference between chromatographs of the different quadruplex structures showing no specificity.

In the effort to recover all of the library members, the beads were washed with an organic solvent (95% MeCN, 5% H_2O, 0.1% TFA) in an attempt to disrupt any electrostatic interactions between the DNA and positively charged monomers. Once again, we did not see improved recovery of the monomers. It was also hypothesized that we did not see full recovery of the library members due to incomplete denaturation of the DNA sample. So, in the effort to aid in denaturation, the beads were washed with 1 mM of 18-crown-6. By adding 18-crown-6, we hoped that the potassium ions would preferentially bind to the 18-crown-6 rather than the DNA thus effectively destabilizing the quadruplex structure. A side-by-side comparison was done of heat denaturation at 80 °C versus the addition of 1 mM of 18-crown-6 at °C in the effort to recover library members. As shown in Figure 4.10, the addition of 18-crown-6 at room temperature was not effective at denaturing the DNA and retrieving the bound library members. The DNA was then denatured by treating the streptavidin beads with 1 mM of 18-crown-6 in addition to mild heating at 50°C, but again, we did not see full recovery of the library members.

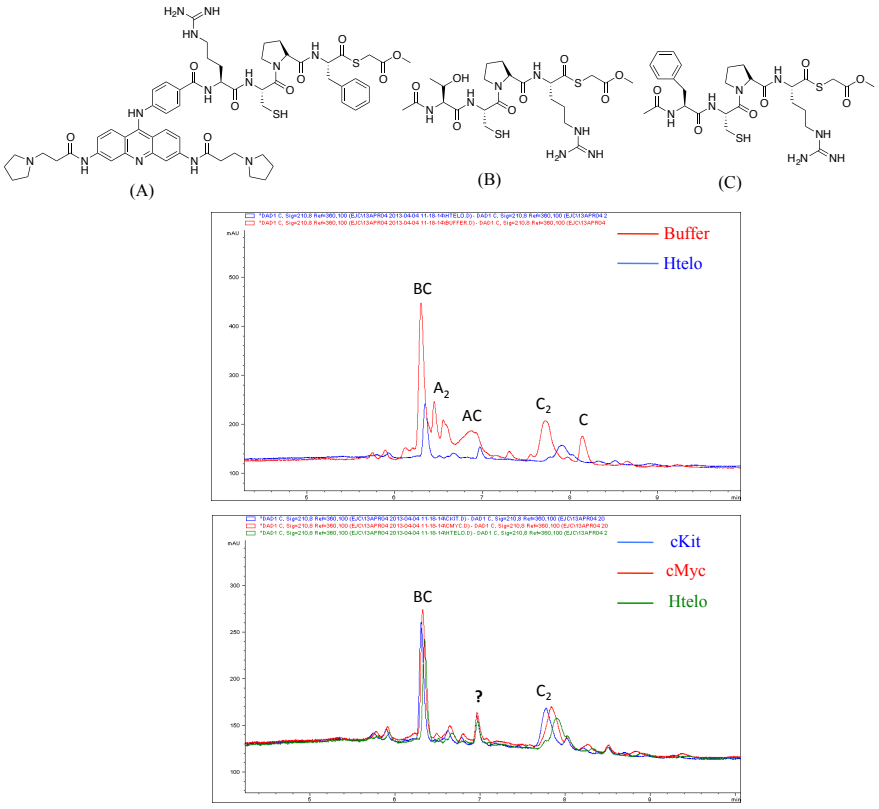

Figure 4.9 Chromatograph of monomers Lig-RCPF-SR, Ac-TCPR-SR and Ac-FCPR-SR screened against quadruplex sequence *Htelo, cKit21* and *cMyc22* at 210 nm. Library equilibrated for 1 hour at 100 µM in 25mM Tris buffer with 50 mM KCl at pH 7.4. Spectra shown are the combination of the supernatant and denature washes.

In a final attempt, streptavidin beads were ordered from a different company due to the possibility that the monomers may still be having unfavorable interactions with the streptavidin resins or impurities contained in the resin. We ordered Streptavidin Magnetic Beads from New England Biolabs. Since the beads were now magnetic, a new protocol had to be developed for working up and isolating the library members. Similar to the method used by Balasubramanian,[3-5] the library of monomers could be mixed with a desired DNA sequence attached to a biotin tag and allowed to equilibrate. The magnetic streptavidin beads could then be added to bind to the DNA. A magnet could subsequently be used to isolate the magnetic beads allowing the supernatant to be pipetted off (Figure 4.11).

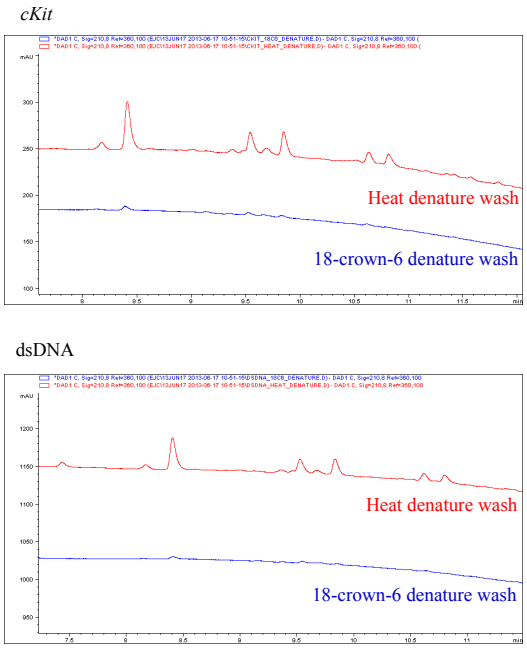

Figure 4.10 Chromatograph of monomers Ac-WCPR-SR screened against quadruplex sequence *cKit21* and dsDNA at 210 nm. Library equilibrated for 1 hour at 250 µM in 25 mM Tris buffer with 50 mM KCl at pH 7.4. Top spectra in both chromatographs is the heat denaturation wash at 80 °C while the in bottom spectrum the denaturation method is the addition of 1 mM 18-crown-6 at 25 °C.

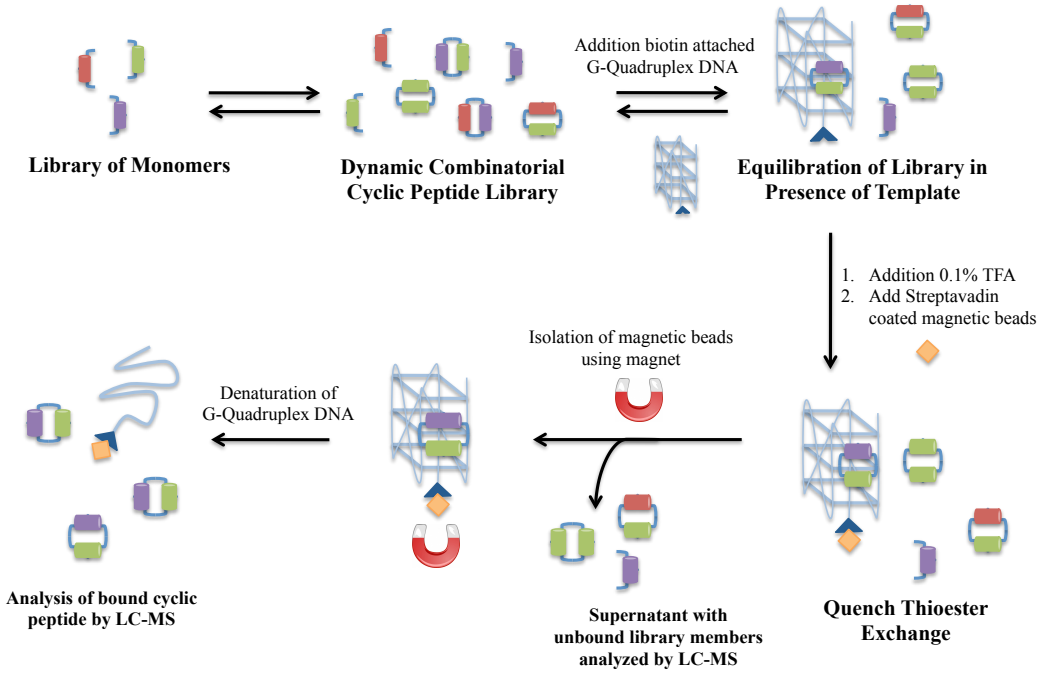

Figure 4.11 Schematic of dynamic combinatorial experiments using streptavidin-coated magnetic beads.

Dynamic combinatorial experiments were carried out under the protocol outlined in Figure 4.11 using the Streptavidin Magnetic Beads from New England Biosystems. The experiments were performed at 250 µM in 25 mM Tris buffer at pH 7.4 with 50 mM KCl. The libraries were allowed to equilibrate in the presence of the sequence *Htelo*. After the supernatant was separated from the oligonucleotides bound to the magnetic beads, the DNA was denatured at 80 °C. Following analysis by LC/MS, it was determined all library members were covered, with the homodimers and heterodimer of monomers Ac-KCPW-SR and Ac-FCPR-SR being released in the denaturing wash (Figure 4.12). Extremely pleased with this result, we repeated the experiment while also screening against the dsDNA sequence. We were once again were able to recover all library members, with the same dimeric species being released in the denature wash when screening against the *Htelo* sequence. Some of the same species were released in the denature wash against the dsDNA sequence, albeit in lower quantities. Since it appeared that these cyclic octapeptide species were binding to the quadruplex structure of *Htelo*, we have begun preliminary binding studies utilizing thermal denaturation and SPR.

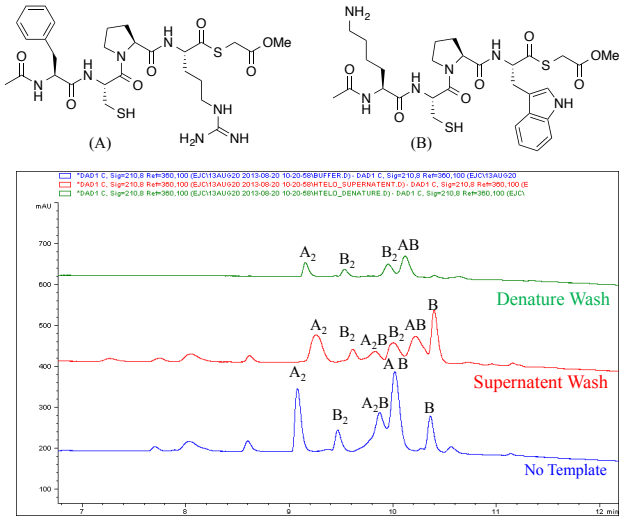

Figure 4.12 Chromatograph of monomers Ac-FCPR-SR and Ac-KCPW-SR screened against quadruplex sequence *Htelo* at 210 nm. Library equilibrated for 1 hour at 100 µM in 25mM Tris buffer with 50 mM KCl at pH 7.4 at 25°C.

