Spring 5-18-2018

Development of LC-MS for the Identification and characterization of non-adjacent DNA Photoproduct Formation in G-quadruplex Forming Sequences

Claudia Posadas
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Chemistry

Development of LC-MS for the Identification and Characterization of Non-Adjacent DNA Photoproduct Formation in G-quadruplex Forming Sequences

by

Claudia G Posadas

A Thesis Presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

May 2018
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Dedication

This thesis is dedicated to my parents Sara and Felipe Posadas, my true heroes always and forever, my wonderful husband Juan A. Sanchez, my beautiful son Armando Sanchez-Posadas, my brother Juan F. Posadas, and all my family that supported me during this time: Sonia Posadas, Xochitl Posadas, Xavier Posadas, Giovanni Posadas, Alicia Ruiz-Duncan, Robert Duncan and Maria Ruiz. A special thanks to my friend Suzanne DeMarco for her guidance, her support and friendship of many years. I cannot overstate my gratitude to them, as my success is a direct product of their endless love, encouragement, and support.

Acknowledgments

It is my pleasure to thank those who made this thesis possible. I would like to express my sincere thanks to my wonderful advisor Professor John-Stephen A. Taylor. It is such an honor for me to work with him over the past 2 years at Washington University. I owe my deepest gratitude to him for his guidance, advice, critiques and constant encouragement. I thank Dr. Taylor for providing me the unique opportunity to work on this project and giving me ample room to explore research possibilities and work with his wonderful group. I would like to thank Dr. Robert Blankenship and Dr. Kevin Moeller for serving on my defense committee. I appreciate all of the advice and guidance I have received from Dr. Kevin Moeller during the past years.

I owe many thanks to Dr. Taylor's wonderful current and former members: Chen Lu, Biao Xiang, Dr. Kesai Wang and Innocent Harelimana for their fantastic friendship and unconditional help and for sharing their expertise with me during experimental designs and data analysis. I would also thank my colleagues at Pfizer Laura Bass, Keith Davis, Roberto Rodríguez, Olga Friese, Aparna Deora, and Justin Sperry for their support and guidance during the development and interpretation of new analytical methods which allowed me to finish this thesis.

I would also like to thank the faculty and staff members of the Department of Chemistry for their help over the last 5 years. I would specially like to acknowledge Dr. Rachel Dunn, Dr. Richard Loomis, Angie Mahon and Barbara Tessmer.

Claudia G Posadas

Washington University in St. Louis

May 2018

Abstract

Development of a New LC-MS Method for the Identification of *anti* Cyclobutane Pyrimidine Dimers Formed in G-quadruplex Forming Sequences

by

Claudia Posadas

Master's Degree in Chemistry

Washington University in St. Louis, 2018

Dr. John-Stephen A. Taylor, Chair

Ultraviolet light is well known to induce cyclobutane pyrimidine dimers (CPD) and pyrimidine (6–4) pyrimidone photoproducts in duplex DNA, which interfere with DNA replication and transcription. Recently, a new class of DNA photoproducts known as *anti* cyclobutanepyrimidine dimers have been discovered, which form in G-quadruplex forming sequences in solution. G-quadruplex structures have been proposed to form in human DNA telomeres and certain promoters *in vivo* but evidence for their existence has been lacking. Since *anti*-cyclobutante pyrimidine dimers have been shown to form in G-quadruplex forming sequences, their formation in irradiated human cells could be used to confirm the existence of G-

quadruplexes *in vivo*. Methods have been developed for assaying these DNA modifications, but it hasn't always been easy to confirm their identity or characterize their structures.

The goal of this thesis was to further investigate the formation of anti-thymine dimers in human telomeric sequences under various conditions known to favor or disfavor G quadruplex formation. Central to this goal was the development of a high resolution LC MS/MS method for separating and identifying all the photoproducts. Part one of this thesis, describes the development of a new LC method to better separate the CPD-containing nuclease P1 (NP1) digestion products of DNA and the utilization of the new method to analyze the NP1 products of UV irradiated telomeric DNA isolated with the previous HPLC method. The second part of this thesis focused on directly coupling the LC method to MS/MS mass spectrometry to simultaneously separate and identify the nuclease P1 digestion products of UV irradiated telomeric DNA in Li+. These products were then compared against the products formed in the presence of K+. In the course of these studies new hexameric digestion products corresponding to $(pTTA)_2$ were also identified which could contain two CPDs dimers.

Chapter One: Introduction

1. DNA structures

DNA is the core of genetic information in all living cells. It was in April 1953, when Watson and Crick published the Nature paper in which they illustrated the base pairing of DNA double helix (21). B-form DNA is the dominant right-handed helix in the biological state. DNA can also adopt a variety of unconventional non-B-form structures under a variety of conditions as well as non-helical secondary structures such as cruciform, hairpin structures and G-quadruplexes (19, 20). The DNA double-helical family members include A-DNA, B-DNA and Z-DNA, while other secondary structures such as cruciform retain the AT and GC base pairing between complementary strands, G-quadruplexes do not involve GC base pairs and their formation requires base pair disruption to allow the free G-rich strand to fold back upon themselves to form the four-stranded structure. The structures of G-quadruplex DNA are influence by several factors such as sequence, topology, temperature and protein binding (19). Extensive studies have been carried out to monitor the existence of these secondary structures using a wide variety of biochemical and biophysical techniques. The A-DNA structure is adopted by both double-stranded DNA and RNA or hybrid DNA-RNA duplex structures (19). The

chemical synthesis of defined oligonucleotide sequences, allowed the determination of high resolution crystal structure, the first of which showed the unexpected occurrence of a left-handed helical structure with alternating glycosidic bonds in *anti* and *syn* conformations. This was termed Z-DNA, which can be generated in regions of alternating GC at high ionic strength (31). Later, it has been proposed that Z-DNA could be involved in gene regulation.

1.1 The history of G-quadruplexes

The self-assembly of guanylic acid was first noted in the early twentieth century by the German chemist Bang who observed the gel formation at high guanosine 5'-monophosphate (GMP) concentrations and defined pH conditions (22). Later studies by Gellert of 5'-GMP fibers by X-ray diffraction showed that tetrameric guanine residues were arranged in a co-planar hydrogen-bonded array (Figure 1) (23). In 1978, Miles and Fraser showed that the stability of G-quartets depends on the positive ion that interacts with the central oxygen atom. Blackburn affirmed that repetitive guanine rich tracts related to the telomeric DNA sequence can form a G-quadruplex structure (24, 26). These studies suggested that certain G-rich DNA sequence may form in living cells. Strong evidence for the existence of G-quadruplex structures in vivo was shown by staining Stylonichia lemnae macronucleo with high affinity quadruplex specific antibodies Sty49 (47). To date

extensive studies have been carried out to explore the native state of unusual DNA secondary structures concerning their configuration, thermodynamic and kinetic stability under physiological conditions.

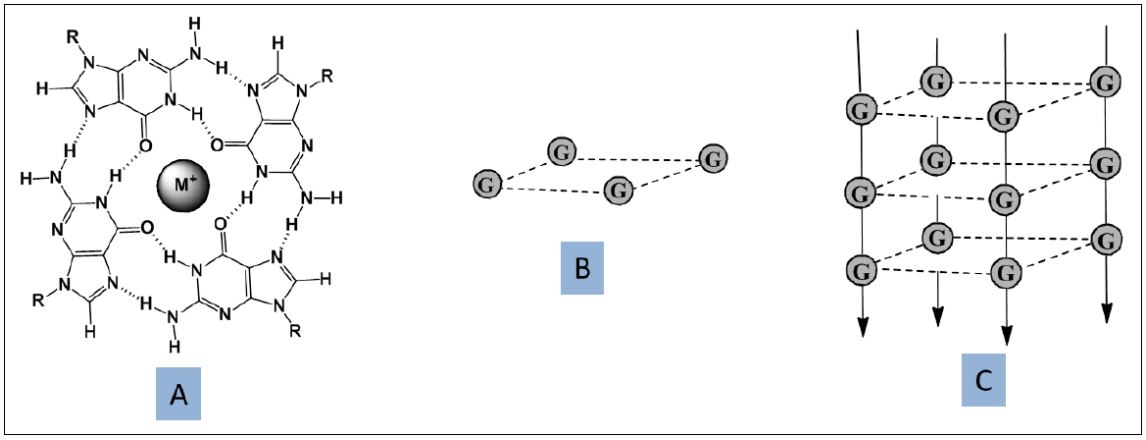


Figure 1: G quadruplex Structure A) chemical Structure of a G-quartet with a central metal ion B) G-quartet as a basic building block of a G-quadruplex C) Topology of intermolecular G-quadruplex consisting of three G-quartets

1.2 G-quadruplex Structures

As discussed above, guanine-rich DNA sequences can self-assemble to form four-stranded structures known as a G-quadruplexes under physiological conditions. These structures are composed of stacks of two or more guanosine quartets; a set of four guanine nucleotides associated with each other in a square Hoogsteen hydrogen-bonded array. Each guanine acts as donor and acceptor of two hydrogen bonds (N1 to O6 and N2 to N7). The association of N7 of guanine bases in the Hoogsteen paired G-quartet assembly protects them from methylation by dimethyl

sulfate (DMS); allowing a chemical discrimination of G-quadruplex formations and other DNA secondary structures.

The four strands constituting an inter- or intramolecular G-quadruplex structure can have four different arrangements. The four strands oriented in the same direction are designated parallel (Figure 2A); three strands oriented in the same direction and one strand oriented in the opposite direction designated (3+1) also known as Hybrid 1 (Figure 2B); two strands oriented in the same direction and two strands oriented in the opposite direction designated anti-parallel (Figure 2C); Two strands oriented opposite of each other and the other two strands oriented in the opposite direction (up-down-up-down) also designated anti-parallel (Figure 2D).

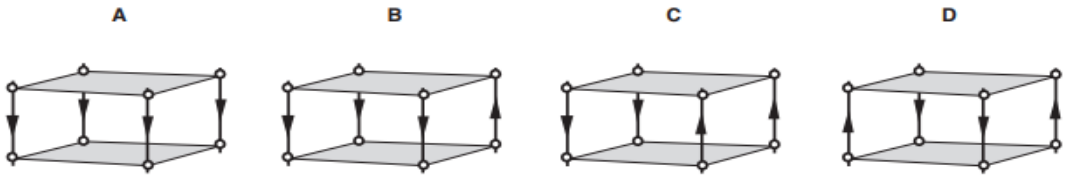


Figure 2: Four types of G-tetrad foundations: A) Parallel G-tetrad, B) (3+1) G-tetrad, C) antiparallel G-tetrad; D) Antiparallel G-tetrad.

Monovalent cations such as K^+ and Na^+ stabilize the structures by interacting with the central electronegative carbonyl O6 atoms of the G-quartet. The most stabilized G-quadruplex structure is generated by intercalating K^+ between two adjacent G-quartets (25). These stacked G-quartets overlap along the four stranded helical structure and the stacks are joined together by the sugar phosphate backbone (48). It is important to state that this configuration does not usually form in the

absence of metal cations (36). G quadruplex structures can be formed by the folding of G-rich single stranded DNA or association of two or four independent G-rich sequences.

1.3 Intramolecular G-quadruplex topology

G-Rich sequences with the potential to adopt intramolecular quadruplexes are comprised of four consecutive runs of guanines that are separated by three loop regions with different lengths and sequences. The non-participant bases in the intramolecular assembly of G-quadruplexes form loops at different lengths and sequences. All the G-quadruplexes have a set of core G-quartet structural features and the differences in these structures are the loops which differ in size and composition. These loops can connect two parallel or anti-parallel adjacent strands creating four possible loop arrangements. A number of studies have shown how loop sequence and length can regulate the structural conformation and stability of G-quadruplex structures (27, 28). These loops can connect two parallel or anti-parallel adjacent strands creating four possible loop arrangements (29, 30).

The type of G-quadruplex structure adopted by a sequence can have different topologies depending on the orientation of the strands and the glycosidic bond angles of the guanines that participate in the assembly of the G-quartet which may have a *syn* or *anti*-conformation. Among them the basket type, which is an

antiparallel-stranded intramolecular structure having one central diagonal loop and two edgewise loops. Hybrid types are a mixture of antiparallel/parallel stranded structures which are formed by different combination of loop arrangements. The hybrid structures have two lateral loops and one propeller loop (Figure 3).

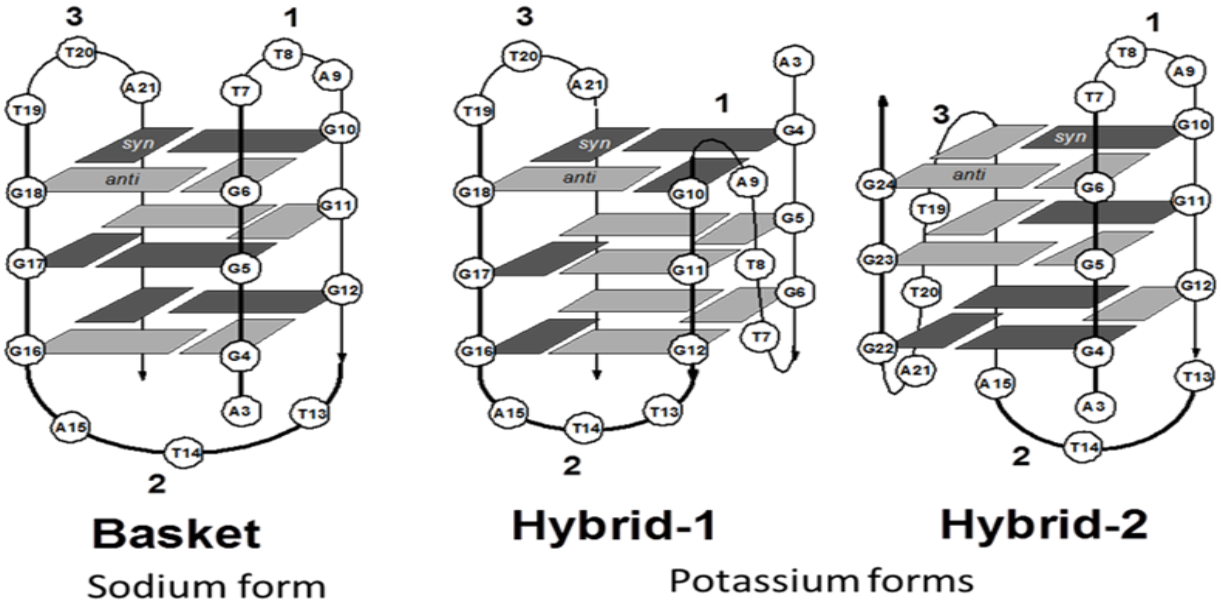


Figure 3: G-Quadruplexes are formed of stacks of square G-quartets, with guanines at each corner surrounding a positive ion.

G-quadruplex forming sequences can adopt many different structures under different ionic conditions. The nature of the monovalent cations can regulate the topology of unimolecular DNA quadruplexes, but usually does not affect the topology of bimolecular DNA quadruplexes (49). For G-quadruplexes it's been shown that an

anti-parallel arrangement appears to be favored in Na^+ solution while a parallel arrangement is generated in K^+ (31, 32).

1.4 G-quadruplex Polymorphism

The classical DNA double helix has two grooves; major and minor. G-quadruplexes have four grooves. The variability of quadruplex structures is related to the flexibility of the glycosidic bond. Attraction or repulsion between the sugar and the base of the nucleoside generates syn- and anti-conformations respectively (Figure 4).

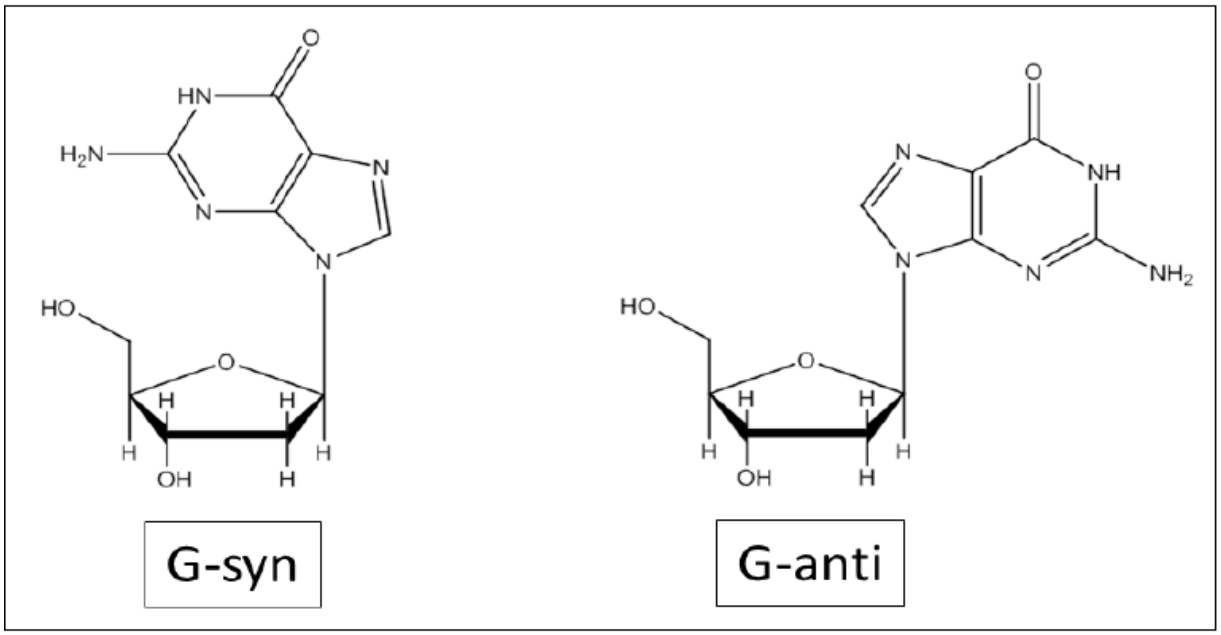


Figure 4: Orientation around the glycosidic bond, where the guanine bases adopt both syn and anti-conformations.

DNA exists as a double helix and the two strands are held together by Watson and Crick base pairing. G-quartet contains eight Hoogsteen hydrogen bonds in a cyclical arrangement, so there are on average two hydrogen bonds per base. Monovalent

cations especially K^+ and Na^+ but not Li^+ (34) are required for the G-quadruplex conformation and stability. The K^+ ion has a large molecular size which is stacked between two adjacent guanine quartets, holding them together and forming a stronger quartet than when Na^+ ion is used due to the smaller molecular size (35). Monovalent cations stabilize G-quadruplexes in the following order: K^+> Na^+> Rb^+> $NH4^+$>Cs^+>Li^+ (36). The oligodeoxynucleotide, d[AGGG(TTAGGG)3] (Tel22), adopts a basket structure in solutions containing Na^+, and in the presence of K^+, the predominant intracellular cation, the hybrid-type structures are favored over the basket (Figure 5).

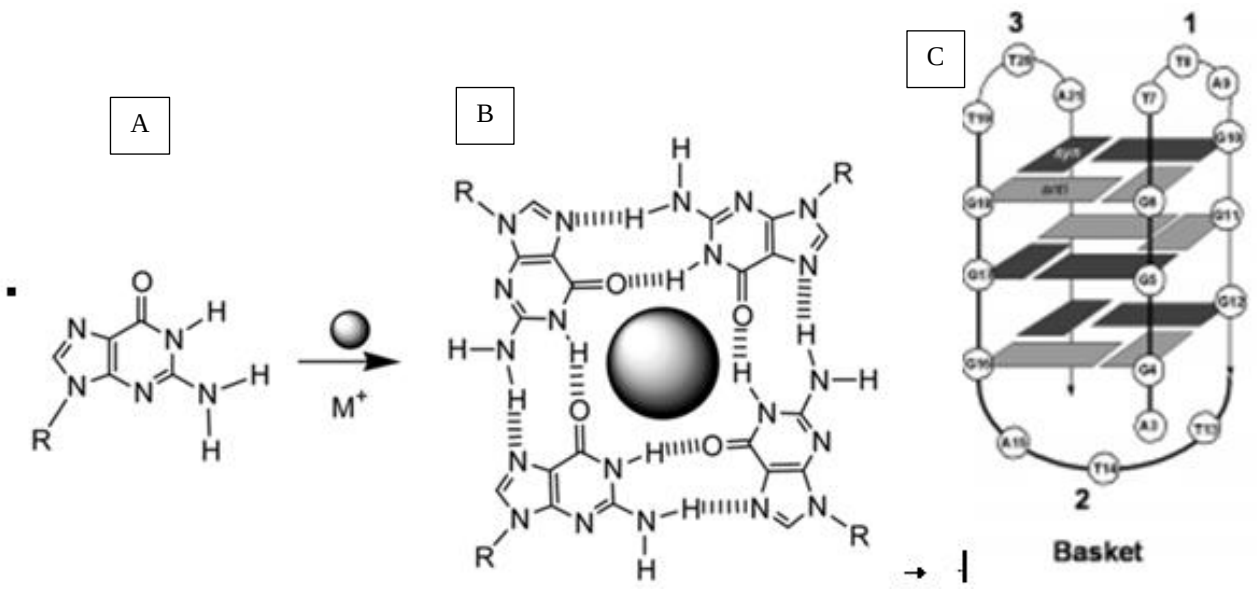


Figure 5: A) The building blocks of G-quadruplexes are G-quartets that arise from the association of four guanines into a cyclic arrangement stabilized by Hoogsten hydrogen bonding (N1–N6 and N2–N7). B) The planar G-quartets stack on top of one another, forming four-stranded helical structures. G-quadruplex formation is driven by monovalent cations such as Na+ and K+.

1.5 G-quadruplex forming sequences in vivo

Studies have revealed repetitive guanine sequences, with the potential to form G-quadruplexes within biologically important regions of the eukaryotic genome including telomeric DNA, oncogenic promoter sequences (38) and nuclease hypersensitive regions (33). The accumulation of these potential quadruplex-forming sequences throughout the genome raised the prospect that quadruplex structures may play an important role in cellular processes which include replication, transcription, translation and impeding telomerase activity (50) leading to the conclusion that these promoter regions can regulate gene expression at the

transcription levels. This implies that there may be a correlation between potential G-quadruplex structure sequences and specific gene function.

The telomere is a specialized nucleoprotein complex at the extreme ends of all eukaryotic chromosomes (24). The telomeres of eukaryotic chromosomes have unique structures that include short nucleotide sequences present as tandem repeats. The telomeric repeat sequence varies between different organisms but all have the general G-rich sequence $d[T_{1-3}\text{-}(T/A)\text{-}G_{3-4}]_4$. Telomeres stabilize the termini of linear chromosomes and protect them from exonuclease degradation and recombination and stop them from being recognized and processed as DNA breaks. In normal (non-cancerous) human somatic cells, telomeric DNA is composed of approximately 10 kilobase pairs of ds DNA sequences that contain tandem repeats of the sequence $(TTAGGGG)_n$ (26). These sequence repeats can form G-quadruplex structures through Hogsten hydrogen bonds in the presence of monovalent cations such as K^+ and Na^+.

In mammals the ends of linear chromosomes can form a protective cap-like structure to maintain chromosome integrity from damage by fragmentation and fusion. Telomeres make it possible for cells to divide and acts as a cellular marker for apoptosis when the sequence has been shortened to a critical length (Figure 6). Several capping proteins have been characterized involving DNA binding proteins (39, 40). Strong evidence for the existence of G-quadruplex *in vivo* comes from the

observations made of telomeres in the macronuclei of ciliated protozoa Stylonychia lemmae specifically bound to quadruplex-specific antibodies (47). The study demonstrated that antiparallel four stranded DNA structures were presented in the macronuclei (51).

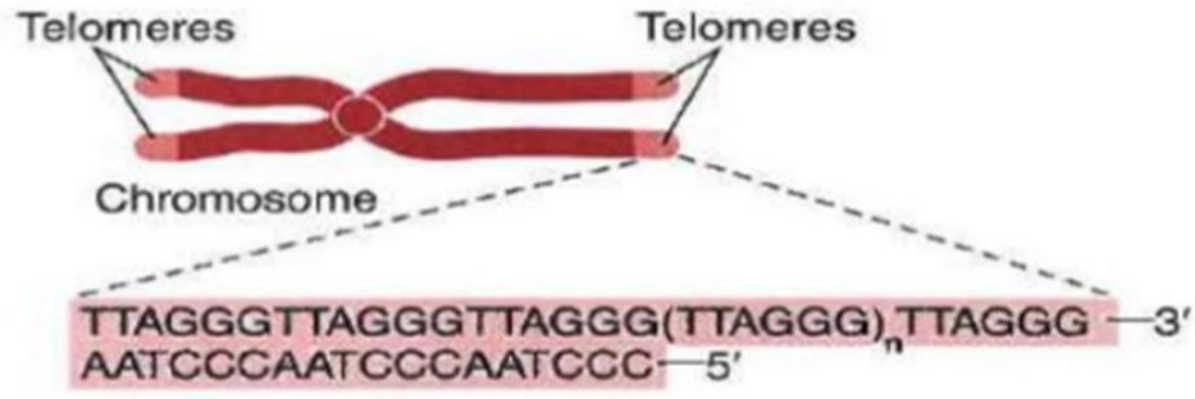


Figure 6: Telomere: Protective cap at the end of the chromosome strands.

