Probing the Chemical Space for Non-Promiscuous Nucleic Acid Binders

by

Jorge Sandoval

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
At the University of Michigan
2023

Doctoral Committee:

Associate Professor John Montgomery
Assistant Professor Sarah Keane
Associate Professor Thomas Minehan
Associate Professor Jayakrishnan Nandakumar

Jorge Sandoval

jorgsand@umich.edu

ORCID iD: 0000-0003-4357-6862

Acknowledgments

I would like to start by thanking Dr. Thomas Minehan, who has been an enduring support, mentor, and friend from my time as an undergraduate and well into my graduate work, thank you, from the bottom of my heart. I would also like to thank the Chemical Biology program, the University of Michigan, and especially Dr. Anna Mapp for the wonderful opportunity to grow as an individual scientist among a great community of scientists. I'd like to thank my committee members for their mentorship and for pushing me to grow as a scientist as well. I would especially like to thank Dr. John Montgomery for his incredible patience and support during the most challenging moments I had during this time. I would also like to thank Dr. Edward Blumenthal for his kindness and for sparking my interest in chemical biology.

To my wife, Julie, thank you for always supporting me and for always encouraging me to strive for more. To my children, Kai and Ronin, thank you for bringing light back into my life during an incredibly dark time. And to My family, especially my father, who came from nothing and built an amazing life for me and my siblings, and mother, who didn't live to see this day; I dedicate this to you, and I hope that I can use everything I've learned during this time to do good and make you proud.

Table of Contents

List of Abbreviations

AA – amino acid

AAV – adeno-associated virus

AG – aminoglycosides

ACNU – Nimustine

ASPGR – Asioglycoprotein receptor

AGO2 – argonaute2

ASO – Antisense oligonucleotide

BPs – binding proteins

bps – base pairs

B-Raf – rapidly accelerated fibrosarcoma kinase B

cat-ELCCA – catalytic- enzyme-linked click chemistry assay

CCG – center for chemical genomics

CD – circular dichroism

CT DNA – calf thymus DNA

cMyc – cellular myelocytomatosis

DDD – drew Dickerson dodecamer

DCM – dichloromethane

DGCR8 – DiGeorge Syndrome Critical Region 8

DNA – Deoxyribonucleic Acid

dsDNA – double-stranded DNA

EMA – European Medicines Agency

FDA – Food & Drug Administration

FID – flame ionization detector

FRET – Fluorescence resonance energy transfer

GalNAc – N-acetylgalactosamine

G/C/A/T/U – Guanidine/ Cytosine/ Adenine/ Thymine/ Uracil

GWAS –genome-wide association studies

h/ min – hour/ minutes

hATTR / TTR– transthyretin/ gene

HR – homologous recombination

HMGA2 – High Mobility Group AT-Hook 2

HR-MS –High-resolution mass spectrometry

HFIP- Hexafluoroisopropanol

HRP – horseradish peroxidase

HTS – high throughput screening

ITC – Isothermal titration calorimetry

IC50 – half maximal inhibitory concentration

Kd – dissociation constant

LCA – Leber congenital amaurosis

LNP – Lipid Nanoparticle

LRC –long-range correlations

NMR – Nuclear magnetic resonance

NRPS- PKS – non-ribosomal peptide synthetase- polyketide synthase

NFAT1 – nuclear factor of activated T cells 1

NHS – N-Hydroxysuccinimide

NHEJ – non-homologous end joining

NP/ NPE – natural products/ extracts

MGBs – minor groove binders

mTet – methyl tetrazine

miRNA – MiR21-miR21- microRNA21

PAZ – PIWI, AGO, Zwille protein domain cluster

PDCD4 –Programmed Cell Death 4

PEG – polyethylene glycol

PTEN – Phosphatase and tensin homolog

PCGs –protein-coding genes

PIWI – P-element Induced WImpy testis in Drosophila

RNA – Ribonucleic Acid

RT-PCR – Reverse transcription-quantitative polymerase chain reaction

RPE65 – retinal pigment epithelium

RISC –RNA-induced silencing complex

siRNA/ mRNA – silencing RNA/ messenger RNA

SAR – structural-activity relationship

SPPS – solid phase peptide synthesis

SNAc – N-acetylsteamine

STAT6 – Signal transducer and activator of transcription 6

SPR – Surface plasmon resonance

TCO – transcyclooctene

TALENs – Transcription Activator- Like Effector Nucleases

TLC – thin layer chromatography

TE –thioesterase

TMP1 – Vacuole Membrane Protein 1

tRNA – transfer RNA

UV- Vis – ultraviolet visibility spectroscopy

VEGFR – Vascular Endothelial Growth Factor A

ZFN – Zinc Finger Nucleases

List of Figures

List of Schemes

Abstract

Our understanding of DNA, RNA & RNA motifs has catapulted in the last few decades, and as we have begun to better understand their role in disease, we've seen an influx of therapeutic tools & technologies that are actively changing how we understand these molecules and how our overall knowledge translates into novel tools or therapeutic strategies. To further our knowledge of nucleic acid biology, novel probes were developed for reassessing the targetability of select nucleic acid motifs. In particular, the two nucleic-acid-based target sites studied were: the major groove pockets of select DNA sequences and a select library of pre-miRNAs.

Corrupted DNA sequences or dysregulation of DNA transcription has been implicated in almost every disease, despite being highly conserved and regulated. DNA's localization in the nucleus makes it incredibly difficult to make DNA-targeted probes & therapeutics. The biological events at the major groove of DNA are of particular interest because several regulatory proteins and transcription factors associate at specific major groove unit sequences. Through the rational design of 2nd generation probes, we were able to derivatize our crystal violet library and confirm a previous observation that an alkene or alkyne linker is necessary for major groove binding. Alkyl & aliphatic linkers were found to be strongly associated in the minor groove, and thereby also observing that not all crystal violet analogs discriminately associate with the major groove. The molecules described do not show significant levels of sequence selectivity, however, a mild affinity

for tighter major groove sequences from bulky trimeric analogs might reveal a structural basis for achieving some level of shape recognition and subsequently, sequence selectivity.

MicroRNAs (miRNAs) are a conserved, non-coding class of RNAs that regulate more than 60% of protein expression via RNA Induced Silencing Complex (RISC). Dysregulation of miRNA expression has been linked to diabetes, obesity, cancer, cardiovascular, neurodegenerative, and other diseases. These correlations have made probing miRNAs an attractive target for therapeutic innovation. However, the development pre-miRNA specific probes have remained poorly explored largely due to their inherent challenges as targetable substrates. By employing a novel screening technique, catalytic-enzyme linked click chemistry assay (cat-ELCCA), and an activity-guided fractionation strategy, we were able to isolate and identify a series of surfactin-based analogs, the hermicidins, and their novel ability to bind premiR21 and inhibit Dicer mediated maturation. These probes may prove to be useful tools in studying the processes regulating pre-miRNA biogenesis. There are currently FDA-approved, surfactin-based drugs on the market, therefore, these findings may provide the basis for a novel therapeutic strategy by modulating dicer-mediated pre-miRNA maturation. Furthermore, our efforts in the total synthesis of similar surfactin-based probes were successful, providing preliminary insight into structural features necessary for their inhibitory activity. Furthermore, novel molecules capable of inhibiting or modulating Dicer activity are briefly described and could serve useful in the study of Dicer-specific studies.

Altogether, these preliminary findings provide some key insight into structural elements necessary for major groove binding & identify a class of macrocycles capable of

interfering with dicer-mediated maturation. Ideally, the results from this work can be further explored to further our understanding of how nucleic acids interact with their surroundings and elucidate the mechanisms by which nucleic acids can be exploited for therapeutic purposes.

Chapter 1 Significance of Nucleic Acid Targets and Their Targetability

It has been several decades since Roblin and Friedmann first suggested disease could be treated by replacing dysfunctional genes with functional gene copies[1]. Now, technologies that enable the delivery of nucleic acid therapies are at the forefront of medicine, ushered in by the COVID-19 pandemic[23], and targeting several other genetic diseases via gene therapy is becoming accepted by the United States Food & Drug Administration (FDA) and the European Medicines Agency (EMA). Therefore, it is important to reflect on the current landscape of Deoxy- and Ribo- nucleic acid-based therapies.

1.0 Biological Significance of DNA

The human genome consists of roughly 20000 protein-coding genes as well as 22000 RNA and RNA motif encoding genes[4]. Since the completion of the human genome project, genome-wide association studies (GWAS) have been used to identify common genetic variants linked to disease and disorders[5]. Studies revealed a wide variety of genetic variations and prompted the development of “P4”; predictive, preventative, personalized, and participatory medicine, a practice that is focused on personalizing medicine based on the genetic make- up[5]. Before we understood DNA as we do today, scientists used to believe that DNA was a long polymer. Thanks to the incredible efforts of many scientists, we now understand that there is an extremely complex system at work that is responsible for encoding the very basis of life. DNA is highly conserved and

shielded in the nucleus and it is responsible for the storing and expression of all biological and hereditary information as well as binding to the transcriptional machinery to do so.

1.1 Sequence-Dependent Polymorphism of DNA

The most common and simple conformation of DNA in living cells is double-stranded (dsDNA) B- DNA, that is, right-handed helices consisting of two complementary polynucleotide chains; connected by a network of hydrogen bonds between A-T and G-C base pairs. The primary structure of DNA is determined by these base pairs(Figure 1C), whose bonding energies are considerably different, -23.8kcal/mol for G:C and -11.88 for A:T[6]. Consequently, the polymorphic characteristics of dsDNA are imparted by the order and abundance of these base pairs in any given sequence, playing a critical role in the physiochemical properties of the overall DNA helix and they are responsible for the coding of both RNA molecules and proteins[7]. To elaborate, the structural code is not determined by one base pair, but rather by a series of successive

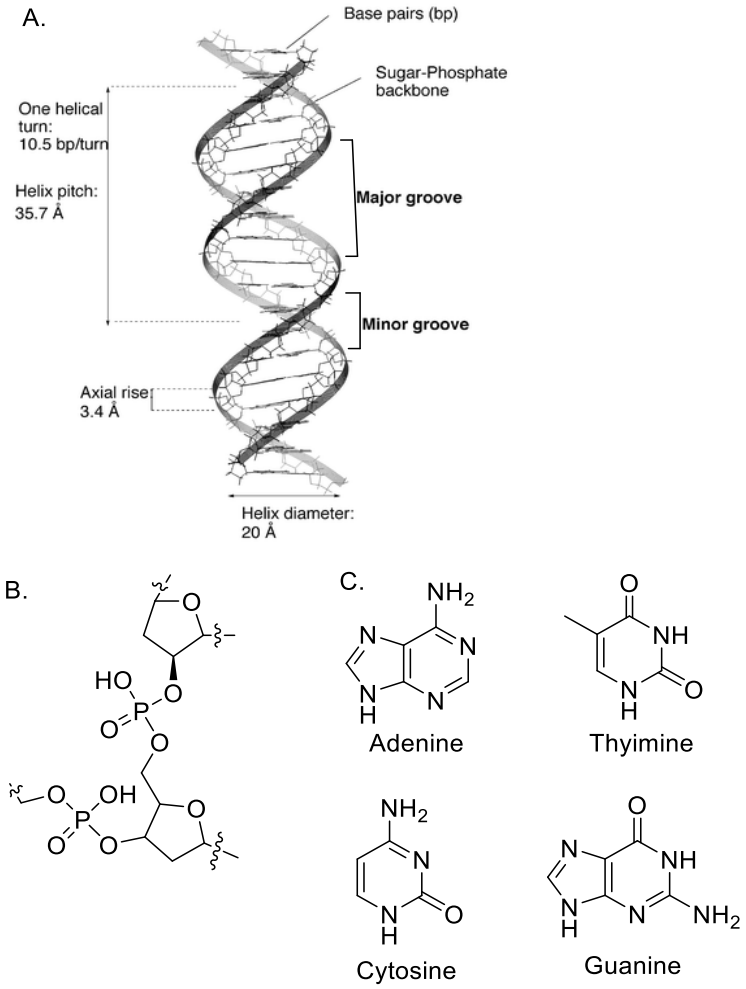


***Figure 1. DNA Structure & Substructures**. **A.** General Structure of ds B- DNA, highlighted are the difference in groove size**. B.** Structure of Phosphate-Sugar backbone. **C.** Canonical structure of bases. https://www.toppr.com/ask/question/explain-the-double-helix-structure-of-dna-with-a-labelled-diagram/*

base pairs[8], and the structure of the DNA helices as well as the width and depth of the major and minor grooves are determined by the sum of the base pairs in any given sequence; mainly those that are A:T rich or G:C rich[9–1112].

According to Hud et al., the secondary structure of DNA can be influenced by cation localization, based on the "electrostatic collapse model"[13]. They identified cation localization in the minor groove of A- track sequences and proposed that the narrowness in B-from DNA is a result of these electrostatic interactions between the phosphate groups in the minor groove and the cations[14–16]. Therefore, the narrowing of the major groove may be a response to cation localization and results in a decreased distance between phosphate- phosphate backbone in the major groove[17,18].

Studies employing A- track, G- track, and generic DNA were used to determine the effects of sequence makeup and cation localization on the secondary structure of DNA. Briefly, A- track sequences were made up of a minimum of four bps, containing only AA or AT base steps, (i.e. AAAAA, AAAAT, AATTT) including homopolymer duplex poly(dA)·poly(dT). It is important to point out that monovalent and divalent cations typically localize deep in the minor groove[14,15,19–21]. Furthermore, evidence suggests that the major groove of A-track DNA lacks cation localization[21–23]. This preference for localization within the minor groove is likely attributed to the electronegative adenine N2 and the thymine O_2 that line the minor groove of A-track DNA[19]. Similarly, they attribute the lack of cation localization in the major groove to the electropositive amino proton at the center of every AT bp. G- track DNA, however, is defined as being rich in GC base steps (ie. GGGCC, CCCCGGGG, CCCGCGGG, GGGCCCC, including poly(dG)·poly(dC), which localize cations within the major groove, and not in the minor

groove. They stipulated that G-tract sequences had to have GGG-rich base steps for max cation localization. Furthermore, G- tract sequences may be unfavorable for minor groove cation localization, a consequence of mild electropositive protons from the guanine amino group. They define generic DNA as all other possible sequences of DNA, such as canonical B- form helices, which contain a mixture of base pairs and include any sequence not defined as A- or G- tract sequences. The degree of cation localization in either groove of generic DNA can vary along the helix due to local variations from the base composition and the sequence itself, however, for their discussion, they assumed that cations occupy both grooves equally. They rationalize that this even distribution of cations between both grooves is responsible for the canonical B-form of generic DNA, whereas in A- and G- track sequences, cations in their respective groove drive the groove in which they are bound to narrow by pulling the phosphate groups in closer and effectively curving the duplex[24,25]. Furthermore, even cation distribution in generic DNA results in a wider minor groove, however, these shifts in the duplex's shape are subtle and require the use of ultrahigh-resolution crystal structures to observe them[21,22,26]. Interestingly, studies done on generic flanked GGGCCC sequences had shown that G-tract axial bending of the helices increased within higher cation concentrations, driving cations into the major groove, providing some evidence that sequence-directed curvature may be connected to cation localization[27,28]. Rouzina and Bloomfield suggested that the connection between cation localization, B- to A- form transitions in G-tracts is likely due to the narrowing of the major groove around the cations could be considered a B- to A-form transition[17]. Furthermore, Olsen et al. revealed axial bending of both B-A junctions and overall bending of the helical axis in a model with A-form DNA flanked by B-form

helices at both ends; illustrating how axial bending may occur at junctions between canonical B-form and other helical forms[29]. Based on these results Hud and Plavec proposed that cation-induced helical axis bending of G-tract GGGCCC might be a direct result of the progression from B- to A- form structural continuum as ionic strength increases, meanwhile, the flanked generic DNA sequences remained mostly unaffected by cation levels[27,28]. This model provides insight that relates the G-tract sequence-directed curvature to sequence-specific localization via the cation-dependent B- to A-form transitions. Similarly, sequence-directed curvature of A-tracts was proposed to bend in, towards the localized cations in the minor groove, by closing the minor groove in around them with maximal axial curvature at the A- tract center where cations are easily localized[13–15]. Despite the limitations of the electrostatic collapse model, other studies have continued to provide support for the sequence-directed curvature of sequences such as GGGCCC, which are driven by the propensity to localize cations and undergo B- to A-A-form transitions. To summarize, G-tract and A-tract sequences differ from generic sequences as they localize cations in the major and minor grooves, respectively. G- tracts transition from B- to A- form, opposite to A-tracts, which are resistant to transition. Furthermore, the A-form helices of G- tracts have a wide and flat minor groove, whereas the A- tract B*-form helix contains a very narrow major groove. Finally, generic DNA-flanked G-tract sequences can lead to axis bending toward the major groove, and A- tract sequences bend towards the minor groove. These highlight how different DNA- tract sequences can be understood in terms of the “tug-of-war” activity for cation localization, and more importantly, taken together with the sequence configuration, effects the overall

secondary structure of DNA and should be considered when developing sequence-specific major groove probes[3012].

Sequences with the widest major groove consist of AT-rich sequences, likewise, the narrowest minor grooves are typically made up of more GC-rich sequences. Typically, B*-form A- tracts contain a minor groove width of 2.8Å, and major groove with a width of 13.9Å and A-form G-Tract typically have a width of 11.2Å and major groove width of 2.4Å, whereas canonical B-form Generic DNA display minor and major groove widths of 5.9Å and 11.2Å, respectively[31]. Furthermore, spatial studies revealed the importance of the major groove since most transcription factors interact with the major groove rather than the minor groove[9]. This space provides access to essential replication and transcription machinery; and can potentially be probed by small molecules to study DNA biology further. Therefore, it is important to consider the binding modes of DNA, which will be discussed in section 1.3. The dysregulation of this machinery or errors in the genetic code can lead to several diseases. To a greater or lesser extent, our genetic build plays a role in almost all diseases; therefore, it is important to evaluate if these novel therapeutic targets are pharmacologically relevant.

1.2 Role of DNA in Cancer and Other Diseases

The wide spectrum of genetic variation has been driven by millennia of genetic mutations. These mutations can lead to genetic disorders, accounting for 80% of rare disorders[4]. As DNA replicates, several random mistakes can occur. These mutations lead to a variation in alleles, which are transferred down generations, and can sometimes lead to diseases. One of the most classical examples of inherited disease is sickle cell anemia, an A to T single point mutation in the beta-globin encoding gene, which changes a glutamic acid to

valine and results in the deformed sickle-like erythrocytes[32,33]. There are several mechanisms by which altered DNA can lead to disease be it hereditary or environmental. The GWAS revealed that trait-linked genetic information was stored in the non-coding regions, likely linked to enhancers; sequences that regulate local tissue gene expression[3435]. Enhancer hijacking is a result of structural variation that induces the enhancer to act on genes that are not its native targets, a phenomenon well documented in the development of cancer and other diseases[363734]. Because of this, there is a plethora of work into understanding DNA, and its various promoters and enhancers as well as the transcription factors that regulate them. These efforts have led to several FDA-approved treatments, many of which are small molecule-based therapies already. Thus, the successful identification of molecules that bind the major groove of a DNA could make the major groove a potential pharmacologically relevant target.

Small molecule strategies have genetic roots; however, they are designed to inhibit channels that are involved in disease development and progression. Drugs such as Sorafenib (figure 1.2A); a multikinase inhibitor that binds B-Raf, Raf-1, and VEGFR, and limits their enzymatic activity[38]. Another example is to disrupt the DNA damage response (DDR) network with Olaparib (figure 1.2B), which renders the cell unable to repair and restore itself[39,40]. However, none of the current therapies directly bind DNA, and with a rise in drug resistance, it is important to find probes that can be developed into new therapies in conjunction with therapy delivery technologies which have recently seen a breakthrough in their development.

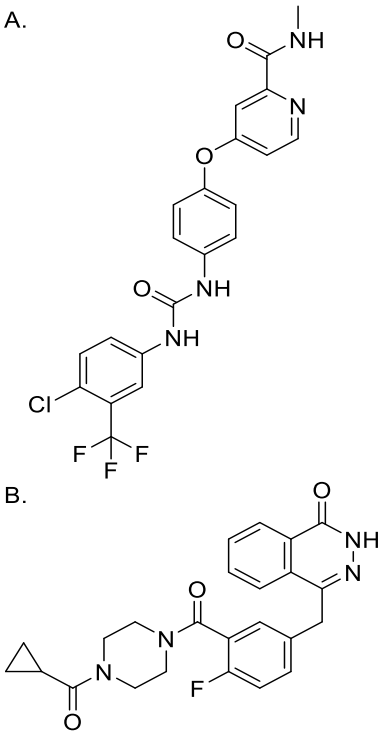


Figure 1.2 Relevant FDA Approved Small Molecule Drugs. ***A.*** *Sorafenib* ***B.*** *Olaparib*

1.3 Binding Modes of DNA

To probe the potential of pharmacologically relevant probes, we must review the binding modes by which small molecules bind DNA, covalent and non-covalent. Covalent binders irreversibly bind DNA, leading to permanent damage and ultimately, cell death. Non-covalent binders are typically reversible and classified as major groove binders, minor groove binders, and intercalators (figure1.3A). Intercalators insert themselves between the adjacent base pairs of a helix(figure1.3B). This base stacking is detrimental to the stability of DNA, as it leads to a significant structural change that damages the cells' cell

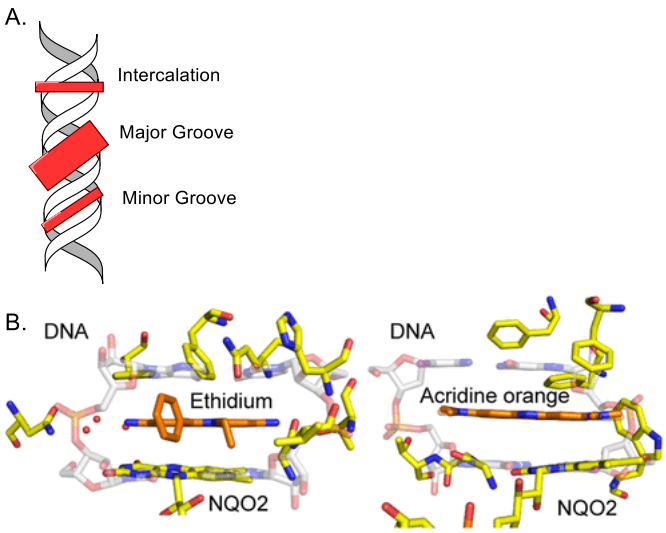


Figure 1.3 Binding Modes of DNA. A*. General schematic of DNA's binding modes.* ***B****. Intercalative stacking between bases by known intercalators, ethidium and acridine orange. Biochemistry 2015, 54, 51, 7438–7448*

cycle and functional machinery, thereby killing it[41]. This has been the reason for developing intercalators as cancer therapies, however, these drugs have proven simultaneously dangerous and lethal to healthy cells. Sequence selectivity among intercalators is impossible due to their π- stacking tendencies(figure1.3B).