C. Conclusions and Future Directions

Recent results have suggested that dimeric species of the peptide monomers bind with some affinity to the *Htelo* sequence specifically over duplex DNA. Initial binding studies have been carried out by performing thermal denaturation experiments at 295 nm however; no change is seen in the presence or absence of the library members. This could be due to the weak binding interaction of the DNA and library members or that the cyclic peptide provides no additional thermodynamic stability to the secondary DNA structure. Future binding studies can be carried out using surface plasmon resonance (SPR) as a more direct means to determine the binding affinity of the cyclic peptide species. Following a fragment based lead discovery, the small cyclic peptides that bind with weak affinity to G-quadruplex DNA can be combined with the acridine ligand to ideally provide a lead with even higher binding affinity. Therefore, binding experiments can be carried out with the hits identified above containing only acetyl-capped monomers in addition to the cyclic peptide hits conjugated to the acridine ligand.

The dynamic combinatorial experiments proved to be more difficult to carry out than initially believed. The type of streptavidin beads used was very important in order to prevent unwanted interaction of the library members with the beads. Additionally, more recent data suggests that the temperature at denaturation is also important. Another theory for incomplete recovery of the library members was reaction of the DNA with the thioester bond. It is possible this occurs at elevated temperatures, close to 80 °C. When the experiments were carried out at a slightly lower denaturation temperature of 70 °C, all members of the library could consistently be fully recovered. However, we were unable to conclusively prove or disprove this theory. Having established a successful protocol for the DCC experiments, additional libraries can be screened for selectivity in the future, including those containing the acridine ligand.

D. Experimental Procedure

i. Oligonucleotides

5'-biotinylated oligonucleotides were purchased from Integrated DNA Technologies at HPLC-purified quality. Oligonucleotide samples were taken up in 1 mL of MilliQ purified water and UV/Vis was used to determine concentrations at 260 nm. Concentrations were determined at 95 °C to ensure that oligonucleotides were completely denatured. The concentration was determined using the molar extinction coefficient provided by Integrated DNA Technologies for each oligonucleotide sample. 250 μM stock solutions were made by transferring 150 nmol of each oligonucleotide to an eppendorf tube, freeze drying the sample, and reconstituting the oligonucleotide in 600 μL of 25 mM Tris buffer at pH 7.4 with 50mM of KCl. In the case of the dsDNA, each of the complementary strands was freeze dried together. Before use in the DCLs, the oligonucleotides were annealed by heating to 95 °C for 5 min, followed by slow cooling to room temperature.

Gene	Sequence
Htelo	/5Biosg/GTT AGG GTT AGG GTT AGG GTT AGG GTT AGG
cKit-21	/5Biosg/CGG GCG GGC GCG AGG GAG GGG
cMyc-22	/5Biosg/GGC ATA GTG CGT GGG CGT TAG C
NRas	/5Biosg/TGT GGG AGG GGC GGG TCT GGG TGC
dsDNA	/5Biosg/GGC ATA GTG CGT GGG CGT TAG C CCG TAT CAC GCA CCC GCA ACG G

Figure 4.13 Sequences of biotinylated oligonucleotides used in the dynamic combinatorial experiments.

ii. Circular Dichroism

CD measurements were performed on a Chirascan Plus Circular Dichroism Spectrometer. CD data was obtained for each nucleotide sample at the concentration indicated in 25 mM Tris at pH 7.4 with 50 mM KCl. Scans were performed at 25 °C from 320 – 220 nm (Figure 4.14).

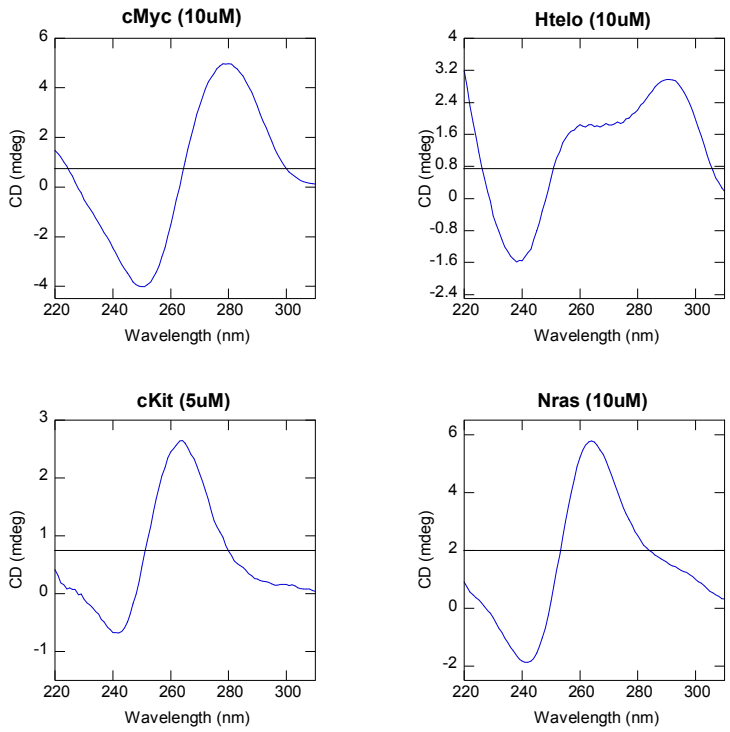

Figure 4.14 CD data for annealed sequences of *Htelo, cMyc, cKit,* and *Nras*. CD data obtained at the concentration indicated in 25 mM Tris at pH 7.4 with 50 mM KCl. Scans were performed at 25 °C from 320 – 220 nm. Sequences *cKit* and *Nras* are parallel folded structures as indicated by a maximum at 260 nm and a minimum at 240 nm. The sequence *cMyc* is an anti-parallel folded structure as indicated by a maximum at 290 nm. *Htelo* is a mixed parallel and anti-parallel folded structure indicated by two maximums at 260 nm and 290 nm.

Thermal denaturation experiments were performed using the same buffer and concentrations as indicated. Measurements were taken between 25 and 94 °C. The melting curves were normalized to show the fraction folded using Eq. 4.1 where Θ_0 is MRE, Θ_D is the MRE for the fully denatured oligonucleotide, and Θ_F is the MRE for initially folded oligonucleotide.

$$Fraction\ folded = \frac{\Theta_O - \Theta_D}{\Theta_F - \Theta_D}$$ Eq. 2.2

iii. DCC Experiments

Monomer building blocks were stored as 5 mM stock solutions. 5 μL (25 nmol) was removed and freeze dried in an eppendorf tube for the DCC experiments. Each DCL was prepared at a final volume of 100 μL and contained 250 μM of template (where appropriate) and 250 μM of each peptide monomer in 25 mM Tris at pH 7.4 with 50 mM KCl. The libraries were left to stand for 6 hours. The mixtures were then acidified by adding 100 μL of a 0.1% TFA solution in water in order to halt thiol-thioester exchange. DNA templates were removed before UV-HPLC-MS analysis. For each DCL, the library mixture was added to 1 mL of streptavidin-coated magnetic beads (Streptavidin Magnetic Beads, 4 mg/mL suspension, New England Biolabs, Inc.) that had been washed repeatedly with 0.1% TFA in water (5×200 μL). The beads and library mixture were incubated for 15 minutes, the beads were retained in the vessel using a magnet, while the supernatant was pipetted off. The beads were washed with a solution of 0.1% TFA in water (3×200 μL) at room temperature followed by a hot solution of 0.1% TFA in water (incubation at 70 °C for 15 min, 3×200 μL). The solvent was removed in vacuo. The residue was redissolved in a solution of 0.1% TFA in water (50 μL) and analysed by UV-HPLC-MS.

References

1. Corbett, P.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. –L.; Sanders, J. K. M.; Otto, S. Dynamic combinatorial chemistry. *Chem. Rev.* **2006,** *106,* 3652-3711.

2. Sherman Durai, C. R.; Harding, M. M. Targeting nucleic acids using dynamic combinatorial chemistry. *Aust. J. Chem.* **2011**, *64,* 671-680.

3. Ladame, S.; Balasubramanian, S. Templated ligand assembly by using G-quadruplex DNA and dynamic covalent chemistry. *Angew. Chem. Intl. Ed.* **2004**, *43,* 1143-1146.

4. Whitney, A. M.; Balasubramanian, S. Targeting nucleic acid secondary structures with polyamides using an optimized dynamic combinatorial approach. *Angew. Chem. Intl. Ed.* **2005**, *44,* 5736-5739.

5. Wieter, J. J-L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. Exploring the differential recognition of DNA G-quadruplex targets by small molecules using dynamic combinatorial chemistry. *Angew. Chem. Intl. Ed.* **2008**, *47,* 2677-2680.

CHAPTER 5: INTRODUCTION
SIGNIFICANCE OF HISTONE MODIFICATION AND TRIMETHYLATED LYSINE

A. Significance of This Research

Epigenetics is the study of changes in the genetic expression that do not involve altering the underlying nucleotide sequence.[1] As such, epigenetics has become a major challenge in understanding the mechanisms of how genetic information is controlled. Accumulating evidence has shown that post-translational modifications (PTMs) are a key component in controlling gene expression.[2] A number of different modifications have been documented, including phosphorylation of serine and threonine, methylation of lysine and arginine, and acetylation of lysine, which act as chemical switches to stimulate or inhibit protein-protein interactions.[3-5] More specifically, we are interested the methylation of lysine at the N-terminal tail of histones, which is a covalent modification shown to be intricately involved in DNA packaging and may directly control levels of gene expression.[6] Therefore, there is a great effort to understand how recognition domains within the effector proteins that read these covalent modifications. Gaining a greater understanding for the noncovalent interactions that dictate biomolecular recognition will provide better insight into the histone lysine methylation effect and will also aid in the development of inhibitors of these interactions.