A great deal of research has been focused on understanding the structural diversity in the human telomere G-quadruplex structure. Early studies show that the single repeat d(TTAGGGT) human telomere sequence can form a propeller parallel-stranded G-quadruplex in K^+ (Hybrid I) (41). The oligodeoxynucleotide, d[AGGG(TTAGGG)$_3$] (Tel22), adopts a basket structure in solutions containing Na^+, and in the presence of K^+ the hybrid-type structures are favored over the basket and chair structures (5). Tel26, d[AAA(GGGTTA)$_3$GGGAA] was found by NMR to adopt a mixture of Hybrid 1 and Hybrid 2 structures in K^+.

Other extensive studies have been done to characterize DNA secondary structures in the promoters of genes. Studies have identified the presence of G-quadruplex forming sequences in the promoter and gene regulatory regions of several proto-oncogenes (42). These proto-oncogenes are involved in growth and proliferation and proximal promoter regions are TATA-less and contain several G/C rich regions.

1.6 DNA Photoproduct Formation in Human G-quadruplexes Forming Sequences.

UVB irradiation overlaps the upper end of the DNA absorption spectra and is the range mainly responsible for skin cancer through direct photochemical damage

to DNA. The majority of the photoproducts arise from a photoreaction between two pyrimidines, the structure and stereochemistry of which depend on the conformation of the two bases at the time of irradiation. Irradiation of double stranded B form DNA under physiological conditions mainly produces *cys-syn* cyclobutane pyridine dimers (CPDs), pyrimidine (6-4) pyrimidone photoproducts, and their Dewar valance isomers (2, 52). Other rare UV photoproducts, including the spore photoproduct, TA*, A=A are also well known. If not repaired, the photoproducts may interfere with DNA replication or transcription (Figure 7). If bypassed by a DNA polymerase, they may lead to mutations. Non-adjacent photodimers of both intra-strand and inter-strand types are much rarer. Intra-strand-type non-adjacent dimer form when one or more nucleotides between two pyrimidines become extrahelical, owing to the formation of single strand DNA or a slipped structure allowing the two pyrimidines to photodimerize in a colinear arrangement to form non-adjacent CPDs. These types of non-adjacent dimers effectively shorten the DNA template, and have been implicated in the formation of UV-induced deletion mutations (17).

Anti CPDs are expected to form between two separate strands, or loops in DNA and were first detected under non-physiologic conditions of desiccated or ethanolic DNA. Later a cis-anti TT CPD was found to form from a short single stranded oligodeoxynucleotides under acidic conditions. The product was identified

by mass spectrometry by first digesting the DNA with Nuclease P1 to produce a tetrameric product comprised of the CPD attached at the 3'-ends to the flanking nucleotide. The observation of non-adjacent CPD formation suggested that it might be possible that these anti CPDs could form between the loops of G-quartet structures in telomeres and promoters (18, 19). Indeed irradiation of human telomeric DNA sequences was found to produce a variety of *anti* CPDs by a combined enzymatic LC-MS/MS assay (52). Most surprisingly, *anti* CPD formation was found to occur in the presence of K^+ which favors hybrid G-quadruplexes that would appear to be incapable of forming anti-CPDs because they do not contain any adajacent lateral loops (49). In contrast, anti CPDs were not detected in Na^+, which would have appeared to favor formation of anti-CPDs because of the presence of two lateral loops. A proposal suggesting that a form 1 two tetrad G quadruplex was the photoreactive conformation was later shown unlikely (45), and a more recent paper has suggested that a reverse Hoogsteen hairpin may be involved (19,20, 10).

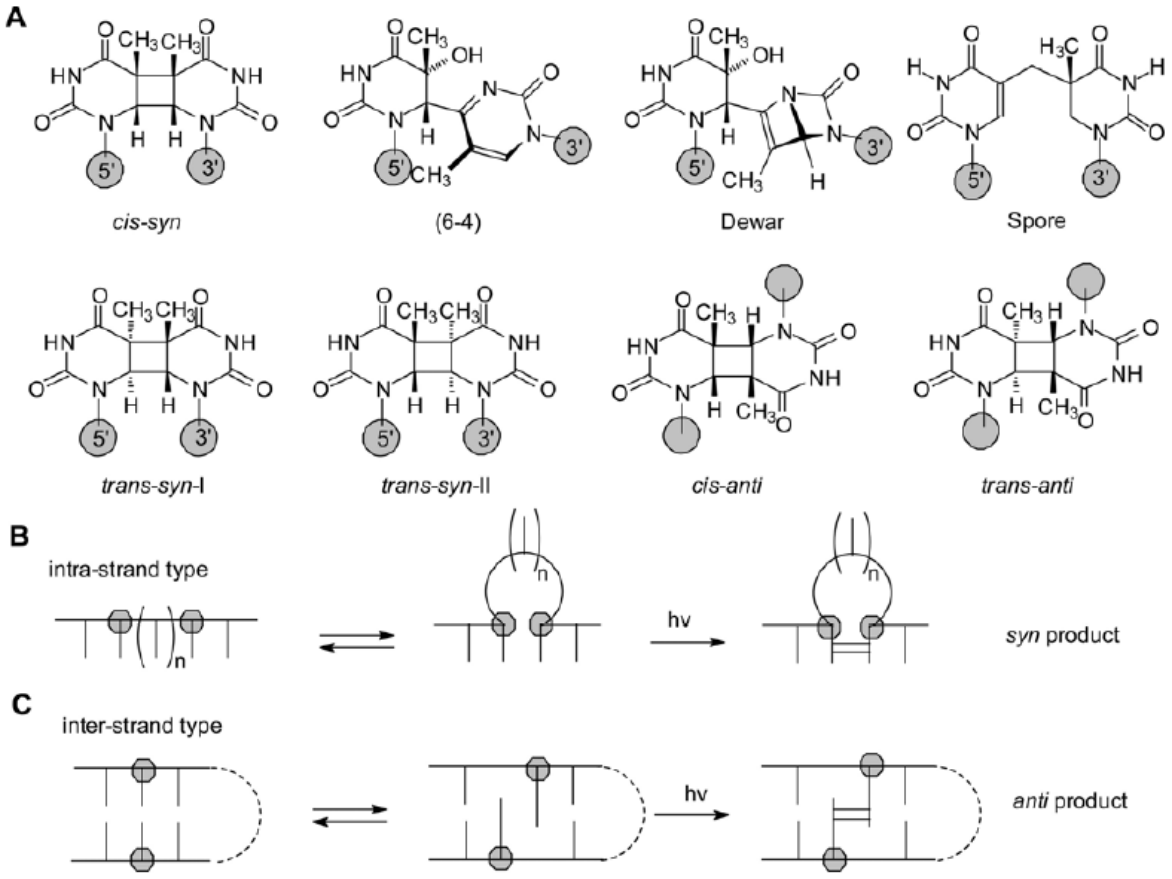


Figure 7: Photochemistry of DNA. A) Structures of the major type of photoproducts. B) Intra- and C) interstrand-type photoreactions, both of which lead to non-adjacent photoproducts, except when n=0 for B, which results in an adjacent photoproduct.

1.7 Thesis

The goal of this thesis was to further investigate the formation of anti-CPD forming in human telomeric sequences under various conditions known to favor or disfavor G quadruplex formation. Central to this goal was the development of a high resolution LC MS/MS method for separating and identifying all the photoproducts. Part one of this thesis, describes the development of a new LC method to better separate the CPD-containing nuclease P1 (NP1) digestion products and utilize it to analyze the NP1 products of UV irradiated telomeric DNA isolated with the previous HPLC method. The second part of this thesis focused on directly coupling the LC method to MS/MS mass spectrometry to simultaneously separate and identify the nuclease P1 digestion products of UV irradiated telomeric DNA in Li+. These products were then compared against the products formed in the presence of K+. In the course of these studies a new hexameric digestion product was corresponding to $(pTTA)_2$ was also identified and consistent with the presence of two anti-CPDs.

Chapter Two

Our interest in the photochemistry of the G-quadruplex forming 26-mer sequence d(AAAGGGTTAGGGTTAGGGTTAGGGAA) was because previous irradiation of Tel26 with 312 nm UV light, appeared to produce two major photoproduct-containing products following NP1 degradation, the *cis, syn* T=TA CPD from adjacent thymines, and the trans, anti T(A)=T(A) CPD formed between two non-adjacent thymines. We also wanted to investigate what products formed in Li^+ solution which does not support quadruplex formations.

Nuclease P1 coupled LC-MS/MS assay.

The photoproducts that are formed in human telomeric sequences are characterized by enzymatic digestion with nuclease P1. NP1 has both endonuclease activity and 3'-phosphatase activity. It has a binding pocket that is specific for the bases in DNA and, upon binding to the 5'-base, catalyzes hydrolysis of the phosphodiester bond to the 3'-side to yield deoxynucleotide with a 3'-hydroxyl group. Because the bases in dypyrimidine photoproducts are covalently linked together, NP1 cannot bind to either base of the photoproduct and consequently cannot cut on the 3'-side of either nucleotide of the photoproduct. As a result Py=PyN trinucleotides are produced from adjacent CPD photoproducts, and Py(N)=Py(N) tetranucleotides are produced from

non-adjacent anti CPDs (Figure 8). The NP1 digestion products of the DNA containing the tri and tetranucleotides are separated from mononucleotides by reverse phase HPLC, and either subsequently or simultaneously analyzed by MS/MS mass spectrometry. To optimize the separation, various solvent systems, ramps, and chromatography columns were tested. The separated products were identified by their mass and fragmentation patterns.

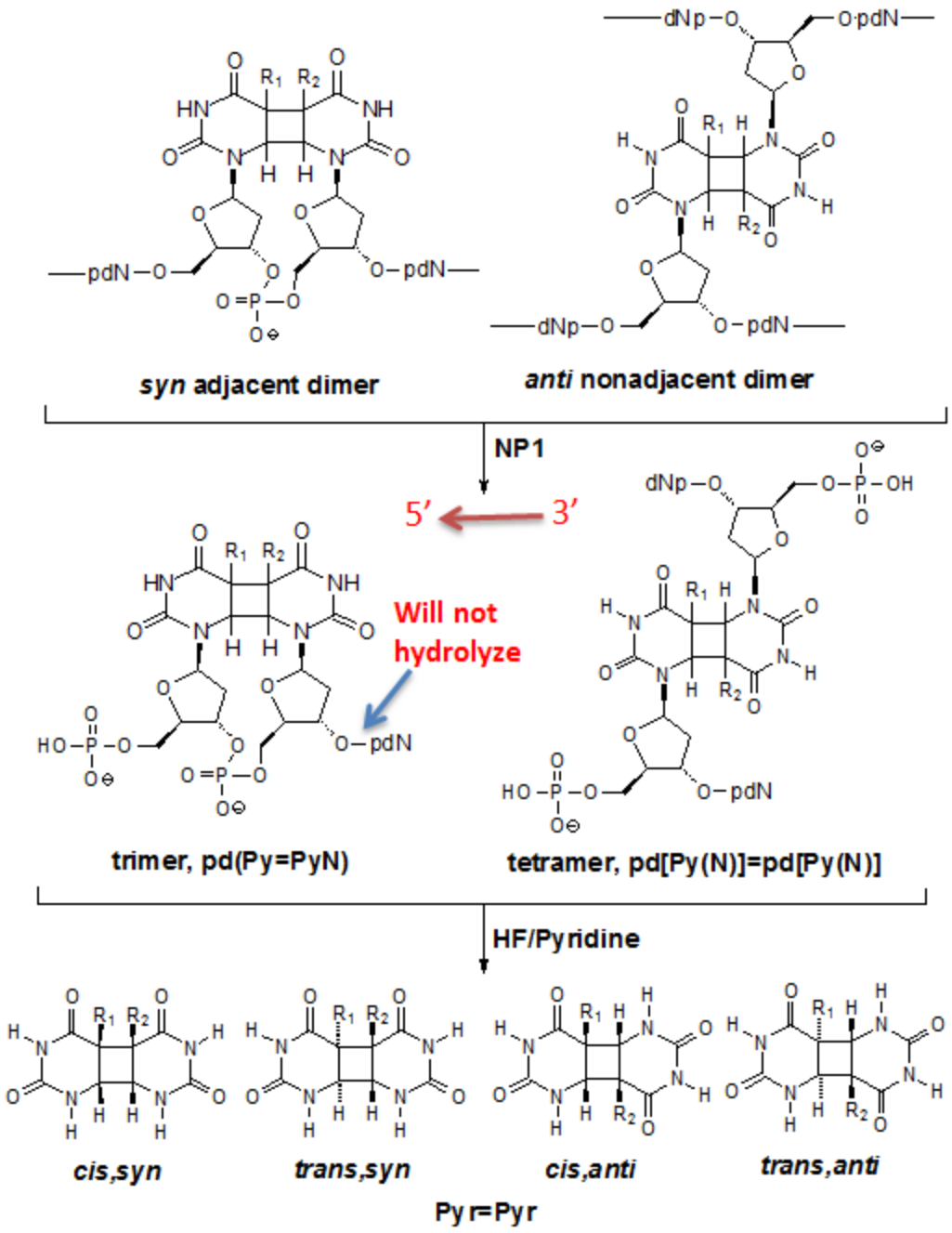


Figure 8: NP1 digestion of photoproducts containing DNA. NP1 is unable to hydrolyze the 3' side of a photodimerize base.

Irradiation and NP1 Digestion of the G-quadruplexes.

The G-quadruplexes were irradiated and digested by Chen Lu at the Department of Chemistry at Washington University. A 50 uM solution of Tel26 G-quadruplex in 150 mM LiCl was irradiated with UVB light for 2.5 h on ice. The sample was then subjected to NP1 enzymatic digestion for at least 36 h at 37°C. The photoproduct containing digestion fragments were separated via reverse phase HPLC and monitored by UV absorption at 260 nm under the conditions shown in Figure 9. The following was summarized from published data of the original HPLC method that was developed at Dr. Taylor's lab: the mononucleotides elute during the first 20 minutes followed by the elution of trinucleotides around 30 minutes and the tetranucleotides around 40-43 minutes. The 260 nm UV absorbance of the T=TA trinucleotide digestion photoproduct results from the presence of 3'-adenine, while the absorbance of the T(A)=T(A) or T(T)=T(A) tetranucleotide digestion photoproducts are the result of two 3' adenines or a 3'-adenine and a 3'-thymidine. The CPDs have very little absorption at 260 nm.

.

	HPLC settings					
	Time	% MPA	%MPB	Flow Rate		
	0	100	0	1 ml/min		
	3	80	20			
	53	80	20			
Run time	53					
	System	Gold BioEssential HPLC with a model 125 binary gradient pump and a Model 168 Diode array detector by Beckman Coulter				
	MPA:	100% (50 mM Trietylammonium acetate, pH 7.5				
	MPB:	50% acetonitrile in 50 mM triethylammonium acetate pH 7.5				
	Column:	Waters X-Bridge column C18, 4.6x 75 mm, 2.5 um, 135-Å pore size				

Figure 9: Original RP-HPLC method settings, Dr.Taylor's lab.

Three fractions containing the photoproduct tri- and tetranucleotides were collected by Chen Lu at Washington University using the method as shown in Figure 9**.** Fraction 1 was collected around 29.1 minutes, Fraction 2 was collected at 41.7 minutes and Fraction 3 was collected at 42.8 minutes respectively.

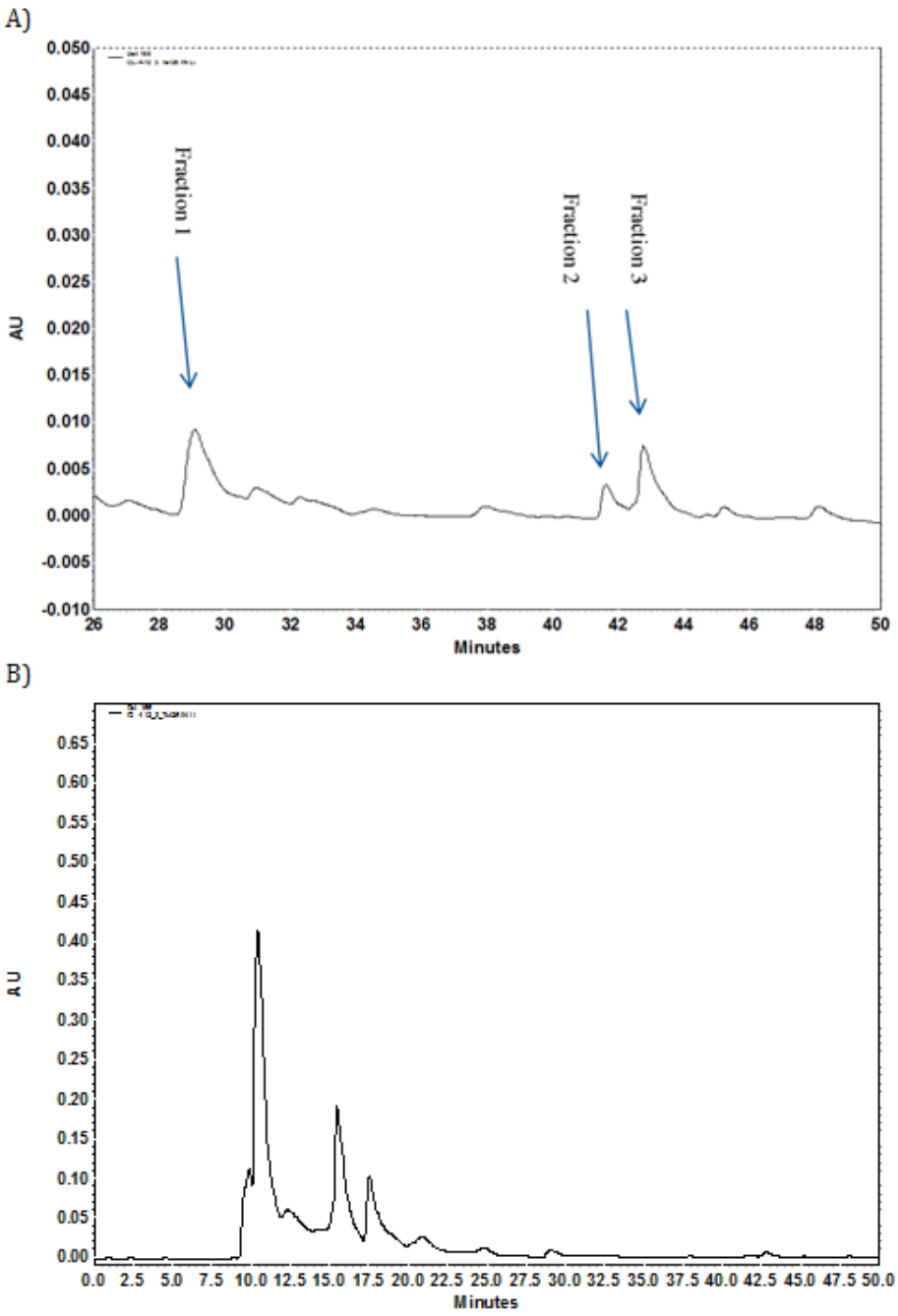


Figure 10: HPLC Chromatograph trace provided by Dr. Taylor's lab displaying fractions collected A) Zoom in B) Zoom out.

Analysis of Samples Obtained by the Original RP-HPLC Method with an New Method

My primary aim was to establish a new LC method with which I could separate and identify the NP1 degradation products with good separation and resolution. RP HPLC was carried out on a 2.1x150 mm Waters Acquity UPLC High Strength Silica (HSS) T3 column, with particle 1.8 µm particle size on a 1260 Infinity Quaternary LC System from Agilent Technologies. I started the development investigating a number of buffer systems and gradients and settled on the buffer and gradient system shown in Figure 11. Np1 digestion for the parent ODN afforded major products corresponding to monomers eluting at 3.95 (pG), 4.18 (pT), 6.63 (pA) and 30.71 (dA) minutes when using the HPLC method. The analysis of fraction 1 that was collected at 29.1 minutes with the original LC method; eluted between 6-8 minutes with the new LC method. Interestingly, this fraction was resolved into two peaks; one of which eluted around 6.57 minutes and second peak eluted around 7.82 minutes as shown in Figure 12.

HPLC settings				
	Time	% Mobil Phase A	%Mobil Phase B	Flow Rate
	0	100	0	0.2
	14	100	0	
	80	78	22	
	82	0	100	
	84	0	100	
	86	100	0	
Total Run time	105			
	MPA:	100% Ammonium Formate		
	MPB:	50/50 (v/v) Ammonium Formate + Acetonitrile		
	Column:	Waters Acquity UPLC HSS T3, 1.8 um 2.1 x 150 nm Silica Based bonded Pore size 100 Å		
	Part#	186003540		

Figure 11: Reverse phase HPLC gradient. MPA: 10 mM aqueous ammonium formate; MPB 50%/50% 10 mM aqueous ammonium formate and acetonitrile.

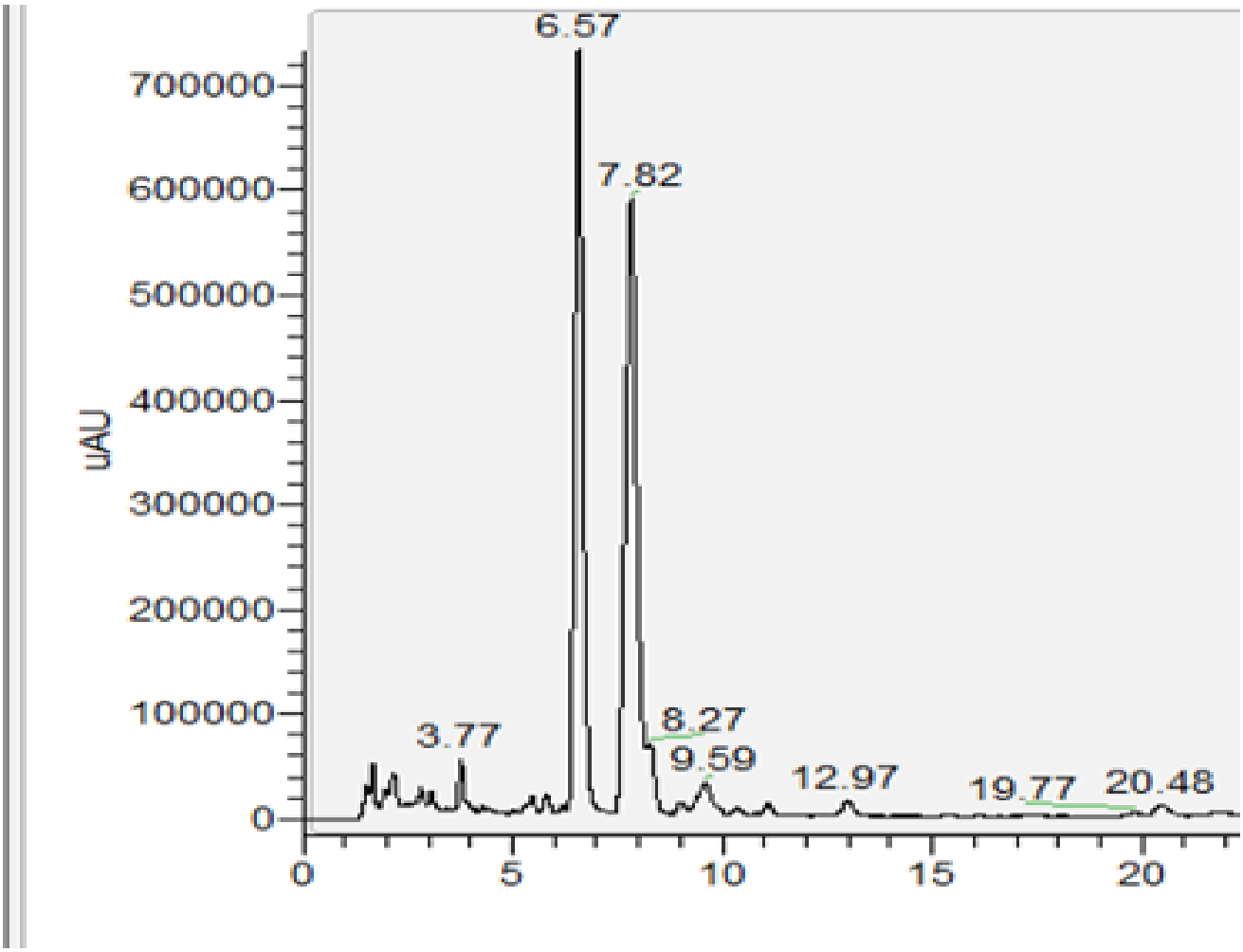


Figure 12: Analysis of Fraction 1 collected by Chen Lu at Washington University in Dr. Taylor's lab.

The analysis of fraction 2 with elution time of 41.7 minutes in the original method resulted in one peak by the new LC method with an elution time of 33.4 minutes as shown in Figure 13.