Major and minor groove binders can coordinate themselves effectively into the sometimes narrow or wide grooves of the DNA helix. As we know, the population of A:T and G:C pairs affect the structure of DNA, making DNA non-canonical; or polymorphic. This structural polymorphism is the key to developing probes with specificity as it allows for variability in the size, shape, and depth of the groove. Minor groove binders are typically small crescent-shaped aromatic molecules that selectively interact non-covalently in the minor groove. Specifically, Netropsin, a natural product with antibiotic properties that specifically targets AT-rich sequences located in the minor groove of dsDNA, and distamycin, a polyamide antibiotic that also binds the minor groove figure 1.4)[4243]. Dervan et.al.

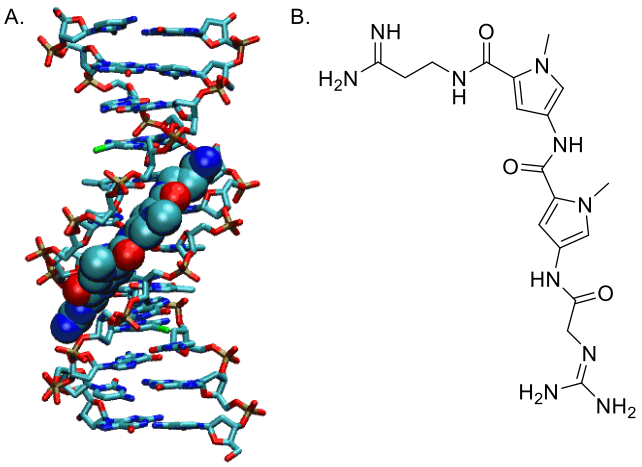


Figure 1.4 Minor Groove Binding. A*. Netropsin, a well-known MGB, bound to minor groove.* ***B*** *Structure of Netropsin. https://upload.wikimedia.org/wikipedia/commons/8/84/Netropsin_DNA_bound.gif*

carried out NMR and X-ray crystallography studies of DNA complexes with these molecules and revealed how the crescent-shaped polyamide molecules achieved sequence specificity. Particularly these molecules were associated with the minor groove of AT-rich sequences and they found that bi-furcated hydrogen bonding between the amide hydrogens of the N- methylpyrrolecarboxamides and the N3 & O_2 of adenine and thymine, respectively, aided in the orientation of these small molecules into the minor groove[43]. Duocarmycin and its derivatives are another set of natural products that have been shown to display antitumor properties that act via AT- sequence selective alkylation of an N3 adenine in the minor groove[44]. Based on the polyamide structure of netropsin and distamycin, Dervan and colleagues developed a series of polyamide hairpin molecules made up of imidazole and pyrrole, tethered by an (R)- and (S) 3,4-diaminobutyric acid (β- amino- γ- turn) (figure1.5). The results of these synthetic molecules demonstrated an increased binding affinity and sequence selectivity in the minor groove, capable of inhibiting androgen receptor-mediated gene expression in cell culture assays[45–5051–53]. In addition to the pyrrole-imidazole GC specific polyamide pair, Dervan and colleagues further developed an analogue with a novel terminal aromatic pair, 2-hydroxy-6-methoxybenzamide/ 1-

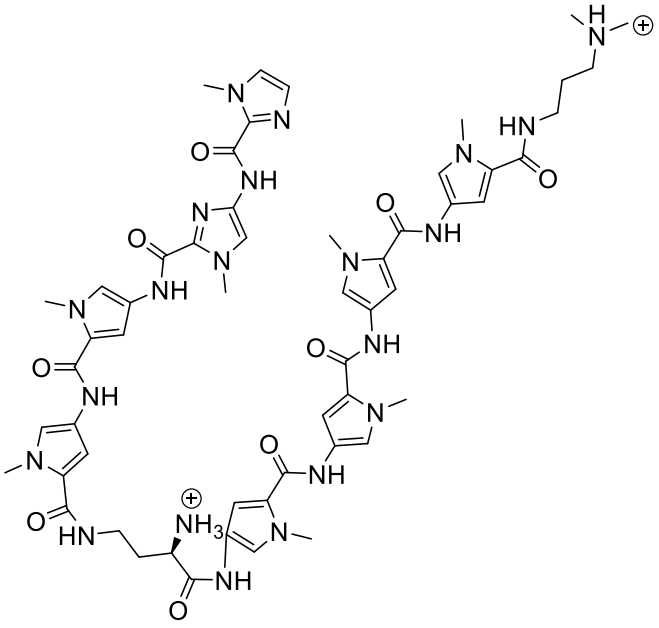


Figure 1.5 Synthetic Polyamide Hairpin*. Associates in minor grove and forms H bonding interactions with local bps. J. Am. Chem. Soc. 2008, 130, 21, 6859–6866*

methylpyrrole, capable of distinguishing T:A from A:T base pairs and both C:G and G:C base pairs within the minor groove[54,55].

The minor groove has proven targetable and provides a broad range of pharmaceutical potential, and because of this, a major effort has gone into developing minor groove binders for therapeutic evaluation[56]. Despite their specificities, these molecules demonstrate several off-target effects, because they are so small, they are prone to intercalation and generally bind several other targets in cells[57]. These provide evidence that sequence selectivity is achievable, therefore, designing major groove probes may open the door to a new set of molecules that allow for specificity and avoid the pitfalls of the previous two binding modes. To date, a useful major groove model probe is yet to be identified for the rational design of small molecules that can bind a 4- or 5- bp sequence with high discrimination within the entire genome. The lack of sequence-specific DNA recognition within our current chemical genome prompts the need for the discovery and development of novel molecules capable of doing so.

1.4 Biological Significance of miRNA

The RNA superfamily, miRNAs are a class of non-coding RNAs that govern post-transcriptional gene regulation[58,59]. More than 2500 miRNAs have been discovered since their discovery in 1993[60,6162]. Since then, our knowledge of miRNAs has remained rather stale, possibly due to the lack of technology to effectively study them. However, recent advances in technology have allowed us to unearth critical correlations between disease and miRNA dysregulation.

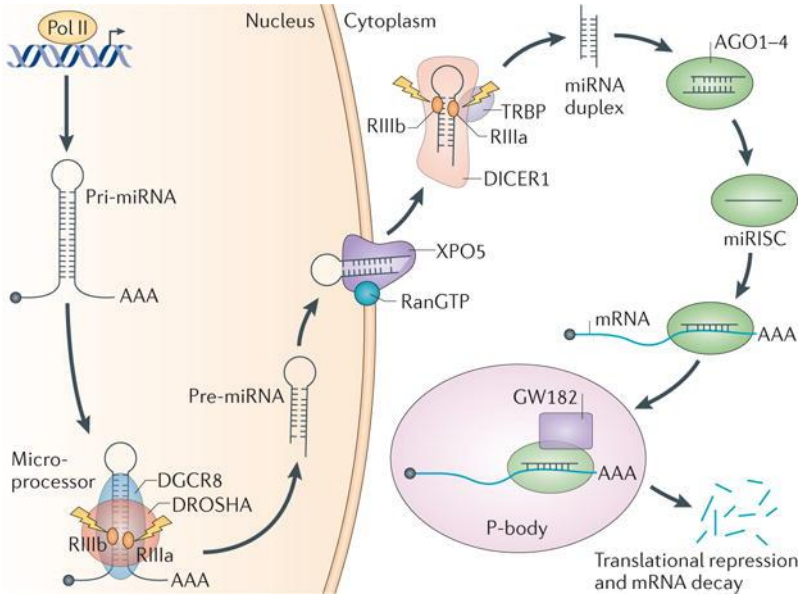


Figure 1.6 The canonical Biogenesis of miRNA. primiRNAi is transcribed by RNA Pol II, then processed by Micro-processor into premiRNA; followed by Dicer maturation and subsequent RISC formation in cytoplasm. Nat Rev Cancer. 2015 Jun;15(6):321-33.

1.5 miRNA Biogenesis

The canonical biogenesis of miRNAs (Figure 1.6) begins with the local transcription of a long 1 kb miRNA coding sequence called primary miRNA (pri-miRNA) by RNA polymerase II. This transcript is then processed by Drosha, a ~160 kDa nuclear RNAse III endonuclease that forms a complex called Microprocessor with its cofactor, DGCR8[63–65]. Precise recognition and cleavage of a pri-miRNA by the Microprocessor complex is crucial, as it both cleaves and folds the pri-miRNA into a ~65nt precursor miRNA (pre-miRNA)[66 67,68]. Once processed, the pre-miRNA is exported to the cytoplasm via a GTP-dependent Exportin 5 (EXP5) nuclear protein[69–71]. In the cytoplasm, the pre-miRNA is further processed by a ~200 kDa RNAse III endonuclease, Dicer[72,73], to form the ~22 nucleotides mature miRNA duplex, of which one strand, the guide strand, is loaded into Argonaute2 (AGO2). This prompts the recruitment of several other accessory proteins to form the RNA-induced silencing complex (RISC) [74–76 77,78]. Concerning gene silencing activity, the miRNA is responsible for the recognition of the mRNA to be modulated through sequence complementarity[79]

1.6 Dicer

Dicer is a large ~220kDa protein that is critical in the final maturation step of miRNA. The domains that make up dicer consist of a helicase domain, AGO, a double-stranded RNA binding domain, a PIWI and a PAZ domain, two RNAse III domains, and one more RNA binding domain currently being studied[80]. It is theorized that PAZ interacts with the two nucleotide 3' overhang of the pre-miRNA while the loop region orients at the helicase domain; two magnesium-dependent RNAse III domains coordinate ~22nts from the PAZ domain and generates a second nucleotide 3' overhang[81,82]. Every miRNA sequence is thought to be processed based on loop size and length; studies have shown that dicer kinetics are affected by these factors, processing larger non-base paired sequences faster than smaller loops or perfect base pair strands, respectively[83,84]. Also known to affect dicer kinetics are co-active proteins such as Pact and TRBP[85,86]

miRNA	Expression	Target Gene	Disease	Reference
let-7	Down	Ras	Cancer	Roush et. al. 2008
10b	Up	HOXD10	Cancer	Ma et. al. 2010
21	up	PTEN	Cancer	Pfeffer et. al. 2015
34a	Down	BCL2	Cancer	Hermeking et. al. 2010
122	Up, Down	HCV, CCNG1	HCV, cancer	Henke et. al./ Gramantieri et. al. 2007
133a	Down	CDC42	Heart Disease	Thum et. al. 2008
155	Up	TP53INP1	Cancer	Faraoni et. al. 2009
373	Up	LATS2	Cancer	Staedel et. al. 2015
525	Up	ZNF395	Cancer	Pang et. al. 2014
544	Up	MTOR	Cancer	Haga et. al. 2015

Figure 1.7 Examples of studies on the involvement of miR21 and its prevalence in different cancers*. Dysregulation (up/ down) of miRNA transcription has been linked to several diseases and cancers*

1.7 miRNA's Significance and Role in Cancer & other Diseases

miRNAs function primarily as RNA-silencing molecules. Through this activity, miRNAs can regulate almost all cellular and developmental processes. Due to their crucial role as regulators of protein-coding genes (PCGs), their biogenesis and activity are highly regulated, both spatially and temporally. Not considering that non-conserved miRNA binding sites also exist, more than 60% of PCGs have at least one conserved miRNA binding site, which would suggest that most

PCGs are regulated by miRNAs[62]. Dysregulation of these small RNAs has been found to lead to diseases including several cancers (figure 1.7), heart disease, metabolic disorders, neurodegeneration, and viral infection[87,88,8990]. Furthermore, deletion of either RNAse involved in the maturation of miRNA was found to be lethal at the embryonic stage[91]. There are no FDA-approved miRNA-targeted drugs currently on the market, however, their recurring role in propagating human disease has made the targeting of miRNAs incredibly attractive as a novel therapeutic strategy[89].

1.8 Nucleic Acids as Therapeutic Drugs and Targets

Due to the biological importance of DNA and RNA, nucleic acid-based therapies have been the subject of several therapeutic strategies. Nucleic Acid-based approaches fall under two categories, nucleic acids as therapeutic drugs or nucleic acids as therapeutic targets.

1.8.1 Nucleic Acids as Therapeutic Drugs

Utilizing nucleic acids as a drug is an approach that requires a payload to be delivered to the cell, where it will carry out its cellular response, unfettered. Examples of these payloads are not limited to but mainly consist of siRNAs, RNA aptamers, mRNA, and other RNA mimics (figure 1.8). siRNAs are double-stranded RNAs that can orchestrate the formation of the RISC, which can then bind and induce degradation of the target mRNA. RNA aptamers are synthetic RNAs that can bind and inhibit the function of a particular protein. mRNA, single-stranded RNA that binds translational machinery, has also been utilized as a payload to induce the translation of functional proteins. However, the use of RNA mimics has its limitations, largely being that they are degraded via nuclease activity. Therefore, technologies have been developed around this approach.

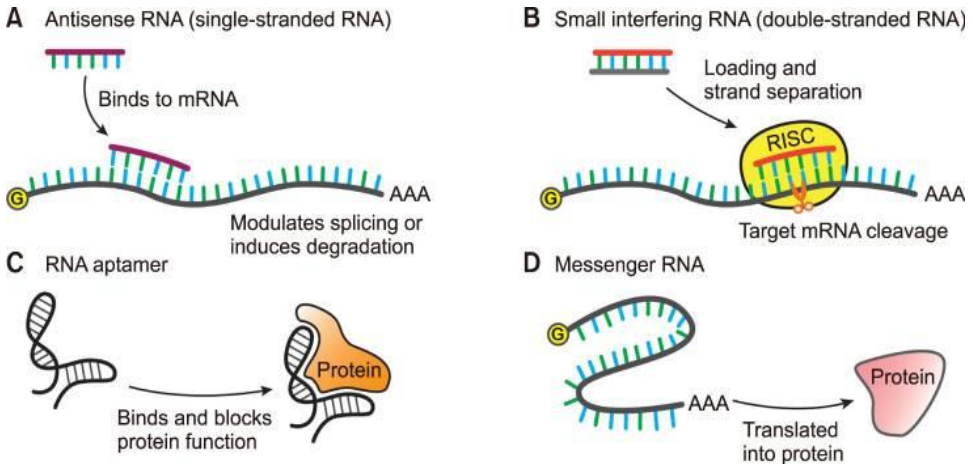


***Figure 1.8** T**ranscription regulation via RNA motif payloads**. **A**. Anti-sense RNA (ssRNA) bind mRNA and silence their translation**. B**. ds siRNA will hydrolyze and activate RISC silencing pathway. **C**. RNA aptamer or ASOs can bind directly to a protein and inhibit their activity. **D**. mRNA induce the expression of favorable proteins. Chonnam Med J. 2020 May; 56(2): 87–93*

1.8.1i ASOs

Antisense Oligonucleotides, or ASOs, are short synthetic nucleic acid sequences that utilize Watson-Crick base paring to modulate gene expression by binding to cellular pre-mRNA or mRNA and regulate the synthesis of their corresponding proteins (figure1.8D)[92,9394]. ASOs consist of synthetic nucleic acids (>5nts) that can regulate protein synthesis by binding to their mRNA counterparts and inducing their degradation. How ASOs function can be through several mechanisms such as steric hindrance of the translational machinery, competitive inhibition, disrupting pre-mRNA processing and splicing; by binding and degrading their target RNA[9596,97]. ASOs can be modified at the sugar, backbone, or nucleobase and these modifications can improve nuclease resistance, modulate pharmacological properties or enhance affinity to the target RNA[98,99].

ASOs have great potential because of their customizability, and their small size (5-10nm) can potentially allow for better tissue penetration. This approach does, however, come with its challenges as ASOs are degraded by nucleases and cleared quickly in vivo[100]. Therefore, first-generation ASOs were heavily designed to prevent degradation; while more recent modifications to ASOs wish to improve specificity, efficacy, and potency for therapeutic purposes.

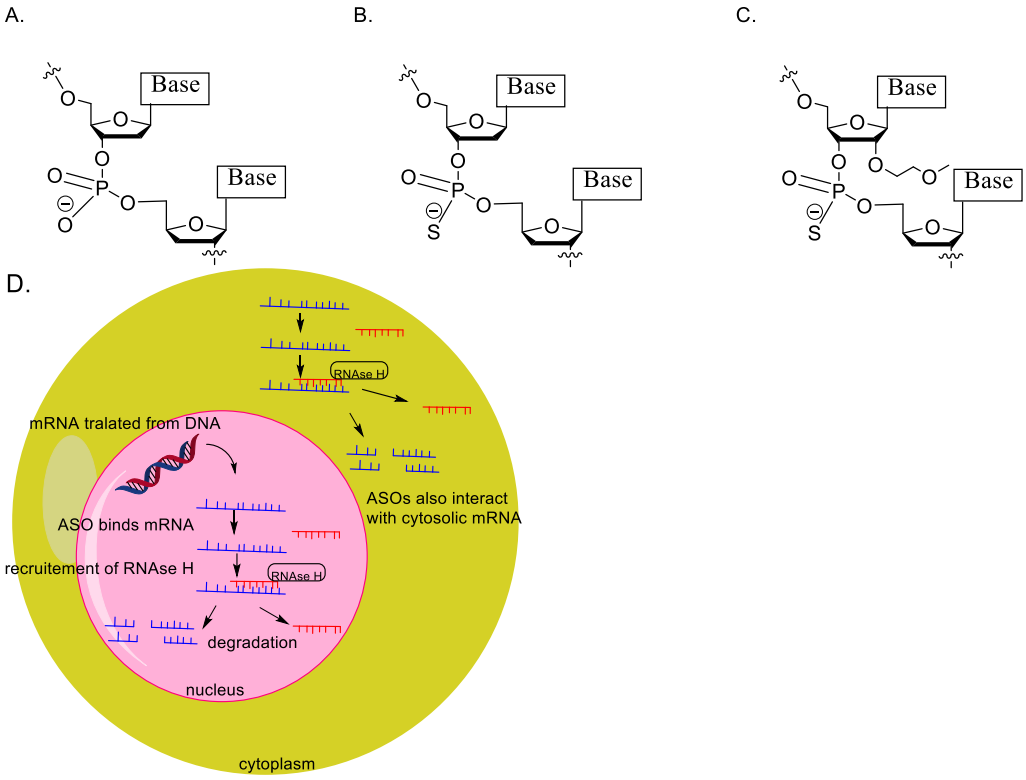


Figure 1.9 ASO modifications & RNase H mediated degradation of mRNA. A*. non modified phosphate backbone of an ASO.* ***B****. Phosphorothioate modified ASO.* ***C.*** *Phsophorothioate and 2'MOE modifications.* ***D****. RNase H mediated degradation activated: translated mRNA is targeted in the nucleus and the cytoplasm. Binding of the target substrate initializes RNase H for degradation.*

Specifically, the phosphorothioate modification was one of the first widely applied modifications which enhanced the metabolic stability of the ASO by replacing a non-linking oxygen along the phosphate backbone, with a sulfur atom[101,102](figure1.9A-C). Furthermore, second-generation ASOs utilize another modification at the 2' position of the sugar. The 2'-O- methoxyethyl (2'-MOE) modification has displayed enhanced nuclease resistance, increased target affinity, and a reduction in cell toxicity[103–106]. ASO stability is critical for their successful application, and although the mechanism of internalization is not fully understood, it is believed that ASOs are internalized via clathrin receptor-mediated endocytosis[101,107–109110]. Once internalized, ASOs can act on their target via different modes, such as RNase H degradation (figure 1.9D) or as seen for the treatment of Duchennes muscular dystrophy (DMD)[105]. DMD arises from mutations in

exon51 in the DMD gene; and ASO PRO-051 functions by binding the mutated exon 51, force skipping the exon being transcribed, resulting in a shorter, active gene[111–115]. This drug however was not given FDA approval in the end.

Previously, the FDA had cautiously approved ASOs because of their poor tissue distribution, and some drugs, notably Formivirsen, have been withdrawn from the market[116–118100]. Recently, however, the FDA has approved at least 4 new ASOs in the last 5 years as COVID-19 and new technologies have forced the scientific community to re-evaluate new and old targets. While still one of the most popular approaches, ASOs remain to have their strategic challenges that will need to be addressed before widespread pharmaceutical adoption of this technology.

1.8.2 Nucleic Acids as Therapeutic Targets

Alternatively, nucleic acids can be used as therapeutic targets. Almost the entire therapeutic market consists of small molecules that have been synthesized or derived from natural products. Genetic engineering tools have also seen a major rise in their potential.

1.8.2i Natural Products and Small Molecules

Despite the exciting new developments in nucleic acid-based therapies, the majority of therapeutic drugs in the market are made up of small molecules, some already being utilized to directly target pathogenic RNAs in patients in a wide range of diseases[119120,121]. Because they are highly structured, DNA, RNA, and RNA motifs form binding pockets

that lend themselves to small molecules binding. There are several synthetic and naturally derived drugs, such as aminoglycosides, macrolides, and tetracyclines that can effectively bind ribosomal RNA or DNA to treat bacterial infections (Figure1.10)

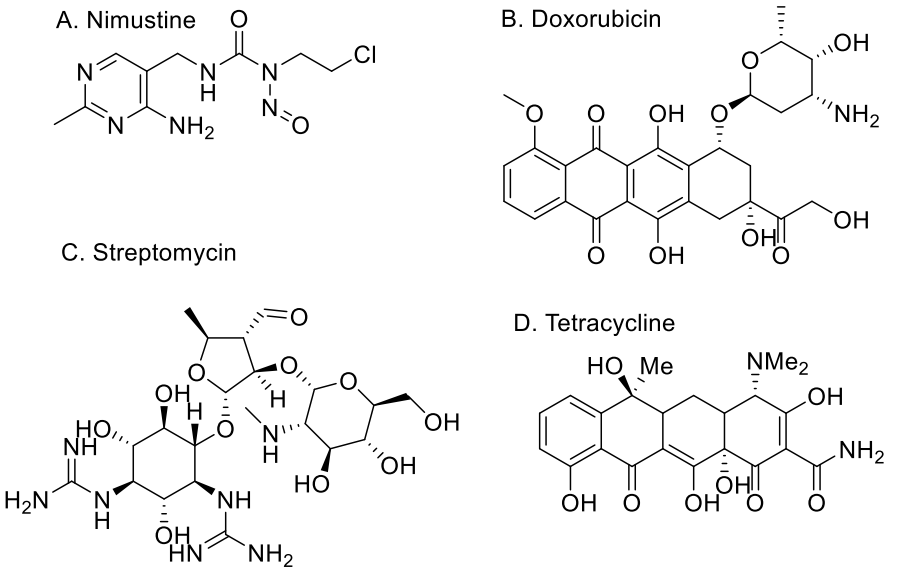


Figure 1.10 Examples of Therapeutically Relevant Natural Products. *Structures of molecules observed to bind the minor and major groove or intercalate.*

DNA is capable of several binding modes, which are discussed in further detail in section 2.3, and they provide us with various opportunities in developing therapeutic strategies. One such therapy is Nimustine (ACNU)(Figure1.10A), an alkylating anticancer drug used for the treatment of malignant gliomas[122]. Although it is not highly selective, nimustine is only one in a major class of alkylating therapeutics and its mechanism of action consists of an entropy-driven complex formation and alkylation event in the major groove of calf thymus DNA[123]. The significant developments that have been made in spectroscopic and biophysical techniques have allowed us to gain insights into Drug-DNA interactions. Non-promiscuous therapeutic major groove binders have not been previously identified; however, this provides evidence for targeting what has previously been an unexplored or poorly reported chemical space.