B. DNA Packaging

The entire genomic DNA is complexed with a mass of proteins, forming a large macromolecular assembly known as chromatin.[7,8] This macromolecular assembly effectively compacts the genomic DNA 10,000-fold and packages it into well-organized higher structures, maintaining accessibility for replication and transcription. The basic building block of the chromatin is the nucleosome, which is comprised of four subunits H2A, H2B, H3, and H4 that dimerize to make up the histone octomer (Figure 5.1). Approximately 147 base pairs of DNA are wrapped around this basic octomeric protein complex. A fifth histone, the linker histone H1, keeps the DNA in place and interacts with the linker DNA, which is the DNA that connects two nucleosomes.[9,10] The entire macromolecular assembly, made up of the histone subunits and the DNA, was found to form a disk-like shape with an approximate diameter of 100Å as determined by X-ray diffraction.[11]

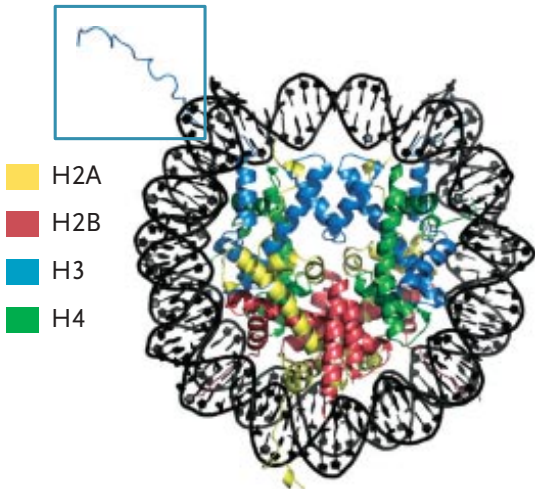

Figure 5.1 Nucleosome structure with DNA wrapped around the histone. Histone colors are indicated and an example of the histone tail is boxed in.[9]

C. Significance of Histone Modification

Epigenetics is the study of relevant modifications to the genome that do not involve change in the nucleotide sequence.[12] Examples of such changes to the genome involve DNA methylation, which can silence gene expression and covalent post-translational modifications (PTMs) of proteins including histones, chromatin remodeling proteins, and transcriptional regulators.[13,14] Both of these examples serve to regulate gene expression without altering the underlying DNA sequence. Consequently, non-genetic factors can cause an organism's genes to behave or express themselves differently. Recently, it has been realized that each of these regulatory mechanisms are interconnected, resulting in additional levels of complexity. Anomalous regulation of these mechanisms have been linked with a variety of diseases such as autoimmune diseases and cancer.[15] Therefore, it is important to increase our understanding of the mechanism and biological impact of these epigenetic regulators so we can gain a better understanding of cellular growth, development and human diseases.

The N-terminus of each core histone protein is largely unstructured, making up 20% of the total amino acids of the histone proteins and protruding from the globular core of the nucleosome. This section of the histone is referred to as the histone tail, and accounts for the majority of variability in post-translation marks between species. These sequences are highly conserved amongst eukaryotes and are made up of mostly positively charged residues resulting in favorable electrostatic interaction with the sugar-phosphate backbone of DNA.[16] The histone tails are known to interact with a variety of proteins, influencing biological processes involving transcription, replication, and chromosome maintenance. A number of different modifications have been documented, including phosphorylation of serine and threonine, methylation of lysine and arginine, and acetylation of lysine amongst others.[3-5]

i. Lysine Methylation on Histones

Methylation is one of the most abundant covalent PTM of the histone and is the focus of this research. The methylation of the histone is one of the most complex and least understood of the covalent modifications, likely due to the numerous levels of possible methylation states of lysine and arginine including mono-, di-, and trimethylation with each methylation state affecting DNA expression differently.[6] Consequently, methylation is site specific and can be associated with both activation and repression of gene expression.

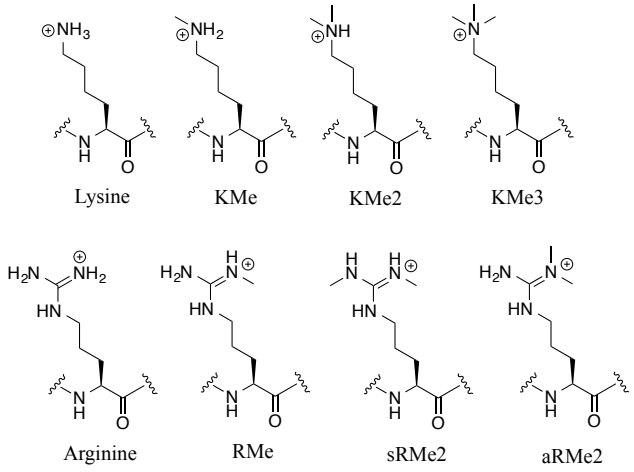

Figure 5.2 Various methylation states of lysine and arginine. Dimethylarginine can exist in the symmetric (s) and asymmetric (a) form.

The methylation of lysine residues within the histones is tightly regulated by methyltransferases (KMTs) and demethylases (KDMs), resulting in maintenance of cell fate and genomic stability. KMTs and KDMs are very specific for a particular residue and the degree of methylation on that residue. Almost all KMTs contain an enzymatic SET-domain, with some acting exclusively as mono-, di-, or trimethyltransferases, whereas others can carry out all three methylastions in the presence of S-adenosylmethionine (AdoMet) as the methyl-donating cofactor.[6] The degree of methylation is dependent on the size of a narrow channel termed the

"methylation" core, while site specificity is dictated by H-bonds to amino acids surrounding the lysine substrate.

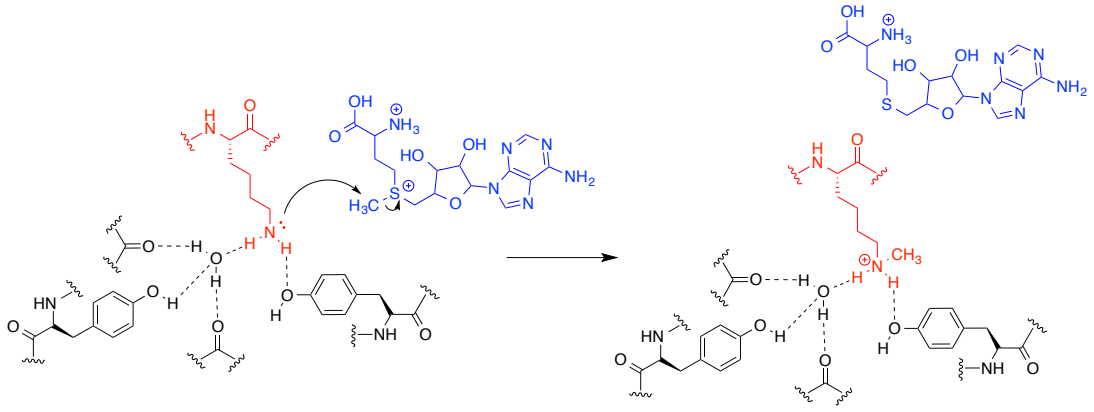

Figure 5.3 Mechanism of lysine 9 methylation catalyzed by the SET7/9 with the cofactor AdoMet. Histone lysine is highlighted in red. AdoMet is highlighted in blue and is converted to S-adenyl homocysteine (AdoHcy). The SET7/9 residues are shown in black.

More recently, understanding regarding lysine demethylation has been greatly enhanced through structural analysis. It was previously believed that methyllysine was an irreversible modification that required phosphorylation of a neighboring amino acid to reverse the effects of the methyl mark. However, structural analysis has identified key interacting partners and amino acids in KDMs that are responsible for their activity and specificity. For example, structures of the JmjC catalytic domain revealed that it consists of a β-barrel structure that coordinates Fe(II) and α-ketoglutarate.[18,19] In addition, KDM4A interacts with the methylated lysine via amino acids within a specific β-sheet in the β-barrel structure and are predicted to impact the degree of demethylation.[20]

Modifications to histone proteins transmit biological information through local binding factors that recognize distinct modifications on specific residues.[21] Numerous reports have described methylated histone lysine binding proteins, also known as 'effector' modules.[22] The chromodomain has been identified in numerous chromatin-associated proteins and has been

shown to be specific for methylated lysines in certain instances. The binding affinities of chromodomains for their respective modified lysines have been shown to be relatively weak (K_D ~ 10^{-4} – 10^{-6}), which are thought to allow for rapid 'on-off' binding.[21] The chromodomain is part of a family of methylated lysine recognition domains also containing tudor, MBT and PWWP that are collectively known as the 'royal family.' All proteins containing domains of the "royal family" are implicated in chromatin and genetic control.[23]

ii. Lysine 9 of Histone H3 Tail

Lysine 9 methylation of histone H3 is catalyzed by a number of different enzymes including human Suv39h, G9a, ESET, and GLP/Eu-HMTase1. The methylation of H3K9 has been linked to the inhibition of H3/H4 acetylation, and H3K4 methylation, all of which are marks of active chromatin, suggesting that H3K9 methylation may define transcriptionally inactive chromatin.[24,25] Cross talk also occurs between H3K9Me3 and other histone modifications, including phosphorylation of H3 serine 10, leading to more complicated regulation in cellular function.

HP1 is a reader protein that is highly selective for H3K9Me3 over other methylated lysine marks, such as H3K4Me3. The HP1 protein contains a chromodomain that is responsible for proper recognition and binding of H3K9Me3. The recognition of H3 trimethylation by the chromodomain of HP1 has been link to transcriptional repression via heterochromatin formation.[26,27,11] Therefore, accurate targeting of HP1 via its chromodomain is necessary for the maintenance of proper chromatin function. The interaction of these two proteins is critical for the control of gene expression and the maintenance of a stable genome during cell development.[28]

Structures of HP1 bound to H3K9 di- and trimethylated peptides reveal an aromatic cage of three residues that binds the methyl ammonium group.[29,30] The structure shows a binding

pocket made up of 5-10 residues of the HP1 chromodomain that participate directly in the binding to the H3 histone tail. The trimethylated lysine sits in the binding pocket, equidistant from the aromatic face of two tyrosine residues and a tryptophan. This strongly suggests that recognition of H3K9Me3 by the aromatic cage of the HP1 chromodomain is governed by cation-π interactions. The cation-π interaction is defined by the attractive interaction between a positively charged moiety and the π electrons of the aromatic ring.[31] Recognition of trimethyllysine via cation-π interaction was further confirmed by studies showing that mutation of each of the aromatic residues to an alanine resulted in a dramatic reduction in binding affinity.[29] Additionally, previous studies in our lab demonstrated that substitution of the cationic methyl ammonium group with a neutral *tert*-butyl group resulted in a binding affinity that was reduced 31-fold.[32] This was a surprising result given the similar size and increased hydrophobicity of the neutral *t*-butyl group compared to the trimethyl ammonium moiety. This study however, further demonstrates the importance of the cation-π interaction in the recognition of trimethyllysine by the HP1 chromodomain.