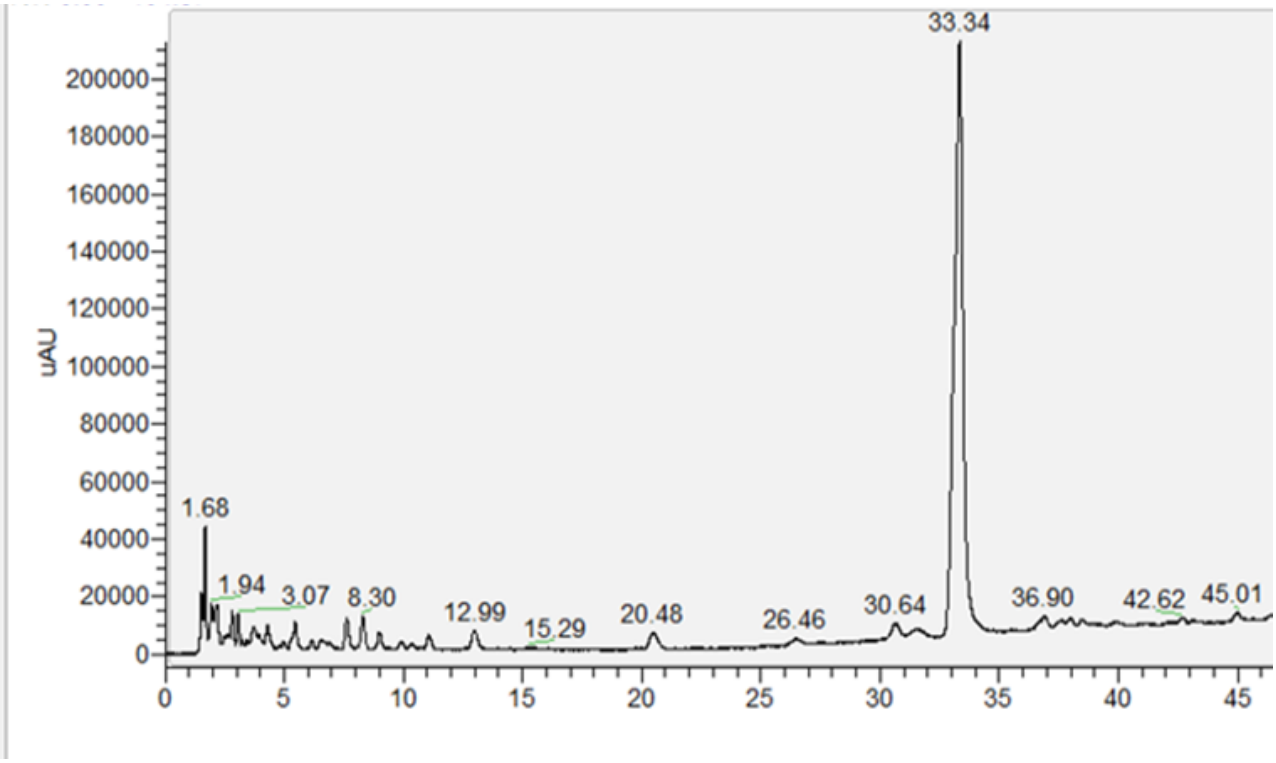


Figure 13: Analysis of fraction 2 collected by Chen Lu at Washington University in Dr. Taylor's lab.

The analysis of fraction 3 that had been collected around 42.8 minutes was resolved into three peaks with the new method. One of the peaks eluted around 33.41 minutes, a second peak eluted around 36.76 minutes and a third and major peak eluted at 38.98 minutes as shown in Figure 14.

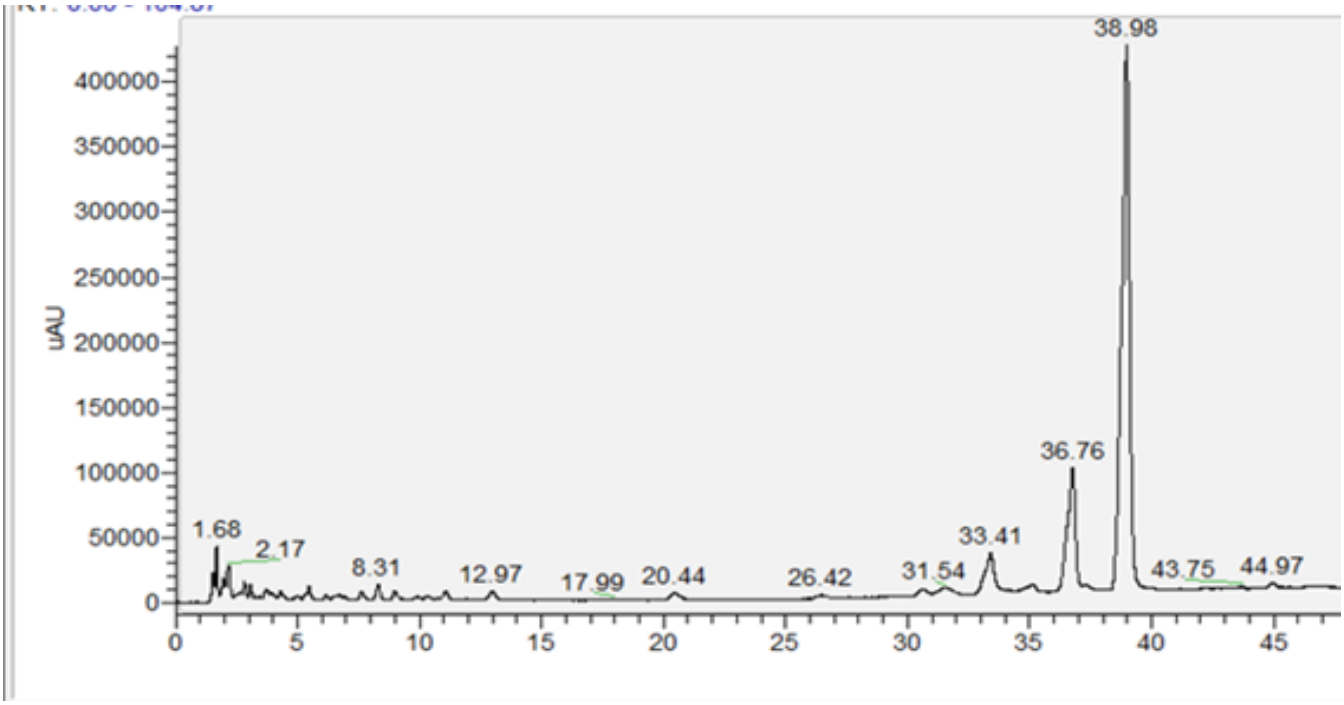
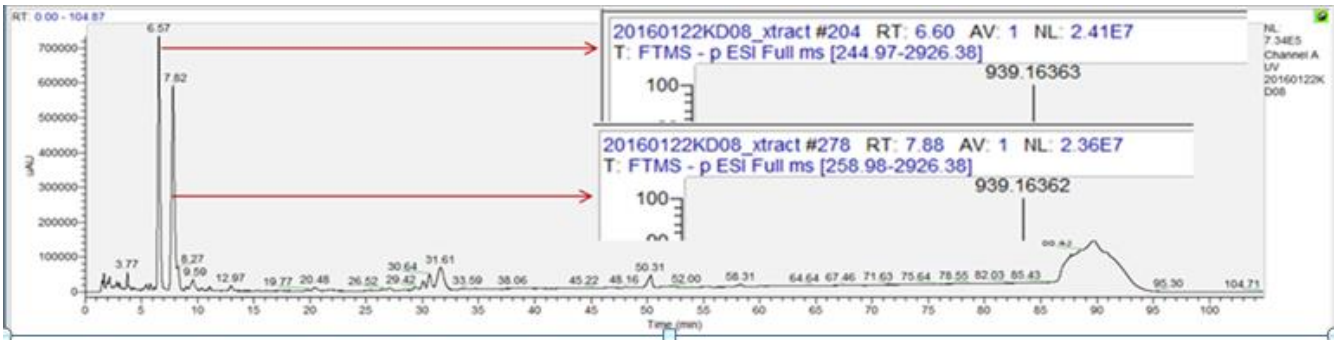


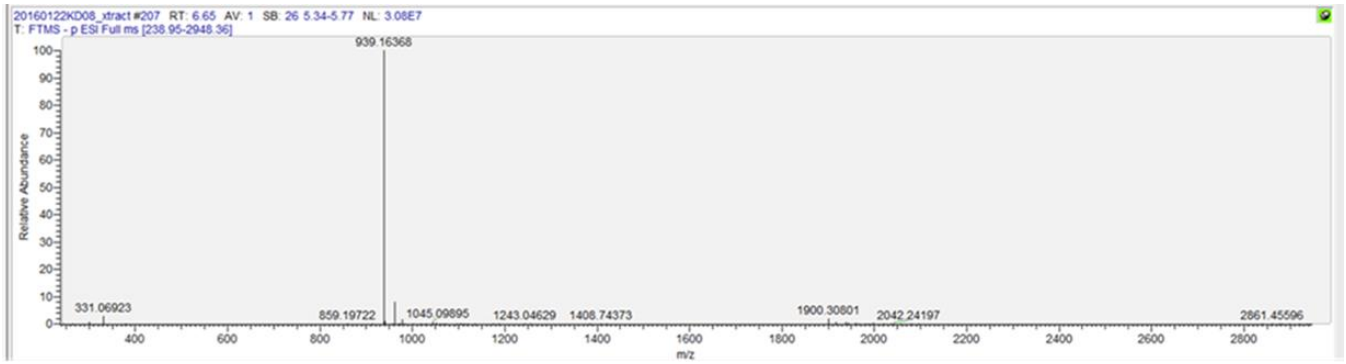
Figure 14: Analysis of fraction 3 collected by Chen Lu at Washington University in Dr. Taylor's lab.

Further analysis of the first fraction by LC-MS is shown in Figure 15. The LC-MS of both of the peaks at 6.65 and 7.90 minutes showed an m/z for of 938.164 corresponding to the $(M-H)^-$ ion of the trinucleotide pT=pTpA, which must correspond to either cis-syn, trans-syn cyclobutane pyrimidine dimers between the two T's or the (6-4) or Dewar photoproducts.

A)



B)



C)

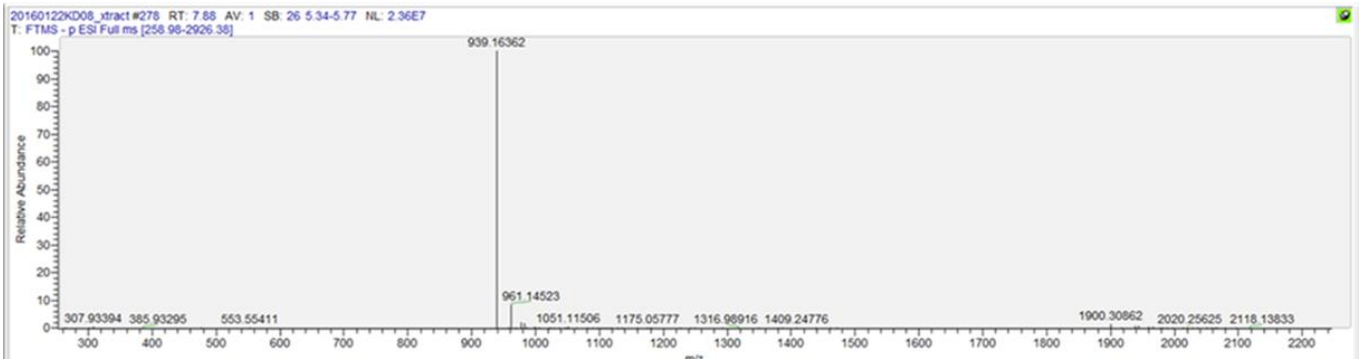
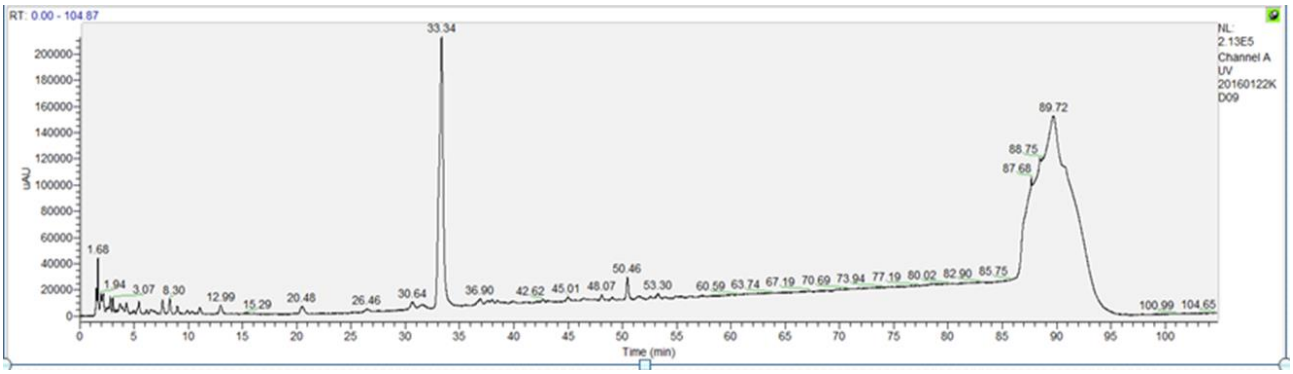


Figure 15: Analysis of fraction 1 by A) LC-MS zoom out B) Full LC-MS extracted and processed data of peak one C) Full LC-MS extracted and processed data of peak 2 in fraction 1.

The analysis of fraction 2 showed which gave a single peak at 33.4 minutes in the new method, as shown in Figure 16. The LC- MS showed the presence of an ion of m/z 1260.209 corresponding to the $(M-H)^-$ ion of a tetranucleotide pT(pA)=pT(pA).

A)



B)

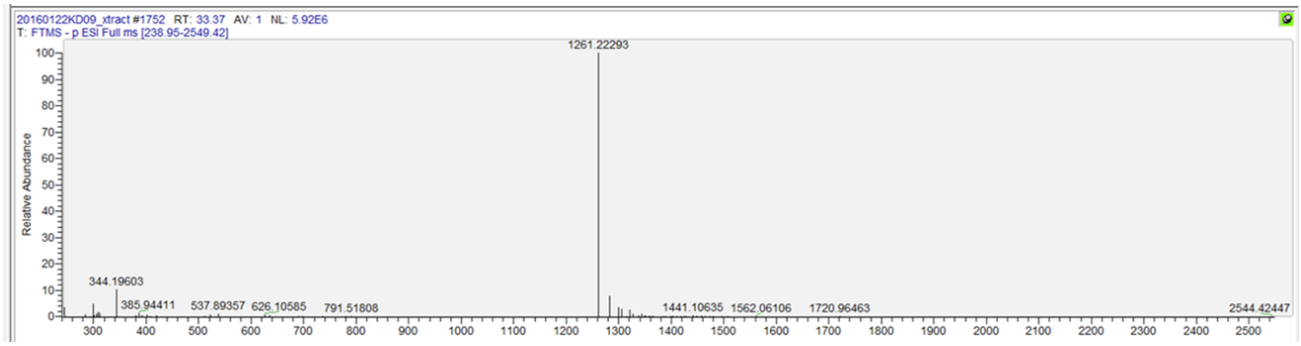


Figure 16: Analysis of fraction 2 by A) LC-MS analysis of peak one in fraction 2. B) Full LC-MS extracted and processed data of peak one in fraction 2.

Three peaks eluted from the analysis of fraction 3 by the new LC-MS method. One of the peaks eluted around 33.4 minutes, the second peak eluted around 36.8 and the third peak eluted around 38.98 minutes. Analysis of the first two peaks by LC-MS (Figure 17), gave an m/z of 1269.235 corresponding to the $(M-H)^-$ ion of pT(pA)=pT(pA) or a molecule of mass 1270.235. The analysis of the third peak gave anm/z of 1260.209 corresponding to the $(M-H)^-$ ion of the tetramer pT(pA)=pT(pT) or a molecule of mass 1261.

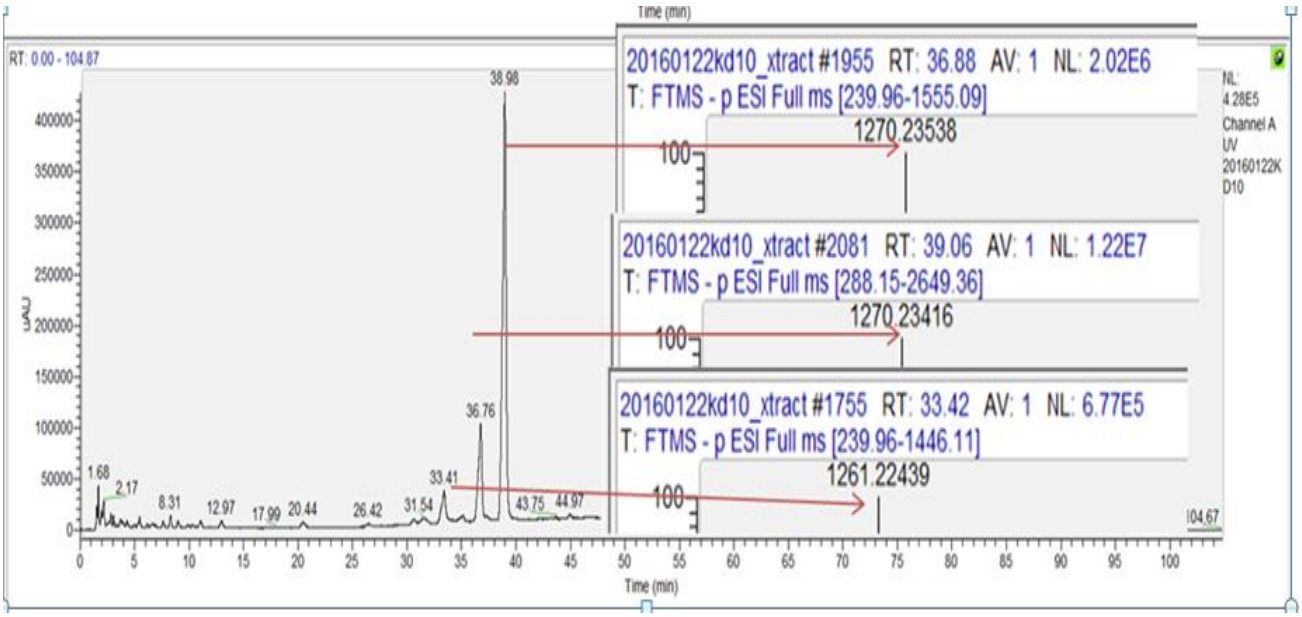


Figure 17: Analysis of fraction 3 by LC-MS.

Comparison of the original with the new RP-HPLC Method

There were some significant differences between the elution times of the nucleotides and the photoproduct digestion products using the original and new RP-HPLC methods. For example, mononucleotides eluted within the first 20 minutes in the original method, wereas the eluted between 3-30 minutes with the new method. Another difference between the two methods is that trinucleotides eluted from at 29.1 minutes in the original method, whereas in the new method the same fraction eluted between 6-8 minutes. Also tetranucleotides that appeared to eluted as a single peak at 41 minutes appeared as three peaks between 31-40 minutes.

Direct LC-MS Analysis of NP1 Digestion Mixtures by the New Method

After the analysis of the mononucleotides and fractions that were provided by Dr. Taylors Lab, we continued with the study of Tel26 in Li^+ and other metal ions.

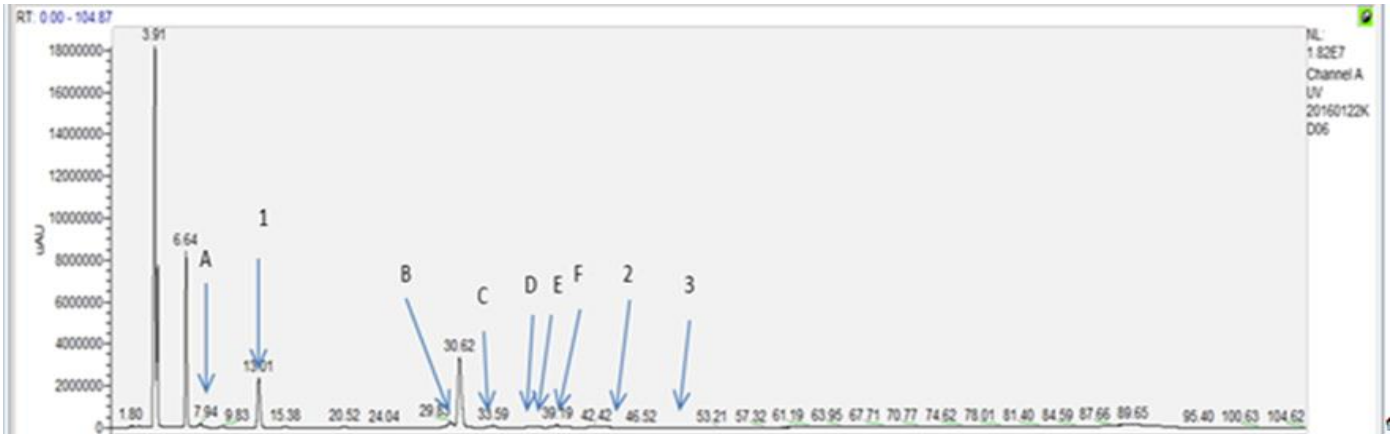
	HPLC settings					
	Time	% MPA	%MPB	Flow Rate		
	0	100	0	0.2 ml/ min		
	14	100	0			
	80	78	22			
	82	0	100			
	84	0	100			
	86	100	0			
Run time	105					
	System	1260 Infinity Quaternary LC System by Agilent Technologies				
	MPA:	100% Amonium Formate				
	MPB:	50% Amonium Formate + 50% Acetonitrile				
	Column:	Waters Acquity UPLC HSS T3, 1.8 um 2.1 x 150 nm Silica Based bonded Pore size 100 Å				
	Part#	186003540				

Figure 18: LC-MS and MS/MS system, column and mobile phase settings.

Oligodeoxynucleotides (ODN) were purchase from Integrated DNA Technologies, Inc. (IDT). Tel26 was prepared in 100 mM Li^+ and heated at 95°C for 5-10 minutes and then was rapidly cooled down in ice. UVB irradiation was performed at 270-400 nm with peak intensity at 312nm on ice with UVB light for about 2-2.5 hours at Washington University by Chen Lu. The analysis of the degradation products was performed by LC using a waters Acquity UPLC HSS T3 silica based column in a

1260 infinity quaternary LC system by Agilent Technologies using the settings shown in Figure 18. The peaks that appeared to correspond to previously identified peaks from the original study were labeled as A, B, C, D, E, F. Other peaks that were not previously identified in the original studies are labeled 1, 2 and 3 respectively as shown in Figure 19. The LC-ESI-MS experiments were carried out in the negative ion mode with the orbitrap mass spectrometer (Q Exactive-Plus). A solution of 10 mM ammonium formate in water and acetonitrile was used as the spray solvent. Each labeled peak in the HPLC trace in Figure 19 was analyzed by LC-MS and MS-MS and compared to previously identified products based on their m/z and fragment ions.

A)



B)

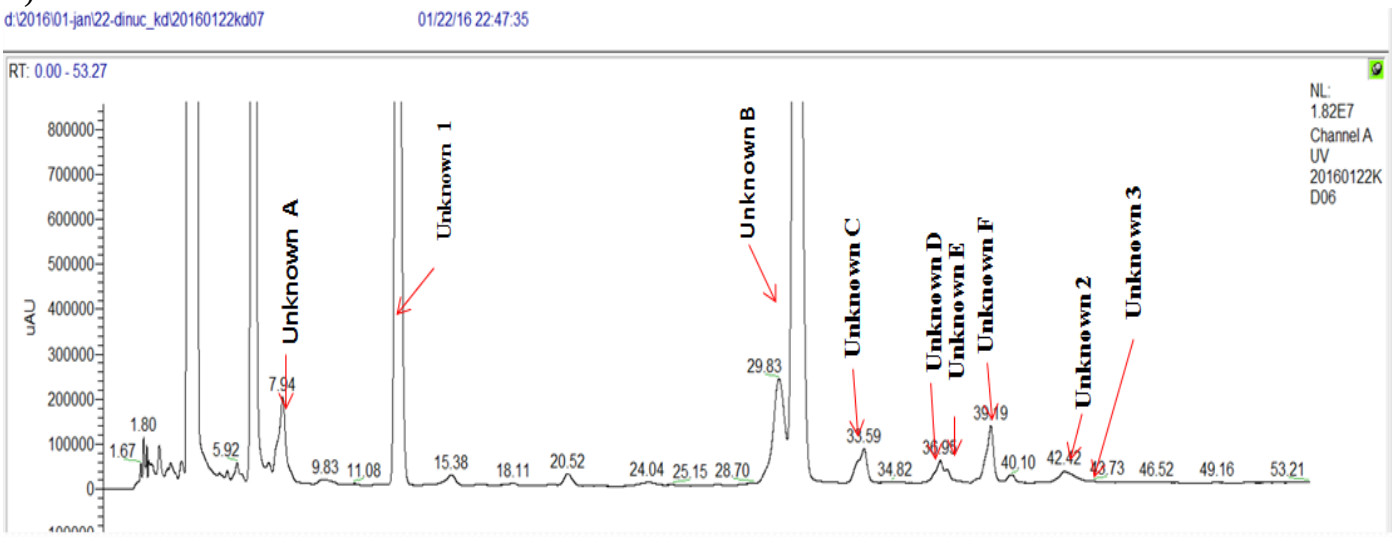
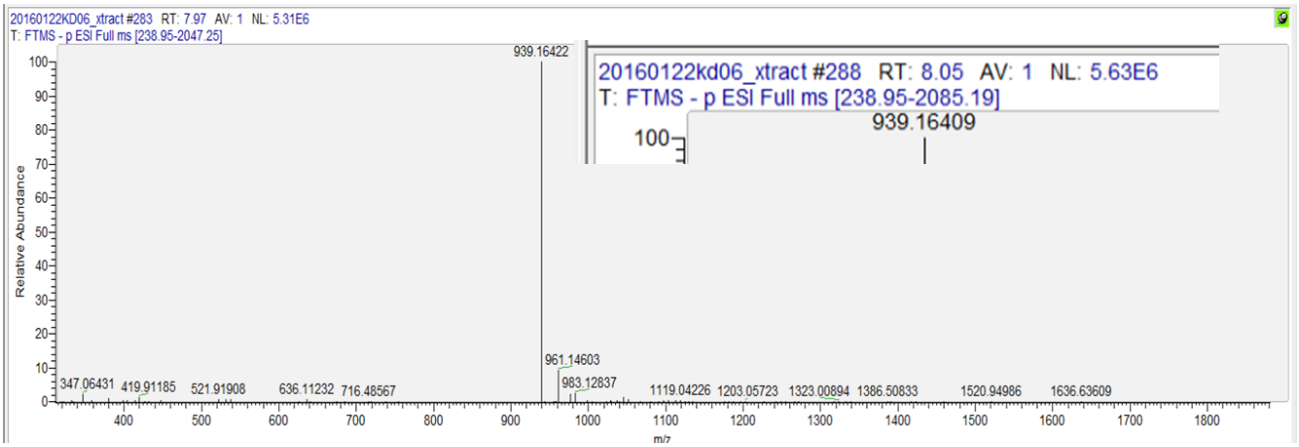


Figure 19: A) Analysis of irradiated and NP1 degraded Tel26 in Li+ at 260 nm (zoom out). B) Analysis of irradiated and NP1 degraded Tel26 in Li+ at 260 nm (zoom in).