Another class of therapeutics is minor groove binders (MGBs)(Figure1.10B). MGBs are typically a planar-crescent-shaped series of hetero and aromatic rings that can non-covalently and reversibly wedge themselves into the minor groove of duplex DNA. Drugs with this MOA include bisamides, polyamides, bis-benzimidazoles, and aureolic acid anti-cancer drugs such as mithramycin, which has been found to regulate gene expression via binding preferentially to a GC-rich, minor groove sequence[124,125]. Some drugs such as Doxorubicin (figure1.10B), an anthracycline that targets the minor groove, are currently used to treat several cancers including ovarian, gastric, multiple myeloma, lung, breast cancer, and others[126–128]. Drugs like doxorubicin have sparked the development of more modern therapeutics, however, several challenges still exist in this space.

Utilizing medicinal chemistry and the working knowledge of moieties that interact with target site sequences, scientists developed another class of therapeutics based on a family of antibiotics, the aminoglycosides (AGs) (figure1.10C). AGs do not act on DNA directly, but rather on its downstream machinery and consist of amino sugars linked sequentially via glycosidic linkages. Streptomycin (Figure1.10C), a first-in-class antibiotic that is FDA-approved, acts by binding irreversibly to the 30S ribosomal subunit, thereby interfering in its interaction with formyl-methionyl-tRNA, disrupting protein synthesis and resulting in cell death[129,130]. Evidence that it is important to consider developing RNA and RNA motif probing tools as another strategy for targeted therapies. Chemical modifications have allowed AGs to be synergistic with their targets, however, despite several improvements; many challenges still exist regarding their therapeutic effectiveness due to high degradation, poor efficacy, and reliance on electrostatic

interactions which makes select sequence affinity difficult[131–133]. As a result, AGs such as amikacin, tobramycin, and gentamycin are prescribed via injection[134].

Indeed, targeting DNA, RNA, and RNA motifs provides us with novel therapeutic targets, however, non-ribosomal RNA remains an untapped or poorly explored therapeutic space, and there are several challenges in identifying novel druggable RNA targets along with their selective binders[135]. Instrumental to the discovery of novel RNA targets and their binding conjugates, has been the development of High throughput screening techniques (HTS), allowing for the rapid screening of key RNA targets associated with disease, pre-miRNA(MiR21) specifically[136,137138–141].

While there is a growing database of molecules that can target and manipulate the biogenesis of RNA and RNA motifs, unfortunately, there is a large amount of chemical and target overlap[136,137]. Therefore, the challenge in this chemical space is the discovery of novel molecules that can bind non-ribosomal RNA (miRNA) or family of RNA with promiscuity. The underlying thesis of my work is to explore natural products for the identification of novel compounds that can interact directly with our nucleic acid target with selectivity.

Currently, major efforts are underway to study DNA & RNA structural motifs, loops, complexes, bulges, etc., and the functional groups that bind them[137,142]. These efforts along with the development of new tools have allowed us to identify novel targets and, in some cases, nucleic acid binders with the potential to modulate cellular activity[140,143–148]. These findings provide evidence for the therapeutic potential of novel nucleic acid targets previously thought undruggable. It should be noted that most RNA-targeted antibiotics don’t follow Lipinski’s rule of five, generally used to ensure the efficacy of a drug, and that

structural optimization is still required for these drugs to ensure selectivity and safety; as they have been found to later act on other targets[149,150].

1.9 Conclusions

An incredible number of advances have emerged to help us develop next-generation nucleic acid-based therapies both as therapeutic molecules and as therapeutic targets for small molecules. Following the COVID-19 global outbreak, another nucleic acid-targeted technology that is emerging and has seen rapid development and FDA approval is mRNA Vaccines. Many mRNA-based vaccines, including those of Moderna and Pfizer, utilize LNP delivery technology, which induces the translation of the viral spike protein and illicit the immune- response and immunity[151,152]. Despite their promise, these technologies still have many challenges to overcome, such as their administration is only via IV and their effectiveness is considerably low due to the nuclease degradation. Small molecules, however, have demonstrated high levels of bioavailability as well as an ability to bind DNA, RNA, and RNA motifs[149]; though lacking in the chemical space are selective pre-miR21 and DNA-Major groove binders that can bind discriminately. Indeed, this is an exciting and transformative moment to study nucleic acids as viable targets and with new drugs already conducted in clinical trials, a new market, based on the targeting of key nucleic acids, is on the horizon.

Chapter 2 Exploring the Structural Demand for the Design of Major-Groove Binding Small Molecule

There has been an increase in the study of major groove binders in recent years. A small number of natural products and synthetic analogs have already been reported that have been observed to target the major groove of DNA. Specifically, a non-covalent probe that can interrupt the replication machinery at the major groove (Figure 2.1) could elucidate a new generation of therapeutically relevant chemical tools for manipulating transcription. Moreover, thermodynamic data regarding the relationship between AT base pairs, their abundance, and DNA structure results in some of the widest major grooves observed, producing a range of potentially targetable binding pockets[12].

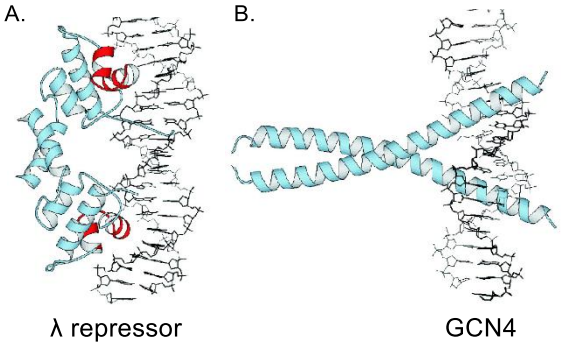


***Figure 2.1 Examples of binding proteins in the major groove**. **A**. λ repressor regulates transcription of immune- relevant genes. **B.** GCN4: a central regulator of protein synthesis. Mol. Cell Volume 8, Issue 5, 21 November 2001, Pages 937-946*

2.0 DNA Substrates

As previously discussed in section 1.2, evidence suggests that the polymorphic characteristic of DNA is attributed to the local sequence composition which may influence the widening of the major groove upon protein binding, especially in tracts with non-alternating AT bps[153,154]. Particularly, a comparison of select oligomers displayed a range of groove widening upon protein binding. Studies comparing the unbound B-form model,

DDD (Drew- Dickerson dodecamer, 5'-CGCGAATTCGCG-3', 13.3Å), to the DNA binding site of NFAT (nuclear factor of activated T cell 1, 5'-GGAAAAA-3'), c-Myc (cellular myelocytomatosis, the "E-box", 5'-CACGTG-3', 17.2Å) and STAT6 (Signal transducer and activator of transcription 6, 5'-TTTCCTAGGAA-3', 19.3Å) revealed a widening of the major groove of 7.9Å, 3.9Å and 6.0Å, respectively; relative to DDD[155–165166,167]. These findings suggest that transcription factor binding sites possess a wider major groove and the rational design of small molecules capable of shape-selective recognition of the transcription binding site could elucidate a useful sequence-selective molecule[160].

Moreover, our DNA targets have been identified as having critical roles in disease development. The upregulation of transcription factor STAT6 is involved in solitary fibrous tumors and dedifferentiated liposarcoma[168,169]. Upregulation of the nuclear transcription heterodimer cMyc/Max is typically associated with neoplasm in the cervix, colon, breast, and in small lung cancer, glioblastomas, myeloid leukemias, and osteosarcomas[160,170,171]. Transcription factor NFAT1 is a frequent culprit in oncogenesis and metathesis, as well as in the development of microenvironments that promote cancer growth, and cancer cell invasion and is overexpressed in glioblastomas as well[172,173]. Therefore, the identification of shape- and/or sequence–selective molecules that can be derivatized would be invaluable for the development of probes to better understand these disease-related genes.

2.1 Development of DNA-Major Groove Binders

Molecules such as azinomycin, nogalamycin, leinamycin (figure2.2A), and a few others, have been found to bind the major groove of DNA, however, most if not all the natural products were found to bind minor and major grooves as well as intercalating, making

them incredibly toxic aswell[174,175]. Therefore, an important consideration when designing sequence-specific major groove binders is that these ligands have polar recognition elements, capable, but not wholly reliant on H- bonding interactions. A positive charge can be associated via electrostatic interactions with the electron-rich groove and the phosphate backbones; however, balance is key to avoid non-specificity. Furthermore, the ligand would need to be flexible or shape-fitting for the major groove and bulky enough to favor major groove binding and avoid binding of the minor groove as well as a preference for B/B*form DNA. Minimally, cell permeability may also need to be considered depending on the delivery method (i.e., IV, LNP delivery, oral).

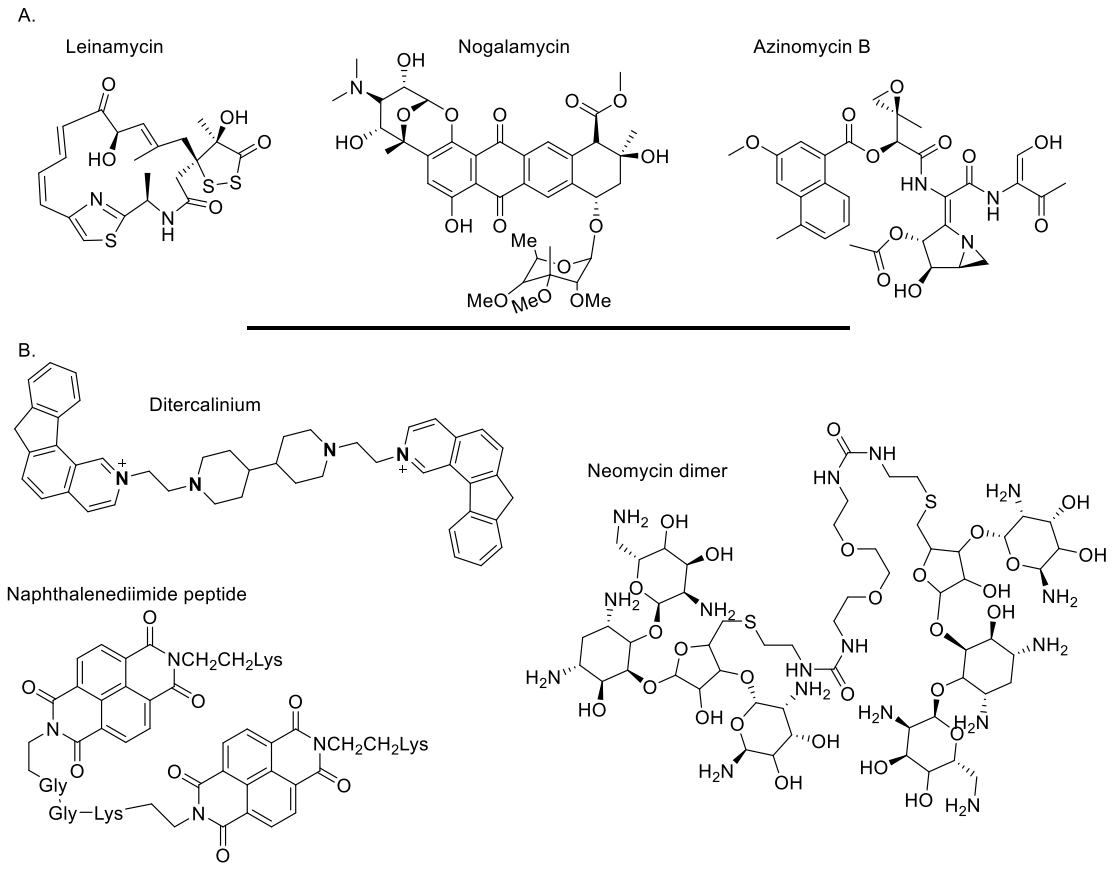
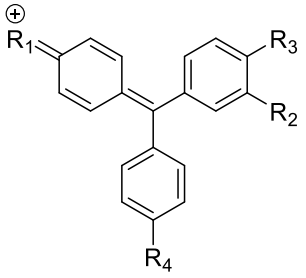


***Figure 2.2 Current Natural and Synthetic Major Groove Interactors**. **A**. Natural products. **B**. Synthetic probes. All observed to intercalate and/or bind the minor groove as well.*

Next have been the 1st generation of synthetic analogs, which for the most part possessed planar structures, and despite having a high affinity for AT-rich sequences, unfortunately, these are capable of intercalation and binding to the minor groove[175]. Among these scaffolds are ditercalinium, naphthalene diimide, and a neomycin dimer (Figure 2.2 B). These studies provided evidence that molecules could be fine-tuned to target major grooves despite their shortcomings. Furthermore, a study by Muller et. al. reported that several dyes, among them Hoechst 33258 and four triphenylmethane dyes, were found to bind tightly with specificity and affinity. Specifically, three of the four triphenylmethane dyes were found to bind preferentially to two adjacent AT pairs when screened against a library of short synthetic duplex oligonucleotide sequences[176].

The triphenylmethane dyes (figure2.3) are a group of histochemical, cationic stains that have been known to bind DNA in a non- intercalative fashion and, due to their positively charged residues, associate along the negatively charged backbone of dsDNA and their bulky structure keeps them from binding the minor groove[177148]. Further studies into these molecules revealed their binding affinities to be ~1x10^{5}M^{-1}, and besides their strong affinity to AT-rich sequences, methyl green was found to bind the major groove specifically[176,177]. Linear and CD studies indicated similar conformations and binding geometries that corresponded to the methyl green- DNA (MG-DNA) complex. Complexes of MG bound to a poly(dA): poly(dT) duplex and a triplex poly(dA): 2poly(dT) were compared. The triplex contained a strand of poly(dT) that runs along the major groove of a duplex, thereby



Fuchsin: $R_1=R_3=R_4=NH_2$; $R_2=CH_3$
Methyl Green: R_1=NMe(Et); R_2=H, $R_3=NMe_2$; $R_4=NMe_3^+$
Mallachit Green: $R_1=R_3=NMe_2$; $R_2=R_4$=H
Crystal Violet: $R_1=R_3=R_4=NMe_2$; R_2=H

Figure 2.3 Structure of Trimethylmethane Dyes. Found to bind to the Major groove and not intercalate.

sterically inhibiting its ability to interact. What they found were similar dichroic spectra between the triplex and MG-DNA complex, confirming its activity as one of the first non-promiscuous major groove binders[177]. Similarly, Wakelin et. al. were able to show those crystal violet associates tightly and preferentially to the major groove of duplex DNA; citing an association constant as 15x tighter over RNA at 0.01 ionic strength and 46x in a salt solution of 0.2M[178]. Furthermore, crystal violet had been previously thought to be susceptible to intercalating, however, several NMR studies have provided evidence of the propeller head configuration associated with the major groove, specifically[179].

2.2 Crystal Violet, Major Groove Probe

The Minehan group has utilized their knowledge of major groove binding to design novel probes that target the major groove of DNA. In one of their earlier efforts, they synthesized derivatives of azinomycin B and reported on a scaffold to have a binding constant of $2.42x10^{7}M^{-1}$ with an affinity for the major groove of similar sequences in a library of synthetic homologues[180]. The molecule with the highest affinity was composed of a central N-triethyl tether, and at each end was the fused ring chromophore of azinomycin(figure 2.4)[180]. A combined approach utilizing viscosity, CD, and competitive binding assays revealed that the tethered molecules assisted in coordinating along the major groove, however, though the tether associated with the major groove, the chromophore often intercalated the DNA backbone[180]. These results provided evidence that their strategy of utilizing the tether to occupy a larger space could improve binding, however, the azinomycin pharmacophore was prone to intercalation.

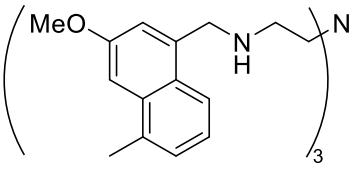


***Figure 2.4 Azinomycin Derivative**. Planar tri- bisbenz azinomycin analogue with N-triethyl linker*

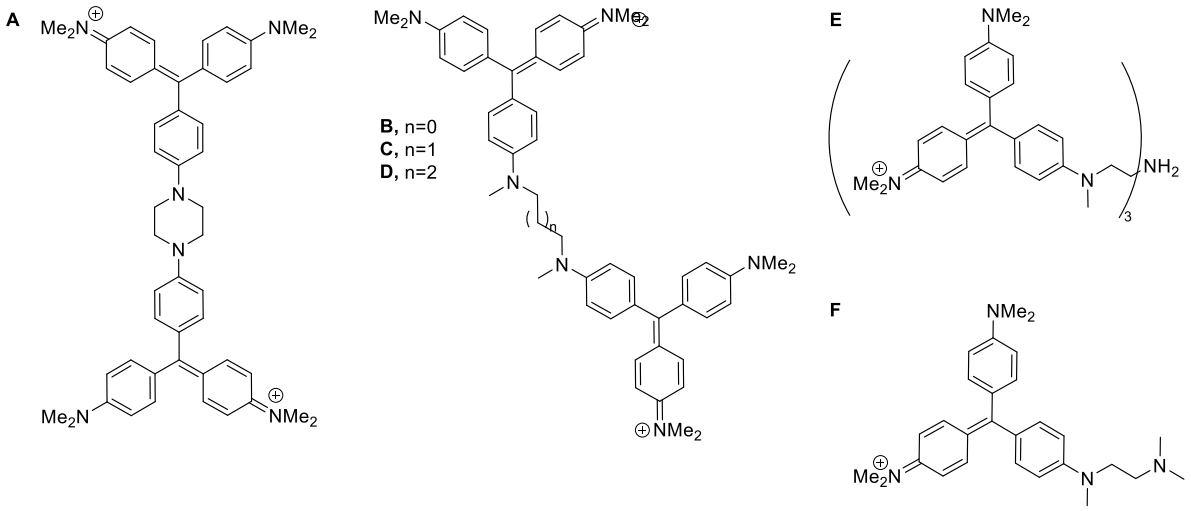


Figure 2.5 1st Generation Crystal Violet Major Groove Probes. *Preliminary SAR screen. Designed to study A) rigidity B, C,D) optimal alky linker length E) size & charge tolerance F) monomeric & linker only****. Bioorganic Chemistry (2019) 83 297-302***

Having observed the higher occupancy in the major groove with the incorporation of a tether, the Minehan group then decided to apply the same concept to crystal violet, a triphenylmethane dye that had previously shown an association constant of K_a= $1.5x10^5M^{-1}$[1178]. Nunez et. al. designed dimeric and trimeric ligands (Figure 2.5E like their previous azinomycin scaffold, which was expected to bind to the major groove. Another set of scaffolds was designed to probe for the optimal distance between charges and DNA interactions by varying non-flexible and flexible tethers as well as the distance between dye units (Figure 2.5A-D). A monomeric scaffold (Figure 2.5F) was made to evaluate the benefits of having the di- and trimeric tether-bound scaffold. Their results demonstrated sub-micromolar binding affinities that correlated strongly with the ionic composition of the buffer. Interestingly, the scope of electrostatic interactions was investigated by varying the NaCl concentrations in an EtBr displacement assay. A release of 2-3 counter ions was observed after incubating with their ligand, which indicated that the dye units were ionically interacting with the DNA backbone[148].

Fluorescent intercalator displacement assays revealed that the trimeric ligand bound exponentially tighter than the others, $16.2\pm3.7x10^{7}M^{-1}$, suggesting that the steric bulky scaffold binds discriminately in a shape selective manner[148]. Competitive binding assays and viscosity studies confirmed the binding mode to be major groove and not intercalation or minor groove binding[148]. Furthermore, molecular docking studies were conducted using Autodesk Vina software (figure 2.6)[181]. Loading the Drew Dickerson dodecamer (5'-CGCGAATTCGCG-3')(PDB 436D), the researchers were able to model their ligand binding the duplex DNA, and in its lowest energy binding mode, and the model demonstrates the efficient association of the nitrogen linker to the central AT-rich region via hydrogen bonding[20]. Interestingly, their model illustrates how well the triphenylmethane units span the width of the groove to engage in ionic interactions across the phosphate backbone[148].

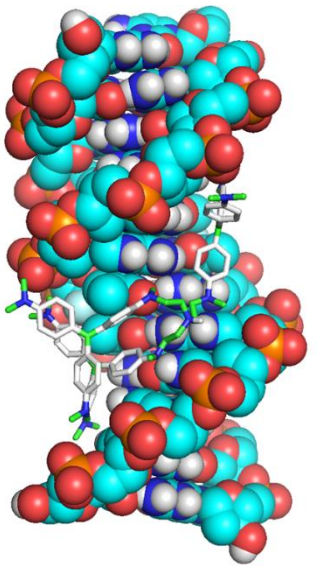


Figure 2.6 Molecular Docking of Drew Dickerson bound to 1st Gen. *ligand, E (fig.2.5). Ligand associates along major. Bioorganic Chemistry 83 (2019) 297–302*

2.3 Research Objective

The completion of the human genome project prompted the GWAS and since then our knowledge of DNA biology has grown exponentially. DNA and its subsequent transcription are responsible for over 20000 proteins and over 22000 more RNA and RNA motifs. Dysregulation or corrupt DNA sequences are often associated with several diseases and cancers. Despite the growing database of DNA biology from molecular to epigenetic, there remains an ever-growing list of unknowns, and among them; how to actively manipulate their activity. Useful probes could be used to study and understand

critical transcription interactions that happen at the major groove and can potentially be used to drive an inquiry into the major groove as a therapeutic target via small molecules.