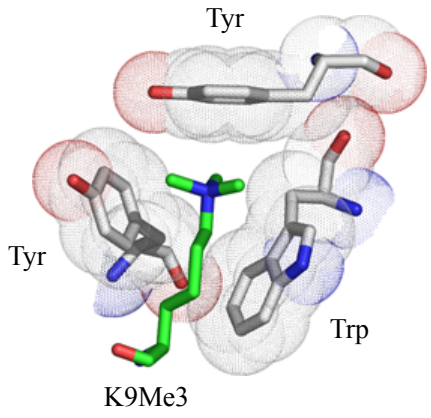

Figure 5.4 Aromatic cage of HP1 chromodomain formed by two tyrosines and one tryptophan bound to trimethyllysine.

D. Purpose of This Work

Lysine can exist in three distinct methylation states under the control of highly specific methyl transferases or demethylases. These methylation states serve to turn on specific protein-protein interactions with partners that specifically recognize the methylated side chain. These interactions are implicated in diverse cellular processes, but are most closely associated with gene regulation pathways and dysregulation leads to a number of diseases, including some types of cancers. More recently, it has been documented that these effector proteins engage methylated amino acids via an aromatic cage. Recognition and affinity is mainly derived from cation-π interactions between the positively charged cationic side chain and the electron rich π surfaces of nearby aromatic rings. It is the goal of this work to study how much the cation-π interaction contributes to the binding of the HP1 chromodomain to trimethylated lysine 9 of histone 3 (H3). This interaction can be quantified by incorporation of fluorinated derivatives of the aromatic amino acids responsible for the binding of H3K9Me3 to the HP1 chromodomain. The cation-π interaction between the aromatic pocket of the HP1 chromodomain and H3K9Me3 can be revealed by incorporation of a series of fluorinated amino acid analogues. A linear correlation between binding affinity and the calculated magnitude of the cation-pi interaction of those groups indicates a cation-π interaction. By carrying out these studies, we hope to gain more information regarding the noncovalent interactions responsible for the biomolecular recognition of H3K9Me3 by aromatic pocket of the HP1 chromodomain. Because many reader proteins for methylated lysine have an aromatic cage in their binding pockets, findings from the investigation of the HP1 chromodomain will provide broad insight into this class of proteins.

References

1. Khorasanizadeh, S. *Cell* **2004**, *116,* 259-272.

2. Kouzarides, T. Chromatine Modifications and Their Function. *Cell* **2007**, *128,* 293-705.

3. Langan, T.A. Histone Phosphorylation: Stimulation by Adenosine 3',5'-Monophosphate. *Science* **1968**, *162,* 579-580.

4. Allfrey, V. G.; Faulkner, R.; Mirsky, A. E., Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. *Biochemistry* **1964**, *51*, 786-794.

5. Mahadevan, L. C.; Willis, A. C.; Barratt, M. J., Rapid Hisonte H3 Phosphorylation in Response to Growth Factors, Phorbol Ester, Okadaic Acid, and Protein Synthesis Inhibitors. *Cell* **1991**, *65*, 775-783.

6. Martin, C.; Zhang, Y., The Diverse Functions of HIstone Lysine Methylation. *Nat Rev Mol Cell Biol* **2005**, *6*, 838-849.

7. Venter, J. C.; al, e., The Sequence of the Human Genome. *Science* **2001**, *291*, 1304-1351.

8. Widom, J., Toward a Unified Model of Chromatin Folding. *Annu Rev Biophys Bio* **1989**, *18*, 365-395.

9. Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. R., Crystal Structure of the Nucleosome Core Particle at 2.8Å Resolution. *Nature* **1997**, *389*, 251-160.

10. Thoma, F.; Koller, T. H.; Klug, A., Involvement of Histone H1 in the Organization of the Nucleosome and of the salt-Dependent Superstructures of Chromatin. *J Cell Biol* **1979**, *83*, 403-427.

11. Pardon, J. F.; WIlkins, M. H. F.; Richards, B. M., Super-Helical Model for Nucleohistone. *Nature* **1967**, *215*, 508-509.

12. Egger, G.; Liang, G.; Aparicio, A.; Jones, P. A., Epigenetics in Human Disease and Prospects for Epigenetic Therapy. *Nature* **2004**, *429*, 457-463.

13. Tate, P. H.; Bird, A. P., Effects of DNA Methylation on DNA-BOnding Proteins and Gene Expression. *Curr Opin Genet Dev* **1993**, 226-231.

14. Kouzarides, T., Chromatin Modifications and Their Function. *Cell* **2007,** *128*, 693-705.

15. Lu, Q.; Qiu, X.; Hu, N.; Wen, H.; Su, Y.; Richardson, B. C., Epigenetics, Disease, and Therapeutic Interventions. *Ageing Res Rev* **2006**, *5*, 449-467.

16. Marino-Ramirez, L.; Hsu, B.; Baxevanis, A. D.; Landsman, D. The histone database: a comprehensive resource for histones and histone fold-containing proteins. *Proteins,* **2006**, *62,* 838-842.

17. Xiao, B.; Jing, C.; Wilson, J. R.; Walker, P. A.; Vasisht, N.; Kelly, G.; Howell, S.; Taylor, I. A.; Blackburn, M.; Gamblin, S. J., Structure and Catalytic Mechanism of the Human Histone Methyltransferase SET7/9. *Nature* **2003**, *421*, 652-656.

18. Whetstine, J. R.; Nottke, A.; Lan, F.; Smolikov, S.; Chen, Z.; Spooner, E.; Li, E.; Zhang, G.; Colaiacovo, M.; Shi, Y., Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **2006**, *125*, 467-481.

19. Tsukada, Y.-i.; Fang, J.; Erdjument-Bromage, H.; Warren, M. E.; Borchers, C. H.; Tempst, P.; Zhang, Y., Histone Demethylation by a Family of JmjC Domain-Containing Proteins. *Nature* **2006**, *439*, 811-816.

20. Chen, Z.; Zang, J; Whetstine, J.; Hong, X.; Davrazou, F.; Kutateladze, T. G.; Simpson, M.; Mao, Q.; Pan, C. H.; Dai, S.; *et al.* Structural insights into histone demethylation by JMJD2 family members. *Cell* **2006**, *125,* 691-702.

21. Fiscle, W.; Wang, Y.; Allis, C. D. Histone and chromatin cross-talk. *Curr. Opin. Cell. Biol.* **2003**, *15,* 172-183.

22. Daniel, J. A.; Pray-Grant, M. G.; Grant, P. A. Effector proteins for methylated histones: An expanding family. *Cell Cycle* **2005**, *4*, 919-926.

23. Maurer-Stroh, S.; Dickens, N. J.; Hughes-Davies, L.; Kouzarides, T.; Eisenhaber, F.; Ponting, C. P.; The tudor domain 'royal family': tudor, plant agenet, chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **2003**, *28,* 69-74.

24. Santon-Rosa, H.; Schneider, R.; Bannister, A. J.; Sherriff, J.; Berstein, B. E.; Tolga Emre, N. C.; Schreiber, S. L.; Mellow, J.; Kouzarides, T., Active Genes are Trimethylated at K4 of Histone H3. *Nature* **2002**, *419*, 407-411.

25. Schotta, G.; Ebbert, A.; Krauss, V.; Fischer, A.; Hoffmann, J.; Rea, S.; Jenuwein, T.; Dorn, R.; Reuter, G., Central Role of Drosophila SU(VAR)3-9 in Histone H3-K9 Methylation and Heterochromatic Gene Silencing. *EMBO J* **2002**, *21*, 1121-1131.

26. Platero, J. S.; Hartnett, T.; Eissenberg, J. C., Functional Analysis of the Chromo Domain of HP1. *EMBO J* **1999**, *14*, 3977-3986.

27. Bannister, A. J.; Zegerman, P.; Partidge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T., Selective Recognition of Methylated Lysine 9 on Histone H3 by the HP1 Chromo Domain. *Nature* **2001**, *410*, 120-124.

28. Bernard, P.; Maur, J.-F.; Partidge, J. F.; Genier, S.; Javerzat, J.-P.; Allshire, R. C.; Requirement of Heterochromatin for Cohesion at Centromeres. *Science* **2001**, *294*, 2539-2542.

29. Jacobs, S. A.; Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **2002***, 295,* 2080-2083.

30. Nielson, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D. *Nature* **2002**, *416*, 103-107.

31. Dougherty, D. A. The cation-π interaction. *Acc. Chem. Res.* **2013**, *46,* 885-893.

32. Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. Recognition of trimentyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl. Acad. Sci.* **2007**, *104,* 11184-11188.

CHAPTER 6: PROBING THE ROLE OF CATION-π INTERACTIONS USING UNNATURAL AMINO ACIDS

A. Background

The cation-π interaction is an important force in molecular recognition and protein structure.[1] Furthermore, it has been established that the cation-π interaction is an important element in the recognition of the methylated amino acids through interaction with the aromatic pocket of effector proteins.[2-4] The cation-π interaction includes a substantial electrostatic component as the $C^{\delta-}$-$H^{\delta+}$ bond dipoles of the aromatic ring produce a region of negative electrostatic potential on the face of the π system (Figure 6.1a).[5] Electrostatics dictates a natural attraction of cations to the surface. By incorporating fluorinated amino acids, there is a deactivating effect for a cation-π interaction due to electron density being withdrawn from the ring by the fluorine atoms.[6] However, the addition of fluorine atoms results in little perturbation in the sterics of the residue. The energetic contribution of this interaction can be systematically probed through progressive fluorination of the aromatic ring (Figure 6.1b and c), which has an additive effect on the cation-π binding ability of the ring. The purpose of this research is to determine the role of aromatic cage of the HP1 chromodomain in its binding to trimethyl lysine 9 of histone 3 (H3 K9Me3). If the binding of a residue to trimethyllysine is governed by cation-π interactions, there will be a linear correlation between the attenuation of the potential cation-π ability of the residue by fluorination and strength of binding to trimethyllysine.[7,8]

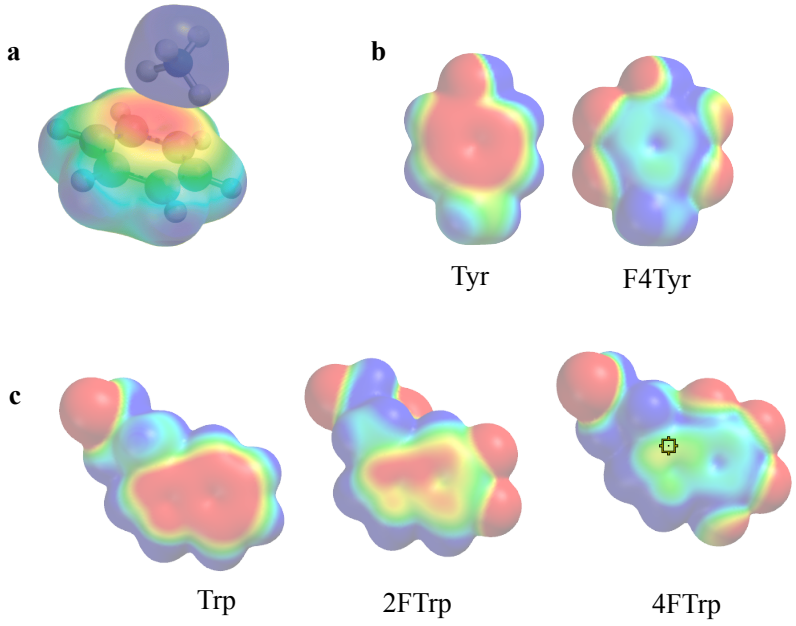

Figure 6.1 Electrostatic potential maps with red regions being the most electron-rich regions of a molecule and blue regions of a map being the most electron-poor regions of a molecule. A) Interaction of benzene with an ammonium cation. B) Electrostatic maps of the sidechains of tyrosine and tetrafluorotyrosine. C) Electrostatic potential maps of the sidechains of tryptophan, difluorotryptophan and tetrafluorotryptophan.