Analysis of Unknown Peak A.

Analysis of unknown peak **A** which eluted at 7.94 minutes (Figure 20A) gave an m/z of 938.16 corresponding to the $[M-H]^-$ of the trimer (T=TA). The $[M-H]^-$ ion gave product ions at m/z 330.06 corresponding to $[pdA-H]^-$ as shown in Figure 20B.

A)



B)

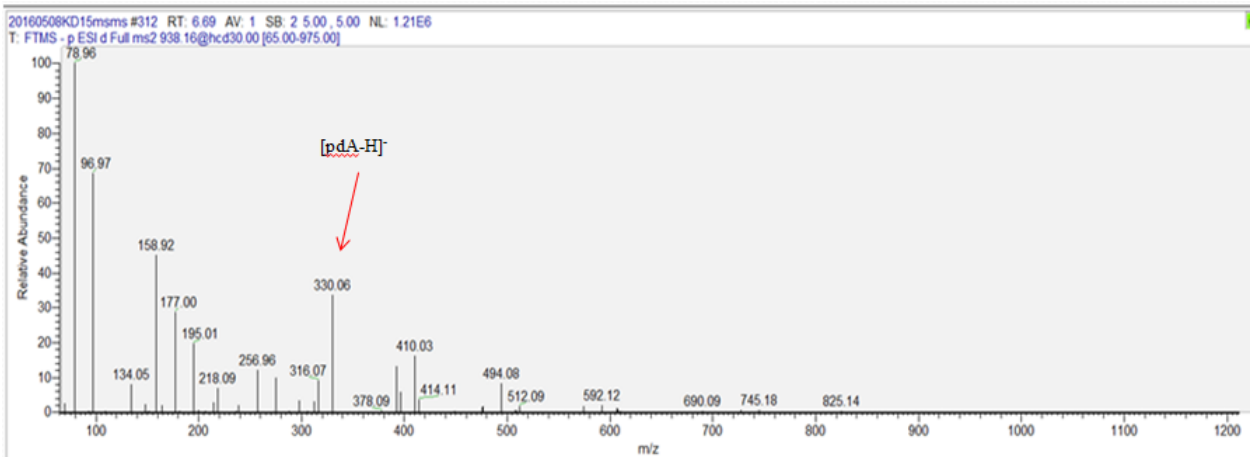
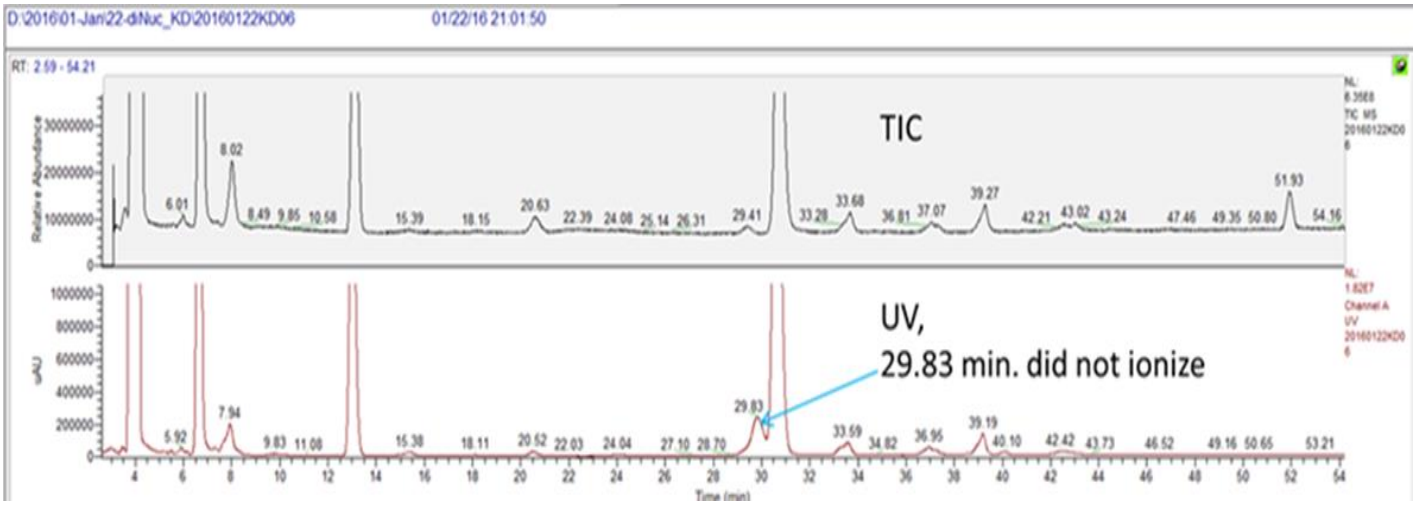


Figure 20: A) Analysis of unknown Peak A by LC-MS and B) MS/MS results for unknown Peak A and product ion analysis.

Analysis of Unknown Peak B.

During early development of the new LC and MS/MS method an unknown peak (unknown B, Figure 19) was observed eluting at 29.83 minutes from in Tel26 irradiated in Li^+ (Figure 21A) which could not be detected by MS in negative ion mode. When the dA at the 5'-terminus of Tel26 was replaced with pdA and irradiated and digested with NP1, the shoulder in unknown peak B was no longer detected as shown in Figure 21B.

A) Analysis of shoulder in peak B



B) LC-MS of Tel26 in Li+ no shoulder in peak B

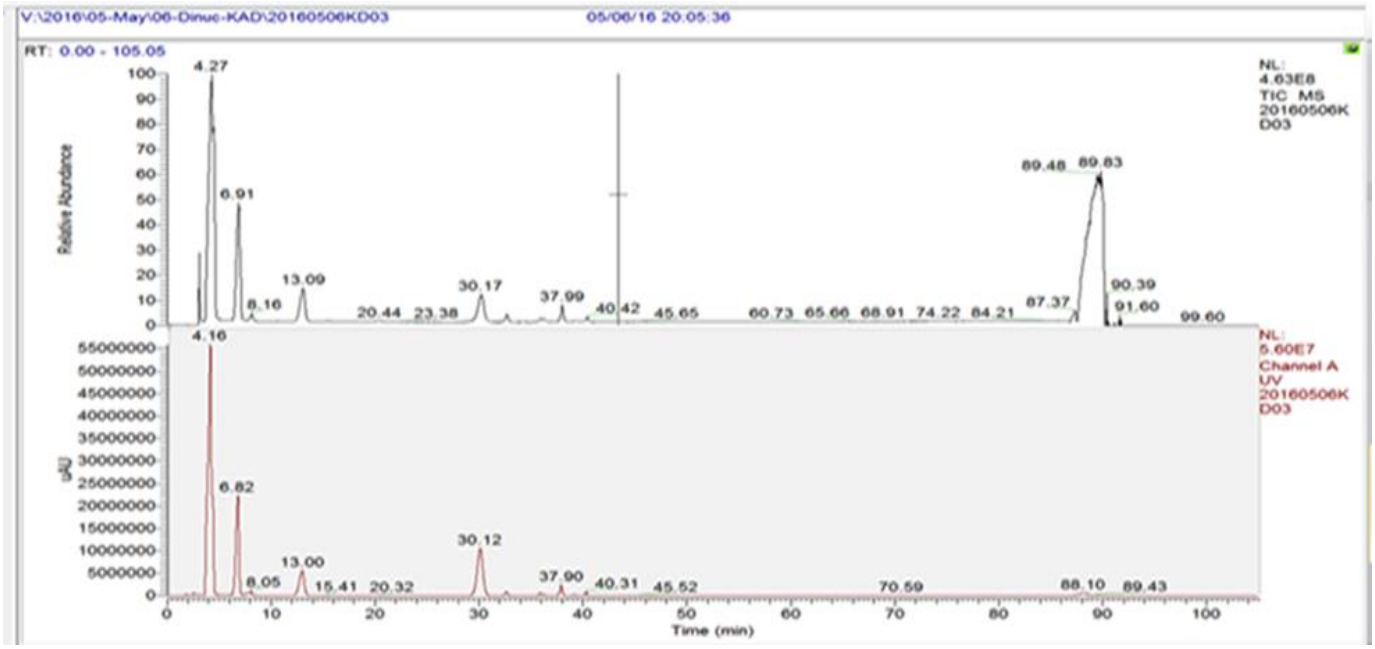
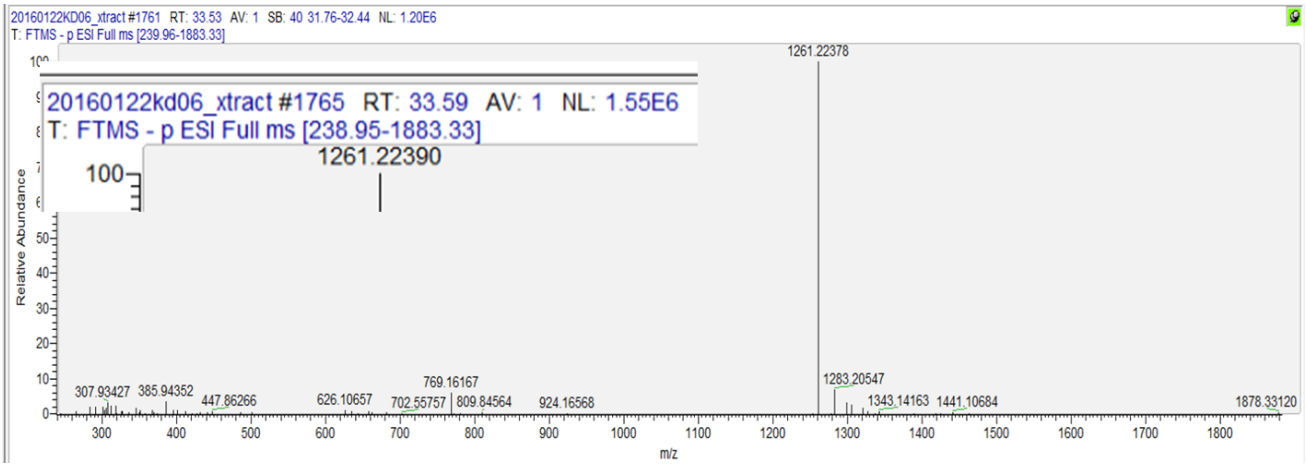


Figure 21: Analysis of unknown B peak by LC-MS ATel26 in Li+ with dA and B) LC-MS without shoulder in peak B.

Analysis of Unknown Peak C.

The analysis of unknown peak **C** eluting at 33.59 minutes, gave an m/z peak at 1260.224 (Figure 22A) corresponding to the $[M-H]^-$ for the tetranucleotide pTp(A)=pTp(T). For the MS/MS analysis the $[M-2H]^{2-}$ at 629.6 was trapped and fragmented (Figure 22B). Product ions were observed at m/z 321 corresponding to [pdT-H]; a product ion was observed at m/z 330.06 corresponding to $[pdA-H]^-$; another product ion was observed at m/z 527.12 corresponding to [pT=pT –HPO3-2H2O –H]$^-$; another product ion was observed at m/z 545.13 corresponding to $[pT=pT-HPO3-H2O-H]^-$; an product ion was observed at m/z 607 corresponding to $[pT=pT -2H_2O–H]^-$; another product ion was observed at m/z 625.09 corresponding to $[pT=pT-H_2O-H]^-$; a product ion was observed at 929.14 $[pT=pT(pT) –H_2O-H]^-$ and a product ion was observed at m/z 938.16 corresponding to $[pT=pT(pA) –H_2O –H]^-$.

A) LC analysis of unknown peak C



B) MS/MS analysis and product ions

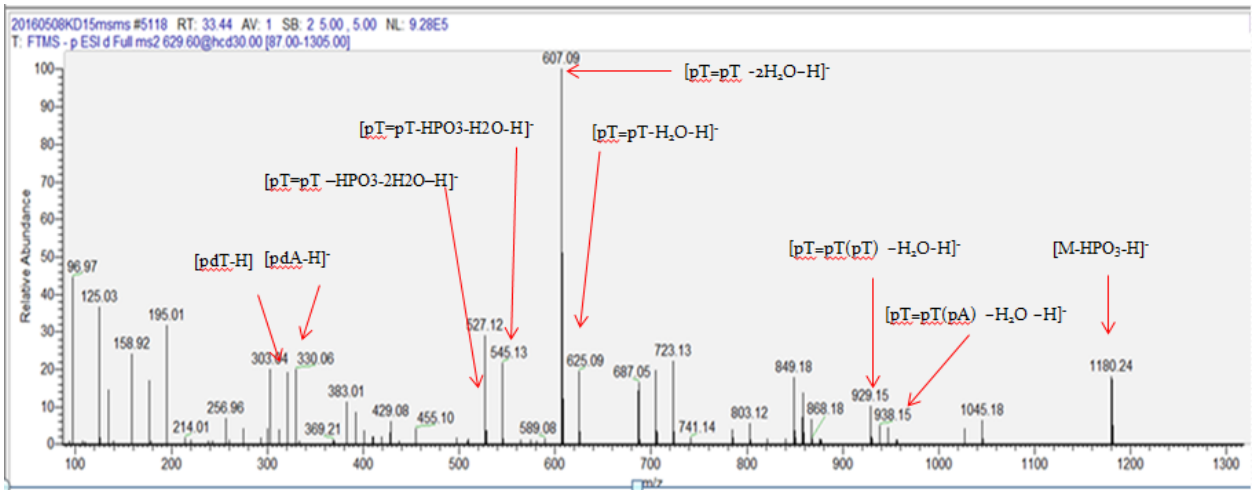
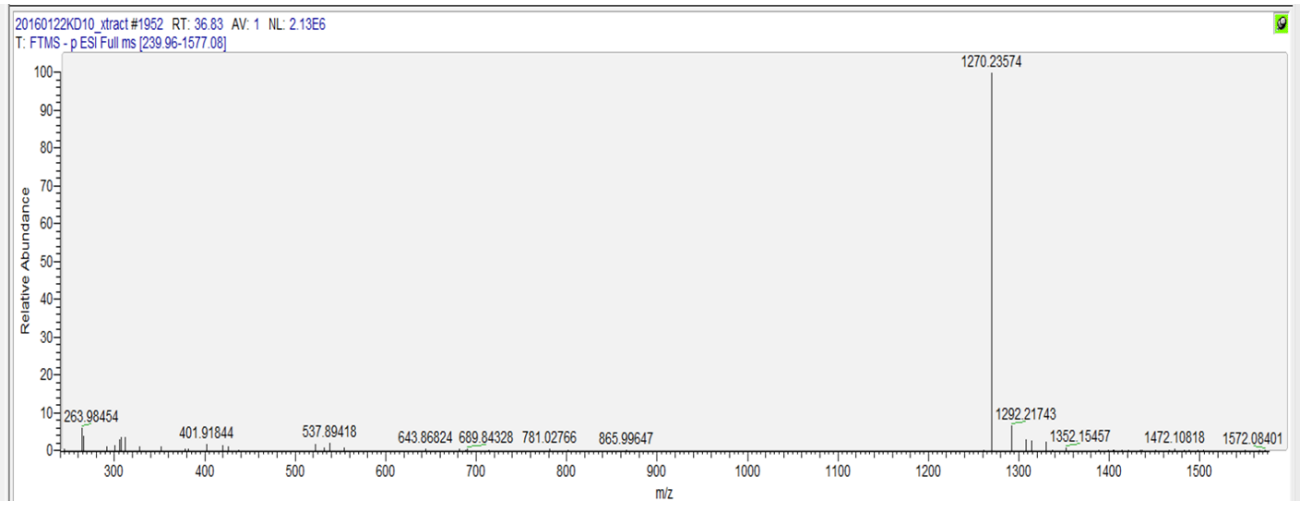


Figure 22: Analysis of Peak C by A) LC-MS and B) MS/MS results and product ions.

Analysis of Unknown Peak D.

The analysis of unknown peak **D** eluting at 36.8 minutes gave an m/z at 1269.235 as shown in Figure 23A corresponding to pT(pA)=pT(pA). The MS/MS experiments were then carried out on $[M-2H]^{2-}$ at 634.11 which was trapped and fragmented (Figure 23B). Product ions were observed at m/z 330.06 corresponding to $[pdA-H]^-$; m/z 527 corresponding to $[pT=pT\ -HPO_3-2H_2O\ -H]^-$; m/z 545.13 corresponding to $[pT=pT-HPO_3-H_2O-H]^-$; m/z 607 corresponding to $[pT=pT\ -2H_2O\ -H\ ]^-$; at m/z 625.09 corresponding to $[pT=pT-H_2O-H]^-$; at m/z 858.19 corresponding to $[pT=pT(pA)\ -HPO_3-H_2O\ -H]^-$ and m/z 938.16 corresponding to $[pT=pT(pA)\ -H_2O\ -H]^-$.

A)LC-MS analysis of unknown peak D



B) MS/MS Analysis and Product ions.

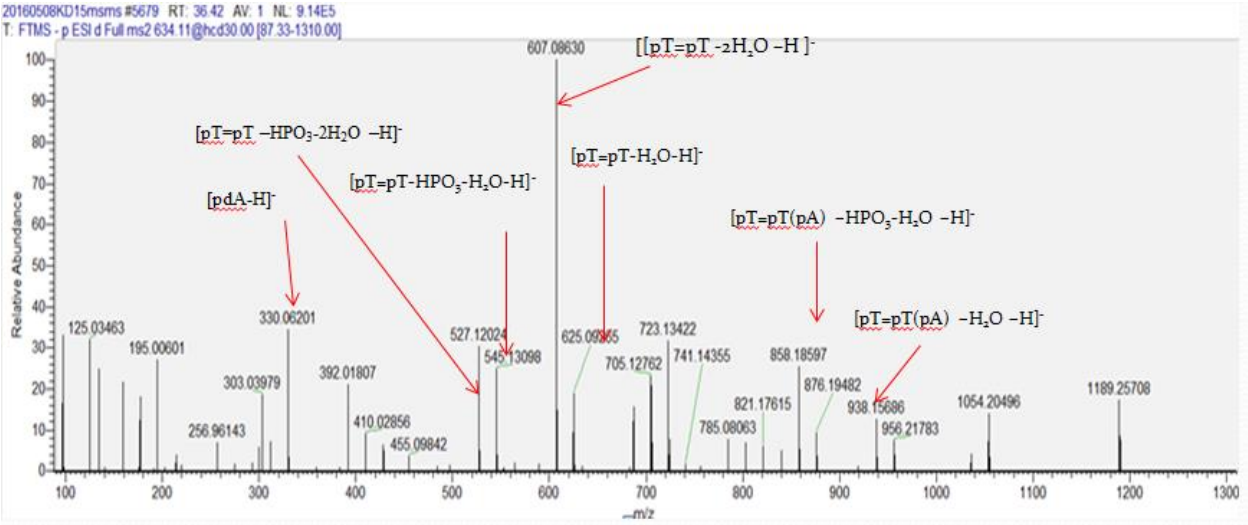
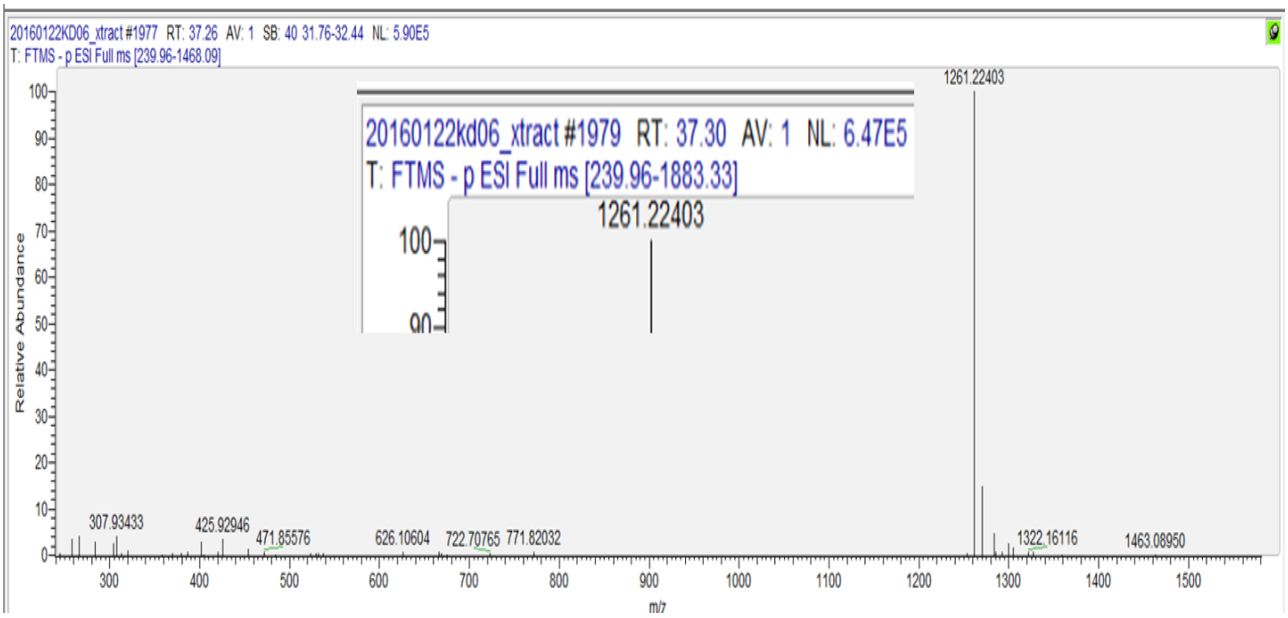


Figure 23: Analysis of Peak D by A) LC-MS and B) MS/MS results and product ions.

Analysis of Unknown Peak E.

The analysis of unknown peak **E** eluting at 37 minutes gave an ion of m/z at 1260.22 as shown in Figure 24A corresponding to $[M-H]^-$ ion of pT(pA)=T(pT). The peak at 37 minutes was then ionized by negative electrospray to give $[M-2H]^{2-}$ of 629.6 which was trapped and fragmented. Product ions were observed at m/z 321.05 corresponding to $[pdT-H]^-$; at m/z 330.06 corresponding to $[pdA-H]^-$; at m/z 607.06 corresponding to [pT=pT $-2H_2O$ –H]; at m/z 849.19 corresponding to [pT=pT(pT) – H_2O-HPO_3-H]$^-$; at m/z 929.14 corresponding to [pT=pT(pT) –H_2O-H]$^-$ and m/z 1180.23 corresponding to [pT(pA)=pT(pT) – HPO_3– H]$^-$.

A) LC-MS analysis of peak E



B) MS/MS analysis and product ions.