The structure of dsDNA as described in section 1.2 renders a minor and major groove, both of which are targetable. While minor groove binders have been developed into therapeutic strategies, targeting the major groove has remained an elusive strategy. A hand full of molecules have been identified to interact with the major groove, however, they were found to also intercalate and bind the minor groove. This work describes the development of crystal violet probes that have demonstrated an affinity for the major groove unit of DNA. Currently, there does not exist any biologically relevant, non-promiscuous set of molecules that can effectively probe a specific major groove unit of a DNA sequence. A critical challenge that has stifled the development of major groove binders has been the lack of model molecules that demonstrated selectivity towards any major grove unit of a DNA sequence. Therefore, the identification of a viable major groove binder that demonstrates selectivity could aid in the development of an entire series of novel probes and further our understanding of DNA biology and its processes. To that end, the Minehan group has developed a series of 1st generation crystal violet derivatives that have displayed an affinity for the major groove unit of CT-DNA. Guided by the findings of the 1st generation molecules, I wanted to investigate the activity of a bulky trimeric scaffold against the select synthetic homologs, and because of DNA's polymorphic nature, we reasoned that we might be able to detect selectivity in some of the widest grooves with these bulkier set of molecules. This work describes a rational design-based approach, followed by the total synthesis and biophysical studies of a set of crystal violet

derivatives that were found to bind the major groove. Specifically, the objective of this work was to:

A) **Design and Synthesize Crystal Violet Analogs to Explore SAR.** First, I needed to design 2nd generation crystal violet derivatives based on the previous SAR results. Secondly, the total synthesis & optimization of ligands might reveal structural moieties that drive major groove binding, shape recognition, or sequence selectivity. Moreover, based on what we know about the relationship between shape & sequence makeup, when designing for sequence selectivity it is worth considering the base pairs and how they may interact with the ligand. Furthermore, useful probes could be used to study and gain insight into how DNA might interact with regulatory proteins and transcription factors. We expect these findings to aid in the development of 3rd generation probes.

B) **Characterize the Interaction of These Molecules with the Major Groove**. To investigate how these molecules are interacting with the DNA substrate, biophysical experiments were employed to discern how they are interacting with each other. Using fluorescence spectroscopy and circular dichroism, we aim to elucidate the mechanism of action of these molecules, specifically, binding affinities and the mode of binding (i.e., major/ minor groove, intercalative). Future ITC studies will also aid in revealing and confirming the mechanism of action. ITC Is a powerful label-free method that is capable of acquiring several thermodynamic parameters of molecular interactions in one experiment which could support our theory and current data. These studies aim to provide insight into the structural

features that drive major groove binding, shape, or sequence selectivity. The results of this work will aid in designing 3rd generation probes.

2. 4 Design Rationale for Crystal Violet Analogues

Guided by the results of the 1st generation scaffolds by Nunez et al., 2nd generation crystal violet probes were designed and synthesized. Based on the SAR previously conducted, molecules that didn't contain an alkene or alkyne in the linker resulted in minor groove binding. At the time, other students were exploring Di- substituted scaffolds so I developed a method to make a similar substrate with two highly constricted alkyne linkers containing a central benzene and two straight arms with rotatable chromophores at the end (figure 2.7). Therefore, to begin exploring the effects of the linker type, I designed my molecules to contain and lack alkyne tethers to compare them to one another as well as to the 1st generation 1E with a flexible tether. Since scaffolds with a flexible tether had been observed to generally associate in the minor groove and the alkene tethers showed strong association towards the major groove, the alkyne linker simultaneously explores how the added rigidity might influence binding behavior. Based on the previous molecular modeling of the 1st generation molecules, our molecules likely orient in an "open scissor" fashion along the backbone in the major groove, further stabilized by ionic interactions along the widespread reach of the rotating triphenylmethane moieties.

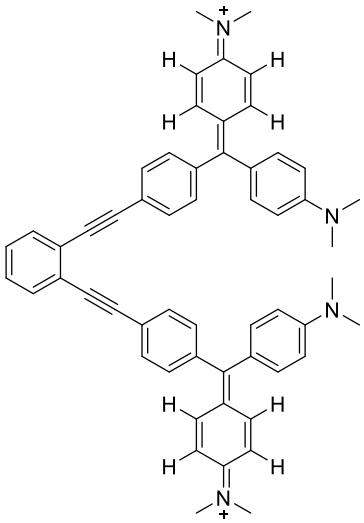


Figure 2.7 Synthesized Violet Derivative: *Dimeric scaffold with alkyne tether,* ***TM1***

1E was the bulkiest and most potent binder from their first-gen molecules, the rest being dimeric or smaller, so the bulkiness of my trimeric compounds could shed light on what the size tolerance might be in some of the major groove units and lead to improved shape selectivity. Similarly, 1E was the only molecule with 3 charges, and they reasoned that downsizing would discourage electrostatic interactions and encourage more selectivity. I wished to evaluate if there was a significant gain in activity or loss in selectivity by increasing the proton count. We also theorized that the bulkier scaffolds might bind and induce sequence-specific expansion, like how transcription factors widen the groove. This rationale led to the design and synthesis of the molecules described in figures 2.7 and 2.8. We expected these molecules to associate in larger major grooves and the rigidity of the arms to orient themselves closely along the base pairs and to make electrostatic contacts with the backbone and H-bonding with the base pairs NH_2 and keto-oxygen for added stability. Importantly, we hoped that the insights generated from these molecules would divulge pertinent information for achieving sequence specificity and the design of 3rd generation probes. The synthetic protocols for the molecules discussed in this section

are outlined in detail in section 2.7. After synthesizing and successfully isolating my molecules, I evaluated their activity via spectroscopic methodologies.

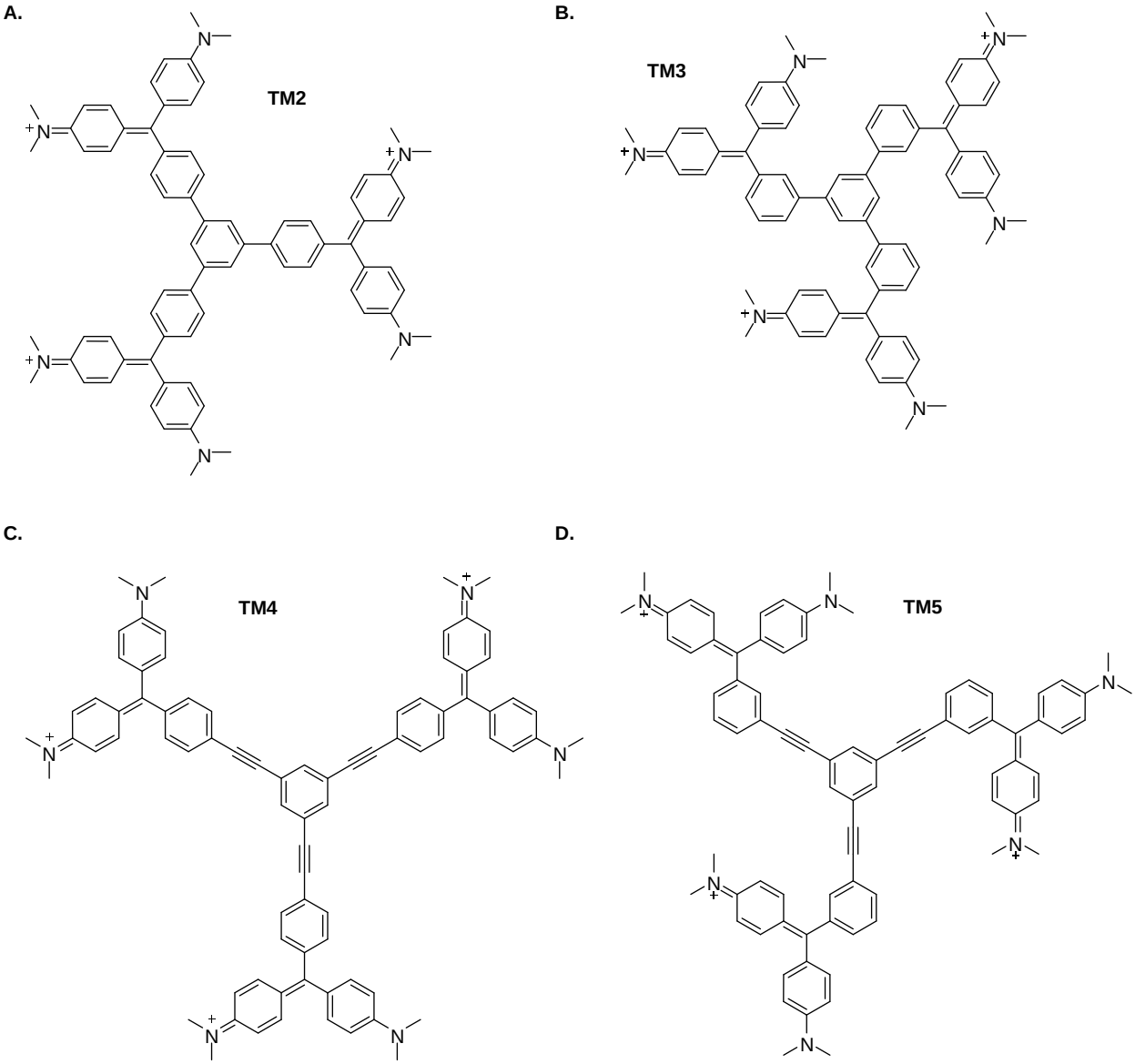


Figure 2.8 Trimeric Crystal Violet probes synthesized. *These were the trimeric molecules designed & synthesized. Rationale described in section 2.4*

2.5 Spectroscopic Studies & Results

2.5.1 UV Vis & Preliminary Fluorescence

With my synthetic compounds in hand, I then used spectroscopic methodologies to gain insight into the binding mode and affinities of our crystal violet derivatives to a series of oligonucleotides. Samples were dissolved in ethanol & UV- Vis was used for monitoring the stability and the concentration of samples, DNA stocks, and dilutions. Experimental details regarding UV-Vis and the calculations of concentrations are outlined in sections 2.8.1-2. Fluorescence spectroscopy was used to quickly probe if any of the scaffolds exhibited affinity towards a particular oligonucleotide.

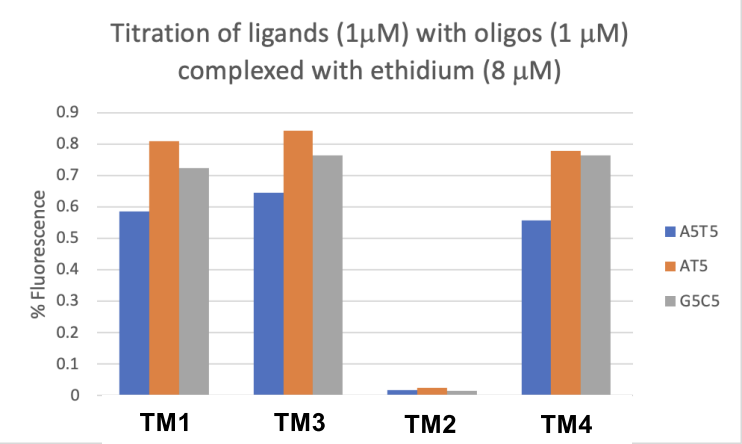


Figure 2.9 *Initial Fluorescence Screen (1μM). Fluorescence of ligands against select oligos. Activity observed for oligos G5C5 & AT5 & A5T5 by TM1 & TM3 look almost identical in ratio, however, TM4 doesn't display the same affinities for AT5 nor G5C5; but rather it displays a degree of selectivity towards A5T5*

Like UV- Vis, fluorescence spectroscopy is a slightly more sensitive method to measure the light emitted from a sample within a given range after absorbing light of higher energy. A series of A-track and G-track oligonucleotides (**A5T5**: 5'-GGGAAAAATTTTTCCC-3'; **(AT)5**: 5'-GGGATATATATATCCC-3'; **G5C5**: 5'-AAAGGGGGCCCCCTTT-3') were then used to obtain preliminary binding data against our molecules and an ethidium displacement protocol was employed to do so[182]. **Results:** I observed that **TM4** displayed a slight affinity for A5T5 among the three studied (figure 2.9 and notably, **TM2** seemed to display a higher affinity overall but no selectivity. However, apparent binding doesn't

always mean target binding, therefore I then wanted to confirm that my molecules were indeed targeting the major groove before moving forward with the titrations.

2.5.2 Circular Dichroism (CD)

CD methodology developed by Greenfield et. al. was used to determine the binding mode of our molecules[183]. CD detects shifts along the DNA backbone as a result of ligand binding; and based on previous studies, negatively induced bands around 310nm correspond to major groove binding, and positively induced bands after 300nm indicate minor groove binding[148,183–186]. The sugars along the backbone induce a signal from 180~260nm, and because each DNA displays a distinct spectrum. The preparation and experimental methods used to produce the following data are discussed in section 2.8.3.

Results: CD spectra (215-330nm) collected for CT DNA in the presence of increasing concentrations of **TM4 and TM5** revealed negative perturbations within the ~310nm region (figure2.10), consistent with major groove binding, not minor groove interactions or intercalation[148,183,185–187]. Interestingly, **TM2 and TM3,** the derivative lacking the alkyne linker (figure2.8) had demonstrated a much higher binding affinity, however, CD spectra indicated a minor groove binding mode. This observation was consistent with previously made observations made by Dr. Minehan and his group for molecules lacking alkene or alkyne linkers in rigid dimeric derivatives (Figure 2.5). In brief, these results provide evidence that the alkene and alkyne tether drive major groove binding. These results may provide insight into structural motifs necessary for shape or sequence selectivity.

Short Discussion: Comparing the CD spectra of my molecules to the control spectra, netropsin, and methyl green, we can confirm that **TM4** & **TM5** bind the major groove

(figure2.10D, E, A, and B, respectively). Netropsin, a minor groove binder, displays positively induced spectra around 310-340nm, and methyl green, a major groove binder, demonstrates a negatively induced spectrum. Furthermore, comparing the spectra of **TM4** & **TM5** to spectra previously obtained by Minehan group members Mickael Khoury and Aren Mirzakhanian, we can observe similar behaviors in binding mechanisms which appear to be dictated by the type of linker utilized. The lack of a linker or an aromatic linker between the dye units, figures2.10C & F respectively, result in minor groove binding; however, the use of an alkene or alkyne linker (figure 2.10 A, B & figure G, H) resulted in spectra that correspond to major groove binding. Interestingly, an aliphatic linker (not shown) demonstrated minor groove binding characteristics.

One theory that might explain this phenomenon is the freedom of rotatability between the different linkers. The rigidity of the alkyne linker may lock the dye units in place, preventing them from rotating into the other binding pockets. Likewise, the alkyl linker could allow for the units to rotate freely into the minor groove. Similarly, the alkene linker could be limiting the rotatability of the units just as the alkyne, however, in the case of the aliphatic linker, one explanation is that the conjugated linker continues to allow for the rotation of units into the minor groove. Subsequent studies will hopefully elucidate the structural factors that drive major groove binding. Considering the data obtained, it appears that an alkyne or alkene linker is required for major groove binding. These findings are interesting indeed and further studies are required to fully characterize this interaction in more detail. Furthermore, ITC studies will soon be underway and would reveal important thermodynamic information regarding any molecular interactions occurring that might

help support our observations that these molecules preferentially bind the major groove as well as their affinity for them.

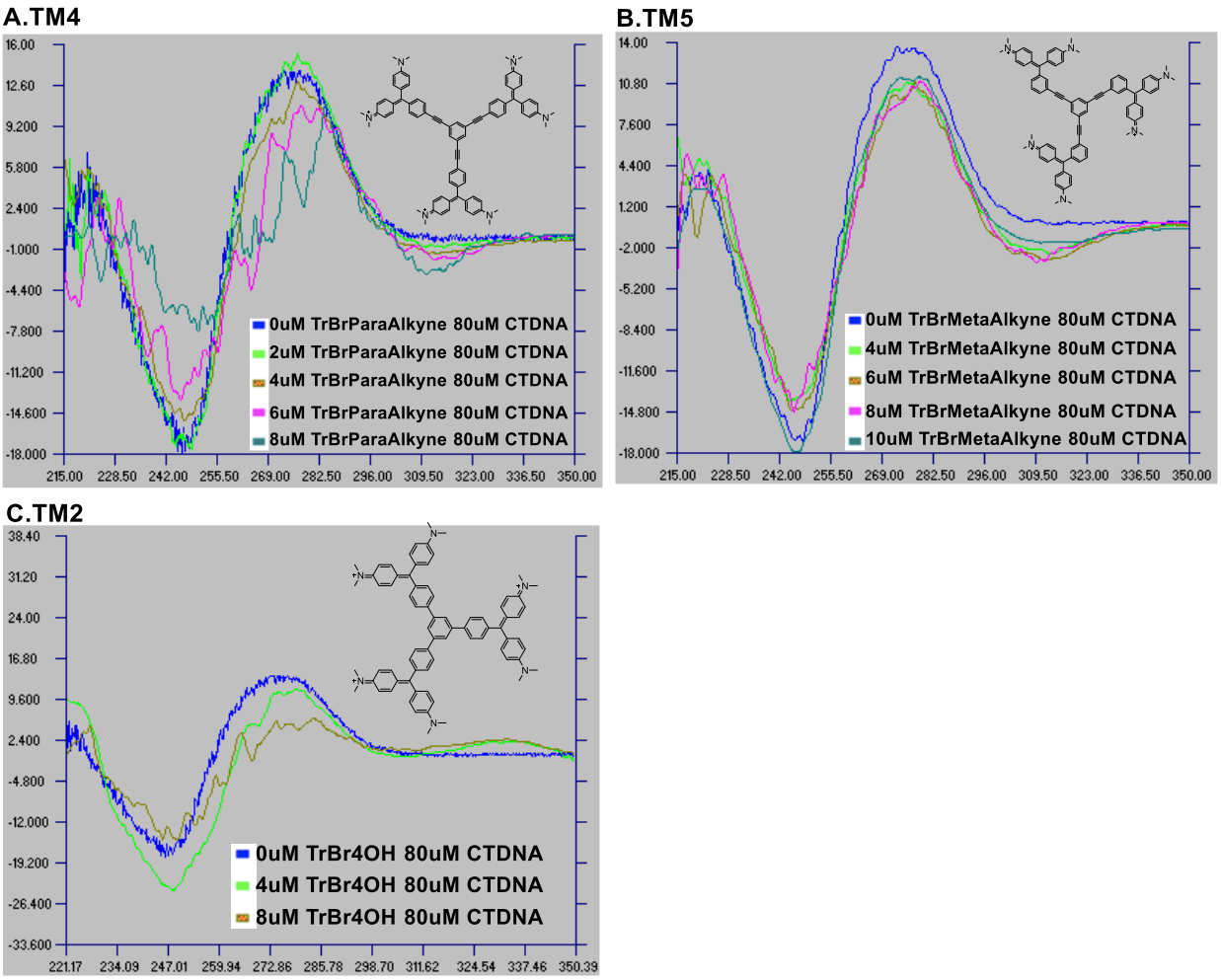
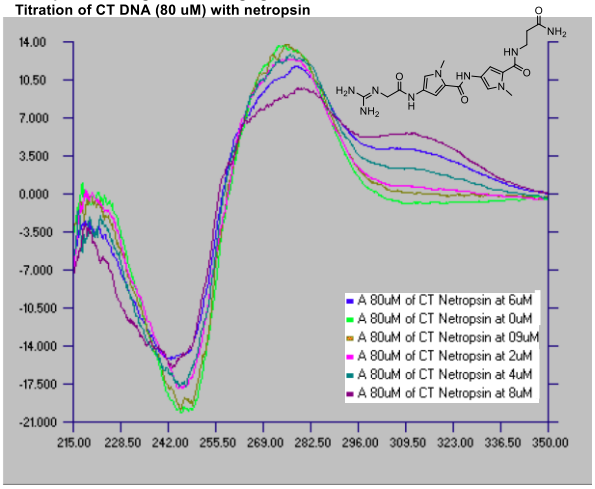
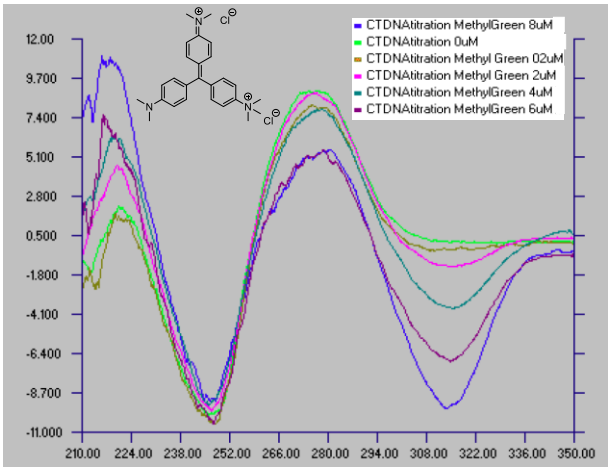


Figure 2.10 CD Spectra of CTDNA and Crystal Violet Derivatives. *Negatively induced perturbations within the ~310nm region in both alkyne- linked (A.TM4 & B.TM5) substrates indicate major groove binding. Positively induced perturbations after 300nm in TM2 indicate MGB.*

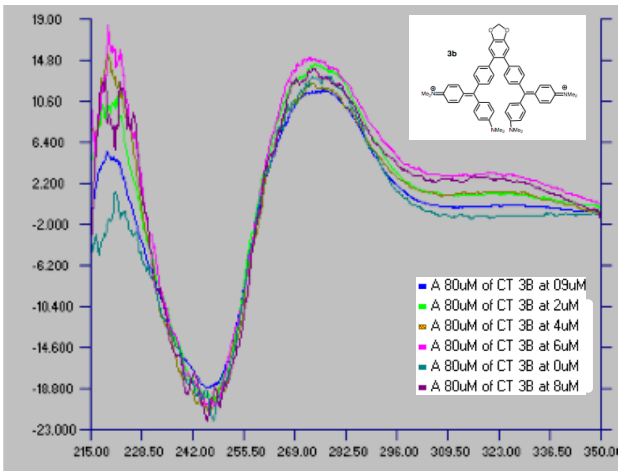
D.
Netropsin - minor groove binding ligand
Titration of CT DNA (80 uM) with netropsin



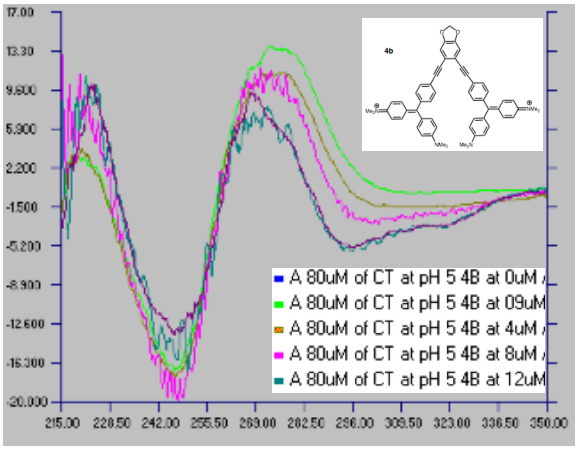
E.
Methyl Green - major groove binding ligand
Titration of CT DNA (80 uM) with methyl green



F.
Titration of CT DNA (80 uM) with compound 3b (minor)



G.
Titration of CT DNA (80 uM) with compound 4b (major)



H. Titration of CT DNA (80 uM) with compound 5b (major)

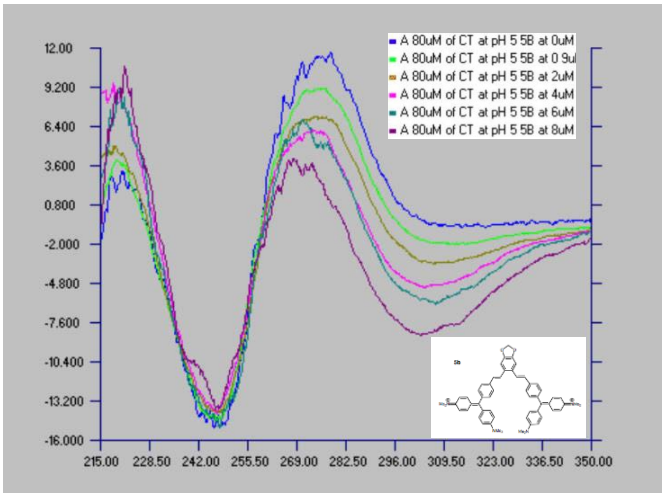


Figure 2.10 (continued)CD Spectra of Controls & other Violet Derivatives*. Negatively induced perturbations at ~310nm region of substrates E, a known major groove binder Methyl Green, G & H indicate major groove binding. Positively induced perturbations after 300nm in D, a known MGB Netropsin, & F indicate MGB.*

2.5.3 Fluorescence

Having gathered preliminary binding mode data for some molecules, fluorescence was employed to evaluate the displacement of ethidium by **TM4** & **TM5** against select hairpin sequences (figure 2.11). The experimental methods are described in section 2.8.2.