Dougherty *et al.* has used this fluorinated amino acid methodology to probe the role of the cation-π interaction in nicotinic acetylcholine receptors (nAChRs).[9] The nicotinic acetylcholine (ACh) receptor is a neurotransmitter-gated ion channel that contains a number of aromatic amino acids that have been identified as contributing to the recognition of acetylcholine. The presence of these aromatic residues at the binding site suggests that cation-π interaction may be involved in binding the ammonium group of acetylcholine (Figure 6.2a). Consequently, the role of the cation-π interaction was quantitated through the incorporation and electrophysiological evaluation of a series of fluorinated Trp derivatives at four sites of the mouse muscle nAChR. All of these sites naturally contained Trp and all had been implicated in binding to ACh. At three of these sites, incorporation of the tetrafluorotryptophan gave EC_{50} values for ACh that were not very different from the wild type, ruling out a strong cation-π

interaction. However, progressive fluorination at position α149 led to a systematic increase in the EC_{50}. Dougherty *et al.* revealed that fluorination at position α149 resulted in a direct correlation between EC_{50} and calculated cation-π ability. This provided compelling evidence that the quaternary ammonium group of ACh makes close contact with the side chain of αTrp-149 in the nAChR via cation-π interactions.

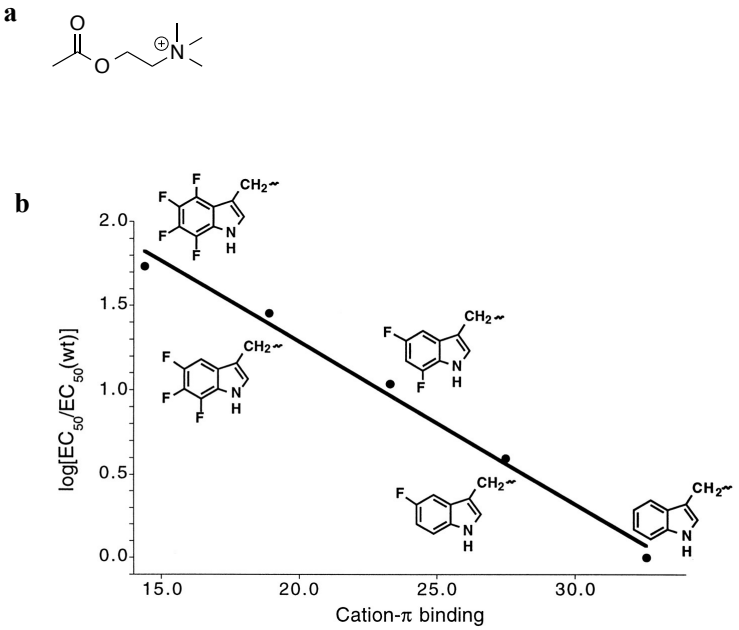

Figure 6.2 A) Structure of acetylcholine. B) Plot of log[EC_{50}/EC_{50}(wild type)] vs. quantitative cation-π binding ability at α149.

B. Peptide Design and Synthesis

The binding of the HP1 chromodomain to trimethyl lysine 9 on the H3 histone tail is an important interaction necessary for the maintenance of proper chromatin function. The deregulation of this interaction has been implicated in a number of disease states. In the effort to further probe the role the cation-π interaction plays in the recognition of trimethyl lysine, we aimed to carry out experiments similar to those done by Dougherty, incorporating fluorinated aromatic acids into the aromatic cage of the HP1 chromodomain. Gaining a greater

understanding for the noncovalent interactions that dictate biomolecular recognition will provide better insight into the histone lysine methylation effect. Additionally, by incorporating fluorinated amino acids into the aromatic binding pocket of the HP1 chromodomain, we can perform assays to determine the influence on the binding of inhibitors designed by the Frye group.

We proposed to accomplish this through the chemical synthesis of residues 22-76 of the HP1 chromodomain. Throughout the chemical synthesis of the HP1 chromodomain fragment, Fmoc protected derivatives of fluorinated amino acids can be site specifically incorporated via traditional solid phase synthesis.

i. Native Chemical Ligation

Despite the fact that much advancement has been made in the field of chemical protein synthesis, the large-scale production of longer peptides (>50 residues) introduces problems not yet resolved by traditional solid phase peptide synthesis. The reiterative nature of solid phase peptide synthesis often results in a complex mixture of deletion products, lowering the overall yield and making the purification and isolation of the desired product difficult. Consequently, the ligation of fully unprotected peptide fragments offers significant advantages over the linear synthesis of longer peptide sequences.

Native chemical ligation has revolutionized the field of peptide chemistry and is the most widely used ligation strategy.[10] Native chemical ligation involves the reaction of unprotected thioesters with N-terminal cysteine peptides (Figure 6.3). The reaction is chemoselective, making it applicable to the ligation of fully unprotected peptides under aqueous conditions. The chemoselective step involves the reversible thiol-thioester exchange of the nucleophilic thiol of the Cys residue located at the N-terminus of the C-peptide and a thioester at the C-terminus of

the N-peptide. The Cys-thioester subsequently undergoes a rapid intramolecular S→N acyl transfer via a favorable five-membered transition state to form a native peptide bond.

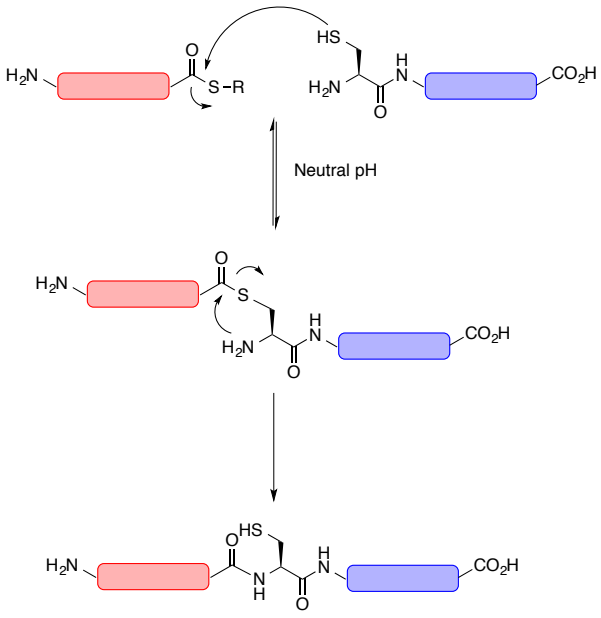

Figure 6.3 Mechanistic aspects of the native chemical ligation.

ii. Synthesis of HP1 Chromodomain

Using the technique of native chemical ligation, we proposed the chemical synthesis of the HP1 chromodomain (residues 22-76) out of the two shorter. We chose to split the sequence between residues 37 and 38 for several reasons. Residue 37 is Gly, which is the preferential terminal amino acid in the thioester fragment due to its greater reactivity than more sterically hindered thioesters. For the C-terminal fragment, we proposed to use SPPS to make the sequence consisting of residues 38-76, in which methionine (Met) at residue 38 was mutated to a cysteine, as shown in Figure 6.4. The proposed sequence is shown in Figure 6.4, with the residues in the aromatic cage highlighted in blue. This site was chosen for mutation due to the location of the residue away from the binding pocket and in a solvent-exposed position, thereby minimizing overall structural changes in the protein.

Peptide	Sequence
HP1 (wt)	Ac-EEYAVEKIIDRRVRKGMVEYYLKWKGYPETENTWEPPENNLDCQDLIQQYEASRKD-NH_2
HP1 22-37 (thioester)	Ac-EEYAVEKIIDRRVRKG-thioester
HP1 38-76 (M38C)	CVEYYLKWKGYPETENTWEPPENNLDCQDLIQQYEASRKD-NH_2

Figure 6.4 Sequence of the wild type HP1 chromodomain, residues 22-76 and ligation sequence with ligation site highlighted in red (M38C). Residues located in aromatic cage are highlighted in blue.

Ligation fragments were synthesized using traditional solid phase peptide synthesis (SPPS). Initial efforts to synthesize both peptide fragments resulted in the aspartimide formation, which is one of the most common undesirable side reactions in Fmoc-SPPS (Figure 6.5).[11] Aspartimide formation is the result of nucleophilic attack from the amide nitrogen of the preceding residue onto the β-carboxy group of aspartic acid (Asp). Normally the β-carboxy group on the side chain of Asp is protected as an ester, thereby behaving as a leaving group in this side reaction. The side reaction can be detected as the $[M-18]^+$ ion by mass spectrometry.

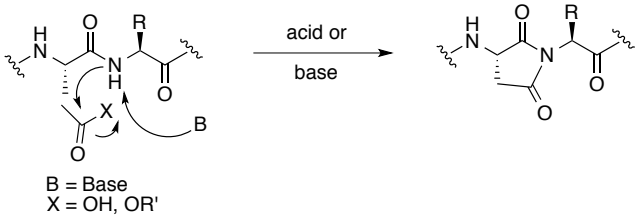

Figure 6.5 Formation of aspartimide.