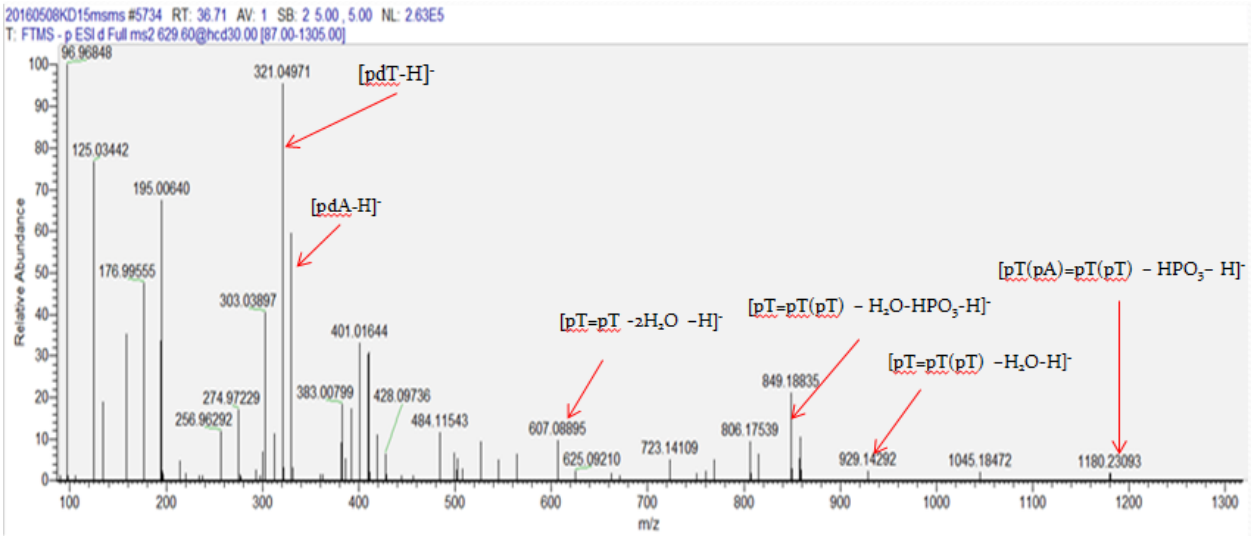


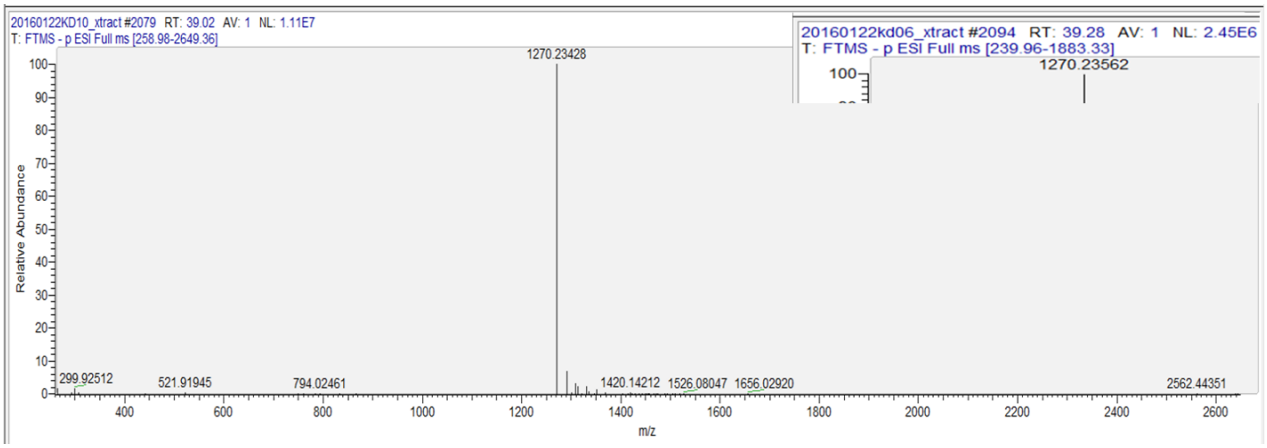
Figure 24: Analysis of unknown Peak E by A) LC-MS and B) MS/MS results and product ions.

Analysis of Unknown Peak F.

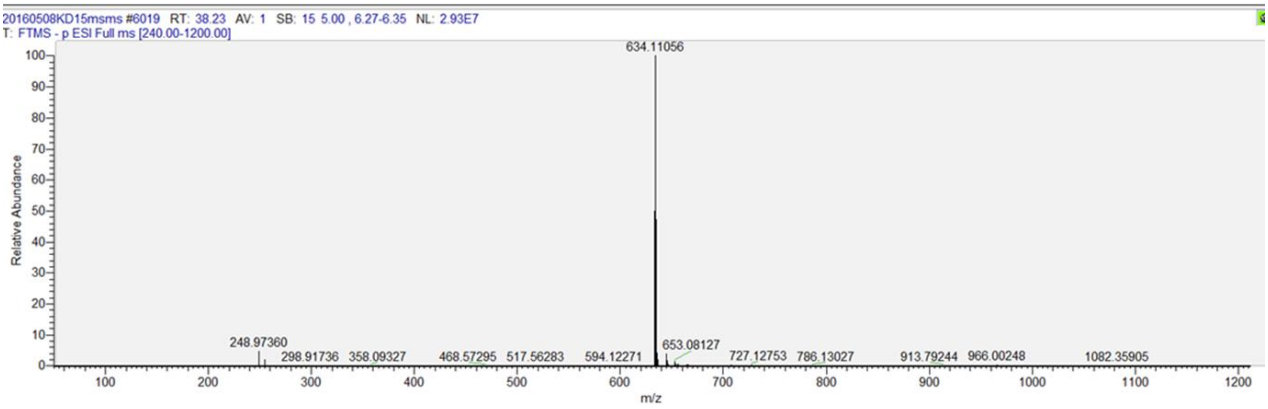
The analysis of peak **F** eluting at 39 minutes gave an m/z ion at 1269.235 as shown in Figure 25A corresponding to the $[M-H]^-$ ion of pT(pA)=pT(pA). The peak at 39 minutes was then ionized by negative electrospray to give $[M-2H]^{2-}$ of 634.11 as shown in Figure 25B which was trapped and fragmented. Product ions were observed at m/z 330.06 corresponding to $[pdA-H]^-$; m/z 527.12 corresponding to $[pT=pT –HPO_3-2H_2O –H]^-$; at m/z 607.06 corresponding to $[pT=pT -2H_2O –H]^-$ and m/z 938.158 corresponding to $[pT=pT(pA) –H_2O –H]^-$.

.

A) LC analysis of unknown peak F



B) Ionization of unknown peak F



C) MS/MS analysis and ion fragments of unknown peak F

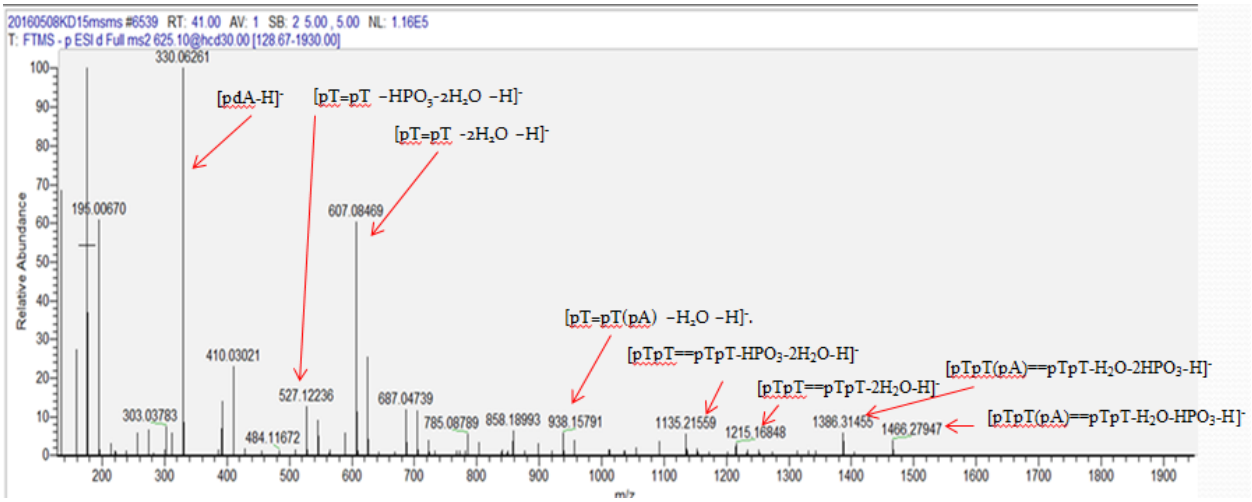
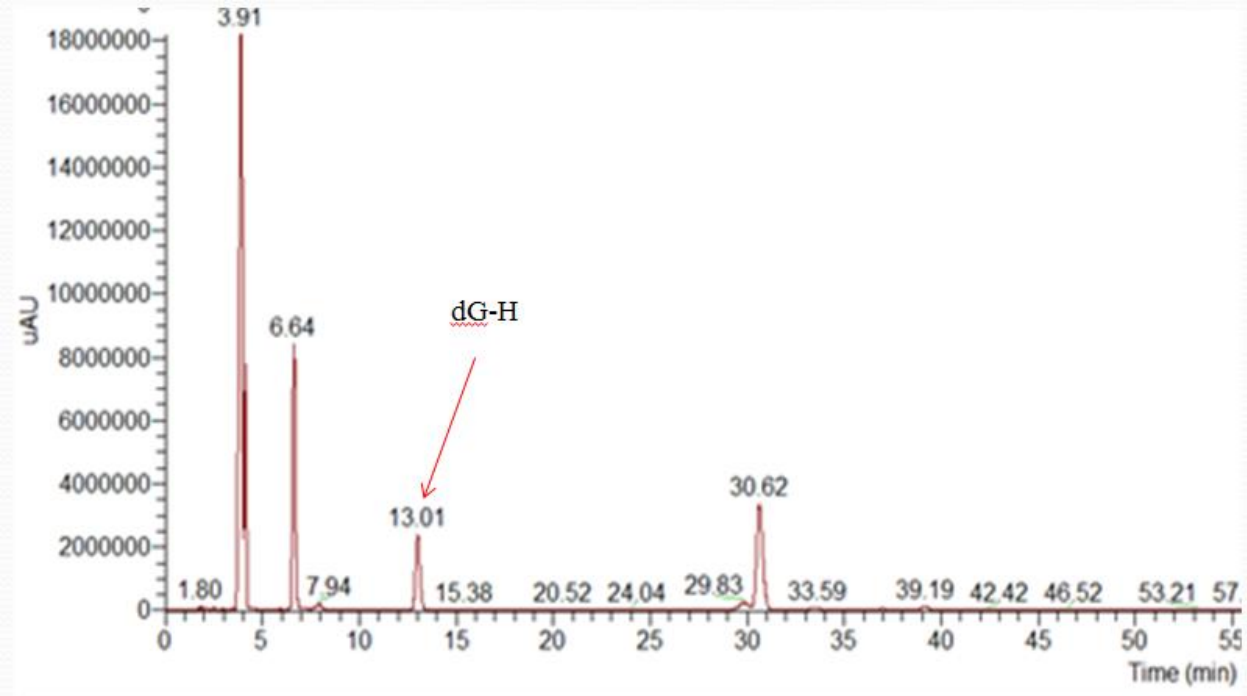


Figure 25: Analysis of unknown Peak F by A) LC-MS and B) and C) MS/MS results and product ions.

Analysis of Peak 1.

The analysis of unknown peak 1 (as seen in Figure 19) eluting between 12 to 15 minutes with a peak at around 13 minutes gave an m/z 266.09746 (Figure 26A) corresponding to the $[M-H]^-$ ion of dG. The m/z ion of of 266.09 was trapped and fragmented to give an ion of m/z 150.04 corresponding to the $[M-H]^-$ ion of guanine (Figure 26B). It is not clear where this digestion product could have come from, since nuclease P1 only produces nucleotide monophosphates, except if a deoxynucleoside is present at the 5'-end, as was the case for the oligodeoxynucleotides having dA at the end. There should not have been any dG at the 5'-end, unless it came from incomplete DNA synthesis.

A) LC-MS of unknown peak 1



B) MS/MS of unknown peak 1

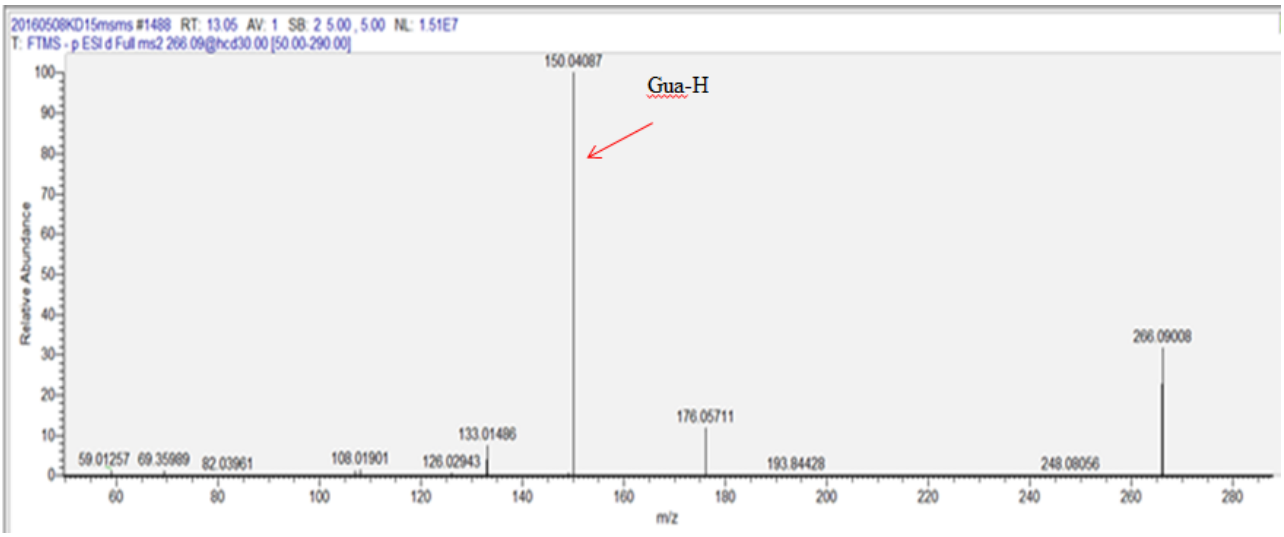
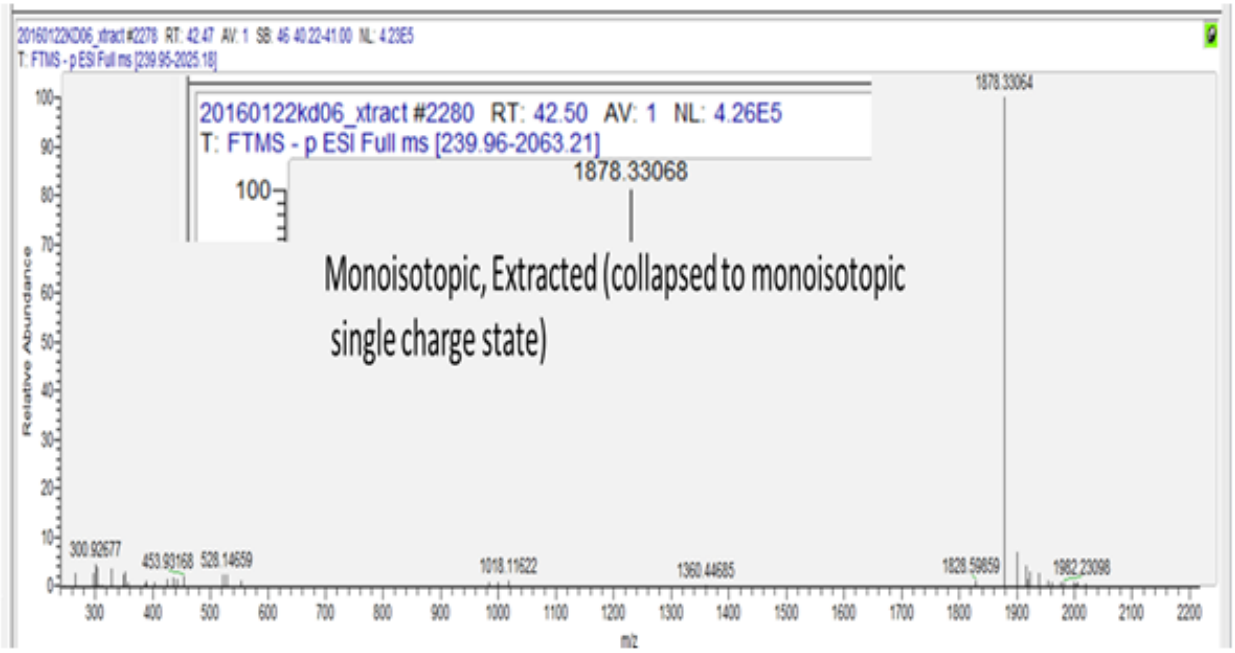


Figure 26: Analysis of peak unknown 1 by LC-MS and B) MS/MS.

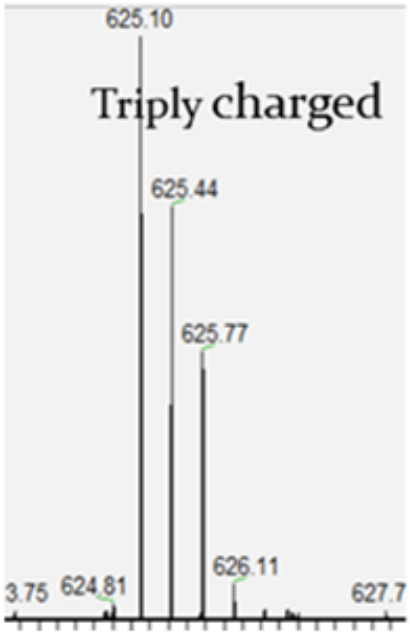
Analysis of Unknown Peak 2.

The analysis of unknown peak **2** eluting between 42.42 minutes gave a m/z ion of 1877.331 as shown in Figure 27A corresponding to a hexamer pTpT(A)=pTpT(A)). The peak at 42 minutes also gave a peak a m/z of 625.1 corresponding to the triply charged ion $[M-3H]^{3-}$ (Figure 27B) and m/z of 938.1524 corresponding to the doubly charged ion $[M-2H]^{2-}$ (Figure 27C) which was trapped and fragmented. Product ions were observed at m/z 330.06 corresponding to $[pdA-H]^-$; and m/z 607.1 corresponding to $[pT=pT\ -2H_2O\ -H]^-$; at m/z 1135.22 corresponding to $[pTpT==pTpT-HPO_3-2H_2O-H]^-$; m/z 1215.17 corresponding to $[pTpT==pTpT-2H_2O-H]^-$; m/z 1386.31 corresponding to $[pTpT(pA)==pTpT-H_2O-2HPO_3-H]^-$ and m/z 1466.28 corresponding to $[pTpT(pA)==pTpT-H_2O-HPO_3-H]^-$. For nuclease P1 to give such a product would require that both T's are forming *anti* photoproducts, and could result from the photodimerization of the two T's flanking the 5'-side of the anti-photoproduct leading to the double dimer pT(pA)==pT(pA) photoproduct. Subsequent hydrogen fluoride (HF) degradation studies by Innocent Harelimana of the Taylor group have provided evidence for the presence of both trans, anti and cis, anti CPDs of thymine in the product, a propose structure of the hexamer is shown in Figure 28.

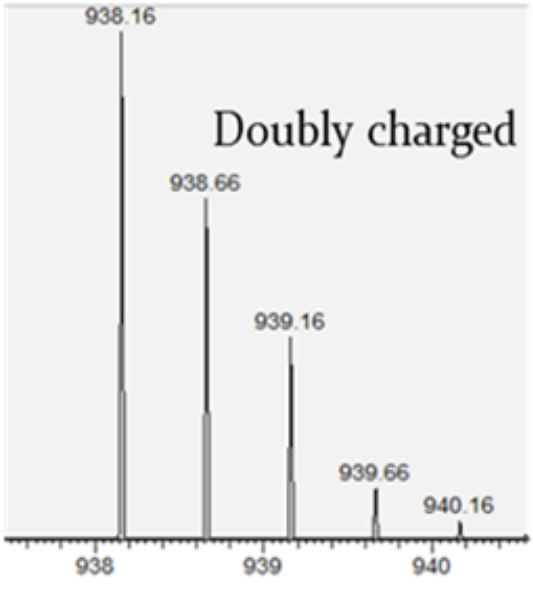
A) LC extracted and processed data for unknown peak 2.



B) Triply charged ions $[M-3H]^{3-}$ of 625.1.



C) Doubly charge ions $[M-2H]^{2-}$ of 938.16.



D) MS/MS data analysis.

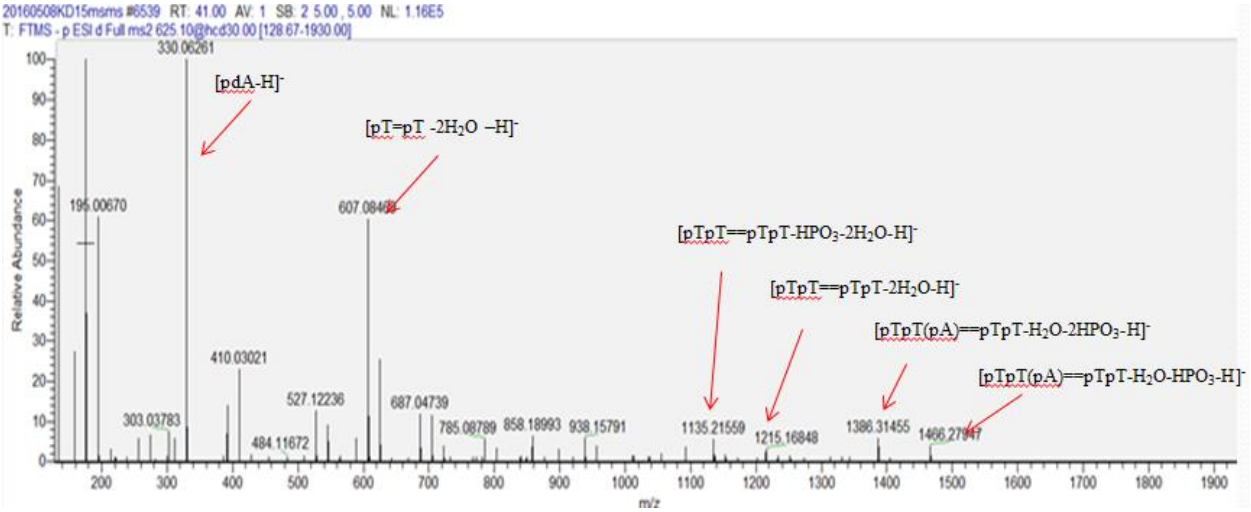


Figure 27: Analysis of unknown peak 2 by A) LC extracted and processed data for unknown peak 2 B) ionization of peak 2 by negative electrospray to give $[M-3H]^{3-}$ triple charge C) Ionization of peak two by negative electrospray to also give $[M-2H]^{2-}$ Doubly Charge D) MS/MS data analysis of unknown peak 2.

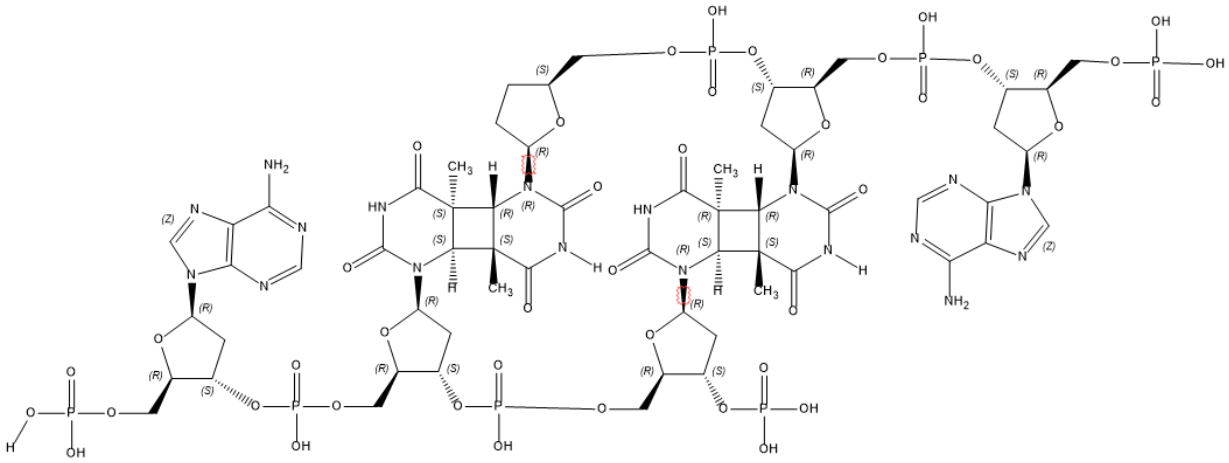
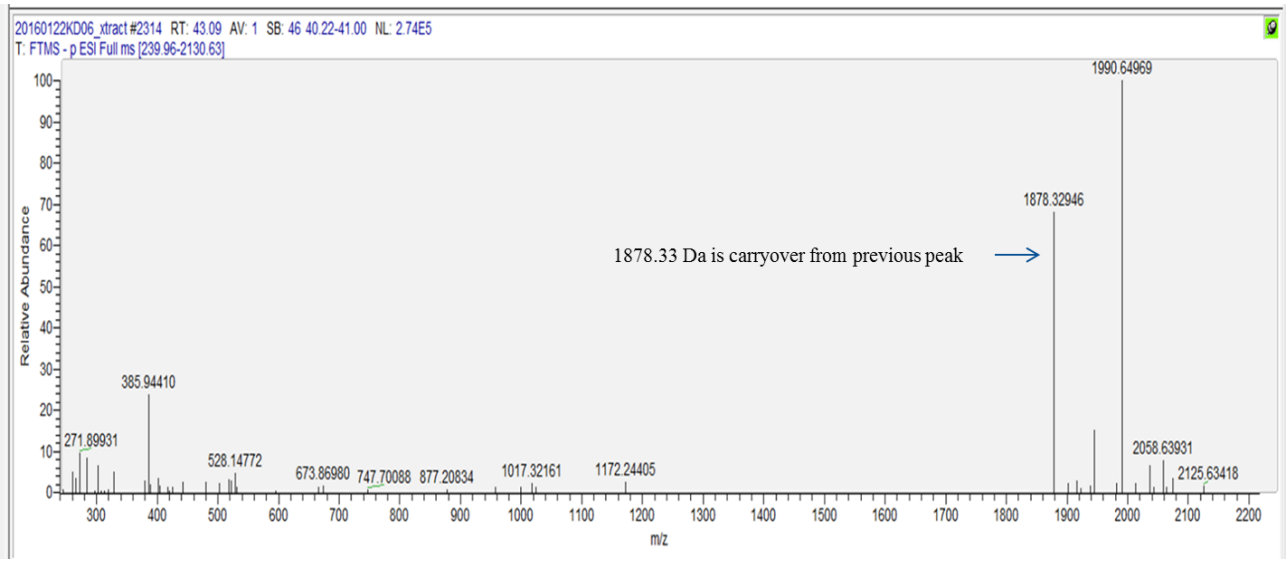


Figure 28: Suggested structure of the double dimer corresponding to a hexamer pTpT(A)==pTpT(A)).