Results: These titrations (figure2.12) revealed that **TM5** had a slight affinity, µM, towards the DDD hp, STAT6 hp, cMyc hp, and the GC hairpin, respectively. However, the molecules did not display a significant level of sequence selectivity. However, these interactions may be shape-recognition driven which could eventually lead to sequence-selective molecules that interact with the base pairs.

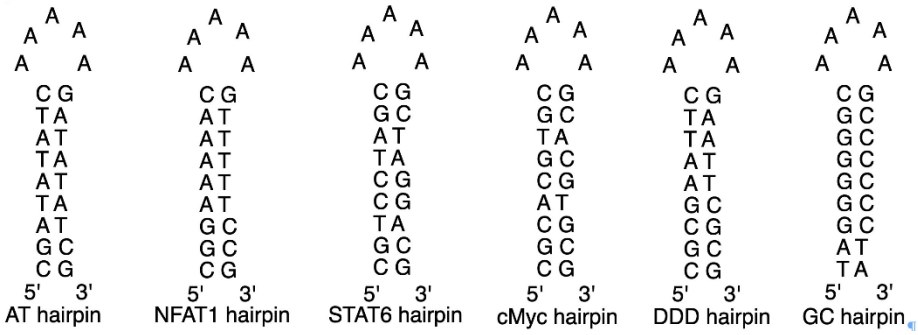
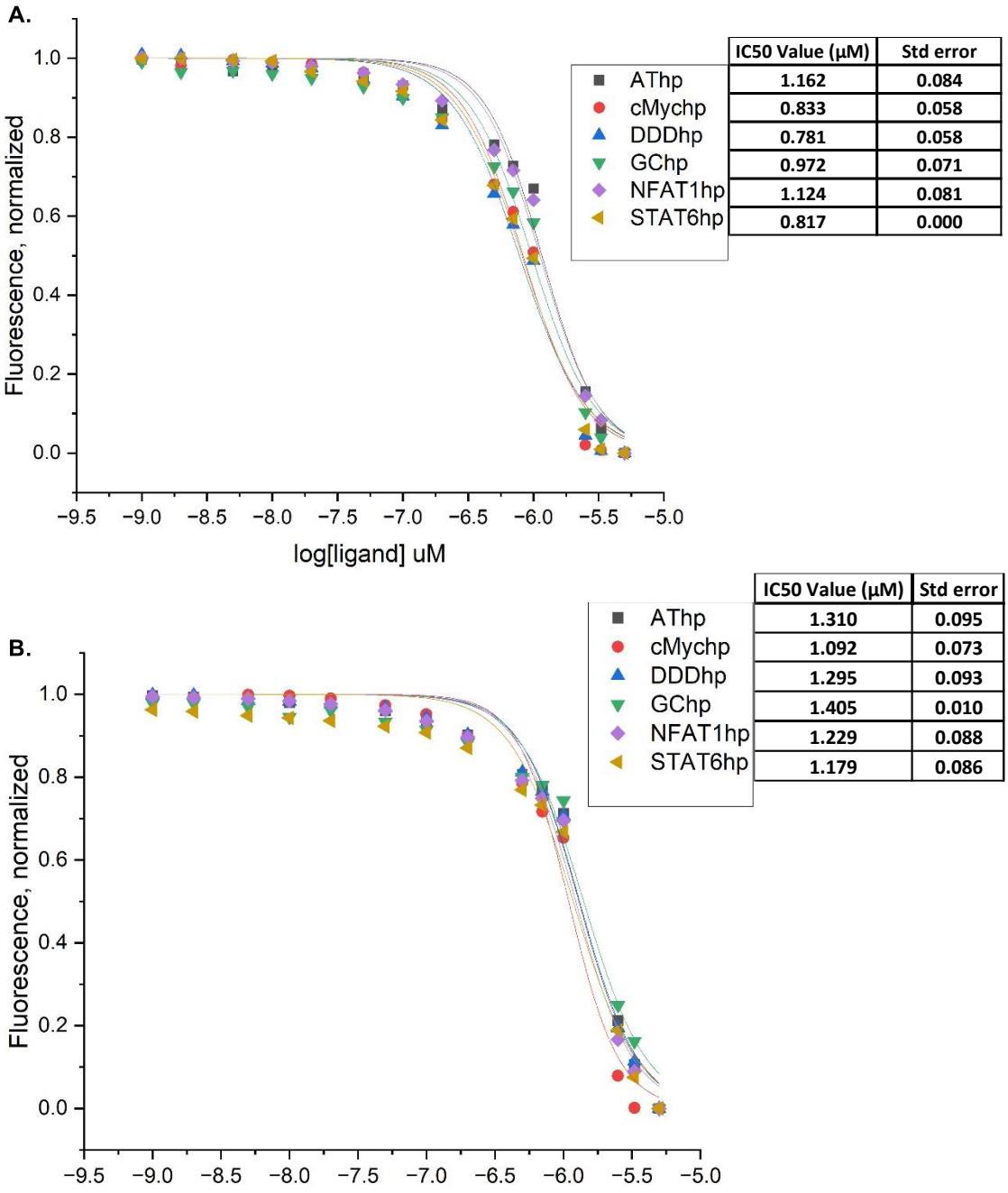


Figure 2.11 *Hairpin Sequences of select oligonucleotides.*

Short Discussion: The preference to GC rich and more balanced oligos came as a surprise since the nucleotide makeup of these hairpins consists of a more balanced mixture of GC:AT sequences, compared to the AT hp and the NFAT1 hp, which are more AT-rich and the methyl green monomer had been shown to have an affinity towards AT-rich sequences as well as having had the wider grooves to accommodate for the bulkier trimeric ligand. This may be an insight into how the rigidity of the molecule can drive shape selectivity for the smaller major grooves despite being bulky.



IC50 Value (µM)	Std error
1.162	0.084
0.833	0.058
0.781	0.058
0.972	0.071
1.124	0.081
0.817	0.000

IC50 Value (µM)	Std error
1.310	0.095
1.092	0.073
1.295	0.093
1.405	0.010
1.229	0.088
1.179	0.086

***Figure 2.12**. Titrations: EtBr displacement with TM4 & TM5 in select synthetic homologs. DNA substrate: 1µM; EtBr: 4µM. Ligand: 0- 5µM. TM5 (A) appears to display a mild affinity for more GC-rich and GC/AT-balanced sequences. However, results for TM4 (B), and TM5 did not display significant levels of selectivity.*

2.6 Conclusions

We hypothesized that the larger tri-substituted scaffolds would occupy the wider grooved hps, however, this was not the case. Moreover, we were able to confirm a structural feature necessary for major groove binding previously observed by the Minehan group. Additional studies are required to investigate how rigidity can drive shape selectivity and how the linker type might drive sequence selectivity in small molecule designs. One possible speculation as to why our molecules associate themselves along the major groove could be associated with the delocalized π orbitals of the alkene/ alkyne tether and its ability to interact with the GC nucleotides. GC nts have an outward-facing NH_2 and keto-oxygen (figure1), both of which are available to interact with the hybridized electron orbitals of the linker, and are further coordinated by the slight electron-deficient dye units as they associate along the DNA backbone. AT nucleotides, however, have no outward NH_2, and only the Keto- Oxygen, perhaps creating a slightly electronegative pocket, giving the non-conjugated linkers enough push to interact with the other binding pockets.

This data taken together with CD data suggests that my molecules might be able to target the major groove of DDD hp with some selectivity over the other synthetic homologs, however, studies are insufficient on their own and more studies are required to investigate how rigidity and other structural features can drive shape selectivity and how sequence selectivity can be designed into small molecule designs that interface closely with the base pairs. ITC will be used to acquire thermodynamic parameters of our molecules interacting with the DNA substrate which will hopefully support our observations that these molecules preferentially bind the major groove as well as their affinity for them.

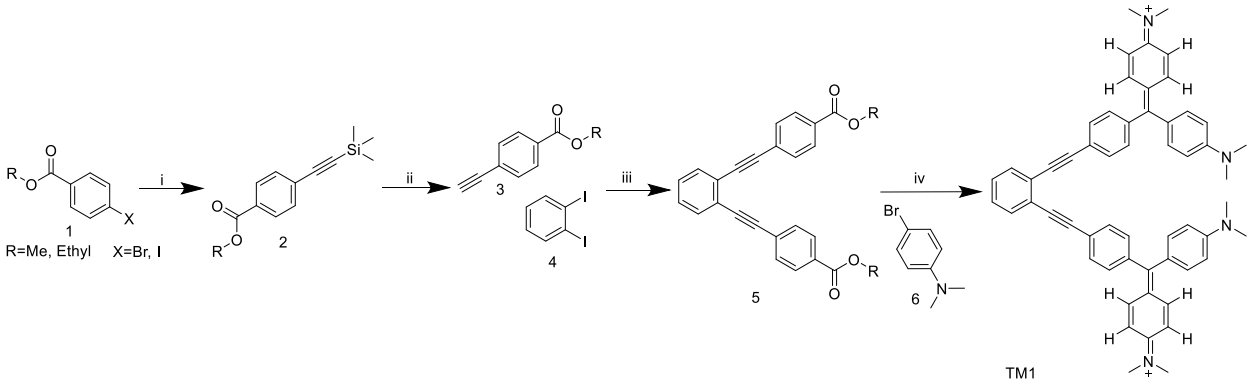
DNA plays an essential role in all biological functions and understanding its biology further could allow us to design better therapeutics or discover novel targets. Although the molecules described did not display high levels of sequence selectivity, this work is an important first step towards developing major groove binders and lays out the foundation for identifying structural features which are critical for driving a preference towards the major versus minor groove binders. To rationally design and develop novel drugs, we must first have a better understanding of the DNA-ligand interactions, and our approach is to develop probes that will give us key insights into efficiently designing those analogs. With a growing interest already underway and FDA-approved drugs already in the market, utilizing small molecules to probe and understand DNA biology to elucidate more of its long-range correlations (LRC) and other pharmacologically relevant targets; there is potential for major-groove selective probes to be built into probes and potentially therapeutic strategies[188,189].

2.7 Methods & Experimental

References in this section pertain to the synthetic methodology implemented. Preliminary NMR spectra are in Appendix: I.

2.7.1a Synthesis General Synthesis of Crystal Violet Derivatives

Scheme 1: Synthesis of TM1:



i. Using previously described sonogashira conditions (***J.Organometallic Chem. Volume 774, 2014, pg 19-25***), a heat-dried Schlenk was loaded with **(1)** (1.5g, 6.60mmol), 0.05eq $PdCl_2(PPh_3)_2$ (0.20g, 0.327mmol), 0.06 eq. CuI (0.075g, 0.393mmol) and 0.1 eq. PPh_3 (0.172g, 0.655 mmol) capped, degassed, and flushed with N_2 (3x). In a separate closed flask was added TEA (32ml) which was subsequently degassed and flushed with N_2 (3x). The TEA was then transferred into the Schlenk reaction vessel. TMS-acetylene was added dropwise over 20min. The reaction was run under inert atmospheric conditions at 60°C. for >12h. This reaction was monitored via TLC. Upon completion, the mixture was cooled and filtered on celite, washed with TEA, and concentrated in vacuo. This crude product was flash columned (Hex:EA10-15%) to isolate intermediate (**2)(75%).**

2.7.1a.(2): ^{1}H NMR (300 MHz, $CDCl_3$) δ 7.95(2H); 7.5 (2H); 3.89 (3H); 0.25 (9H) ppm. Previously described in **J. Org. Chem. 2003, 68, 12, 4862–4869**.

^{13}C NMR (100MHz, $CDCl_3$) δ 166.5, 131.8 (2C), 129.6, 129.3 (2C), 127.7, 104.0, 97.7, 52.2, 0.19 (3C) ppm. Previously described in **Org. Lett. 2011, 13, 5, 998–1000.**

ii. Using a previously described method (**Synthesis 2008(4): 605-609**), Intermediate **(2**) was dissolved in MeOH (20ml) and an excess of K_2CO_3 (≥2eq.) was added and stirred aggressively for 3h at RT. The solvent was removed in vacuo and redissolved and extracted in DCM and water. The organic phase was dried over Na_2SO_4, concentrated in vacuo, and subjected to flash chromatography (Hex:EA 3-5%) to isolate **(3)(80%).**

2.7.1a.(3): ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.01 (2H); 7.56 (2H); 3.93 (3H); 3.24 ppm. Previously described in **Org. Lett. 2011, 13, 5, 998–1000.**

^{13}C NMR (100MHz, $CDCl_3$) δ 166.49, 132.16, 130.19, 129.54, 126.82, 82.87, 80. 18,52.38 ppm. Previously described in **J. Med. Chem. 2010, 53, 3, 994–1003**

iii. In a similar fashion to step (i), sonogashira conditions (***J.Organometallic Chem. Volume 774, 2014, pg 19-25)*** were employed to cross couple substrates **(3)** and **(4).** A heat dried Schlenk was loaded with 1eq. **(4)** (0.14g, 0.438mmol), 0.09eq $PdCl_2(PPh_3)_2$ (0.03g, 0.044mmol), 0.12 eq. CuI (0.010g, 0.053mmol) and 0.14 eq. PPh_3 (0.016g, 0.061 mmol) capped, degassed, and flushed with N_2 (3x). In a separate closed flask was added TEA (2ml) which was subsequently degassed and flushed with N_2 (3x). The TEA was then transferred into the Schlenk reaction vessel. In a separate flask, **(3)** (0.175g, 1.09mmol) was dissolved in degassed (0.5ml) TEA and transferred dropwise to the reaction vessel over 20min. The reaction was run under inert atmospheric conditions at 60°C. for >12h. This reaction was monitored via TLC. Upon completion, the mixture was cooled and filtered on celite, washed with DCM, and concentrated in vacuo. This crude product was flash columned (Hex:EA6-9%) to isolate intermediate (**5)(55%).**

2.7a.5: ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.02 (d, *J*= 9, 4H); 7.61 (d, *J*= 8, 2H) 7.57(m, 6H); 7.34 (m, 2H) 3.92 (s, 6H) ppm. Preliminary NMR data. Previously unreported spectra.

^{13}C NMR (100MHz, $CDCl_3$) δ 166.3, 131.9, 131.4, 129.7, 129.5, 128.5, 127.8, 125.4, 92.9, 91.0, 77.4, 77.1, 76.8, 77.4, 77.1, 76.8, 52.2 ppm. Preliminary NMR data. Previously unreported spectra.

iv. Finally, a previously described method (***Bioorganic Chemistry 83 (2019) 297–302)*** was used to cross-couple **(5)** and **(6)** to synthesize the crystal violet derivatives[148,190]. To a heat-dried shlenck was added 6eq bromo-analine **(6)** (0.583g,2.91mmol) in 2ml of dry THF. The solution was then cooled to -78°C under inert, N_2, conditions. 6eq of n-BuLi (2.0M in hexanes, 1.46ml, 2.91mmol) was then added to the reaction over 20min and stirred for another 15 minutes at -78°C to allow for sufficient Li-Br exchange. In a separate flask, 1eq **(5)** (168mg, 0.486mmo) was dissolved into dry THF (0.5ml) and subsequently transferred dropwise into the reaction vessel and allowed to stir at -78°C for 15m. The solution was then brought to room temperature and allowed to stir for >12h; allowing the substrates enough time to coordinate before being quenched with an excess of 1M HCl (15ml) and stirring for 15min; protonating the Li-coordinated intermediate to give us our crude product. This solution was then extracted into Chloroform (6-8x 50ml) (Product is expected to be lost because the protonated species dissolves into water as well. The organic phases were collected and dried over Na_2SO_4, concentrated in vacuo, and subjected to flash chromatography (Hex:EA) to isolate the bis-crystal violet derivative, **TM1**. The isolation of these products was indeed tedious and

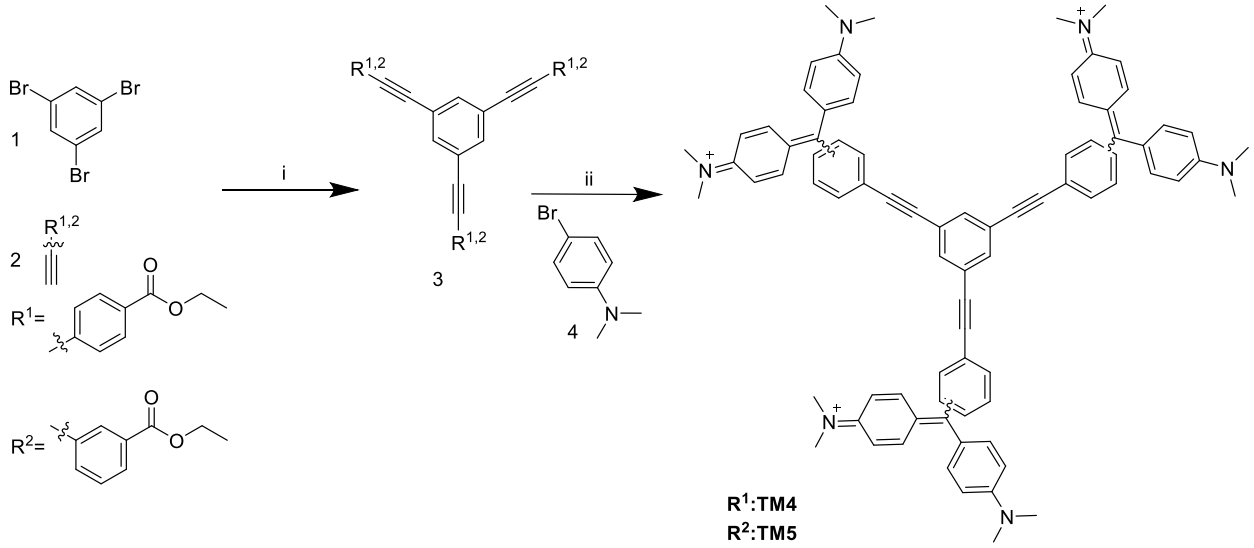
inconsistent as they tended to stick to the column and elute with contaminants, warranting a better purification strategy (25-30%).

TM1: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.71(m,1H); 7.57 (m, 9H); 7.28(m,7H); 7.19(d,*J*=8,1H); 7.13(m, 5H); 6.68(m, 6H); 2.95(m,24H). Preliminary NMR data. Previously unreported spectra.

^{13}C NMR (100MHz, $CDCl_3$) δ 166.8, 141.4, 139.5, 134.5, 130.8, 130.5, 130.1, 128.0, 127.9, 127.8, 77.4, 77.0, 76.7, 52.0. Preliminary NMR data. Previously unreported spectra.

UV (Ethanol): λ_{max} 210nm (Ɛ 66,331) 318 (43,687) 620(174,000)

2.7.1b Scheme 2: Synthesis of Tri-alkyne Substituted Benzene



i. Alkyne-linked- esters **(2)** were prepared following conditions discussed above in **2.7.1a.i-ii**. Followed by another sonogashira reaction (***J.Organometallic Chem. Volume 774, 2014, pg 19-25)***, however, in this case, 1eq of the tri-halogenated

(1) substrate was used, and 3.3eq of the alkyne **(2)** substrate was used to give intermediate (**3)** after isolation (70-80%).

2.7.1b.(3, R^1): ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.05 (d, J = 7.5 Hz, 6H, ArH), 7.71 (s, 3H, 11 ArH), 7.60 (d, J = 7.5 Hz, 6H, ArH), 4.41 (q, *J* = 6.9, 6H, -CH2-), 1.41 (t, *J* = 6.9 Hz, 9H, -CH3). As previously described in **Carbon Volume 87, June 2015, Pages 338-346.**

^{13}C NMR (125MHz, $CDCl_3$) δ NMR 166.2, 134.8, 131.8, 130.5, 129.8, 127.4, 90.5, 90.3, 61.5, 14.6. As previously described in **Carbon Volume 87, June 2015, Pages 338-346.**

2.7.1b.(3, R^2): ^{1}H NMR (200 MHz, $CDCl_3$) δ 8.23 (dd,J=1.6, 1.6, 3H); 8.05 (ddd,J=7.9,1.6, 1.6, 3H); 7.72 (ddd,J=7.9, 1.6, 1.6, 3H); 7.71 (s, 3H); 7.48 (dd,J=7.9,7.9, 3H); 4.42 (q,J=7.1, 6H); 1.43 (t,J=7.1, 9H). As described in **Helvetica Volume 86, Issue 2, February 2003. pg 494-503.**

^{13}C NMR (75MHz, $CDCl_3$) δ 166.16; 135.91; 134.62; 133.06;131.19; 129.89; 128.81; 124.08; 123.36; 89.87; 88.61; 61.39; 14.33 ppm. As described in **Helvetica Volume 86, Issue 2, February 2003. pg 494-503.**

ii. Derivatives **TM4** and **TM5** were synthesized using the protocol described above in section 2.7.1a (***Bioorganic Chemistry 83 (2019) 297–302)***. However, the following equivalents were used: 1eq of intermediate **(3**), 9eq of the bromo-analine **(4**), and 9eq of n-BuLi(20-20%).

TM4: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.68(m,7H); 7.39(m,23H); 7.02(d,J=9,10H); 3.40(s,36H). Preliminary NMR data. Previously unreported spectra.

^{13}C NMR (100MHz, $CDCl_3$) δ 206.9, 156.9, 140.7, 134.8, 131.6, 127.2, 114.0, 77.3, 77.0, 76.7, 41.2, 29.6, 22.2 ppm. Preliminary NMR data. Previously unreported spectra.