A number of factors can affect aspartimide formation. The stability of the Asp residue can vary in the presence of various secondary and tertiary amines. Piperidine, the standard secondary amine applied in the removal of Fmoc in SPPS, gives rise to a considerably higher percentage of aspartimide formation. By using the weaker secondary amine, piperazine, we were

able to help minimize aspartimide formation. Furthermore, the addition of N-hydroxylamine-based additives can contribute to the prevention of aspartimide formation.[12] The acidic properties of N-hydroxylamines help to prevent aspartimide formation through acidic buffering of the solution. The deprotonation of the Asp-X amide backbone is the critical step in cyclization that leads to aspartimide formation. Therefore, when HOBt is used as an additive, conversion to its anion by the base decreases the amount of negatively charged amide backbone nitrogen, which is responsible for aspartimide formation. By using a combination of piperazine and HOBt as an additive in the Fmoc deprotection step, we were able to successfully avoid aspartimide formation.

The fragment containing the N-terminal Cys residue was synthesized easily on CLEAR-amide resin using the Thuramed Tetras synthesizer. The thioester fragment however, proved to be more difficult to synthesize. Similar to the procedure described in Chapter 3, the thioester fragment was synthesized on 2-chlorotrityl chloride resin and elongated using Fmoc-SPPS. For conversion to the thioester, the fully protected peptide must be cleaved from the resin, converted to the thioester, and then deprotected. The cleavage of peptide from the resin proved to be problematic, however. Initially, the fully protected peptide was cleaved 3x 3 minutes with 3 mL of 1% TFA in CH_2Cl_2 followed by formation of the thioester using PyBOP, DIPEA and methylthioglycolate in DMF. Under these cleavage conditions, the resulting product was less polar than expected based on its retention time by HPLC and had a mass significantly larger than expected. It was believed that the *tert*-butyl groups were being deprotected from the carboxylic group of aspartic acid and glutamic acid under the TFA conditions for cleavage. The decrease in polarity was then hypothesized to result due to subsequent conversion of these free carboxylic acid groups to thioesters. In order to avoid this side product, the cleavage conditions were

switched to a milder 1:1:8 mixture of acetic acid, trifluoroethanol and dichloromethane. Under these cleavage conditions however, the protected peptide was insoluble in DMF as well as CH_2Cl_2 and therefore unable to be converted to the thioester. We believed that cleavage with TFA aided in solubility due to the formation of TFA salts in the peptide. To overcome the solubility problem, the peptide was cleaved for shorter periods at 5x 1 minute with 3 mL of 1% TFA in CH_2Cl_2. The shorter washing times ensured that the *tert*-butyl groups remained on the carboxyl groups of the aspartic acid and glutamic acid side chains, resulting in good yield of the desired peptide thioester with no evidence for the higher mass products that had been obtained initially.

C. Future Directions

Following synthesis of both fragments, the native chemical ligation can be carried out.[13] This will be done under buffered conditions at pH 7 in the presence of 4-mercaptophenylacetic acid (MPAA), which acts as a thiol catalyst, and tris(2-carboxyethyl)phosphine (TCEP). TCEP is added as a reducing agent to prevent the formation of disulfides during the native chemical ligation. After the conditions have been determined for the native chemical ligation, we can proceed with the incorporation of fluorinated amino acids into the sequence of the HP1 chromodomain.

The aromatic amino acids present in the aromatic pocket of the HP1 chromodomain are two tyrosine residues and a tryptophan. We are working to incorporate fluorinated derivatives of tryptophan as well as fluorinated derivatives of tryrosine (Figure 6.6). However, there are concerns that the progressive fluorination of tyrosine would perturb the pK_a of the phenolic hydrogen. If the pK_a is altered, progressive fluorination may affect more than just the cation-π

ability of the aromatic ring. To avoid this, we hope that methyltyrosine would serve as a viable substitute (Figure 6.6).

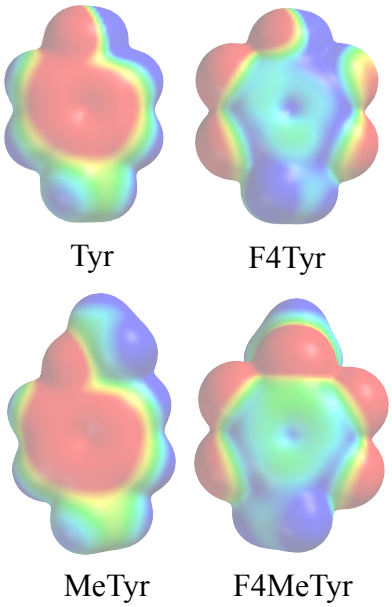

Figure 6.6 Electrostatic potential maps comparing Tyr and MeTyr as well as fluorinated Tyr and MeTyr. Red regions represent the most electron-rich regions of a molecule and blue regions of a map are the most electron-poor regions of a molecule.

Following synthesis of the fluorinated derivatives of the extensive structural and functional studies will be performed. Structural studies can be performed using circular dichroism to monitor the folding of the peptide. And finally, binding studies will be done with all derivatives of the HP1 chromodomain to probe the effect of fluorination on the cation-π ability of the aromatic pocket in binding to trimethyllysine. In collaboration with the Frye lab, we hope to use the information gained in the functional studies for the development of inhibitors for this interaction.

D. Experimental Procedure

Solvents were of HPLC or reagent grade quality and purchased commercially from Sigma-Aldrich. Starting materials were purchased commercially from Fisher Scientific unless otherwise noted and used without further purification.

Synthesis thioester fragment 9-Fluorenylmethoxycarbonyl (Fmoc) protected amino acids and HBTU and HOBt were purchased from Advanced Chemtech. PyBOP was purchased from Novabiochem. 2-Chlorotritylchloride resin was purchased from ChemPep. The glycine was preloaded onto 2-chlorotrityl resin (0.1 mmol). Resin was activated with 1mL of thionyl chloride in 9 mL of CH_2Cl_2 for 1 hour in a peptide flask and bubbled with nitrogen. The glycine was loaded onto the resin with DIPEA (8 equiv) and Fmoc-glycine (2 equiv) in CH_2Cl_2 for one hour, two times while being bubbled with nitrogen. The remaining portion of the peptide was synthesized by automated solid-phase synthesis on a Thuramed Tetras peptide synthesizer (version X94) using Fmoc protected amino acids. Activation of amino acids was performed with HBTU and HOBT in the presence of DIEPA in DMF. Peptide deprotection was carried out in 2% piperazine and 2% HOBt in DMF for approximately 15 min. Double coupling cycles (60 min each) were used for each amino acid coupling step. Peptides were manually acetylated at the N-terminus with 5% acetic anhydride and 6% lutidine in DMF for 30 min. Cleavage of the peptide from the resin was performed in 3 mL of 1% TFA in dry dichloromethane at five times for 1 minute. The filtrate was concentrated by rotary evaporation to obtain the cleavage product as a yellow oil. The product was dried under high vacuum and then was dissolved in 3.6 mL of dry DMF. To that solution was added 40 μL N,N-Diisopropylethylamine (DIPEA), 22 μL methyl thioglycolate, 86 mg PyBOP and the reaction was stirred under N_2 at room temperature for 1 h. DMF was removed under high vacuum to obtain yellow oil. After formation of the thioester the

peptide was globally deprotected using 95:2.5:1:2.5 trifluoroacetic acid (TFA): water: triisopropylsilane (TIPS): ethanedithiol (EDT) for 4 h. TFA was evaporated and cleavage products were precipitated with cold ether. The peptide was extracted into water and lyophilized. It was then purified by reverse-phase HPLC, using a C-18 semipreparative column and a gradient of 0 to 100% B over 50 min, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. After purification, the peptide was lyophilized to powder and identified with ESI-TOF mass spectroscopy.

N-terminal cysteine fragment Peptide was synthesized by automated solid-phase synthesis on a Thuramed Tetras peptide synthesizer (version X94) on CLEAR-amide resin (0.1 mmol) purchased from Advanced Chemtech using Fmoc protected amino acids. Activation of amino acids was performed with HBTU and HOBT in the presence of DIEPA in DMF. Peptide deprotection was carried out in 2% piperazine and 2% HOBt in DMF for approximately 15 min. Double coupling cycles (60 min each) were used for each amino acid coupling step. Peptide was globally deprotected and cleaved from the resin using 95:2.5:1:2.5 trifluoroacetic acid (TFA): water: triisopropylsilane (TIPS): ethanedithiol (EDT) for 4 h. TFA was evaporated and cleavage products were precipitated with cold ether. The peptide was extracted into water and lyophilized. It was then purified by reverse-phase HPLC, using a C-18 semipreparative column and a gradient of 0 to 100% B over 50 min, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. After purification, the peptide was lyophilized to powder and identified with ESI-TOF mass spectroscopy.

References

1. Daze, K. D.; Hof, F. The cation-π interaction at protein-protein interaction interfaces: developing and learning from synthetic mimics of proteins that bind methylated lysines. *Acc. Chem. Res.* **2013**, *46,* 937-945.

2. Nielson, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D. Structure of the HP1 chromodomain bound to histone H3 methylated lysine 9. *Nature* **2002**, *416,* 103-107.

3. Mellor, J. *Cell* **2006**, *126,* 22-24.

4. Sims, R. J.; Reinberg, D. *Genes Dev.* **2006**, *20,* 2779-2786.

5. Dougherty, D. A. The cation-π interaction. *Acc. Chem. Res.* **2013**, *46,* 885-893.

6. Mecozzi, S.; West, A. P.; Dougherty, D. A. Cation-π interactions in simple aromatics. Electrostatics provide a predictive tool. *J. Am. Chem. Soc.* **1996**, *118,* 2307-2308.

7. Tavares, X. D.; Blum, A. P.; Nakamura, D. T.; PUskar, N.; L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Variations in binding among several agonists at two stoichiometris of the neuronal, α4β2 nicotinic receptor. *J. Am. Chem. Soc.* **2012**, *134,* 11474-11480.

8. Beene, D. L; Brandt, G. S.; Zhong, W. G.; Zacharias, N. M.; Lester, H. A.; Dougherty, d. A. Cation-π interactions in ligand recognition by serotoergic (5-HT_{3A}) and nicotinic acetylcholine receptros: The anamolous binding properties of nicotine. *Biochemistry* **2002**, *41,* 10262-10269.

9. Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. From ab initio quantum mechanics to molecular neurobiology: A cation-π binding site in the nicotinic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95,* 12088-12093.

10. Schwarzer, D.; Hackenberger, C. P. R. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angew. Chem. Int. Ed.* **2008**, *47,* 10030-10074.

11. Ruczynski, J.; Lewandowska, B.; Mucha, P.; Rekowski, P. Problem of aspartimide formation in Fmoc-based solid-phase peptide synthesis using Dmab group to protect side chain of aspartic acid. *J. Pept. Sci.* **2008**, *14,* 335-341.