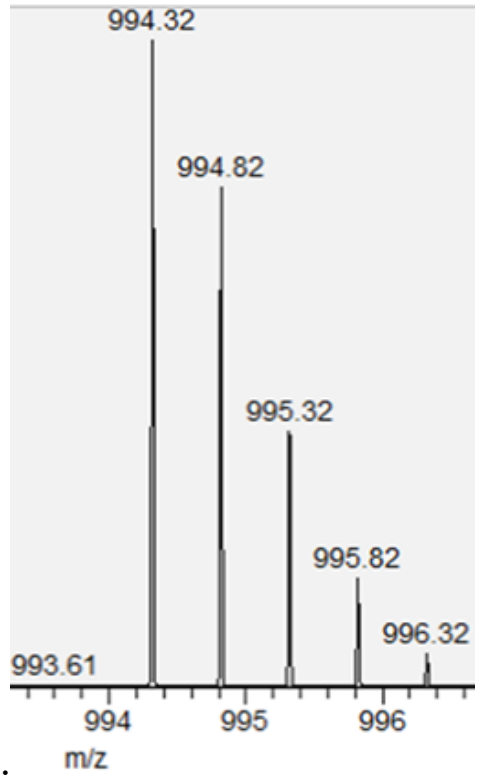
Analysis of Unknown Peak 3.

The analysis of unknown peak 3 eluting between 39 to 43 minutes gave a m/z of 1989.6496 as shown in Figure 29A and does not correspond to any expected hexamer. The peak at 43 minutes also gave a peak a m/z of 994 corresponding to the doubly charged ion $[M-2H]^{2-}$ as shown in Figure 29B which was trapped and fragmented. Product ions were observed at m/z 330.06 corresponding to $[pdA-H]^{-}$ as shown in Figure 29C.

A) LC-MS data extracted and processed



B) Double charge: ionization by negative electrospray to give $(M-2H)^{2-}$ of 994



.

C) MS/MS data

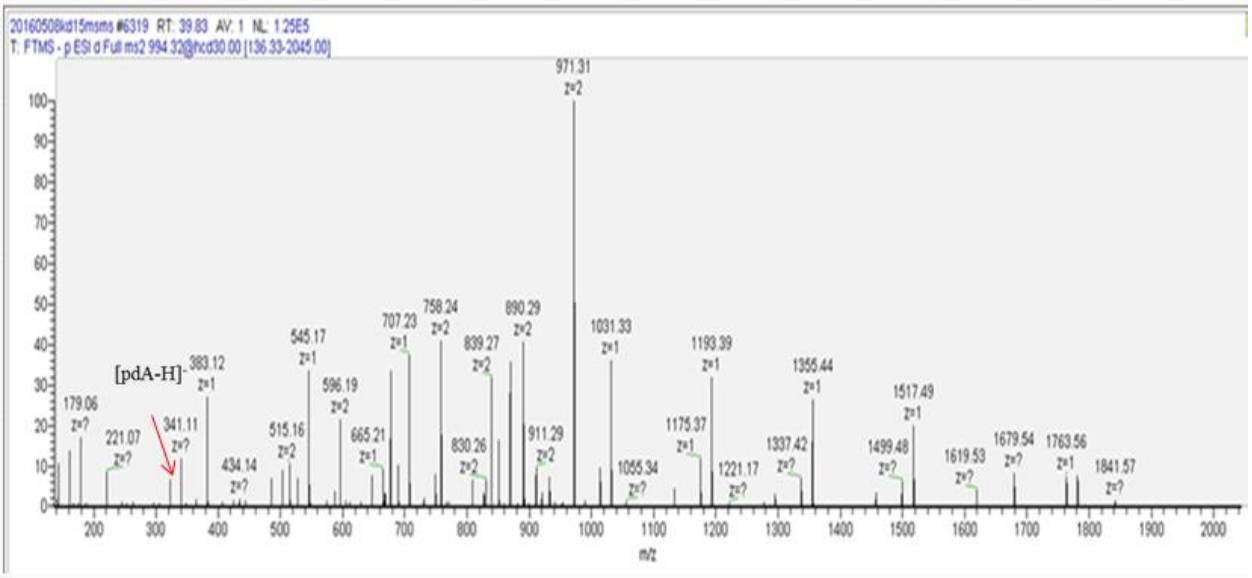


Figure 29: Analysis of unknown peak 3 by A) LC-MS and B) Ionization of unknown peak 3 by negative electrospray to give $[M-2H]^{2-}$ doubly charge C) MS/MS spectrum of unknown peak 3.

Summary of identified peaks:

The summary of the peaks identified in the new LC/MS method are shown in Table 1 below and as they appear in a chromatogram in Figure 30.

TEL26 Li^+			
Unknown	Elution Time	m/z observed	Name
	3.95	346.0565	pG
	4.18	322.0573	pT
	6.63	331.06846	pA
A	7.94	938.1641	pT=pT(pA)
1	13.01	267.097	dG
B	30	297.108	dA
C	33.59	1260.224	pT(pA)=pT(pT)
D	36.88	1269.236	pT(pA)=pT(pA)
E	37.3	1260.224	pT(pA)=pT(pT)
F	39.28	1269.236	pT(pA)=pT(pA)
2	42.42	1877.3306	pTpT(pA)=pTpT(pA)
3	43	1989.65	possible hexamer

Table 1 Summary of identified unknown peaks by LC-MS

All of the nuclease P1 digestion products A, C-F had been previously identified. As mentioned, peak B the front shoulder was remove and B was identified to be dA. Peak 1 was identified as dG, but its source is unknown. The unknown peak 2 m/z of 1877.331 at 42 (Figure 30) corresponding to TT(A)=TT(A) had been previously identified by MALDI analysis of an isolate peak, but no MS/MS data had been obtained. The MS/MS data supports the assignment of a double dimer, since no fragments corresponding to pT were observed. The unknown peak 3 m/z 1989 at 43 minutes will need to be further investigated because does not correspond to any expected hexamer. The peak at 43 minutes also gave a peak a m/z of 994

corresponding to the doubly charged ion $[M-2H]^{2-}$ as shown in Figure 29B which was trapped and fragmented.

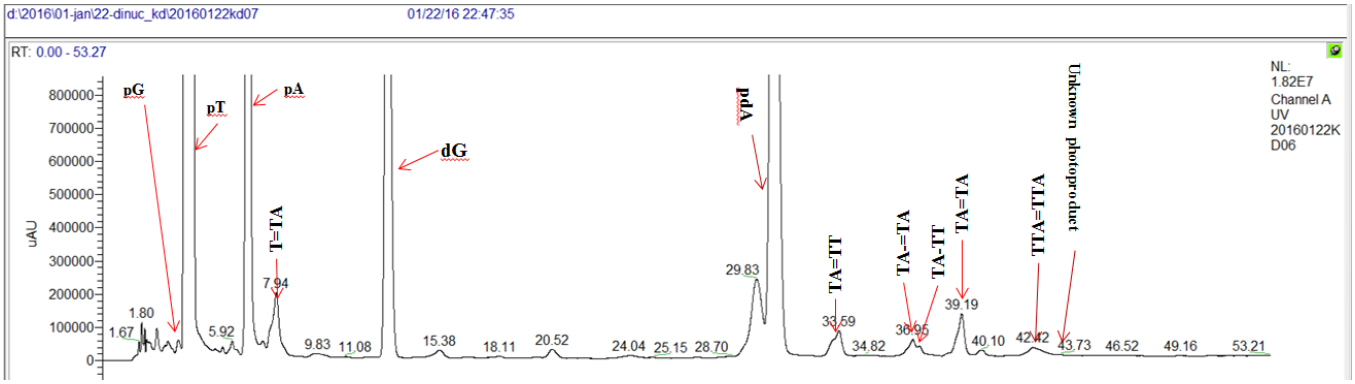
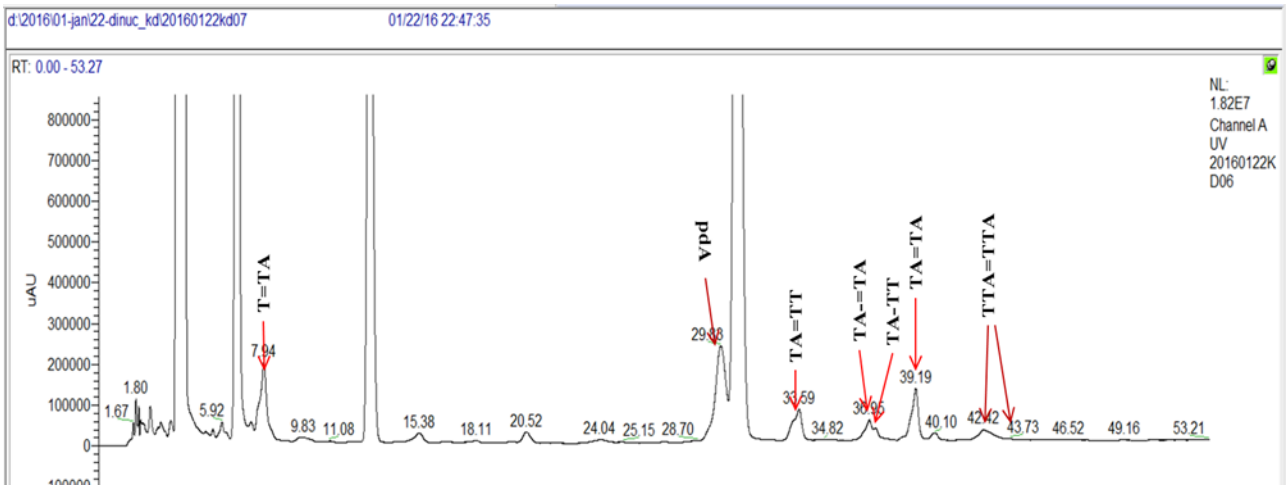


Figure 30: Summary of Photoproduct formation of Tel26 in Li^{+}

Comparison of photoproduct formation in Li+ with K^+.

Photoproduct formation in Tel26 in Li^+ were compared against the products formed in the presence of K^+. It was observed that there were more photoproducts produced in Tel26 in Li^+ (Figure 31A) than in K^+ as shown in Figure 31B, most notably a greater amount to TA=TT product, and the TTA=TTA photoproduct.

A) Tel26 in Li^+ UV Trace



B) Tel26 in K+ UV Trace

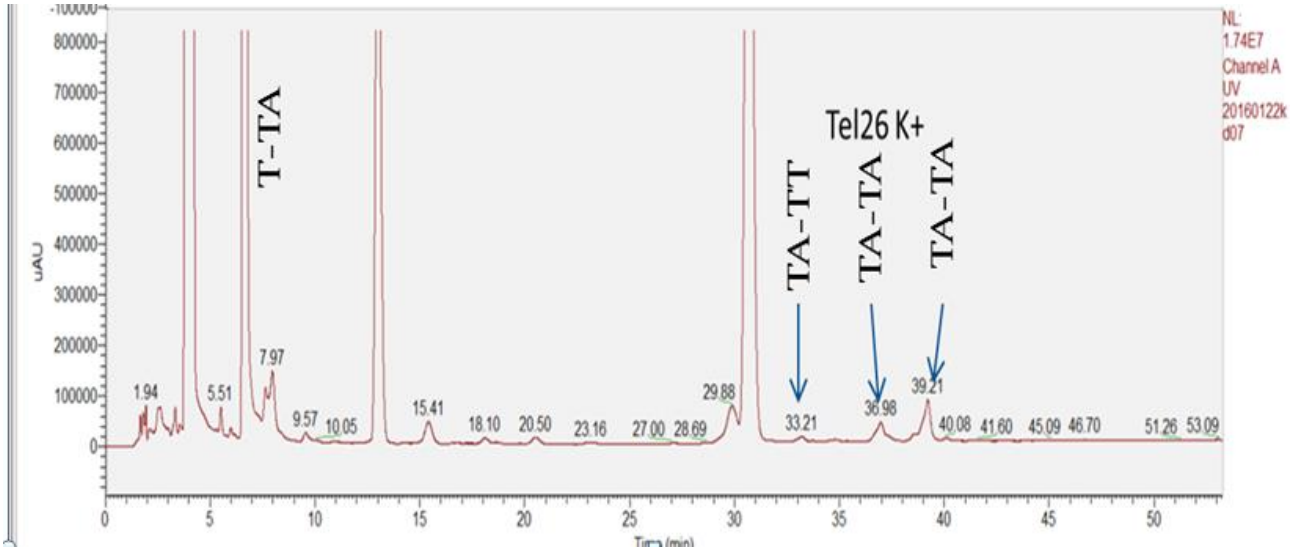


Figure 31: Peak analysis of A) Tel26 in Li+ and B) Tel 26 in K+ UV trace zoom in.

Experimental Section.

2.1 Materials and Methods

Oligodeoxynucleotides were purchase from Integrated DNA Technologies, Inc., (IDT, Coralville, Iowa) and Lifetechnologies.

Nuclease P1 (NP1) from penicillium citrinum, 10 mM ammonium formate and acetonitrile were purchased from Sigma (St. Louis, MO).

Acquity UPLC High Strength Silica (HSS) T3, particle size 1.8 µm 2.1x150 mm from Waters. Milli-Q (18.2 mΩ/cm) water obtained from Milli-Q water purification system.

2.2 Preparation of the Guanine-Quadruplexes

Oligodeoxynucleotides from Integrated DNA Technologies (IDT) were used without further purification. Typically, 50 uM ODN (IDT) in 10 mM Tris-HCl, pH7.5, with 150 mM KCl or LiCl; were heated at 95°C for 10minutes and then rapidly cooled down in ice.

2.3 UV Irradiation of Oligodeoxynucleotides (ODNs)

UVB irradiation was carried out immediately after sample preparation. G-quadruplex samples were irradiated on a bed of ice for 2-2.5 h at a distance of ~ 1 cm from the UVB lamp.

2.4 Digestion of the Oligodeoxynucleotides (ODN) by Nuclease P1

Typically, 1 ul of 1U/uL aqueous NP1 from Penicillium citrinum (Sigma) and 1 uL of 10 mM $ZnCl_2$ were added to 100 uL of 50 μM UVB irradiated sample and digested at 37°C for ≥36 h.

2.5 HPLC separation

HPLC separation and analysis were carried out on 1260 Infinity Quaternary LC System by Agilent Technologies. An Acquity UPLC High Strength Silica (HSS) T3,

particle size 1.8 µm 2.1 x 150 mm from Waters was used for reverse-phase HPLC. A gradient was used at a flow rate of 0.200 mL/min. Method: Mobile phase A (MPA) 10 mM aqueous ammonium formate; Mobile phase B (MPB) 50/50 (v/v) 10 mM aqueous ammonium formate and acetonitrile as shown in Figure 32.

HPLC settings				
	Time	% Mobil Phase A	%Mobil Phase B	Flow Rate
	0	100	0	0.2
	14	100	0	
	80	78	22	
	82	0	100	
	84	0	100	
	86	100	0	
Total Run time	105			
	MPA:	100% Ammonium Formate		
	MPB:	50/50 (v/v) Ammonium Formate + Acetonitrile		
	Column:	Waters Acquity UPLC HSS T3, 1.8 um 2.1 x 150 nm Silica Based bonded Pore size 100 Å		
	Part#	186003540		

Figure 32: Reverse phase HPLC gradient. MPA: 10 mM aqueous ammonium formate; MPB 50%/50% 10 mM aqueous ammonium formate and acetonitrile. Reverse phase HPLC gradient. MPA: 10 mM aqueous ammonium formate; MPB 50%/50% 10 mM aqueous ammonium formate and acetonitrile.

2.6 ESI-Mass Spectrometry (LC-MS) and MS/MS

ESI-MS and MS/MS experiments were carried out in the negative ion mode with the orbitrap mass spectrometer (Q Exactive-Plus) by ThermoScientific. A solution of 10 mM ammonium formate in water and acetonitrile was used as the spray solvent. The spray voltage was 2.5 kV. The capillary voltage and temperature were 25 Voltage and 250°C, respectively. MS/MS experiments were done by using normalize collision energy (NCE). The Mass window for precursor-ion selection was 3.0 m/z units. Approximately 2 scans were averaged for each spectrum. Each scan consisted of 100 microscans with a maximum injection time of 300 microscans. For MS/MS each scan consisted of 60 microscans for full scans with a maximum injection time of 300 microscans.

Chapter 3. Conclusion

Guanine-rich nucleic acids can form G-quadruplexes. These four-stranded complexes repeatedly occur in the human genome, playing an important role in gene regulation and serving as targets of drugs for cancer treatment. Crystallographic and nuclear magnetic resonance studies show the core of a G-quadruplex to be the G-tetrad, a planar assembly of four guanine bases networked via hydrogen bonds. G-tetrads stack one on another, with a cation located between adjacent tetrads in coordinating eight carbonyls of guanine bases for stabilization. Previous studies using biophysical approaches, such as CD, UV spectroscopy have outlined the thermodynamic profiles of the folding and unfolding of quadruplexes. The quadruplex and photoproducts formation varies with the cation species as is shown in this report. UV light is well known to cause adjacent pyrimidine base pairs next to each other in duplex DNA to covalently bond together to form cis, syn cyclobutane pyrimidine dimers (CPDs). More recently, thymines in G quartet structures associated with human telomeres, the sequence at the end of chromosomes, were found to form anti CPDs upon UV irradiation in the presence of K^+ and Li^+, but not Na^+. In part one of this thesis I described the development of a new LC method that better separates the CPD-containing nuclease P1 (NP1) digestion products of UV irradiated telomeric DNA. In the second part I focused on

directly coupling the LC method to MS/MS mass spectrometry and separated and identified the nuclease P1 digestion products of UV irradiated telomeric DNA in Li^+. The separated products were identified by their mass and fragmentation patterns. The photoproduct distribution in the presence of Li^+ was then compared against the products formed in the presence of K^+. Also, in the course of this thesis I have further characterized a new tandem photoproduct of Tel26 in Li^+ that appears to be a double dimer formed between the two flanking TTA segments of the quadruplex corresponding to pTpT(pA)=pTpT(A). This product likely forms from further photodimerization of the anti dimer of pT(pA)=pT(pA). In Figure 30 and Table 1 summarizes all the unknown peaks that were separated and identified by LC and MS/MS during this study. Also there is a summary of all fragment ions identified during this study in Table 2. I was able to observe that Tel26 in Li+ has more photoproducts and isomer formations when compared to Tel26 in K^+ as shown in Figure 31.

Further calculation of the mass accuracy was performed by subtracting the observed mass minus the theoretical mass divided by the observed mass times 1 million (ppm) for m/z accuracy (Table 3) to confirm.

Table 2: Assignments of negative fragment ions from tri (A: T=TA), tetra- (C & E: T(A)=T(T), D & F: T(A)=T(A)) and hexanucleotide (2: TT(A)==TT(A)).

m/z (negative ion)	Ion	Observed for unknown
125.03	Thy-H	
134.05	Ade-H	
150.042	Gua-H	1
241.08	dT-H	
250.094	dA-H	
266.09	dG-H	1
321.05	pdT-H	C, E
330.06	pdA-H	A, C, D, E, F, 2
346.06	pdG-H	
527.12	pT=pT $-HPO_3-2H_2O-H$	C, D, F
545.13	pT=pT-HPO_3-H_2O-H	C, D
607.06	pT=pT $-2H_2O-H$	C, D, E, F, 2
625.09	pT=pT-H_2O-H	C, D
849.19	pT=pT(pT) $-H_2O-HPO_3-H$	E
858.19	pT=pT(pA) $-HPO_3-H_2O-H$	D
929.14	pT=pT(pT) $-H_2O-H$	C, E
938.16	pT=pT(pA) $-H_2O-H$	C, D, F
1180.23	pT(pA)=pT(pT) $-HPO_3-H$	E
1135.22	pTpT==pTpT-HPO_3-2H_2O-H	2*
1215.17	pTpT==pTpT-$2H_2O-H$	2*
1386.31	pTpT(pA)==pTpT-$H_2O-2HPO_3-H$	2*
1466.28	pTpT(pA)==pTpT-H_2O-HPO_3-H	2*

* These peaks were also observed in the MS/MS of the m/z 634 ion of unknown peak F (T(A)=T(A).

Peak Name	Theoretical m/z	Observed m/z	Mass Accuracy ppm
pA	331.0682	331.06846	1
dA	251.1018	297.11	154852
pG	347.06	347.1	115
pT	322.0566	322.1	135
TA-T	938.2	938.2	0
TA-TA	1269	1269.22	173
TA-TT	1260.209	1260.224	12
TT-TT	1252.20578	1252.21109	4
Unknown 1	1877.313	1877.331	10
Unknown 2	1990.08	1990.65	286
Unknown 3	266	266.09	338

Table 3: Summary of the identified peaks mass accuracy.

References

1. J. S. Taylor. Unraveling the Molecular Pathway from Sunlight to Skin Cancer Acc. Chem. Res. 1994, 27 (3), pp 76-82.
2. Douki, T. and Cader, J. (2003) Inter-strand Photoproducts are produced in high yield within A-DAN exposed to UVC radiation. NAR, Vol.31, No.12, 3134-3142.
3. Shim Wook, Ji; Tan, Qiulin and Gu, Li-Qun (2009). Single-molecule detection of folding and unfolding of the G-Quadruplexes aptamer un a nanopore.
4. Phong Lan Thao Tran, Anne De Cian, Julien Gros, Rui Moriyama, and Jean-Louis Mergny (2012) Tetramolecular Quadruplex Stability and Assembly.
5. Su, Dian G. T.; Fang, Huafeng; Gross, Michael L.; Taylor, John S. (2009). Photocrosslinking of human telomeric G-quadruplex loops by anti cyclobutane thymine dimer formation.
6. Wong Alan., Wu, Gang. (2003). Selective Binding of Monovalent Cations to the Stacking G-Quartet Structure Formed by Guanosine 5′-Monophosphate: A Solid-State NMR Study.
7. J. Nový, M. Urbanová. Article Biopolymers, 2007. Setnicka V, Nový J, Böhm S, Sreenivasachary N, Urbanová M, Volka K. Langmuir. 2008 Jul 15;

24(14):7520-7. doi: 10.1021/la800611h. Epub 2008 Jun 14. Molecular structure of guanine-quartet supramolecular assemblies in a gel-state based on a DFT calculation of infrared and vibrational circular dichroism spectra.

8. Nový J, Böhm S, Králová J, Král V, Urbanová M. Biopolymers. 2008 Feb; 89(2):144-52. Formation and temperature stability of G-quadruplex structures studied by electronic and vibrational circular dichroism spectroscopy combined with ab initio calculations.
9. Risitano, Antonina., Fox Keith R. (2004). Influence of loop size on the stability of intramolecular DNA quadruplexes.
10. Smith, Jillian., Lu, Chen., Taylor, John-Stephen. (2014) Effect of Sequence and metal ions on UVB-induced anti cyclobutane pyrimidine dimer formation in human telomeric DNA sequences. Nucleic Acids Research, 42, No. 8 5007-5019.
11. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res., 34, 5402-5415.
12. Vorlícková M1, Chládková J, Kejnovská I, Fialová M, Kypr J. (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. Nucleic Acids Res. 2005; 33(18): 5851–5860.

13. Vorlíčková M1, Kejnovská I, Bednářová K, Renčiuk D, Kypr J. (2012) Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes.
14. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL & Wu JRA (1988) Highly conserved repetitive DNA sequence, (TTAGGG) n, present at the telomeres of human chromosomes. Proc Natl Acad Sci USA 85, 6622–6626.
15. Blackburn EH (2001) Switching and signaling at the telomere. Cell 106, 661–673.
16. Stewart SA & Weinberg RA (2006) Telomeres: cancer to human aging. Annu Rev Cell Dev Biol 22, 531–557.
17. d'Adda di Fagagna F, Teo SH & Jackson SP (2004) Functional links between telomeres and proteins of the DNA-damage response. Genes Dev 18, 1781–1799.
18. Wright WE, Tesmer VM, Huffman KE, Levene SD & Shay JW (1977) Normal human chromosomes have long G-rich telomeric overhangs at one end. Genes Dev 11, 2801–2809.
19. Belmont Philippe, Constant Jean-Francois, Demeunynck Martine (2001). Nucleic acid conformation diversity: from structure to function and regulation. Chem. SOc. Rev., 2001, 30, 70-81.