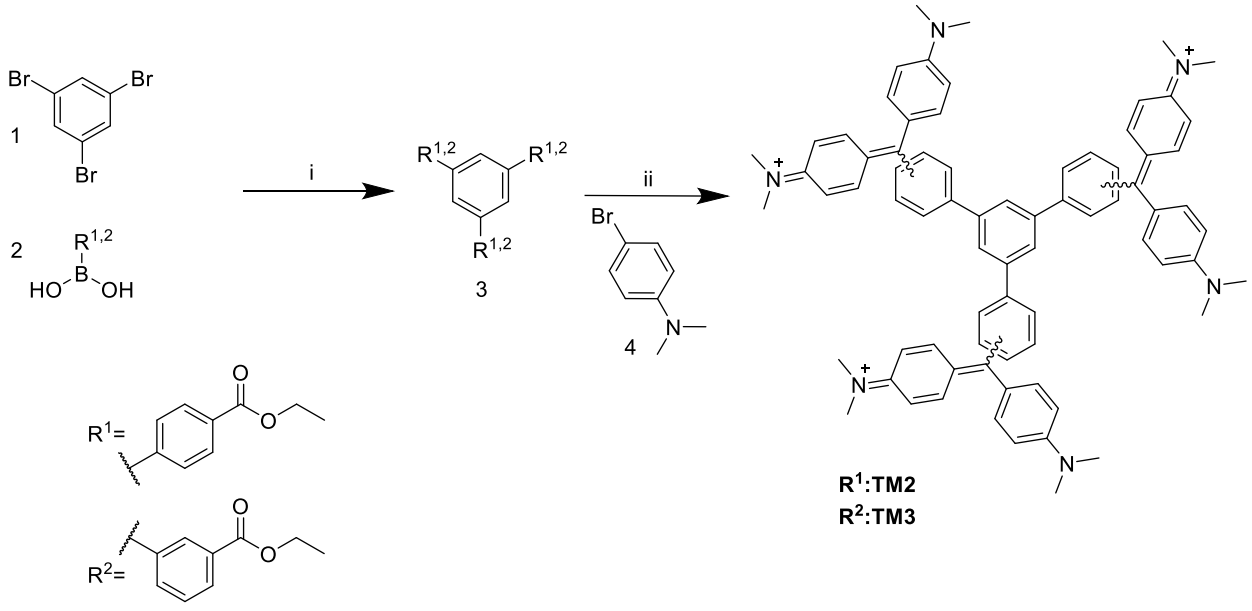
UV (Ethanol): λ_{max} 210nm (Ɛ 252,365) 308 (101,006) 630(261,000)

TM5: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.56(m,3H); 7.47(s,3H); 7.43(s, 3H); 7.41(s, 2H); 7.33(s, 1H), 7.23(s,1H); 7.13(d,J=8, 10H); 7.04(s, 2H); 7.02(s, 2H) ppm. Preliminary NMR data. Previously unreported spectra.

^{13}C NMR (100MHz, $CDCl_3$) δ 149.0, 140.7, 136.8, 130.0, 128.8, 128.4, 127.7, 127.2, 114.1, 111.9, 111.7, 111.5, 77.3, 77.0, 76.7, 64.9, 41.2, 40.58, 40.52, 40.47, 40.41, 13.9. Preliminary NMR data. Previously unreported spectra.

UV (Ethanol): λ_{max} 210nm (Ɛ 154,347) 310 (102,250) 628(261,000)

2.7.1c Scheme 3: Synthesis of Tri- substituted benzene, no linker



i. Following general Suzuki coupling protocol from **Organic Letters 2011, 13, 2, 252-255**, Under N_2 atmospheric conditions, into a heat-dried Schlenk was added 1eq 1,3,5-tribromo-benzene **(1**)(1mmol), 4eq of the boronic ester(**2**)(4mmol), 9eq K_2CO_3 (1.2g, 9.0mmol) and (0.1-0.3%mol) $PdCl_2(PPh_3)_2$ catalyst. Reactants were suspended in a 1:1 solution of Tol:EtOH and stirred at 95°C for >12h. The reaction was then cooled and extracted (H_2O,EA)(3x 20ml). The organic phase was collected and dried over Na_2SO_4, concentrated in vacuo, and columned (Hex: EA) to isolate the tri-substituted ester **(3)**.

2.7.1c.(3, R^1): ^{1}H NMR (500 MHz, CDCl3) δ 8.16 (d, J = 8.4 Hz, 6H), 7.85 (s, 3H), 7.76 (d, J = 8.4 Hz, 6H), 4.42 (q, J = 7.2 Hz, 6H), 1.43 (t, J = 7.2 Hz, 9H) ppm. As described in **JACS Au 2021, 1, 11, 2080–2087.**

^{13}C NMR (125 MHz, CDCl3) δ 166.5, 145.0, 141.8, 130.3, 129.9, 127.4, 126.1, 61.2, 14.5 ppm. As described in **JACS Au 2021, 1, 11, 2080–2087**.

2.7.1c.(3, R^2): ^{1}H NMR (400 MHz, CDCl3) δ 8.32 (m, 3H); 8.09 (d, J= 8, 3H); 7.84 (d, J= 8, 3H); 7.56 (m, 3H); 4.45 (m, 6H) 2.19 (s, 2H); 1.44 (m, 10H) ppm. Preliminary NMR data. Previously unreported spectra.

^{13}C NMR (125MHz, CDCl3) δ 166.2, 144.2, 130.1, 130.0, 128.3, 127.17, 127.10, 77.4, 77.0, 76.7, 61.0, 14.3 ppm. Preliminary NMR data. Previously unreported spectra.

ii. Derivatives **TM2** & **TM3** were synthesized using the protocol described above in section 2.7.1a. iv (***Bioorganic Chemistry 83 (2019) 297–302)***. However, the following equivalents were used: 1eq intermediate **(3**), 9eq of the bromo-aniline(**4),** and 9eq of n-BuLi(20-30%).

TM2: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.95(d, J=8,5H); 7.51(d, J=8,6H); 7.48(d, J=9,13H) 7.37(s,3H); 7.06(d, J=9,12H); 3.42 (s, 36H). Preliminary NMR data. Previously unreported spectra.

^{13}C NMR N/A

UV (Ethanol): λ_{max} 210nm (Ɛ 327,676) 310 (169,128) 624(261,000)

TM3: ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.75(m,8H); 7.57(m,5H); 7.46(m,3H); 7.28(m,3H); 7.21(m,10H); 6.71(m,10H); 2.95(m, 36H). Preliminary NMR data. Previously unreported spectra.

^{13}C NMR N/A

UV (Ethanol): λ_{max} 210nm (Ɛ 1,180,173) 264 (996,545) 624(261,000)

2.7.2 Methods, Materials, and Extinction Coefficient References

A transparent quartz cuvette was used for all spectroscopy experiments.

UV- Visibility Spectroscopy (UV-Vis)

UV-Vis was used to quantify the concentration of our dye stock solutions in EtOH (crystal violet derivative dissolved in 200-500µL) as well as our CT DNA & hairpin substrates. To do so, absorbance data is collected by measuring the amount of light passing through our sample (2ml 100mM Tris-HCl buffer, pH7 (fluorescence)/ pH5.04 (CD)), 2µL sample stock sol. or 15µL CT DNA. Absorbance was scanned between λ: 400-600nm in triplets for our ligand samples and at λ: 260nm for CT DNA. To calculate the concentration, the absorbance was applied to the lambert-beer law equation, which describes the absorbance (A) as being dependent on

optical path length(L), concentration (C), and the sample/wavelength-specific molar extinction coefficient (Ɛ); A= L x C x Ɛ. Therefore, provided the extinction coefficient is known, this formula can be re-arranged to derive the concentration, {C= A / (L x (Ɛ x 2/3 units))}. Extinction Coefficients of our crystal violet derivatives were 87,000AU/unit, therefore we multiplied the molar coefficient by a factor of 2 or 3, depending on how many crystal violet sub-units the compound had. For CT DNA, we used a previously published extinction coefficient of 6600AU/unit. After calculating and assigning a concentration, stocks were used to prepare fresh dilution samples stored long term at -22°C, as they degraded quickly when stored in buffer. The wavelength at which absorbance was observed (i.e. λ:592), was saved for later determining the range in my fluorescence experiments. An HP 8452A Diode Array Spectrophotometer & software were used to collect and translate UV-Vis data.

Fluorescence Spectroscopy: Ethidium Displacement Assay

To collect fluorescence data, fresh, serial dilutions (outlined below) were prepared for concentrations 1mM/100µM/10µM Experiments were started at a volume 2ml of pH7 buffer (100mM Tris- HCl, pH6.8), 4µM EtBr, 1µM DNA substrate. The DNA substrate and ethidium bromide were allowed to incubate for 30min. Then a varying amount of freshly prepared samples to obtain the desired dilution concentration was added. Ethidium bromide experiments were done keeping the DNA substrate constant and increasing the concentration of my ligands. The titrations and the subsequent measurements were done by hand (samples were added and allowed to incubate 3min before readout). The measurements were

collected with a max emission wavelength of 520nm and an excitation wavelength of 590nm; and fluorescence recordings from 550- 660nm. Scan Speed: 200nm/min; Em slit: 10.0; Ex slit: 10.0nm. A Perkin Elmer LS55 luminescence spectrometer & FL win Lab, Spekwin32, and Extract DataV5 software were used to collect and translate fluorescence data. Protocols were derived from Jenkins and Nielson et. al[191,192].

Serial dilutions for titrations: all dilutions were prepared fresh in buffer from a stock solution in EtOH. A 1mM stock (150µL) of each molecule was made. From the 1mM stock, serial 10x dilutions were prepared (100µM and 10µM). In this way, volume from the lowest concentration stock would be added to achieve the lowest desired concentration of ligand to DNA up to the highest concentration from the more concentrated stocks (1µM-5mM).

Fluorescence data were first processed via excel. Ligand concentration was converted to µM and plotted in log form. The fluorescence data were normalized using excel and the following formula: $\mathbf{((X_{\text{fluorescence data point}} - X_{\text{min fluorescence}})/(X_{\text{max fluorescence}} - X_{\text{min fluorescence}}))}$. Plotting was done with one origin software by laying a scatter graph and manually fitting it with a global dose-dependent sigmoidal curve.

Circular Dichroism (CD)

CD is an absorption spectroscopy technique that can measure the difference in absorption between left- and right-circularly polarized light, as a function of wavelength or frequency. CD is particularly well suited for monitoring

conformational changes in the backbone within the region of 230-178nm and it is assumed that the spectrum is a linear combination of CD spectra, each of which contributes to the secondary structure. To collect data, 2 ml (Xmmol, 80µM CT DNA, 100 mM Tris-HCl) of fresh samples were prepared and loaded into the cell. Final ligand concentrations: 2-8µM ligand and 80uM/bp CTDNA. The samples were put into a quartz cuvette and measurements were taken from 350- 215nm with a step resolution of 0.2 nm increments, a speed of 90 sec, and monochromator/ slit wavelength at 466.3nm, and bandwidth at 1nm. The resulting CD spectra were an average of two scans taken at room temperature. To obtain an adequate signal-to-noise, determining the spectra of the secondary structure required 60-90min runs. The raw CD was then processed, smoothed, and inverted since the initial readout is a mirror image of the actual data. Changes in secondary structure as observed from the reference spectra were considered. An AVIV CD 202 Spectrometer and software were used to collect and translate CD data.

Data Analysis, Materials, and extinction coefficient references

All DNA substrates for the biophysical experiments were purchased and prepared as recommended by the seller. Stock concentration and stability/decomposition checks were performed using UV- Vis.

Crystal violet: Adams, E. Q.; Rosenstein, L. J. Am. Chem. Soc. 1914, 36, 1452–1473. **EtB**r: Saucier, J. M.; Festy, B.; LePecq, J.-B. Biochimie 1971, 53, 973-980.

CT-DNA was purchased from CALBIOCHEM. Solutions of CT-DNA were prepared in 10mM Tris-EDTA buffer at pH 5.48 (as described in Jenkins: Jenkins, T.C.

Optical Absorbance and Fluorescence Techniques for Measuring DNA-Drug Interactions. In Methods in Molecular Biology, Drug-DNA Interaction Protocols; Fox, K.R., Ed. Humana: Totowa, 1997; Vol 90, pp.195-217) and gave a 1.78:1 absorbance ratio at 260 nm and 280 nm. DNA and ligand concentrations were determined using 8452A HP Diode Array Spectrophotometer: CT-DNA, e260 = 6600 M-1 cm-1 bp-1

Binding data were processed using Microsoft excel and OriginLab

Chapter 3 The Identification of a Novel Oncogenic pre-miRNA Probes

The understanding of miRNA's critical role in human biology has prompted several studies to attempt to uncover their binding conjugates and how to manipulate their activity. miRNAs activities have garnered a lot of interest in recent years which has prompted studies into characterizing these molecules, the molecules they interface with, and their biological activities. Though several miRNAs of interest have been identified, miR21 and pre-let7 are the most prevalent in disease and the most widely studied miRNAs, respectively.

3.0 RNA Substrates

Of all the miRNAs to study, we decided to screen miRNAs that are associated with the development of a wide range of diseases, primarily miR21, and the highly studied let-7d.

pre- Let7d

The let7 family was among the first miRNAs discovered and coincidentally is also one of the first conserved miRNAs across species[193]. In humans, there are over 10 variations of mature let7 and all of them target key oncogenic proteins such as c-Myc, Ras, and HMGA2 and are involved in cell differentiation; making it a critical biomarker in patient prognosis[194–197].

Let7 is an oncolytic miRNA, therefore, a decrease in let7 results in disease. Its inherent function has forced cancer cells to adopt several mechanisms for suppressing let7's anti-

tumoral activity and creating an undifferentiated cell state. One of which is to disrupt transcription, let7 gene affects DNA methylation; and low transcription is associated with hypermethylation. Another is post-transcriptionally, and it does so by interfering with the biogenesis pathway, resulting in low levels of mature let7d[198]. Furthermore, single nucleotide mutations in the miRNA complementary site result in RISC not being able to identify the target mRNA for degradation[199].

pre- miR21

miR21 is one of the most highly expressed and studied miRNAs since first having been identified in the VMP1 locus of chromosome 7[200]. They are highly conserved miRNAs, expressed locally in basically every cell where they carry out critical regulatory functions that manage health and disease.

Interest has grown through the years due to its high prevalence and causative role in neoplastic and non-neoplastic pathologies by way of its target genes, which are mainly TPM1, PDCD4, and PTEN. miR21 is an oncogenic miRNA, therefore cancer cells have adopted mechanisms by which to high-jack regulatory machinery, effectively streamlining the overexpression of the miR21[201,202].

3.1 Novel in-vitro Assay for Studying miRNA

To probe the potential of small molecules as therapeutics, several technologies have been evolved to study miRNA. Of the in vitro assays, the techniques to note are:

Microarray techniques were developed to study gene expression using populations of RNA. Now, this method has also been refined for purposes such as analyzing the level

of protein expression or gene amplification of DNA. For example, this technology can be used to analyze RNA levels to study changes in gene expression in tumors.

Fluorescence displacement assay (FID) is a fast, high-resolution technique that can evaluate the selectivity and affinity of nucleic acids for proteins and other molecules.

Fluorescence resonance energy transfer, or FRET, is a technique that can accurately detect changes in molecular proximity in a diversity of biological phenomena.

The former techniques are certainly important when getting a full understanding of a molecule/substrate's SAR, but they have their limitations. To quickly screen large libraries (>50,000) of molecules/ natural product extracts (NPEs) for therapeutic value, high throughput screening methodologies have been developed. However, until recently, probing the chemical space for miRNA binders was not possible. Section 3.2 briefly describes a methodology developed in-house for rapidly and reliably screening molecules and their ability to interrupt dicer-mediated maturation.

3.2 cat-ELCCA

The Catalytic Enzyme Linked Click Chemistry Assay (cat- ELCCA) is a method that makes use of click chemistry to overcome many of the pitfalls of typical HTS assays and is discussed in greater detail elsewhere[89,138–140]. Earlier this year Dr. Morten Meldal, Barry K. Sharpless, and Carolyn R. Bertozzi won the Nobel prize in chemistry for their work in developing click chemistry tools; a testament to their significance in pushing assay development forward. Click chemistry negates the need for using expensive antibodies and reagents, making it ideal for large screening campaigns[203].

Cat-ELCCA was developed and refined by Dr. Daniel Lorenz and Amanda Garner to identify and quantify the association of molecules with a pre-miRNA substrate. The design allows for easy interchangeability of substrate and incorporates two modifications(Figure 3.1). To immobilize the RNA substrate and avoid interfering with dicer's activity, a polyethylene glycol linker with a biotin label was attached at the 5' end of the sequence. In the stem-loop of the pre-miRNA substrate, another modification was introduced, an amino allyl uridine capable of reacting with N-Hydroxy succinimide (NHS) to incorporate the click chemistry handle, trans cyclooctene (TCO), which later allows for the quantification of dicer mediated digestion.

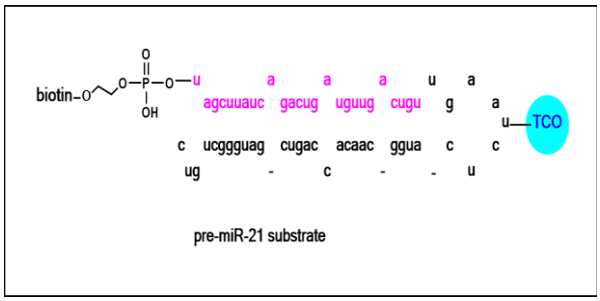


Figure 3.1 Modified pre-miR21 *substrate with the biotinylated. Pre-miR21 with aminoallyl uridine in the stem loop and a polythelene glycol linker with biotin at the 5' end. SLAS Discovery 2018, Vol. 23(1) 47–54.*

A brief description of cat-ELCCAs experimental design:

Making use of another strong bio-ligation combo, cat-ELCCA utilizes streptavidin-coated plates and a 5' biotinylated pre-miRNA substrate which contains a modified uridine bound to the click handle, TCO, located in the stem-loop region. Firstly, the miRNA is incubated overnight to allow for adequate binding to the plate

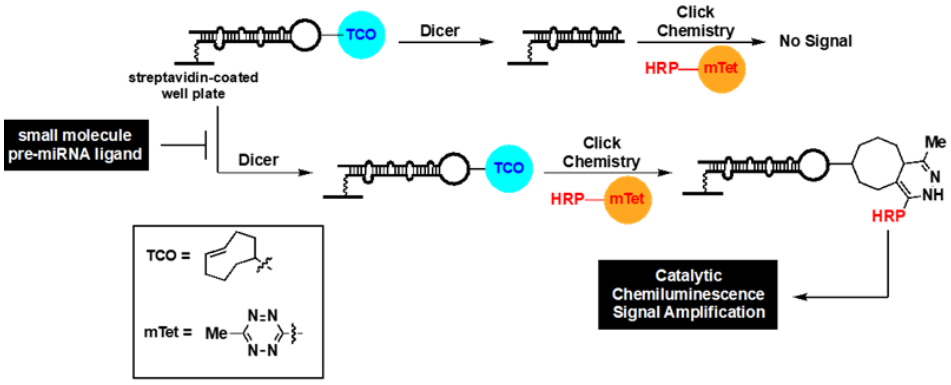


figure 3.2 Experimental Design of cat-ELLCA. *miRNA is immobilized to plate, incubated with ligand, then incubated with Dicer, washed, and incubated with HRP-mtet bio conjugate before a final wash before chemiluminescent readout. Inhibition conserves TCO on the stem-loop for click reaction. Dicer maturation results in loss of TCO bioconjugate and lack of chemiluminescent signal. SLAS Discovery 2018, Vol. 23(1) 47–54. DOI: 10.1177/2472555217717944*

(Figure 3.2). After washing away any unbound pre-miRNA, the bound pre-miRNA is then incubated with the compound or NPE of interest. This is followed by a five-hour incubation with dicer, and finally, after thoroughly washing the wells, a third incubation with a modified horseradish peroxidase containing the complementary click handle, methyl-tetrazine (met). If dicer maturation occurs, the stem-loop along with the TCO will have been cleaved off and no signal will be observed. However, if dicer maturation is inhibited, TCO will remain intact and readily react with mtet to undergo a strain-promoted inverse electron demand Diels- alder click chemistry reaction. Once ligation occurs, the wells are washed once more to remove any unreacted HRP before the addition of HRP substrate, which results in a chemiluminescence amplification readout that correlates directly to pre-miRNA intact.

3.3 Research Objective

With the completion of the human genome project, our understanding of miRNAs and their critical role in human biology have made them an attractive therapeutic strategy, however, to date, it remains an untapped pharmacological target. One of the major challenges in exploiting this strategy is the discovery of molecules that discriminately act on miRNA and the tools to carry out the discovery of novel molecules. The Garner lab developed cat-ELCCA, previously described, to conduct HTS of miRNA against chemical libraries to aid in identifying selective miRNA binders. To date, the FDA has not approved any miRNA-targeted drugs, however, natural products have always offered inspiration in the discovery of novel, therapeutic molecules, another poorly tapped resource regarding RNA. The implementation of cat-ELCCA and natural products to the discovery of novel miRNA-targeted probes resulted in the identification of naturally derived novel miRNA-

targeted probes, the surfactins. However, we referred to our synthesized derivatives as hermicidins. The hermicidins, as we called them, are a class of lipodepsipeptides known as surfactins, a heptapeptide characterized by the presence of a β-hydroxy fatty acid and cyclized via a lactone bond. Surfactins are powerful antiviral, antitumoral, and anti-mycoplasma agents[204]. The synthetic hermicidins and naturally derived surfactins described here and by Robertson et. al., respectively, have revealed previously reported and unreported surfactin molecules (NP1-6) to display a mild affinity towards pre-miR21[147]. Furthermore, SAR studies revealed that these macrocycles were binding with sub-micromolar affinity to miRNA as well as demonstrating activity in a luciferase assay expressing miR21. This work describes our structure-based approach that makes use of medicinal chemistry and other characterizing methods to enrich our understanding of these molecules and their potential as pharmacologically relevant targets. Specifically, the objective of this work was to:

A) **Further Characterize and Confirm the Structure of our Target Hermicidin Molecules.** The first step in developing the hermicidins was to confirm the structure and stereochemistry via total synthesis of the initial target molecule. Despite the promising activity of these molecules, their activity is modest at best. Therefore, the synthetic strategy used will be employed to conduct structural SAR studies by using medicinal chemistry to derivatize the hermicidin scaffold. Furthermore, in collaboration with Dr. Tripathi, a chemoenzymatic synthesis strategy was developed. The result of these studies will be to have confirmed and optimized the surfactin-based probes that can inhibit pre-miRNA maturation.

B) **Discover Additional Lead Molecules that can Manipulate miRNA Activity**.

From the original screen of natural products, other NPEs demonstrated different levels of selectivity and should be followed up on. Furthermore, a chemical library provided by Elli Lilly might also elucidate more molecules that target pre-miRNA via cat-ELCCA. The identification of other miRNA interacting pharmacophores could aid in developing molecules that selectively inhibit premiR21.