12. Subiros-Funosas, R.; El-Faham, A.; Albericio, F. Aspartimide formation in peptide chemistry: occurrence, prevention strategies and the role of *N*-hydroxylamines. *Tetrahedron* **2011**, *67,* 8595-8606.

13. Kent, S. B. H.; Total chemical synthesis of proteins. *Chem. Soc. Rev.* **2009**, *38,* 338-351.

BIBLIOGRAPHY

Allfrey, V. G.; Faulkner, R.; Mirsky, A. E., Acetylation and Methylation of Histones and Their Possible Role in the Regulation of RNA Synthesis. *Biochemistry* **1964**, *51*, 786-794.

Anderson, S.; Anderson, H. L.; Sanders, J. K. M. *Acc. Chem. Res.* **1993,** *26,* 469.

Andrianov, A. M. *Mol. Biol.* **1999**, *33,* 534-538.

Bagaut, A.; Balasubramanian, S. A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **2008,** *47,* 689-697.

Balagurumoorthy, P; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic Acids Res.,* **1992**, *20*, 4061.

Balasubramanian, S.; Neidle, S. G-Quadruplex nucleic acids as therapeutic targets *Curr. Opin. Chem. Biol.* **2009,** *13*, 345−353.

Bannister, A. J.; Zegerman, P.; Partidge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T., Selective Recognition of Methylated Lysine 9 on Histone H3 by the HP1 Chromo Domain. *Nature* **2001**, *410*, 120-124.

Beene, D. L; Brandt, G. S.; Zhong, W. G.; Zacharias, N. M.; Lester, H. A.; Dougherty, D. A. Cation-π interactions in ligand recognition by serotoergic ($5\text{-}HT_{3A}$) and nicotinic acetylcholine receptros: The anamolous binding properties of nicotine. *Biochemistry* **2002**, *41,* 10262-10269.

Bernard, P.; Maur, J.-F.; Partidge, J. F.; Genier, S.; Javerzat, J.-P.; Allshire, R. C.; Requirement of Heterochromatin for Cohesion at Centromeres. *Science* **2001**, *294*, 2539-2542.

Bochman, M. L.; Paeschke K.; Zakian V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nature Rev. Genetics* **2012,** *13,* 770-780.

Brooks, T. A.; Hurley, L. H. The role of supercoiling in transcriptional control of MYC and its importanct in molecular therapeutics. *Nat. Rev. Cancer* **2009**, *9,* 849-861.

Brooks, T. A.; Kendrick, S.; Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **2010,** *277,* 3459-3469.

Burge, S.; Parkinson, G. N.; Hazel P.; Todd, A. K.; Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006,** *34,* 5402-5414.

Busch, D. H. *Incl. Phenom. Mol. Recognit. Chem.* **1992**, *12,* 389.

Campbell, N. H.; Parkinson, G. N.; Reszka, A. P.; Neidle, S. Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc.* **2008**, *130,* 6722-6724.

Capra, J. A.; Paeschke, K.; Singh, M.; Zakian, V. A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae. Plos Comput. Biol.* **2010,** *6,* e1000861

Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry Part A,* Plenum Press: New York, 1990, p. 451.

Casals, J.; Debethune, L.; Alvarez, K.; Risitano, A.; Fox, K. R.; Grandas, A.; Pedroso, E. Directing quadruplex-stabilizing drugs to the telomere: synthesis and properties of acridine-oligonucleotide conjugates. *Bioconjugate Chem.* **2006,** *17,* 1351-1359.

Chen, Z.; Zang, J; Whetstine, J.; Hong, X.; Davrazou, F.; Kutateladze, T. G.; Simpson, M.; Mao, Q.; Pan, C. H.; Dai, S.; *et al.* Structural insights into histone demethylation by JMJD2 family members. *Cell* **2006**, *125,* 691-702.

Corbett, P.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. –L.; Sanders, J. K. M.; Otto, S. Dynamic combinatorial chemistry. *Chem. Rev.* **2006,** *106,* 3652-3711.

da Silva, M. W. NMR methods for studying quadruplex nucleic acids. *Methods,* **2007**, *43,* 264-277.

Daniel, J. A.; Pray-Grant, M. G.; Grant, P. A. Effector proteins for methylated histones: An expanding family. *Cell Cycle* **2005**, *4*, 919-926.

Daze, K. D.; Hof, F. The cation-π interaction at protein-protein interaction interfaces: developing and learning from synthetic mimics of proteins that bind methylated lysines. *Acc. Chem. Res.* **2013**, *46,* 937-945.

Dingley A. J.; Peterson R. D.; Grzesiek, S.; Feigon, J.; Characterization of the cation and temperature dependence of DNA quadruplex hydrogen bond properties using high-resolution NMR. *J. Amer. Chem. Soc.* **2005,** *127,* 1446-14472.

Dougherty, D. A. The cation-π interaction. *Acc. Chem. Res.* **2013**, *46,* 885-893.

Eddy, J.; Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **2006,** *34,* 3887-3896.

Egger, G.; Liang, G.; Aparicio, A.; Jones, P. A., Epigenetics in Human Disease and Prospects for Epigenetic Therapy. *Nature* **2004**, *429*, 457-463.

Emmel, E. A.; Verweij, C. L.; Durand, D. B.; Higgins, K. M.; Lacy, E.; Crabtree, G. R. *Science,* **1989***, 246,* 1617-1620.

Fernandes, P. A.; Ramos, M. J. *J. Chem.-Eur. J.* **2004**, *10,* 257.

Fiscle, W.; Wang, Y.; Allis, C. D. Histone and chromatin cross-talk. *Curr. Opin. Cell. Biol.* **2003**, *15,* 172-183.

Freidinger, R. M.; Veber, D. F.; Perlow, D. S.; Brooks, J. R. *Science,* **1980**, *210,* 656-658.

Ghosh, S.; Ingerman, L. A.; Frye, A. G.; Lee, S. J.; Gagne, M. R.; Waters, M. L. Dynamic cyclic thiodepsipeptide libraries from thiol-thioester exchange. *Org. Lett.* **2010***, 12,* 1860-1863.

Gilbert, H. F. *J. Biol. Chem.* **1997,** *272,* 29399.

Giuseppone, N.; Lehn, J. –M. *J. Am. Chem. Soc.* **2004,** *126,* 11448.

Gonzalez, V.; Hurley, L. H. The c-MYC NHE III_1: function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50,* 111-129.

Guedin, A.; Gros, J.; Alberti, P.; Mergny, J. L. How long is too long? Effects of loops size on quadruplex stability. *Nucleic Acids Res.* **2010,** *38,* 7858-7868.

Haider, S. M.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2003**, *326,* 117.

Hardin, C. C.; Perry A. G.; White, K. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **2000,** *56,* 147-194.

Harrison, R. J.; Reszka, A. P.; Haider, S. M.; Romagnoli, B.; Morrell, J.; Read, M. A.; Gowan, S. M.; Incles, C. M.; Kelland, L. R.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2004**, *14,* 5845.

Hazel, P.; Parkinson, G. N.; Neidle, S. Predictive modeling of topology and loop variations in dimeric DNA quadruplex structures. *Nucleic Acid Res.* **2006,** *34,* 2117-2127

Henderson, E. *et al.* Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **1987,** *51,* 899-908.

Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex forming sequences in *Saccharomyces cerevisiae. Nucleic Acids Res.* **2008,** *33,* 144-156.

Horton, D. A.; Bourne, T. T. Smythe, M. L. Exploring privileged structures: The combinatorial synthesis of cyclic peptides. *J. Comput. Aided Mol. Design,* **2002**, *16,* 415-430.

Hoss, R.; Vogtle, F. *Angew. Chem. Int. Ed. Engl.* **1994**, *33,* 375.

Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. Recognition of trimentyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl. Acad. Sci.* **2007**, *104,* 11184-11188.

Huppert, J. L. Structure, location and interactions of G-quadruplexes *FEBS Journal* **2010,** *277,* 3452-4358.

Huppert, J. L.; Balasubramanian, S. Prevalance of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33,* 2908-2916.

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

Jacobs, S. A.; Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **2002***, 295,* 2080-2083.

Jin, R.; Gaffney, B. L.; Wang, R. A.; Jones, A.; Breslauer, K. J.; *Proc. Natl. Acad. Sci.,* 1992, 89, 8832.

Katz, B; Liu, B.; Cass, R.; Structure based design tools: structural and thermodynamic comparison with biotin of a small molecule that binds to streptavidin with micromolar affinity. *J. Am. Chem. Soc.* **1996,** *118,* 7914-7920.

Kent, S. B. H.; Total chemical synthesis of proteins. *Chem. Soc. Rev.* **2009**, *38,* 338-351.

Khorasanizadeh, S. *Cell* **2004**, *116,* 259-272.

Kouzarides, T. Chromatine Modifications and Their Function. *Cell* **2007**, *128,* 293-705.

Kratzschmar, J.; Krause, M.; Marahiel, M. A.; *J. Bacteriol.* **1989**, *171,* 5422-529.

Ladame, S.; Balasubramanian, S. Templated ligand assembly by using G-quadruplex DNA and dynamic covalent chemistry. *Angew. Chem. Intl. Ed.* **2004**, *43,* 1143-1146.

Lane, A. N.; Chaires, B. J.; Gray, R. D.; Trent, J. O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008**, *36,* 5482-5515.

Langan, T.A. Histone Phosphorylation: Stimulation by Adenosine 3',5'-Monophosphate. *Science* **1968**, *162,* 579-580.

Larsson, R.; Pei, Z. C.; Ramstrom, O. *Angew. Chem. Int. Ed.* **2004**, *43,* 3716.

Lehr, C. –M. *et al. Pharm. Res.* **2006,** *23,* 1031.

Lu, M.; Guo, Q.; Kallenbach, N. R. *Biochemistry*, **1993**, *32*, 598.

Lu, Q.; Qiu, X.; Hu, N.; Wen, H.; Su, Y.; Richardson, B. C., Epigenetics, Disease, and Therapeutic Interventions. *Ageing Res Rev* **2006**, *5*, 449-467.

Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. R., Crystal Structure of the Nucleosome Core Particle at 2.8Å Resolution. *Nature* **1997**, *389*, 251-160.

Mahadevan, L. C.; Willis, A. C.; Barratt, M. J., Rapid Hisonte H3 Phosphorylation in Response to Growth Factors, Phorbol Ester, Okadaic Acid, and Protein Synthesis Inhibitors. *Cell* **1991**, *65*, 775-783.

Marino-Ramirez, L.; Hsu, B.; Baxevanis, A. D.; Landsman, D. The histone database: a comprehensive resource for histones and histone fold-containing proteins. *Proteins,* **2006**, *62,* 838-842.