20. Bacolla Albino, Well Robert D (2004). Non-B DNA Conformations, Genomic Rearrangements, and Human Disease. Journal of Biological Chemistry 279 (46):47411-4.

21. Watson, J. D., & Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. Nature 171, 737–738

22. Bang, I (1910) Examination of the Guanylic acid. Biochem. Z. 26, 293-311.

23. Gellert M, Lipsett M.N., Davies D. (1962). Helix Formation by guanylic acid. Chemistry, 48, 2013-2018.

24. Henderson, E., Hardin, C.C., Walk, S. K., Tinoco Jr., I. and Blackburn, E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular strucutres containing guanine-guanine base pairs. Cell, 51, 899-908.

25. Neidle, S. (2009) The structure of quadruplex nucleic acids and their drug complexes. Curr. Op. Struct. Biol., 19, 239-250.

26. Williamson, J.R., Raghuraman, M.K. and Cech, T. R. (1989) Monovalent cation-induced structure of telomeric DNA: the G quartet model. Cell, 559, 871-880.

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

28. Risitano, A. and Fox, K. R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. Nucleic Acids Res., 32, 2598-2606.

29. Rachwal, P.A., FIndlow, I. S., Werner, J. M., Brown, T. and Fox, K. R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. Nucleic Acid Res., 35, 4214-4222.

30. Patel, D. J., Phan, A. T. and Kuryavyi, V. (2007) Survey and Summary Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. Nucleic Acid Res., 35-7429-7455.

31. Wang, Y. and Patel, D. J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. Strucutre, 1, 263-282.

32. Parkinson, G. N., Lee, M. P. and Neidle, S. (2002) Crystal Structure of parallel quadruplexes from human telomeric DNA. Nature. 417, 876-880.

33. Huppert, J. L., Bugaut A., Kumari, S. and Balasubramanian, S. (2008) G-quadruplex: The beginning and end of UTRs. Nucleic Acid Res., 36, 6260-6268.

34. Simonsson, T. (2001) G-quadruplex DNA Structures-Variations on a Theme. Biol. Chem., 382, 621-628.

35. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. and Neidle, S. (2006) Survey and summary quadruplex DNA: sequence, topology and structure. Nucleic Acid Res., 34, 5402-5415.

36. Pagano, B., Mattia, C. A. and Gaincola, C. (2009) Applications of Isothermal Titration Calorimetry in Biophysical studies of G-quadruplexes. Int. J. Mol. Sci., 10, 2935-2957.

37. Rawal, P., Kummarasetti, V. B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S. K and Chowdhury, S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. Genome Res., 16, 644-655.

38. Palumbo, S. L., Mammott, R. M., Uribe, D. J., Krotova-Khan, Y., Hurley, L. H. and Ebbinghaus, S. W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulation of promoter activity. Nucleic Acid Res., 36, 1755-1769.

39. Yang, Q., Zheng, Y. L. and Harris, C. C. (2005) POT1 and TRF2 cooperate to maintain telomeric integrity. Mol. And Cell. Biol., 25, 1070-1080.

40. Giannone, R. J., McDonald, H.W., Hurst, G. B., Shen, R. F., Wang, Y. and Liu, Y. (2010) The protein network surrounding the human telomere repeat binding factors TRF1, TRF2, and POT1. PLoS One, 5, el2407.

41. Wang Y. and Patel, D. J. (1992) Guanine residues in d(T2AG3) and d(T2G4) from parallel-stranded potassium cation stabilized G-quadruplexes with antiglycosidic torsion angles in solution. Biochem., 31, 8112-8119.

42. Sun, D., In vitro footprinting of promoter regions within supercoiled plasmid DNA. Methods Mol. BIol., 613, 223-233.

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

44. Luu, K. N., Phan, A. T., Kuryavy, V., Lacroix, L. and Patel, D.J. (2006) Structure of the Human Telomere in K+ Solution: An Intramolecular (3+1) G-Quadruplex Scaffold. Journal of the American Chemical Society, 128. 9963-9970.

45. Lim, K. W., Amrane, S., Xu, W., Mu, Y., Patel, D.J., Luu, K. N., and Phan, A. T. (2009) Structure of the human telomere in K+ solution: A Stable Basket-type G-quadruplex with only two G-tetrad layers. Journal of the American Chemical Society, 131, 4301-4309.

46. Rich, A., Nordheim, A. and Wang, A.H.J. (1984). The chemistry and biology of left-handed Z-DNA. Annu. Rev. Biochem., 53, 791-846.

47. Schaffitzel, C., Berger, I., Posberg, J., Hanes, J., Lipps, H.J. and Plückthun A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemmae macronuclei. Proc. Natl. Acdd. Sci. USA., 98, 8572-8577.

48. Davies, J. T. (2004) G-Quartets 40 Years Later: From 5'-GMP to molecular biology and supramolecular chemistry. Angew. Chem.. Int. Ed., 43, 668-698.
49. Schultze, P., Hud, N.V., Smith, F.W., and Feigon, J. (1999). The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by Oxytricha nova telomere repeat oligonucleotide d(G4T4G4). Nucleic Acid Res., 27, 3018-3028.
50. Lipps, H.j., and Rhodes, D. (2009). G-quadruplex Structures: in vivo evidence and function. Cell Biology, 19, 414-422.
51. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. Nature Chemistry, 5, 182-186.
52. Wang, Y., Taylor, J.S. and Gross, M. L. (2001) Isolation and mass spectrometric characterization of dimeric adenine photoproducts in oligodeoxynucleotides. CHem. Res. Toxicol. 14,738-745.

Appendix

Previous studies have discovered that photocrosslinking of T's can occur in structures other than quadruplexes providing evidence that implicates antiparallel reverse-Hoogsteen base-paired hairpins as the photoreactive conformation leading to the trans, anti T(A)=T(A) CPD. These results suggested that reverse Hoogsteen hairpin structures co-exist with G-quadruplexes.

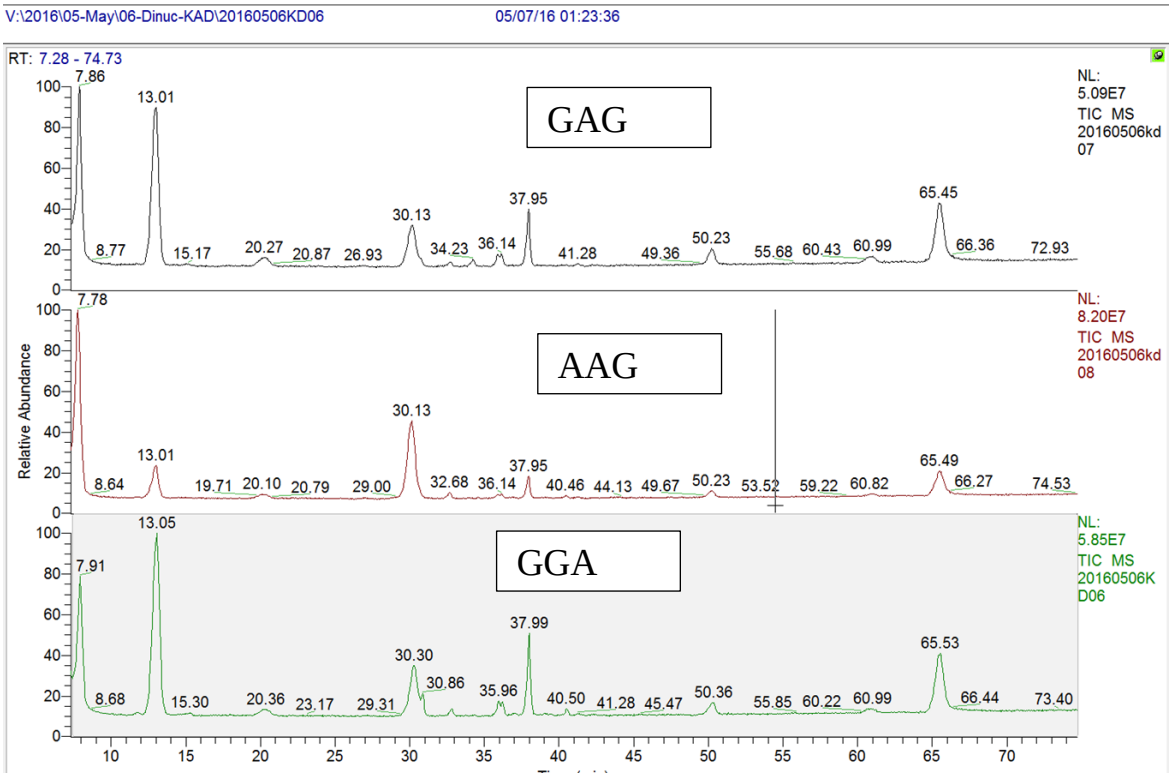
The following sequences of Tel21 (Cl-4-98, 19Apr2015) were prepared in potassium solution. Tel21 GGA; Tel21GAG; Tel21AAG and Tel21_3' were prepared in Dr. Taylor's lab at Washington University by Chen Lu. The sequences were then analyzed and processed by Claudia Posadas using the new LC and MS-MS methods.

C.

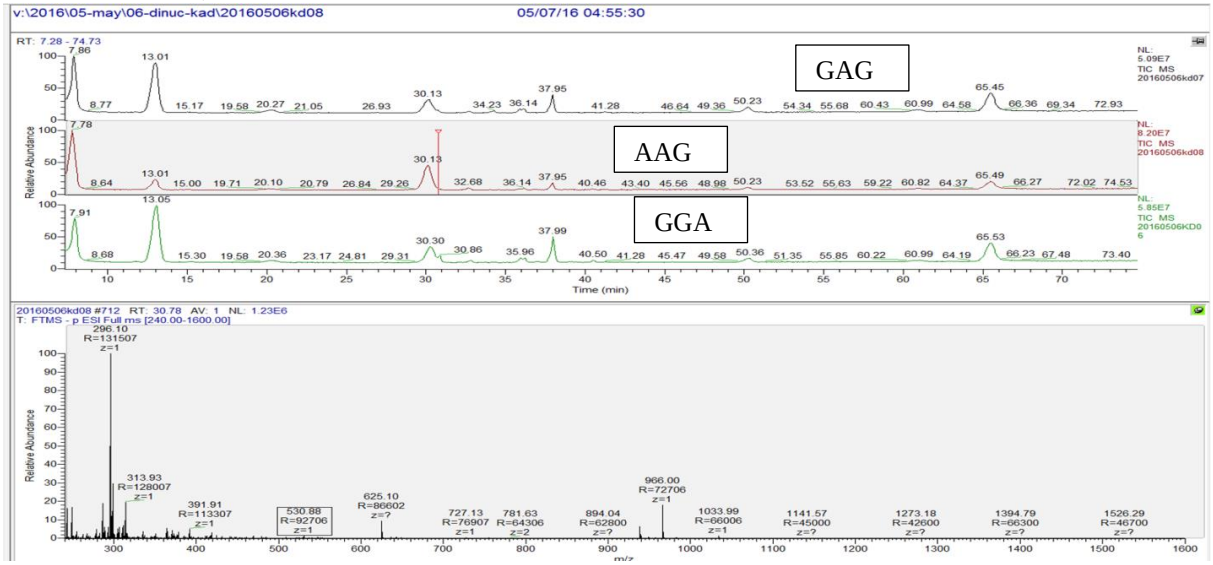
	X		Loop 1		Loop 2		Loop 3		Y
	1	4	7	10	13	16	19	22	
Tel26	AAA	GGG	TTA	GGG	TTA	GGG	TTA	GGG	AA
Tel22	A	GGG	TTA	GGG	TTA	GGG	TTA	GGG	
NF3		GGG	TTA	GGG	TTA	GGG	TTA	GGG	T
Tel15			TTA	GGG	TTA	GGG	TTA		
Tel15U²			TTA	GGG	TUA	GGG	TTA		
HP15WC			TTA	CGG	TTA	GCG	TTA		
Tel21_3'alt		GGG	TTA	III	TTA	GGG	TTA	III	
Tel21_3'con		GGG	TTA	GGG	TTA	III	TTA	III	
Tel21_5'alt		III	TTA	GGG	TTA	III	TTA	GGG	
Tel21_5'con		III	TTA	III	TTA	GGG	TTA	GGG	
Tel21_mix		IGI	TTA	GIG	TTA	IGI	TTA	GIG	
HP21GGA		GGA	TTA	GGA	TTA	AGG	TTA	AGG	
HP21GAG		GAG	TTA	GAG	TTA	GAG	TTA	GAG	
HP21AAG		AAG	TTA	AAG	TTA	GAA	TTA	GAA	
HP21WC		GCG	TTA	GGG	TTA	GGG	TTA	CGC	

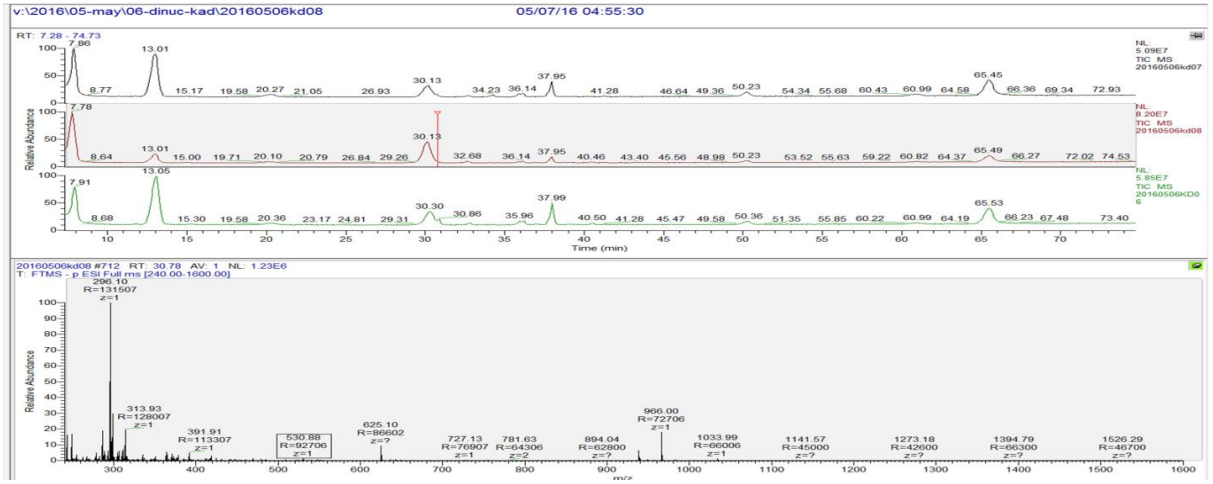
Figure 1. Tel21 sequences provided by Dr. Taylor's Lab at Washington University in St. Louis.

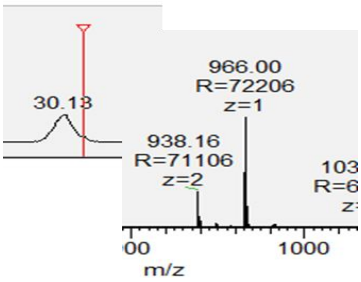
1. LC-MS Trace of GAG, AAG, GGA



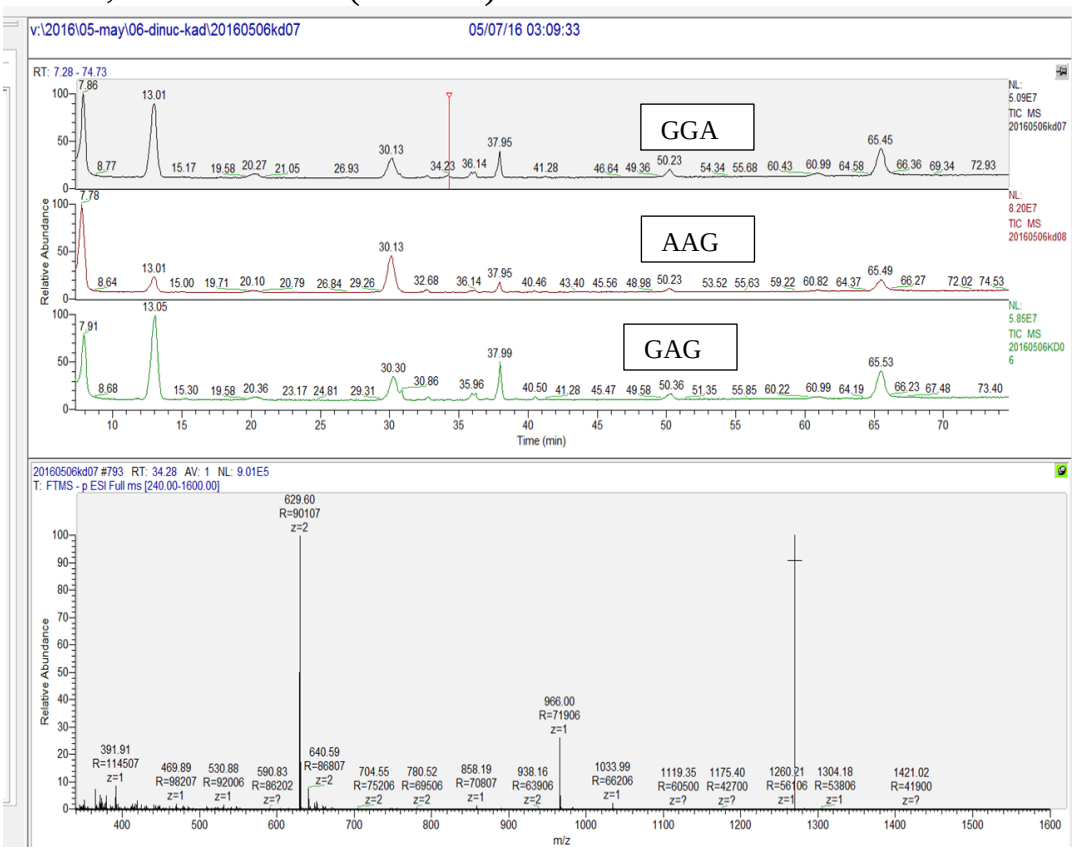
2. LC-MS and MS/MS Trace of GAG, AAG, GGA looking at peak with retention times at 30.78 and 30.86 minutes
   938.16, R=71106 z=2
   966.00, R=72206 z=1



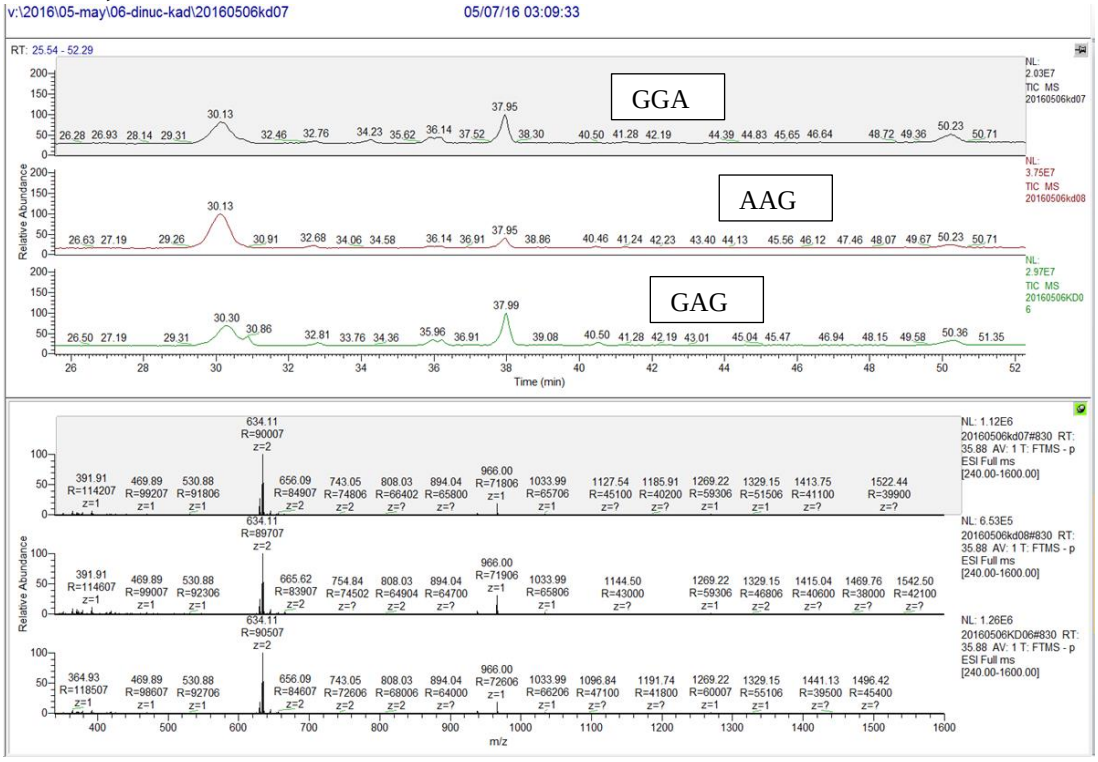




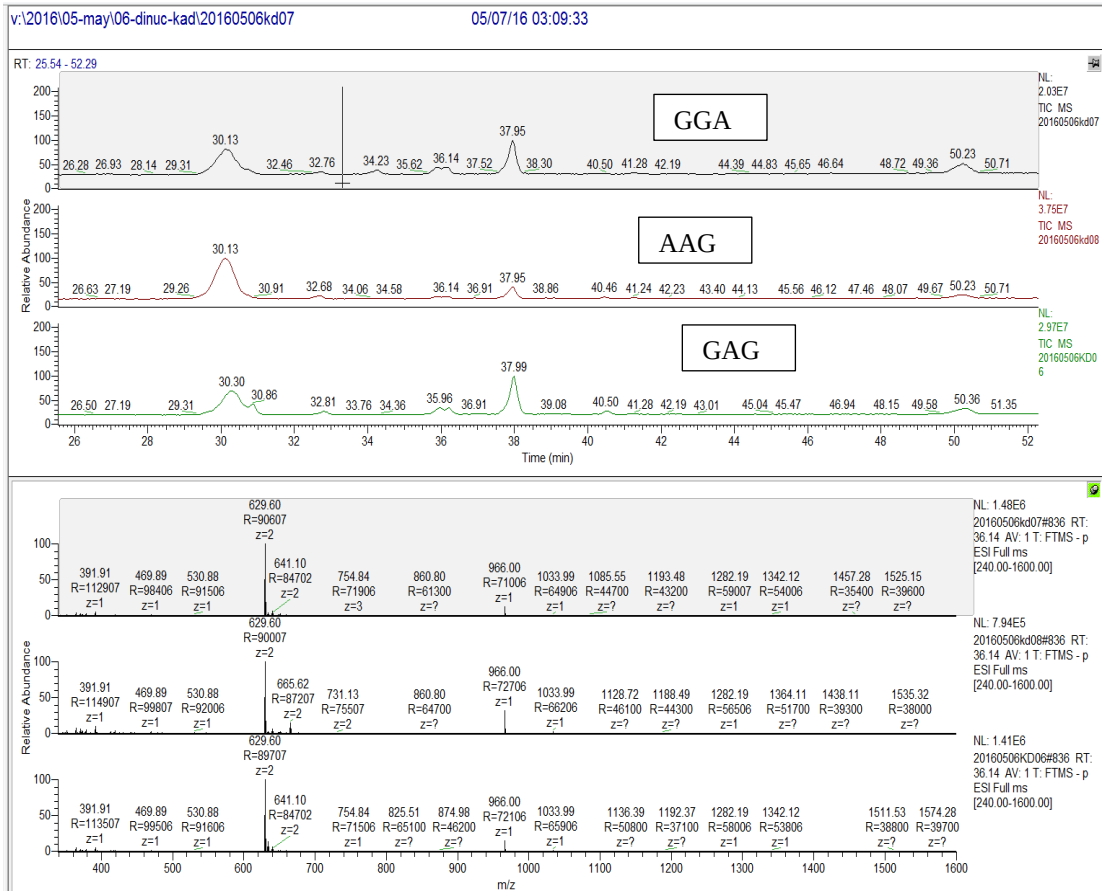
3. LC-MS and MS/MS Trace of GGA, AAG, GAG looking at peak with retention times at 34.28 minutes
   629.60, R=90107 z=2 (tetramer)