C) Further Characterize by Collecting Biophysical and Cellular Activity Data.

Despite demonstrating a crude binding affinity for pre-miRNA, several unknowns regarding how these molecules interact with the miRNA substrate remain. Therefore, another objective is to characterize how the NPE surfactins and the synthetic analogs are interacting with the miRNA substrate and to establish a mechanism of action. Cellular activity profiles will also be conducted to further understand the activity of these surfactins.

3.4 cat-ELCCA, A Platform for the Discovery of pre-miRNA Binders

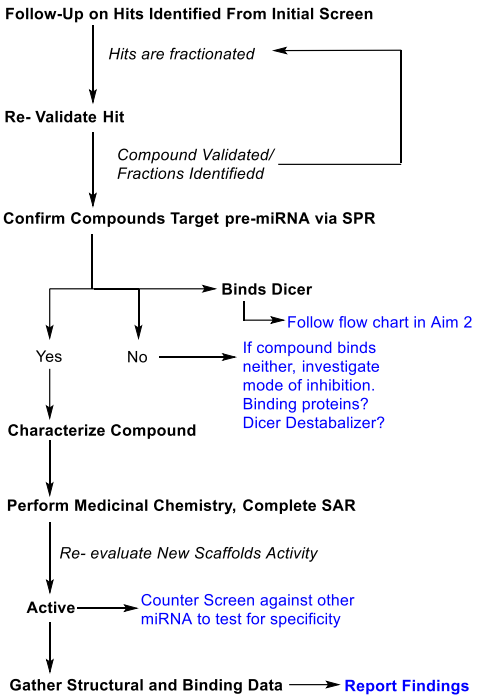
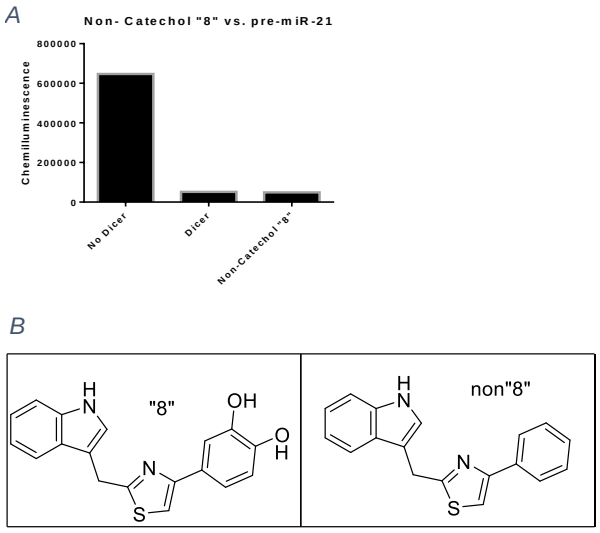


***Figure 3.3 General work-flow of hit follow up**. Hit validation prompts characterization of lead molecule and interaction with substrate. SAR includes biophysical and biochemical assays*

Previously published work by my colleague and friend, Dr. Daniel Lorenz provided evidence that cat-ELCCA is a useful tool for probing pre-miRNAs[89,140]. The HTS campaign was conducted at the University of Michigan's Center for Chemical Genomics (CCG) and cat-ELCCA was the method utilized for screening a library of ~85,000 commercial small molecules and natural product extracts to identify potential pre-miR-21-selective inhibitors. Upon counter-screening against a second miRNA, pre-let-7d, it was revealed that natural products may be a source of selective inhibitors of a single miRNA. This discovery was exciting, as this screen revealed a chemical space that remained to be explored. Follow-up on these hits followed the workflow in figure 3.3.

3.5 Inhibition of Dicer Mediated Maturation of miRNA via Small Molecules



***[3]Figure 3.4 Small Molecule: Catechol**. A. cat-ELCCA of non- catechol containing, non-8. Controls: No Dicer has no enzyme, therefore no cleavage of the stem loop so chemiluminescence is expected. Dicer, contains enzyme as observed by the loss of signal. Non "8" does not exhibit inhibition. B. Structure of "8" and non-"8"*

From the initial screening campaign, I followed up on two of the hits that appeared to be binding the miRNA. One of the first compounds to emerge from our screen was referred to as "8" figure 3.4B. Having pulled the structure from the libraries index, we then attempted to purchase the molecule. However, the compound was not commercially available, so we purchased a similar substrate without the catechol moiety, non-catechol "8" figure 3.4B. To quickly assess the activity of this scaffold I ran a digestion gel assay and observed a loss in activity (figure 3.4 A). Later, cat-ELCCA studies confirmed loss of activity, therefore we theorized that the inhibition was derived from the catechol moiety. I then formulated a synthetic route to confirm the structure and activity of compound 8 and several other derivatives. Concurrently I was screening a collection of tetracyclines that had been identified to inhibit dicer-mediated maturation in the same screen.

The activity of the tetracyclines was confirmed via digestion gel and cat-ELCCA (figure 3.5), however, another method was needed to investigate how our small molecules were interacting with the substrates. In collaboration with Dr. Erin Gallagher, select molecules were studied using surface plasmon resonance (SPR), a label-free, sensitive method for studying molecular interactions such as binding events. SPR data revealed that neither set of molecules was interacting with either ligand to any significant degree, therefore, another mechanism was taking place. To investigate the mechanism of action further, we screened a selection of tetracyclines, included was CMT-3, a tetracycline that lacks any antibacterial properties possibly due to their lack of RNA binding[205]. SPR studies by Dr. Gallagher confirmed that CMT-3 does not interact with RNA, however, cat-ELCCA studies I conducted for all tetracyclines indicated varying degrees of inhibitory activity, like the compound “8” scaffolds.

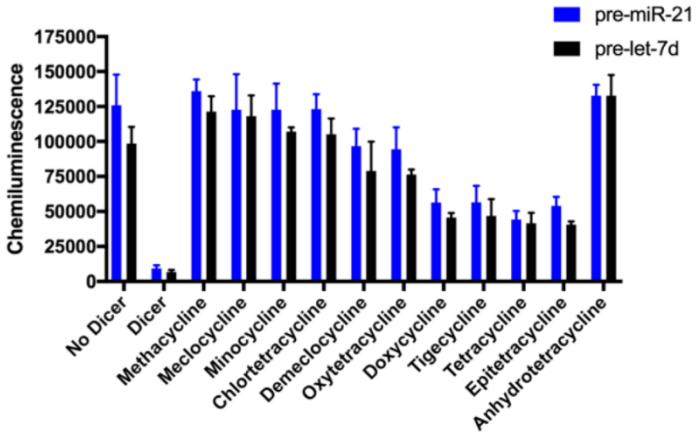


Figure 3.5. cat-ELCCA of Tetracyclines *and their apparent inhibition of Dicer mediated maturation of pre-miRNAs at 1.0mM. ACS Medicinal Chemistry Letters (2019) 10(5) 816-821*

Further investigation suggests that despite their apparent ability to inhibit pre-miRNA maturation, these molecules were chelating magnesium ions in the assay buffer solution, an ion that is critical for dicers' coordination of the RNAse III domains and proper functionality. To probe this phenomenon, in collaboration with Dr. Tanpreet Kaur, we synthesized an acetylated oxytetracycline (figure 3.6A) that was designed to prevent metal-ion chelation[206]. The results of these experiments exhibited a loss of activity, further supporting the theory that metal coordination is necessary for tetracycline-based dicer inhibition, and not their role in RNA binding (figure 3.6 B). Controls without enzyme (lane 1) remained unprocessed just as the lane with the suspected metal coordinator, oxytetracycline (lane 3). However, lane 4, with the Ac. Oxytetracycline exhibited similar processing as the control with miRNA and Dicer. These findings correlated to previous publications citing these molecules as dicer inhibitors and Mg^{2+} chelators[207,208206]. The full list of molecules and their characterization are available in **ACS Medicinal Chemistry Letters (2019) 10(5) 816-821**. These findings highlight the importance of thoroughly characterizing interactions using both biochemical and biophysical methods. These molecules could be used to provide insight into dicer biology and its role in human disease and taken together provide support for the use of cat-ELCCA as a powerful tool to screen for viable targets and interactions and should be used concurrently with other characterization methods.

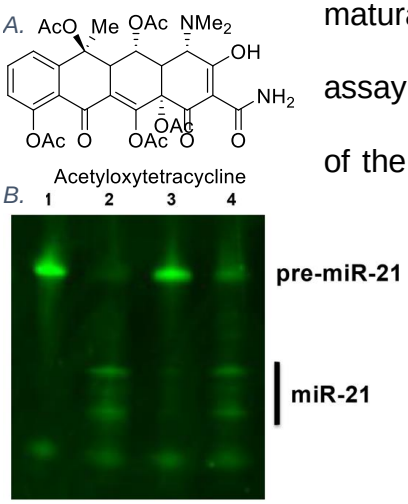


Figure 3.6 Acetyloxytetracycline Results *A. Acetyl- protected oxytetracycline.* ***B****. Inhibitory activity in a Dicer digestion assay.* ***Lane 1*** *= pre-miR-21 (500 nM) -Dicer;* ***Lane 2*** *= pre-miR-21 + Dicer;* ***Lane 3*** *= pre-miR-21 +Dicer and oxytetracycline (1.0 mM);* ***Lane 4*** *= pre-miR-21 + Dicer and the acetylated oxytetracycline (1.0 mM). ACS Medicinal Chemistry Letters (2019) 10(5) 816-821*

3.6 Targeting let7d-Lin28 Protein Interaction

Additionally, Dr. Lorenz screened several molecules against we well studied pri & pre-miRNA-Protein interaction, pr-let7d- Lin28. Lin28 is a key example of a modulating RBP; Lin28 associates with pre-let7d at the terminal stem-loop and inhibits Drosha and Dicer-mediated maturation. I was able to follow up on and quickly derivatize a set of sulfonamide molecules (figure 3.7). What we found was that our molecules interrupted the lin28-let7d interaction, however, SPR data by Dr. Gallagher revealed that these molecules were not binding the RNA substrate, therefore, more studies are necessary to evaluate the mechanism- of- inhibition for these molecules. The full list of molecules and their characterization is found in **ACS Medicinal Chemistry Letters (2018) 9(6) 517-521**. Moreover, this work provides further evidence for the use of cat-ELCCA in the discovery of novel targets and can potentially reveal novel targetable interactions and binding proteins.

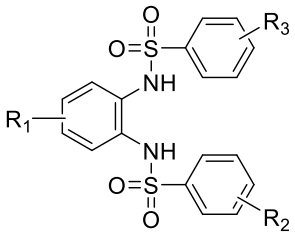


Figure 3.7 General structure of Sulfonamides. *Highly derivatized set of Line28 probes. ACS Medicinal Chemistry Letters (2018) 9(6) 517-521*

3.7 Discover of Bioactive Surfactins as Potential pre-miRNA Probes

Natural product-derived drugs have proven themselves to be some of the most powerful therapies, therefore, we wanted to include the NPE library at UofM's CCG. Following up on the screen of over 32,000 NPEs, I followed up on 22 extracts that we selected, and were regrown in collaboration with Dr. David Sherman's Group. The NPE library was made up of unknown mixtures, therefore, active extracts required activity-guided fractionation before isolating the active molecule (Appendix:II). This was done in

collaboration with Dr. Ashootosh Tripathi, the Director of the University of Michigan's Natural Products Discovery Core in the Sherman Group.

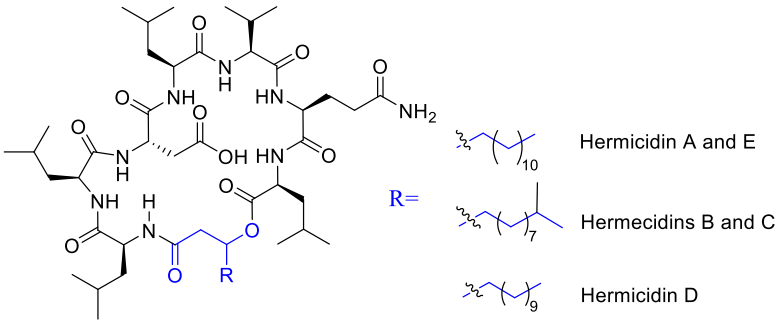
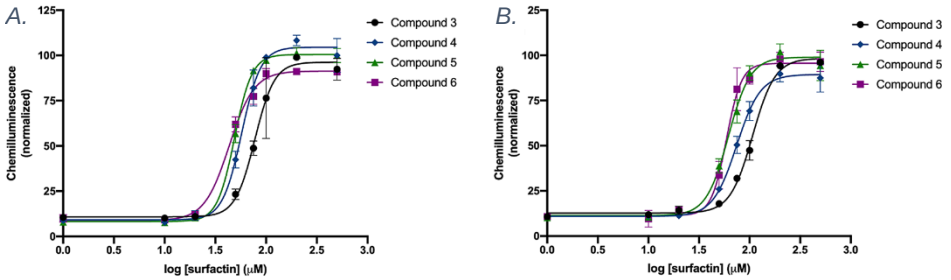


Figure 3.8 Structure of Isolated Surfactins A-E. *ACS Medicinal Chemistry Letters (2021) 12(6) 878-886*

Of the 22 extracts regrown, two strains: 91085 and 82379, demonstrated nanomolar affinity towards pre-miR21. We decided to prioritize our efforts on extract 82379, as it had demonstrated 3-fold selectivity for pre-miR21 over pre-let7d. After several rounds of activity-guided fractionation, we were able to purify, isolate and identify several target molecules. To elucidate the structure of these molecules, high-resolution mass spectroscopy (HRMS) studies were conducted and revealed HR ESI-MS $[M+H]^+$ ion peaks of m/z 1008.6610, 1008.6649, 1022.6786, and 1036. 6941 g/mol and the corresponding chemical formulas $C_{51}H_{91}N_8O_{12}$, $C_{52}H_{93}N_8O_{12}$, and $C_{53}H_{97}N_8O_{12}$, respectively. After this, Dr. Tripathi performed 1D- and 2D- NMR experiments to determine the planar structures of our isolated products. Additionally, Marfay's analysis was utilized for the chemical derivatization of our molecules to assign absolute stereochemical confirmation[209]. These studies revealed a class of macrocyclic lipodepsipeptides belonging to a non-ribosomal peptide synthetase- polyketide synthase (NRPS-PKS) originating from Playa Hermosa, Costa Rica. Unlike other RNA targeting molecules, our molecules, the surfactins (Figure 3.8), were uniquely intriguing as they did not contain any aromatic or charged residues, which can very often lead to non-specific

binding[210–214]. The surfactins belong to a larger family of macrocycles that have been studied for their therapeutic potential[215,216].



C.

Surfactin	IC_{50} (μM)		K_d (μM)	
	pre-miR-21	pre-let-7d	pre-miR-21	pre-let-7d
1	>200	>200	NT	NT
2	>200	>200	NT	NT
3	71–84	97–120	0.38 ± 0.07	0.3 ± 0.2
4	52–63	70–79	0.43 ± 0.03	2.5 ± 0.06
5	48–50	54–69	0.73 ± 0.03	1.4 ± 0.05
6	38–48	54–65	0.78 ± 0.08	1.2 ± 0.05

***Figure 3.9 Biochemical activity of hermicidins**. (A,B) cat-ELCCA of Dicer mediated maturation of miR21 and Let7d, respectively. Table C) IC_{50} values from cat-ELCCA, 95% confidence and K_d values as measured by SPR. NT: not tested. ACS Medicinal Chemistry Letters (2021) 12(6) 878-886*

Having obtained the isolated products and their chemical structures, I then proceeded to confirm their activity and strategize a synthetic pathway not only to confirm the structure but also to then derivatize our scaffold. The synthetic efforts of these molecules are described in section 3.10.1. SPR data collected by my collaborator, Dr. Gallagher, was able to confirm that pre-miRNA binding was the mechanism of inhibition and determine K_d values between 380-800nM (figure 3.9C) To determine and investigate the relationship between binding affinity and inhibition, next, I measured IC_{50} values of surfactins C-E utilizing cat-ELCCA and each demonstrated antagonistic, dose-responsive curves with a range of IC_{50} values for different pre-miRNAs (figure 3.9A, B, C). There were some observed discrepancies between the binding and inhibition that had been previously

observed in peptide-based probes and we attributed this to the complex nature of dicer interactions[143,217].

Next, we then wanted to investigate where our molecules were interacting with the RNA substrate. Previous binding studies had revealed a secondary structural shift occurring in the loop region; therefore we designed chimeric pre-miRNAs which were made up of a pre-let7d loop on a pre-miR21 stem and pre-miR21 apical loop on a pre-let7d stem and I screened their activity via denaturing gel and cat-ELCC assay[218,219]. Binding was not observed in the let7d loop/ miR21 stem chimera, possibly due to improper folding, however, we did observe binding between the mir21 loop/ let7d stem chimera. These findings along with the previous data collected by my colleagues and me supported our theory, the surfactins are likely to be interacting with the loop region, and the selectivity observed for premiR21 is likely due to its terminal loop motif. We hypothesized that the cyclic peptides may be acting through shape recognition and interfacing at the stem-loop, similar to how RNA- binding proteins (Lin28) regulate miRNA maturation[220,221]. These findings were very exciting for us as not many molecules with these unique properties had been reported[222–224]. [1]

In collaboration with another lab member, Arya Menon, we then employed a previously reported luciferase reporter assay to gain insight into how this activity might translate into cells[225–227]. Delightfully, the results of these studies garnered a statistically significant 2.2± 0.5- fold increase in firefly luciferase activity from the miR21 reporter[147]. These findings corresponded to the inhibition of miRNA-mediated gene silencing. The full

characterization of the isolated natural products discussed, and experiments can be found in **ACS Medicinal Chemistry Letters (2021) 12(6) 878-886**. Fueled by these encouraging results I was eager to begin the synthesis & optimization of the hermicidins.

3.8 A Synthetic Journey into Hermicidins

The target hermicidins were seven amino acids in length and contained a chiral β-hydroxy acid linker (figure 3.8). Guided by structure-based design, I utilized solid-phase peptide synthesis, and organic and medicinal chemistry principles to synthesize and optimize our macrocyclic scaffolds. However, first I needed to resynthesize the isolated peptides to confirm their structure and stereochemistry. Luckily, the total synthesis of other depsipeptides had been previously described and aided in developing my synthetic strategy[228–231].

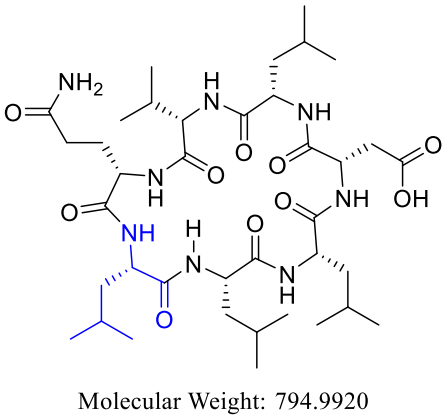


Figure 3.10 Macrocycle lacking the β-Hydroxy fatty acid. *Made to evaluate the importance of the depsie linkage.*

3.8.1 Initial Strategy

As a starting point, I wanted to determine if the β-hydroxyl linker was of any importance, using the known sequence of our depsipeptide, I synthesized a macrocyclic peptide lacking the depsie- ester (Figure3.10). and found that nearly all, if not all activity had been lost (Appendix IV), indicating that one or both, the depsi- portion or the long aliphatic hydrocarbon chain, were the active moieties.

Using my working knowledge of sequence order and stereochemistry, I attempted my first synthesis of hermicidins. The synthesis of the hermicidins involved several challenges which would require an orthogonal approach to cleave the linear peptide from the resin, cyclize, and deprotect without disrupting the pre-incorporated depsie- bond. An approach described by Papini et al. would allow for the sequential synthesis of the linear peptide, the incorporation of the depsipeptide, cleavage from the resin, and a head-to-tail cyclization[232]. However, because of the need for concentrated TFA, standard rind amide SPPS was not suitable, so I opted for the more acid-labile resin, chlorotrityl resin. Due to the reactivity of the depsie-ester towards some SPPS reagents, this approach required that the depsi- linkage be incorporated just before cleavage and cyclization. Therefore, I synthesized a monomer by functionalizing a leucine to contain the depsie-ester linker and the terminal alloc protecting group (figure 3.11A).

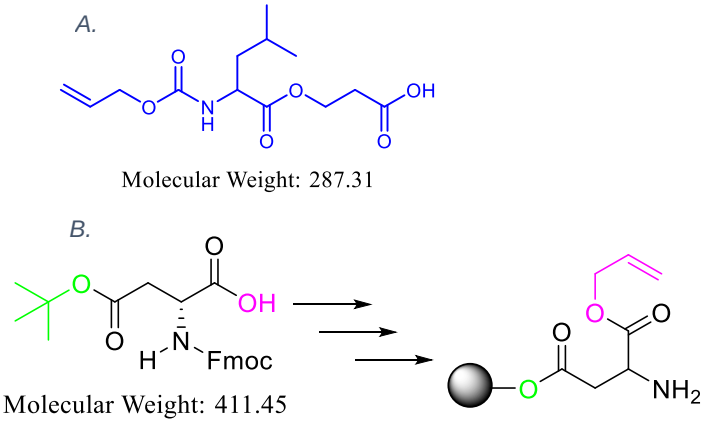


Figure 3.11 Strategies for Incorporation of Depsi-ester Linkage.
A) *Modified Leucine Monomer for Head-to-tail cyclization. J. Pep. Sci. (2004) 10(4) 218-228*
B*) Modified Aspartic acid for Head-to-side chain cyclization. Tetrahedron Letters (1992) 33(32) 4557-4560*

Simultaneously I was developing a strategy in which cyclization would occur on- resin in a head-to-side chain fashion. Ideally, this would address one foreseeable issue with the former strategy, polymerization rather than cyclization. Here, I utilized the former monomer and similar SPPS methodology to build the peptide, however, first I modified an aspartic acid so that the acid sidechain associated with resin, leaving the AAs alloc-protected carboxyl- end open for deprotection and cyclization (figure 3.11B).

Unfortunately, both approaches were unsuccessful as I was unable to deprotect alloc successfully, and rather than troubleshooting for too long, I searched for another.

3.8.2 Final Synthetic Strategy of Hermicidins

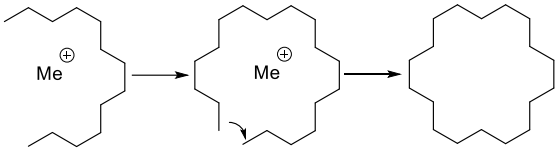


***Figure3.12. Ion Templated approach**. Ion serves as a template for growing macrocycle.*

At that time, Johnston et. al. had recently reported on the synthesis of macrocycles using mitsunobu conditions coupled with an ion-templated approach (figure 3.12)[233]. This strategy made use of Fmoc solid phase peptide synthesis, 2-chlorotrityl resin, HFIP for milder cleavage, and a Na^+ ion template that had been reported to promote desired macrocycle size and required me to synthesize the chiral β-hydroxy linker since not many compatible beta-hydroxy linkers were available for purchase. Using this method, I was able to conduct preliminary experiments to prepare a hermicidin derivative and test its activity via cat-ELCCA, although my preliminary experiments were not further validated. The synthetic strategy is outlined in 3.10.1.