Martin, C.; Zhang, Y., The Diverse Functions of HIstone Lysine Methylation. *Nat Rev Mol Cell Biol* **2005**, *6*, 838-849.

Masiero, S.; Trotta, R.; Pieraccini, S.; De Tito, s.; Perone, R.; Randazzo, A.; Spada, G. P.; *Org. Biol. Chem.* **2010**, *8,* 2683-2692.

Maurer-Stroh, S.; Dickens, N. J.; Hughes-Davies, L.; Kouzarides, T.; Eisenhaber, F.; Ponting, C. P.; The tudor domain 'royal family': tudor, plant agenet, chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **2003**, *28,* 69-74.

Mecozzi, S.; West, A. P.; Dougherty, D. A. Cation-π interactions in simple aromatics. Electrostatics provide a predictive tool. *J. Am. Chem. Soc.* **1996**, *118,* 2307-2308.

Mellor, J. *Cell* **2006**, *126,* 22-24.

Monchaud, D.; Teulade-Fichou, M. P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* **2008**, *6*, 627-636.

Mootz, H. D.; Marahiel, M. A. *J. Bacteriol* **1997**, *197,* 6843-6850.

Murat, P.; Singh, Y.; Defrancq, E. Methods for investigating G-quadruplex DNA/ligand interactions *Chem. Soc. Rev.* **2011,** *40,* 5293-52307.

Neidle, S. Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **2010,** *277,* 1118-1125.

Nielson, P. R.; Nietlispach, D.; Mott, H. R.; Callaghan, J.; Bannister, A.; Kouzarides, T.; Murzin, A. G.; Murzina, N. V.; Laue, E. D. Structure of the HP1 chromodomain bound to histone H3 methylated lysine 9. *Nature* **2002**, *416,* 103-107.

Oganesian, L; Graham, M. E.; Robinson, P. J.; Bryan, T. M. Telomerase recognizes G-quadruplex and linear DNA as distinct substrates. *Biochemistry* **2007,** *46,* 11279-11290.

Pardon, J. F.; WIlkins, M. H. F.; Richards, B. M., Super-Helical Model for Nucleohistone. *Nature* **1967**, *215*, 508-509.

Perry, P. J.; Read, M. A.; Davies, R. T.; Gowan, S. M.; Reszka, A. P.; Wood, A. A.; Kelland, L. R.; Neidle, S. *J. Med. Chem.* **1999***, 42,* 2679.

Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. Structure of two intramolecular G-quadruplexes formed by natural human telomerase sequences in K+ solution. *Nucleic Acids Res.,* **2007**, *35,* 6517-6525.

Phan, A.T.; Kuryavyi, V.; Patel, D. J. *Curr. Opin. Struct. Biol.* **2006**, *16*, 288.

Platero, J. S.; Hartnett, T.; Eissenberg, J. C., Functional Analysis of the Chromo Domain of HP1. *EMBO J* **1999**, *14*, 3977-3986.

Qin, Y.; Hurley, L. H. Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochemie* **2008,** *90,* 1149-1171.

Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M. Parkinson, G. N.; Todd, A. K.; Ladame, S. Balasubramanian, S.; Neidle, S. Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **2005**, *127,* 10584-10589.

Rawal, P. *et al.* Genome-wide prediction of G4 DNA as regulatory motifs; role in *Escherichia coli* global regulation. *Genome Res.* **2006,** *16,* 644-655.

Redman, J. E.; Granadino-Roldan, J. M.; Schouten, J. A.; Ladame, S.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. Recognition and discrimination of DNA quadruplexes by acridine-peptide conjugates. *Org. Biomol. Chem.* **2009***, 7,* 76-84.

Ruczynski, J.; Lewandowska, B.; Mucha, P.; Rekowski, P. Problem of aspartimide formation in Fmoc-based solid-phase peptide synthesis using Dmab group to protect side chain of aspartic acid. *J. Pept. Sci.* **2008**, *14,* 335-341.

Santon-Rosa, H.; Schneider, R.; Bannister, A. J.; Sherriff, J.; Berstein, B. E.; Tolga Emre, N. C.; Schreiber, S. L.; Mellow, J.; Kouzarides, T., Active Genes are Trimethylated at K4 of Histone H3. *Nature* **2002**, *419*, 407-411.

Schotta, G.; Ebbert, A.; Krauss, V.; Fischer, A.; Hoffmann, J.; Rea, S.; Jenuwein, T.; Dorn, R.; Reuter, G., Central Role of Drosophila SU(VAR)3-9 in Histone H3-K9 Methylation and Heterochromatic Gene Silencing. *EMBO J* **2002**, *21*, 1121-1131.

Schultes, C. M.; Guyn, B.; Cuesta, J.; Neidle, S. *Bioorg. Med. Chem. Lett.* **2004***, 14,* 4347.

Schwarzer, D.; Hackenberger, C. P. R. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angew. Chem. Int. Ed.* **2008**, *47,* 10030-10074.

Sen, D.; Gilbert, W. Formation of parallel four-stranded complexes by gunine-rich motifs in DNA and its implications for meiosis. *Nature* **1988,** *334,* 364-366.

Shay, J. W.; Wright, W. E. Role of telomeres and telomerase in cancer. *Seminars Cancer Biol.* **2011,** *21,* 349-353.

Sherman Durai, C. R.; Harding, M. M. Targeting nucleic acids using dynamic combinatorial chemistry. *Aust. J. Chem.* **2011**, *64,* 671-680.

Shirude, P. S.; Okumus, B.; Ying, L. M.; Ha, T.; Balasubramanian, S. Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene *c-kit. J. Am. Chem. Soc.* **2007,** *129,* 7484-7485.

Shirude, P. S.; Okumus, B.; Ying, L.; Ha, T.; Balasubramanian, S. Single-molecular conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit. *J. Am. Chem. Soc.* **2007,** *129,* 7484-7485.

Sims, R. J.; Reinberg, D. *Genes Dev.* **2006**, *20,* 2779-2786.

Smith, F.W.; Feigon, J. *Nature*, **1992**, *356,* 164.

Subiros-Funosas, R.; El-Faham, A.; Albericio, F. Aspartimide formation in peptide chemistry: occurrence, prevention strategies and the role of *N*-hydroxylamines. *Tetrahedron* **2011**, *67,* 8595-8606.

Sun, D.; Hurley, L. H.; The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.* **2009,** *52,* 2863-2874.

Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. *J. Med. Chem.* **1997,** *40*, 2113.

Sundquist, W. I.; Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989,** *342,* 825-829.

Tate, P. H.; Bird, A. P., Effects of DNA Methylation on DNA-BOnding Proteins and Gene Expression. *Curr Opin Genet Dev* **1993**, 226-231.

Tavares, X. D.; Blum, A. P.; Nakamura, D. T.; PUskar, N.; L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Variations in binding among several agonists at two stoichiometris of the neuronal, α4β2 nicotinic receptor. *J. Am. Chem. Soc.* **2012**, *134,* 11474-11480.

Thomas, F.; Koller, T. H.; Klug, A., Involvement of Histone H1 in the Organization of the Nucleosome and of the salt-Dependent Superstructures of Chromatin. *J Cell Biol* **1979**, *83*, 403-427.

Todd, A. K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005,** *36,* 144-156.

Troitskaya, L. A.; Kodadek, T. Peptides as modulators of enzymes and regulatory proteins. *Methods,* **2004,** *32,* 406-415.

Tsukada, Y.-i.; Fang, J.; Erdjument-Bromage, H.; Warren, M. E.; Borchers, C. H.; Tempst, P.; Zhang, Y., Histone Demethylation by a Family of JmjC Domain-Containing Proteins. *Nature* **2006**, *439*, 811-816.

Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *Science,* **2009**, *325,* 73-77.

van Dijk, L.; Bobbert, P. A.; Spano, F. C. *J. Phys. Chem. B,* **2010**, *114*, 817.

van Mourik, T.; Dingley, A. J. Characterization of the monovalent ion position and hydrogen-bond network in guanine tetrads by DFT calculations of NMR parameters. *Chemistry* **2005,** *7,* 6064-6079.

Venter, J. C.; al, e., The Sequence of the Human Genome. *Science* **2001**, *291*, 1304-1351.

Whetstine, J. R.; Nottke, A.; Lan, F.; Smolikov, S.; Chen, Z.; Spooner, E.; Li, E.; Zhang, G.; Colaiacovo, M.; Shi, Y., Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **2006**, *125*, 467-481.

Whitney, A. M.; Balasubramanian, S. Targeting nucleic acid secondary structures with polyamides using an optimized dynamic combinatorial approach. *Angew. Chem. Intl. Ed.* **2005**, *44,* 5736-5739.

Widom, J., Toward a Unified Model of Chromatin Folding. *Annu Rev Biophys Bio* **1989**, *18*, 365-395.

Wieter, J. J-L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. Exploring the differential recognition of DNA G-quadruplex targets by small molecules using dynamic combinatorial chemistry. *Angew. Chem. Intl. Ed.* **2008**, *47,* 2677-2680.

Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. Monovalent cation induced structure of telomeric DNA: the G-quartet model. *Cell* **1989,** *59,* 871-880.

Woll, M. G.; Gellman, S. H. *J. Am. Chem. Soc.* **2004**, *126,* 11172-11174.

Wong, H. M.; Payet, L.; Huppert, J. L. Function and targeting of G-quadruplexes *Curr. Opin. Mol. Ther.* **2009,** *11,* 146-155.

Wu, Y.; Brosh, R. B. G-Quadruplex nucleic acids and human disease. *FEBS Journal* **2010,** *277,* 3470-3488.

Xiao, B.; Jing, C.; Wilson, J. R.; Walker, P. A.; Vasisht, N.; Kelly, G.; Howell, S.; Taylor, I. A.; Blackburn, M.; Gamblin, S. J., Structure and Catalytic Mechanism of the Human Histone Methyltransferase SET7/9. *Nature* **2003**, *421*, 652-656.

Xodo, L. E.; Cogoi, S. G-Quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nuc. Acids. Res.* **2006**, *34,* 2536-2549.

Xue, L.; Ranjan, N.; Arya, D. P. Synthesis and Spectroscopic Studies of the Aminoglycoside (Neomycin)-Perylene Conjugate Binding to Human Telomeric DNA *Biochemistry* **2011,** *50,* 2838-2849.

Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **1991,** *350,* 718-720.

Zakian, V. A. Telomeres: the beginning and ends of eukaryotic chromosomes. *Exp. Cell Res.* **2012,** *318,* 1456-1460.

Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. From ab initio quantum mechanics to molecular neurobiology: A cation-π binding site in the nicotinic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95,* 12088-12093.