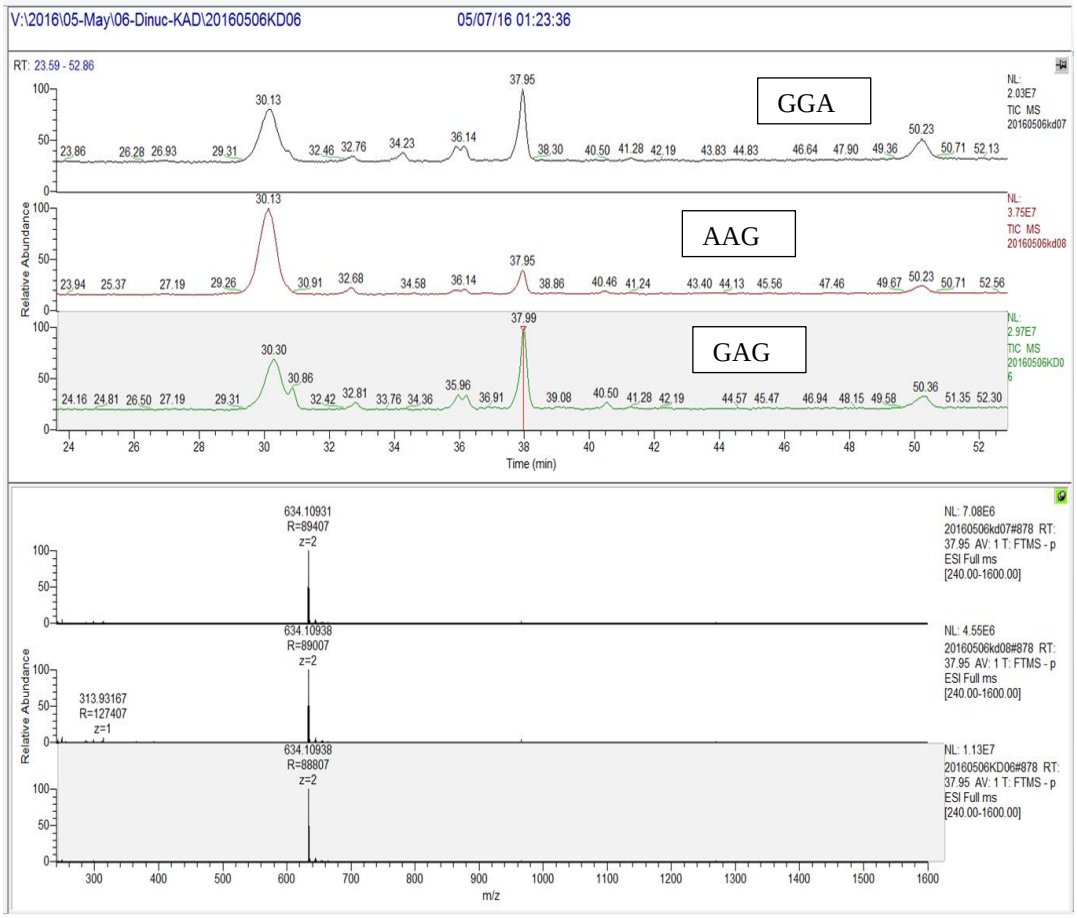
4. LC-MS and MS/MS Trace of GGA, AAG, GAG looking at peak with retention times at 35.88 minutes.
   634.11, R=90007 z=2



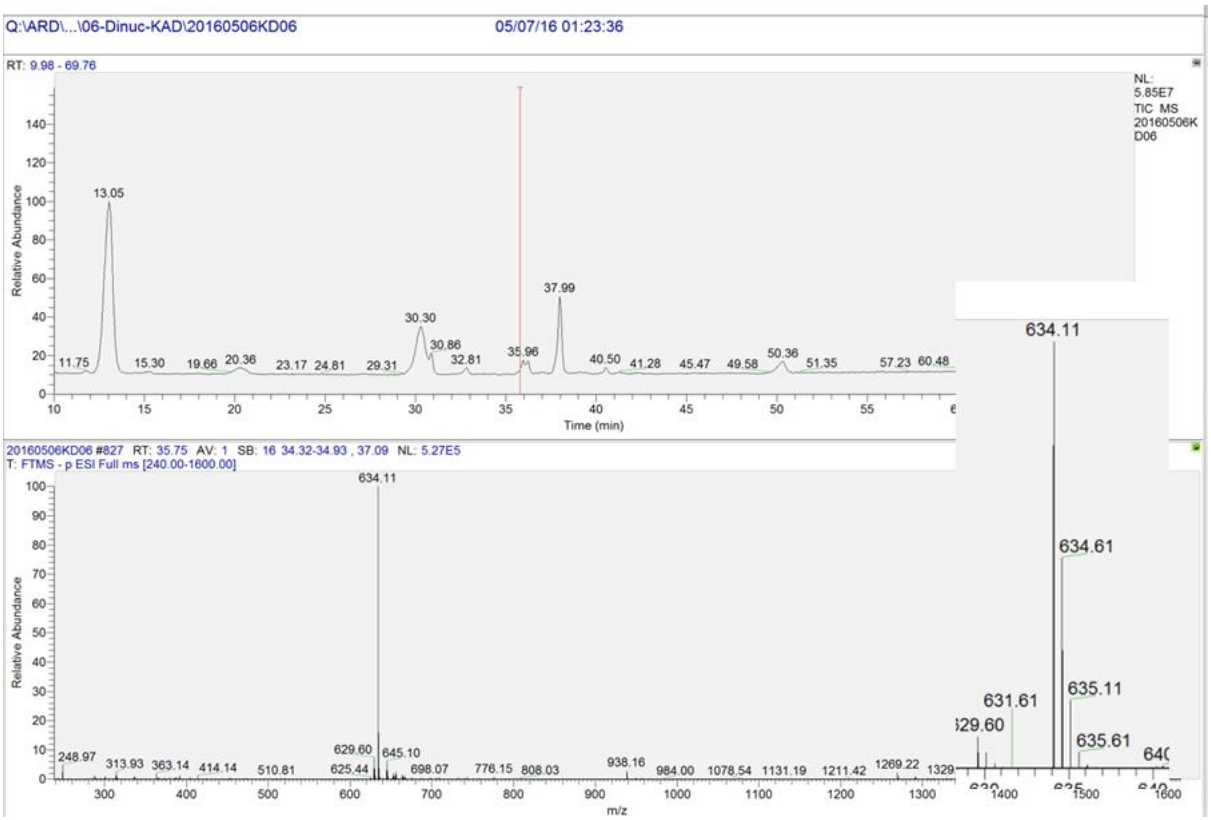
5. LC-MS and MS/MS Trace of GGA, AAG, GAG looking at peak with retention times at 36.14 minutes.
   629.60 R= 90607 z=2



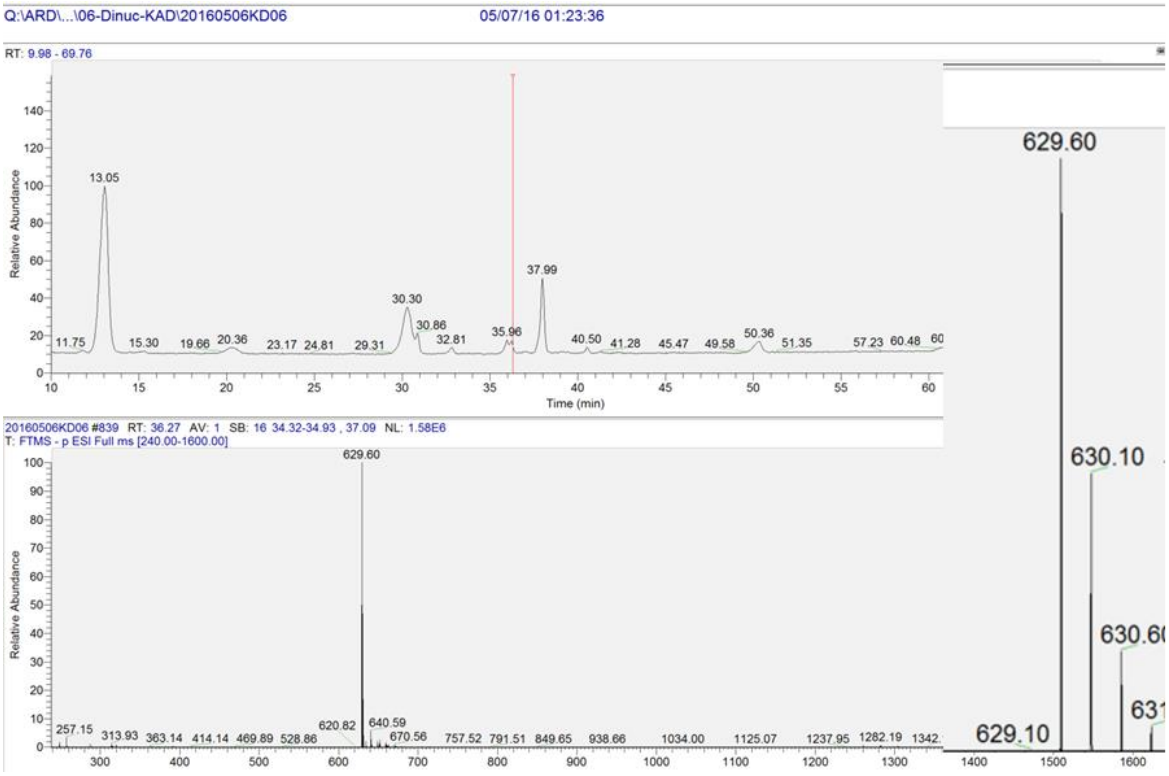
6. LC-MS and MS/MS Trace of GGA, AAG, and GAG looking at peak with retention times at 37.95 minutes.
   634.10931 R= 10931 z=2



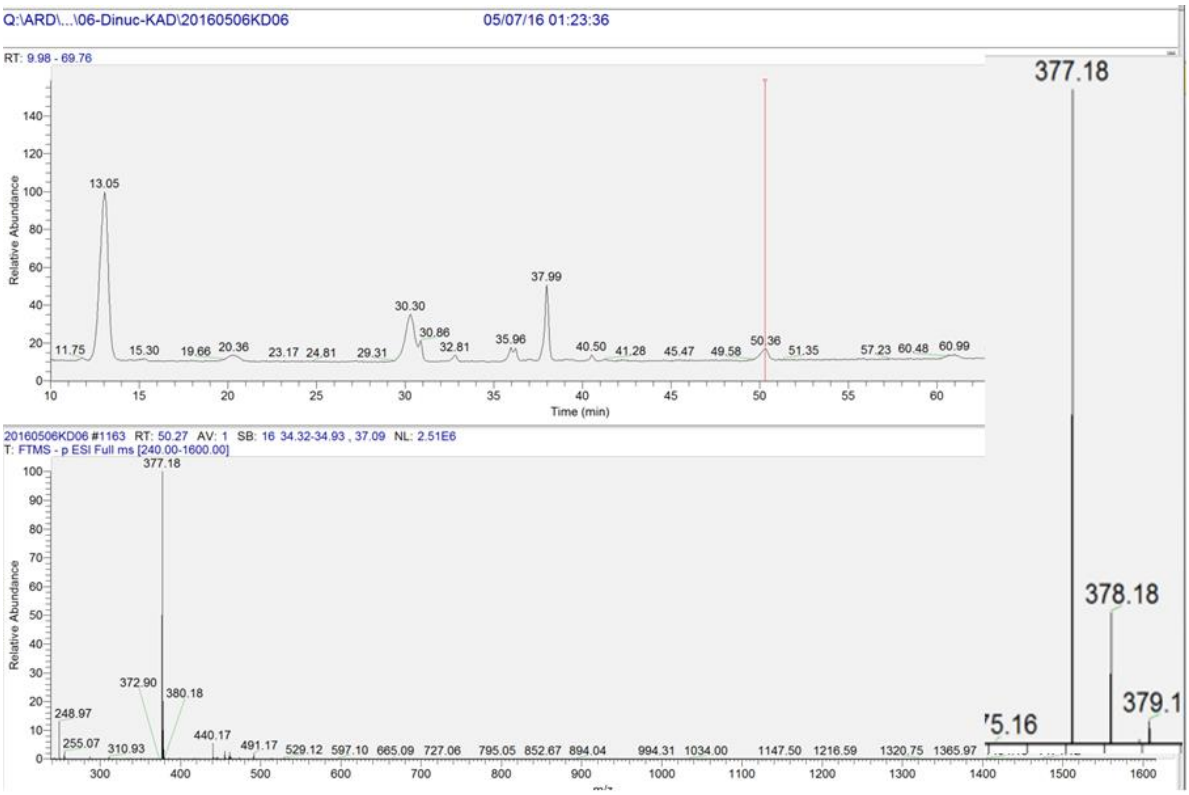
7. LC-MS and MS/MS Trace of GGA, AAG, and GAG looking at peak with retention times at 35.75 minutes.



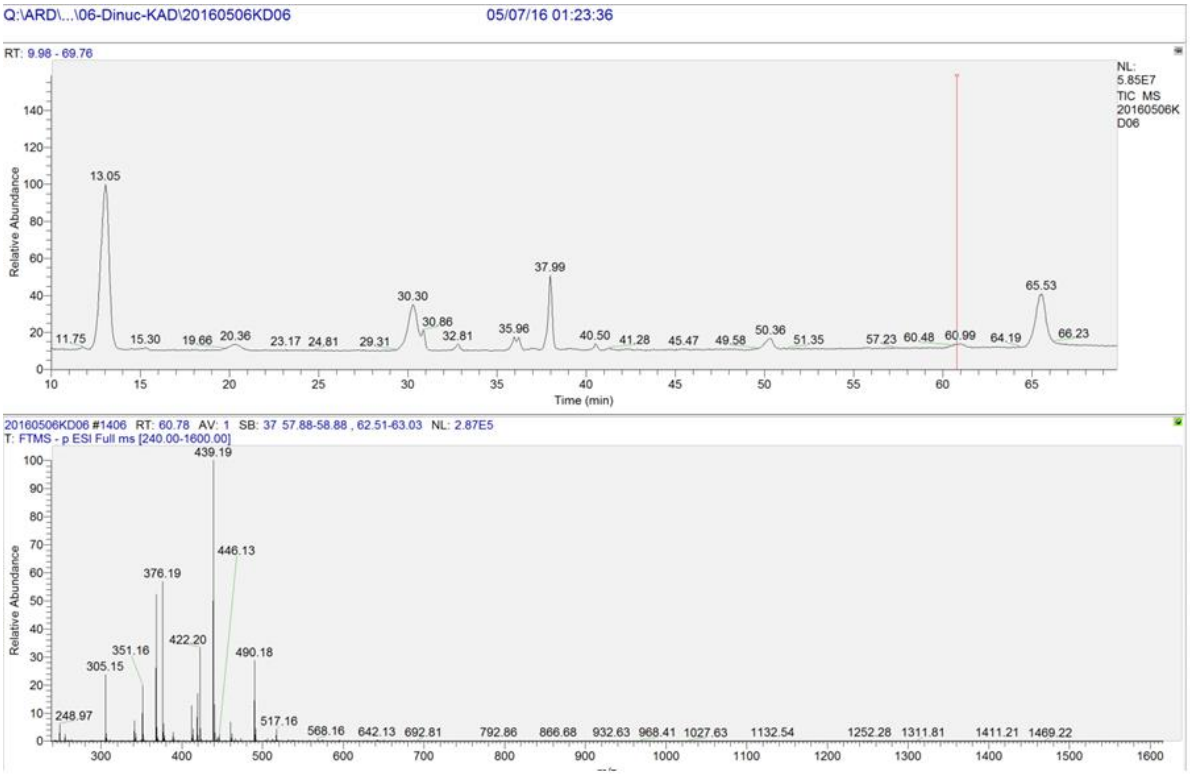
8. LC-MS and MS/MS Trace of GGA, AAG, and GAG looking at peak with retention times at 36.27 minutes.



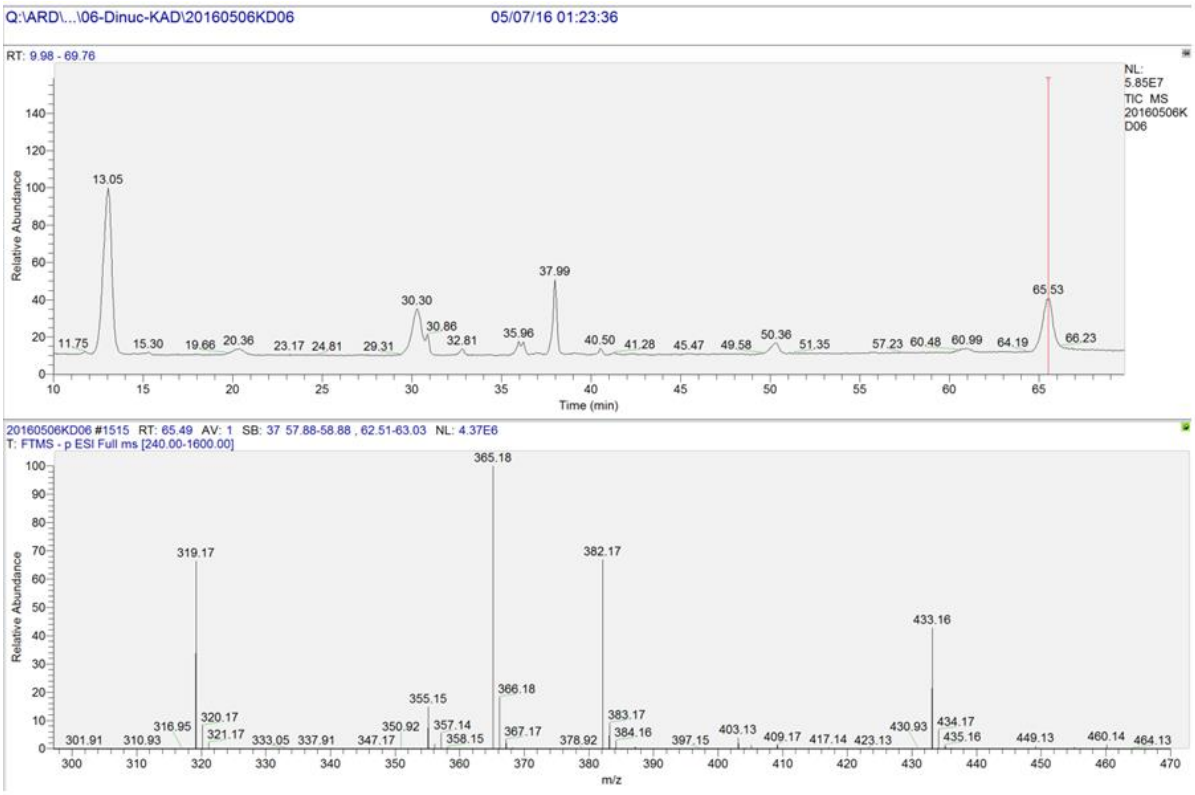
9. LC-MS and MS/MS Trace of GGA, AAG, and GAG looking at peak with retention times at 50.27 minutes.

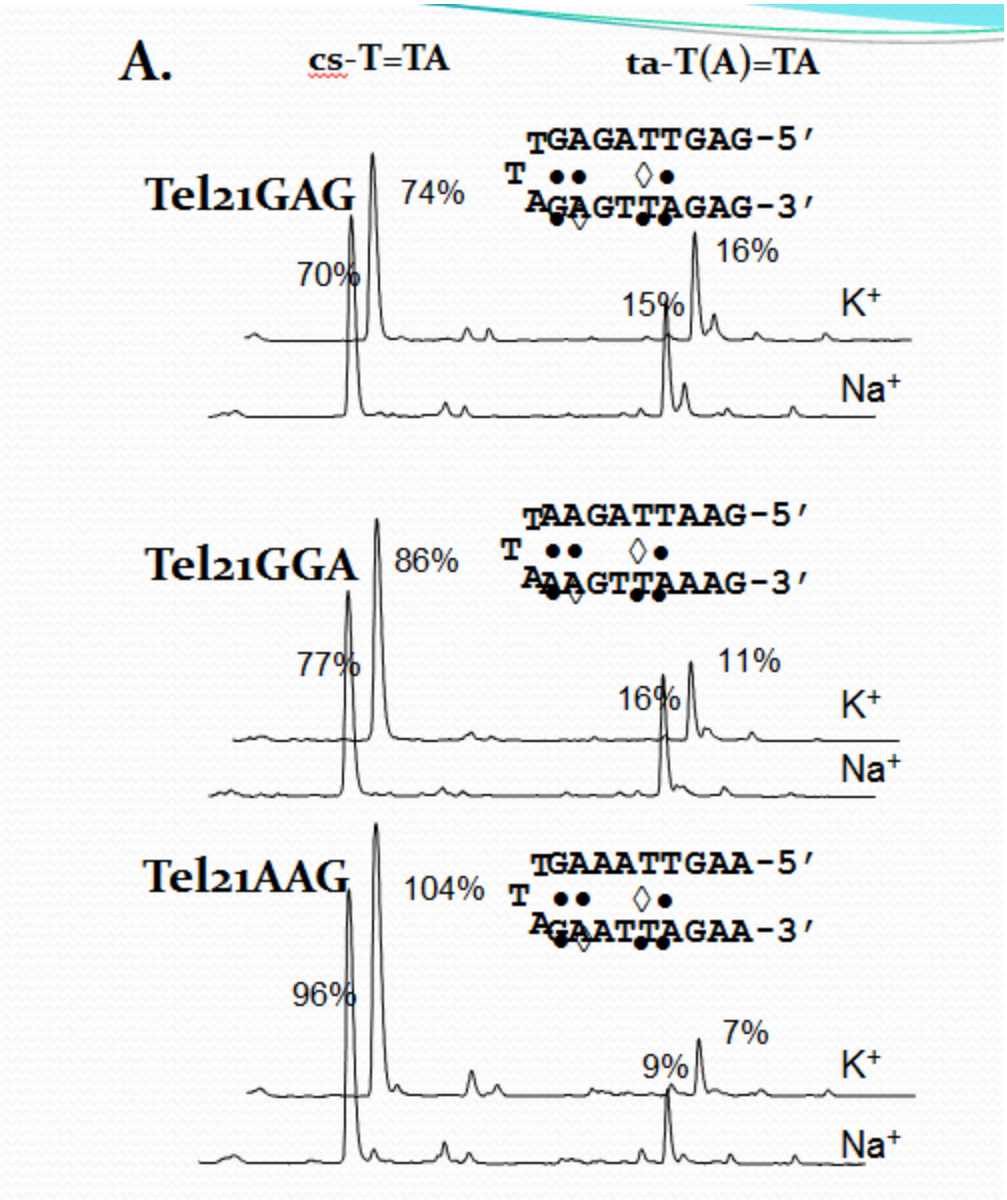


10. LC-MS and MS/MS Trace of GGA, AAG, and GAG looking at peak with retention times at 60.78 minutes.



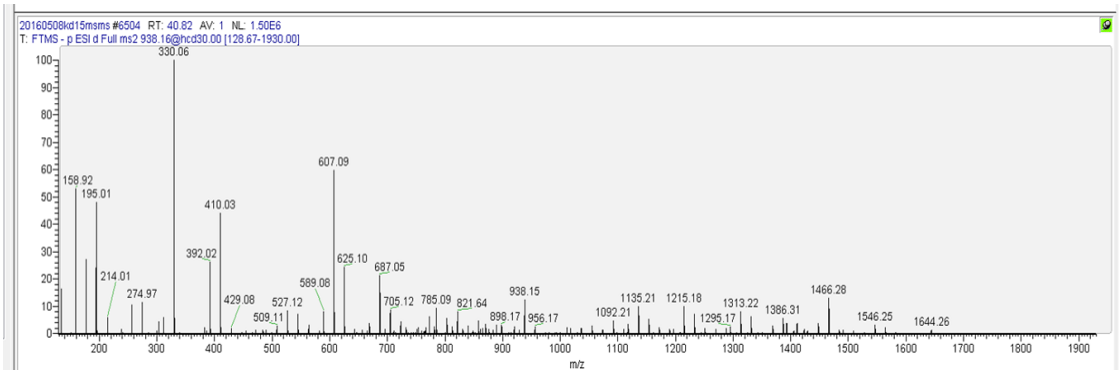
11. LC-MS and MS/MS Trace of GGA, AAG, and GAG looking at peak with retention times at 65.49 minutes.



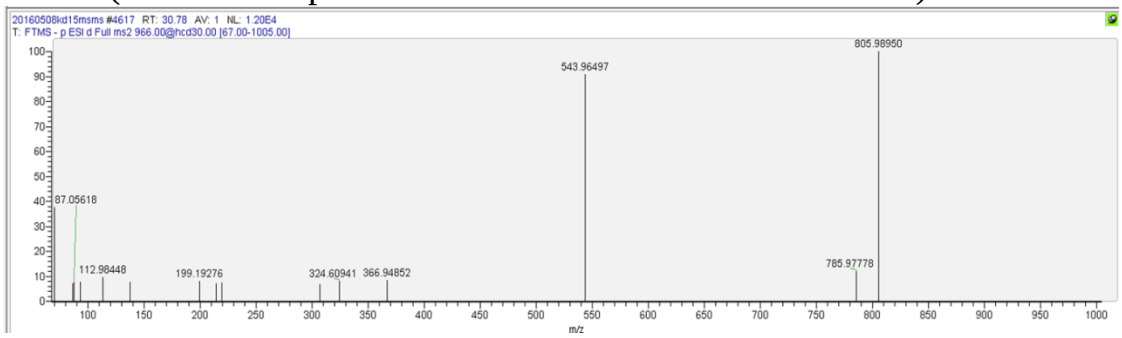


Sequence and image provided by Washington University-St. Louis from Dr. Taylor's lab.

12.MS/MS Trace of pTel26 Li+ of 938 m/z peak at 41 minutes (isomer)



13.MS/MS Trace of p Tel26 Li+ of 966Da no real UV peak, right after peak at 31min (trace of 938 present at 31 min but too low for MSMS).



14.MS/MS Trace of p Tel26 Li+ of 1990Da peak at 39.83 minutes.