3.9 Conclusions

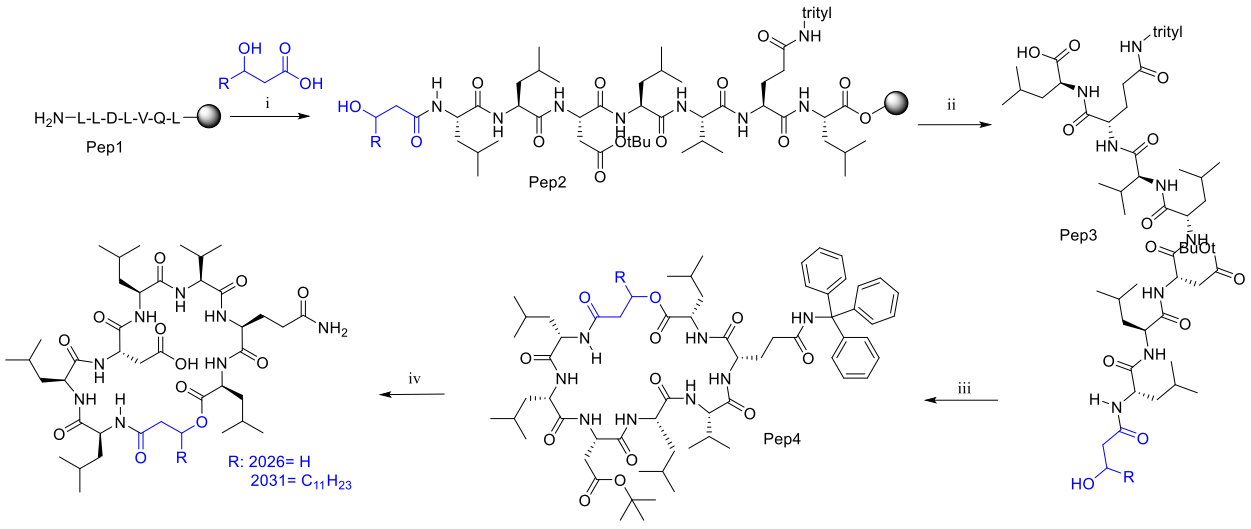
For the collective efforts of my collaborators and me; along with the implementation of innovative technologies such as cat-ELCCA, fate allowed us to successfully identify the surfactins for the first time as inhibitors of miR21-mediated gene silencing. Though these molecules had previously been characterized to some extent, their activity towards pre-miRNA and inhibition of dicer-mediated maturation could prove to be model molecules that can be optimized into selective pre-miRNA probes. These molecules had been previously characterized as a group of molecules named surfactins and their synthesis

had been previously published via SPPS and other synthetic strategies similar to the ones described here in this work[234]. How these molecules are interacting, however, needs to be studied further, but we theorize that these molecules are interacting like with the stem-loop of the miRNA substrate, similar to binding proteins such as Lin28. I hope that this strategy will serve as a guide in the continued development of these molecules. Collectively, these findings are all very encouraging and provide evidence for a probe that can target a single miRNA selectively and will require more extensive studies if they are to be developed into a therapeutic strategy.

3.10 Methods and Experimental

Methods and results described should be considered preliminary. NMR data is unavailable, and experimental are provided to illustrate procedures but not to claim proof of structure and purity. The full characterization of the naturally derived surfactins discussed in this chapter were conducted by Dr. Tripathi and can be found at **ACS Med. Chem. Lett. 2021, 12, 6, 878–886**.

3.10.1 Scheme 4: Mitsunobu Coupled Ion-Templated Macrocyclization



i. The complete synthesis of **(Pep1)→** hermicidins **2026 & 2031** was prepared following a procedure outlined in the **J. Am. Chem. Soc. (2018) 140(13) 4560-4568**, which combines standard SPPS and an ion-templated macrocyclization approach described in section 3.9.2. The linear heptamer was sequentially made with a majority of natural enantiomer (L)-AAs, interestingly however, the AA Asp/[Q] was the only (D)-AA incorporated into the peptide. The β-Hydroxy acid linkers used to make 2031 and 2030 were purchased as Hydroxy myristic acid and 3-hydroxy propanoic acid, respectively, and were incorporated into **(Pep1)** using standard coupling protocols, however, this coupling was carried on overnight (>12h).

ii. Next, the resin was treated with a solution of 20% HFIP in DCM (10ml) and shaken for 1hr, twice to fully cleave the peptide off and render linear **(Pep3)**. The solvent was then removed in vacuo and the crude residue was used as is. HFIP reagent had recently been published at the time and has since seen more adoption as an efficient, milder solution for cleaving fully protected peptides[235–237].

iii. Into a dry flask flushed with N_2, 2.5eq. of template salt sodium perchlorate or sodium borotetrafluorate, 1eq. of linearized protected **(Pep3)**, 6 eq. PPh_3 and 10% DMF in toluene (0.02M) was allocated. 5eq. DIAD was then added over the course of 1h and stirred at RT for 24h. The solvent was removed, and the remaining residue was purified in a short silica plug to remove some of the unwanted by-products such as TPPO from the previous reaction. No more than 1% MeOH, in DCM, was used to remove byproducts. Then a quick incremental

increase to 20%MeOH to flush out the peptide quickly as it would stick to the column. The concentration of active fractions yielded our protected macrocycle **(Pep4).** The Na-salts were selected because these particular ions promote the appropriate heptamer lactonization that we wanted[233,238].

iv. Macrocycle **(Pep4**) was then treated with a milder solution of TFA (10%) in DCM for 2h to remove the acid labile-protecting groups. The solvent was then removed in vacuo and purified via RP-HPLC. Purification of these molecules proved to be quite challenging, perhaps requiring an enantiomeric column might help resolve the mixture of co-eluents. Excitingly, I was however able to isolate enough **(2030**) and **(2031)** for testing over the of several preliminary experiments.

3.10.2 pre-miRNA Substrates

The following sequences were ordered from Dharmacon:

Pre-let-7d RNA Sequence:

5'-Biotin-(18-atom spacer; hexaethyleneglycol)-AGAGGUAGUAGGUUGCAUAGUUUUAGGGCAGGGA-(5-aminoallyl uridine)-UUUGCCCACAAGGAGGUAACUAUACGACCUGCUGCCUUUCU-3'

pre-miR-21 RNA Sequence:

The following sequence was ordered from Dharmacon:

5’-Biotin-(18-atom spacer; hexaethylene glycol)-UAGCUUAUCAGACUGAUGUUGACUGUUGAA-(5-aminoallyl uridine)-CUCAUGGCAACACCAGUCGAUGGGCUGUC-3’

3.10.3 2-Chlortrityl Chloride Resin SPPS Protocol

Resin Blooming:

Peptides were synthesized sequentially using Solid phase protein synthesis protocols on 2-Chlortrityl Chloride Resin. Reaction vessels were sterile fritted syringes that were capped and shaken gently during reactions. This makes it easy to wash the resin and set up the next reaction by hand with efficiency. To bloom the resin and remove any byproducts on the resin, the resin was first washed with DCM, DMF, and MeOH (3x 10ml each) and was washed similarly after any deprotection or coupling reaction. The resin (1.0g, 0.82 mmol/g) bloomed in DCM (10ml) for 30m before the first AA loading.

Loading, Fmoc Deprotection, and Coupling:

Loading to the Trityl resin was lowered by utilizing excess resin and by only coupling 1.5 eq of the initial AA. Any unreacted amines were end-capped using a solution of DCM/MeOH/DIPEA (17:2:1) (2x 2min) so that they wouldn’t interfere with the remainder of the peptide synthesis. The first AA (1.5eq) is coupled for 2hr in the presence of 8 eq. DIPEA and 3eq. HBTU. General subsequent couplings were carried out twice for 20 min for each AA using a mixture of 3eq. HBTU, 3eq. HOBt, 8eq. DIPEA and 3eq. Fmoc-XAA-OH in DMF.

Fmoc Deprotection was carried out using 20% piperidine in DMF (2x 15min) and was washed with DMF and DCM.

Cleavage

Cleavage of the peptide from the resin was conducted by treating the resin with a solution of TFA/phenol/thioanisole/TIPS (83:6.25:4.2:4.2:2). The resulting peptide was then precipitated in ice-cold water. This method was too aggressive for the hermicidins; therefore I adopted the use of 20% HFIP in DCM for the complete cleavage of protected peptides.

Purification

Peptides were synthesized using RP-HPLC with 0.1% formic acid in ACN or MeOH. However, because a good method was never developed for the purification of the hermicidins, this warranted a look into other HPLC reagents, but I was unable to follow up on these HPLC methods. Any pure peptides were then dissolved in acetic acid and water (1:1), lyophilized, and used for our binding studies.

3.10.4 cat-ELCCA

Preparation of RNA-Click Conjugate

TCO, mtet, and NHS-ester of 4-pentyoic acid were purchased or prepared as follows: RNA substrate (5.0µL of the 1.0mM RNA stock in 100mM pH8 phosphate buffer with a final concentration of 5nmol) was incubated with the NHS-ester (5.0µL of 10mM stock in DMSO; 50nmol) for 1h at RT. The product was then purified via precipitation in sodium acetate (1.1µL of a 3M solution at a pH of 5.2) and cold ethanol (60 ml). Centrifugation at 14,000 RPM for 4°C gave us our RNA conjugate in a pellet that was then stored at -80°C as a 1.0mM (100mM phosphate buff., pH8) stock solution.

HRP-mTet and HRP preparation

2.5mg HRP was dissolved in PBS7 (185µL, 100mM PBS buffer, pH7, 150nM CaCl) and was coupled with 14.2µL 100mM Tet-NHS in DMSO. This was shaken for 3h, followed by centrifugation to remove excess reactants. The HRP-met conjugate was stored at 4°C.

Preparation of HRP-N3

The HRP conjugate was prepared by using previously published methods and developed for our purposes by Dr. Lorenz[139,239].

Cat-ELCCA HTS Assay Protocol

384-well plates coated with streptavidin (pierce 15407) are first washed with 50µL (3x) of sodium PB7 (100mM) using a Biotek 405 ELX plate washer. Biotinylated pre-miRNA (5µL, 500nM final) was then dispensed by a multidrop combi reagent dispenser. Plates were then centrifuged for a minute at 1k RPM. Plates were then tape sealed and incubated overnight at 4°C. The next day the plates are washed with (3x 50µL PB7) followed by the addition of 5µL of the dicer digest buffer (20mM Tris, 1mM fresh DTT, 12mM NaCl, 2.5mM $MgCl_2$ and 4.5% DMSO). The plates were again centrifuged, and stock compounds were then added (50nL of 5mM DMSO stock, 25µM final) to the wells with a Sciclone liquid handler with a V&P pin tool. DMSO was added to the control wells. This was followed by a 15m incubation at RT before the addition of Dicer, 5µL buffer containing 217µg/nL of the enzyme (108µg/ml dicer, 5% glycerol, and 0.01%Tritonn X-100 final). A dicer buffer lacking dicer was added to the positive controls. After centrifugation, the plates were sealed and incubated at 37°C for 5h. Plates were then washed (3x 50µL) with PB7. At this point, mTet-HRP in PB7 (1µL, 750nM final) was added

to the wells and the plates were centrifuged, sealed, and incubated for 2h at RT. The plates were then washed again (3x 50µL) with wash buffer (2mM imidazole, 260 mM NaCl, 0.5mM EDT, 0.1% Tween-20, pH7.0) and were then washed again with PB7 (3x 50µL). SuperSignal West Pico (25µL) was then added to the wells and incubated for 5 min. before using the PHERAstar plate reader using LUM plus module to detect the chemiluminescence signal[140].

Cat-ELCCA by Hand

Using the HTS protocol, cat-ELLCA was done using a hand pipette and the plates were not centrifuged. Biotek Cyation3 was used for the detection of chemiluminescence.

Dicer purification

Dicer prepared and purification prepared as reported by Dr. Lorenz[240].

Dicer Digestion

Digestion assays were done in 10uL. RNA (500nm final) and our substrate of interest in buffer (20mM Tris-HCl, 7.4, 12nM NaCl, 2.5 mM $MgCl_2$, 40 U/ml RNase out, 1.0mM fresh DTT) was treated for 3h with Dicer (1µL, 1.3mg/ml) at 37°C. The resulting digests were analyzed with 12.5% TBE- Urea gel and were visualized with SYBR Gold.

Data Analysis

PHERAstar plate reader was used to detect chemiluminescence. Data was collected in triplicates. All raw data was then processed and analyzed on GraphPad Prism.

Chapter 4 Conclusions and Future Directions

The work put forward in chapters 2 & 3 highlights the use of chemical biology to study nucleic acids. The isolation, identification, and synthesis of the hermicidins outlined in chapter 3 may prove to be the groundwork of miRNA-based inhibitors. The crystal violet probes developed in chapter 2 lend us insight into how we might make the major groove units of DNA druggable. The novel scaffolds provide us with new insights into manipulating biology at different critical stages. These studies, however, are indeed in their infancy and require further development to be considered as therapeutic strategies.

4.0 Further Development of Dicer, miRNA, and miRNA Probes

A. Dicer & miRNA Inhibitors

The work in chapter 3 outlines why the miRNA pathway is critical as well as how we were able to discover and develop novel drug targets. Identifying and characterizing novel molecules that can selectively act on miRNA or Dicer is paramount to understanding how these molecules interact with each other and with other binding proteins, such as Lin28[241]. Understanding these fundamental interactions could help us design selective molecules in the future and so are the tools to help us study these interactions. Concurrently, expanding our library can also help us further understand these interactions. In-vivo and in-vitro cell (RT-qPCR) experiments should also be considered to determine the applicability of these molecules.

B. Medicinal Chemistry, Biophysical and in-vitro Studies of Hermicidins

The surfactins/ hermicidins discussed warrant further structural confirmation via NMR (**2026 & 2031**), in-vivo, and in-vitro cell experiments, medicinal chemistry, and SAR studies to determine the scope at which these molecules can be applied. Furthermore, the derivatization of the macrocycle is dependent on the derivatization of the β-hydroxy acid. To that end, I was working on the synthesis of the β-hydroxy acid monomer but was unable to finish the work, therefore, this approach (**Chem. Commun., 2009, 5763–5765; J. Am. Chem. Soc., 2004, 126 (26), pp 8126–8127**) should be followed up on if the synthetic strategy described here for the macrocycle is to be employed. Beyond structure-based optimizations, further structural studies should be conducted, such as microRNA microarray analysis via global analysis assay (renilla- luciferase assay), GeneChip technology, SPR, ITC, and NMR spectroscopy to have a full understanding of these interactions and their limitations.

C. Chemoenzymatic Synthesis of Hermicidins

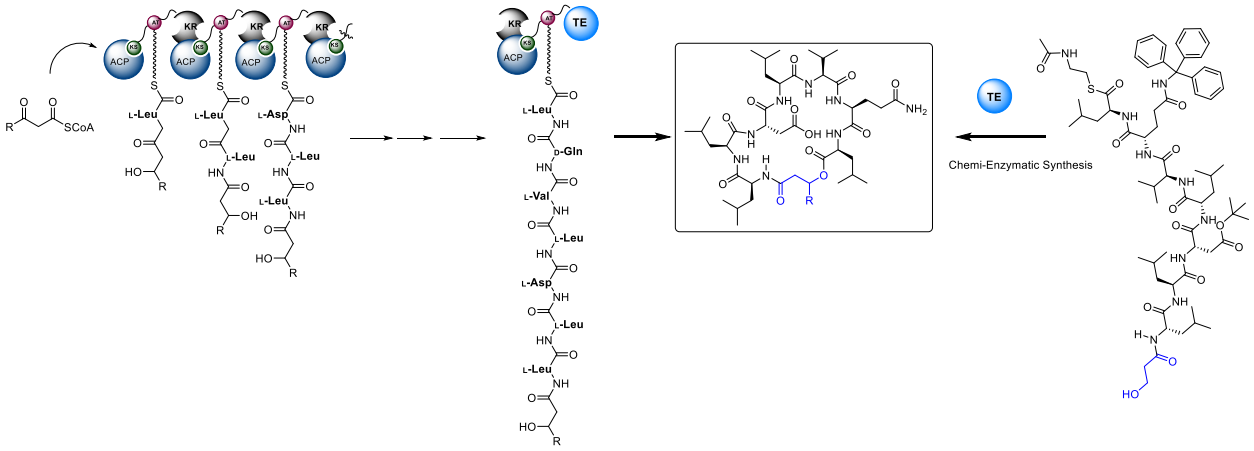


Figure 4.1 Chemo-Enzymatic Synthesis of Hermicidins. A) *Step- wise synthesis of macrolides via PKS- NRPS cluster.* ***B)*** *Our approach: SNAc-peptide conjugate cyclized via isolated TE*

An approach that I had begun but also was unable to follow up on was the chemo-enzymatic synthesis of our hermicidin analogs. Despite my success in synthesizing some hermicidin scaffolds, cyclization, purification, and derivatization of the β-hydroxy linker continue to be a challenge[242]. Therefore, utilizing a biosynthetic approach could result in the rapid derivatization of the hermicidins. The polyketide synthase-non ribosomal peptide synthetase (PKS- NRPS) is a cluster of enzymatic proteins that coordinate the synthesis of macrolides (Figure 4.1A)[243]. This set of enzymatic proteins was found to produce pikromycin in streptomyces and studies have suggested that they have a high tolerance for acting on and cyclizing un- natural substrates[244–246]. The peptide is built in a step-wise fashion, once complete, the peptide is released and cyclized by the thioesterase (TE) domain. These enzymes have been widely studied and approaches have been developed to isolate these enzymes, including the TE domain, for chemoenzymatic synthesis[247–249]. We planned to follow this approach in collaboration with Dr. Tripathi and the Sherman Group. To that end, efforts had gone into synthesizing the hermicidin with the H-substituted β- hydroxy linker with an N-acetylsteamine (SNAc) tag for TE recognition[245,246]. A follow-up/ optimization of this substrates' compatibility with this enzyme is warranted once Dr. Tripathi has isolated viable TE.

D. cat-ELCCA on Elli Lilly Compounds

Another set of compounds that were waiting to be screened via cat-ELCCA was a library provided by Elli Lilly.

4.1 Further Development of DNA- Major Groove Probes

A. Further biophysical & in-vitro Studies

The studies discussed in chapter 2 provide us with strong evidence that targeting the major groove units of DNA is possible. The limit of these scaffolds described here remains to be determined as we have other experiments that will be underway soon. Further binding assays against controls such as netropsin & methyl green are required to further support their binding preference. Further experiments can include ITC and SPR which would reveal important thermodynamic and binding information.

B. Medicinal Chemistry Based on Current Data

Based on the data discussed in chapter 2, medicinal chemistry on the TrBrXX scaffolds is necessary to investigate the limitations of these linkers (Alkyne vs. Alkene) and the orientation (meta/Para) of the crystal violet moieties. Based on these, and previous findings, an alkyne or alkene linker is required, however, more studies are required to establish optimal linker and linker length. Furthermore, the 2,6 position is the best for major groove binding of the NMe groups and alkyl linkers have been strongly associated with minor grove binding. These findings should be considered when designing 3rd generation probes.

C. Truncated, Chimeric Oligonucleotides

Similar to the experiments when we attempted to investigate how the surfactins interacted with the miRNA substrate, developing chimeras or truncating the oligos that were studied may reveal shape-selective binding behavior and importantly, critical sequence tracks related binding behavior. One reason for the association of our molecules in the major

groove may be driven by the hybridized electron orbitals of the alkene/ alkyne linker via electrostatic interactions with the outward-facing NH2 and keto-oxygen of the GC nts. AT nucleotides, however, have no outward NH2, only the keto-oxygen, perhaps creating a slightly more electronegative pocket that is not as stable an interaction with the hybridized linker. Therefore, a study into how the base pairs might be interacting with our ligands individually and as a whole within a sequence could reveal how to achieve some level of sequence selectivity.

4.2 Concluding Remarks

Indeed, we are on the cusp of a therapeutic renaissance as we continue to gain a more detailed understanding of DNA & RNA biology, and how to modulate their activity. Critical to these advancements is technology such as cat-ELCCA, as well as the discovery of novel molecules and druggable targets. To expand our ability to target previously undruggable units of DNA & RNA, in chapters 2 & 3 I provided evidence for two sets of probes that displayed some degree of selectivity, though minimal, and in-vitro activity in the case of the hermicidins, which could lay the groundwork for the development of DNA major groove or pre-miRNA modulators. The discovery of molecules that can discriminately bind to a specific major groove or pre-miRNA could be invaluable and open the door to an entirely new therapeutic space.

Appendix

Data in appendix should be considered preliminary.

I. Preliminary NMR Data of Crystal Violet Derivatives

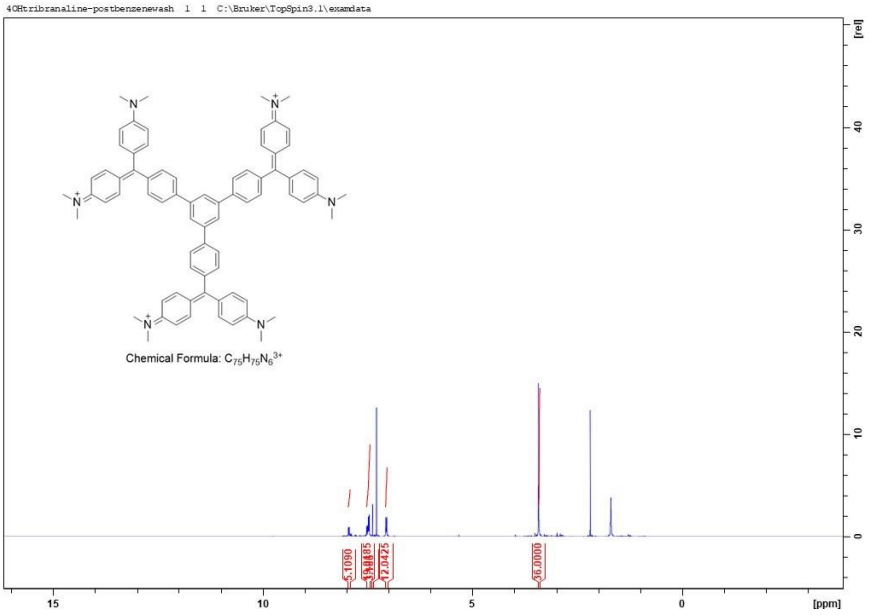
II. Preliminary NMR Data: Walkthrough of Activity Guided Fractionation of NPEs

III. Dicer- Digestion Gel Assay of Initial Activity Confirmation

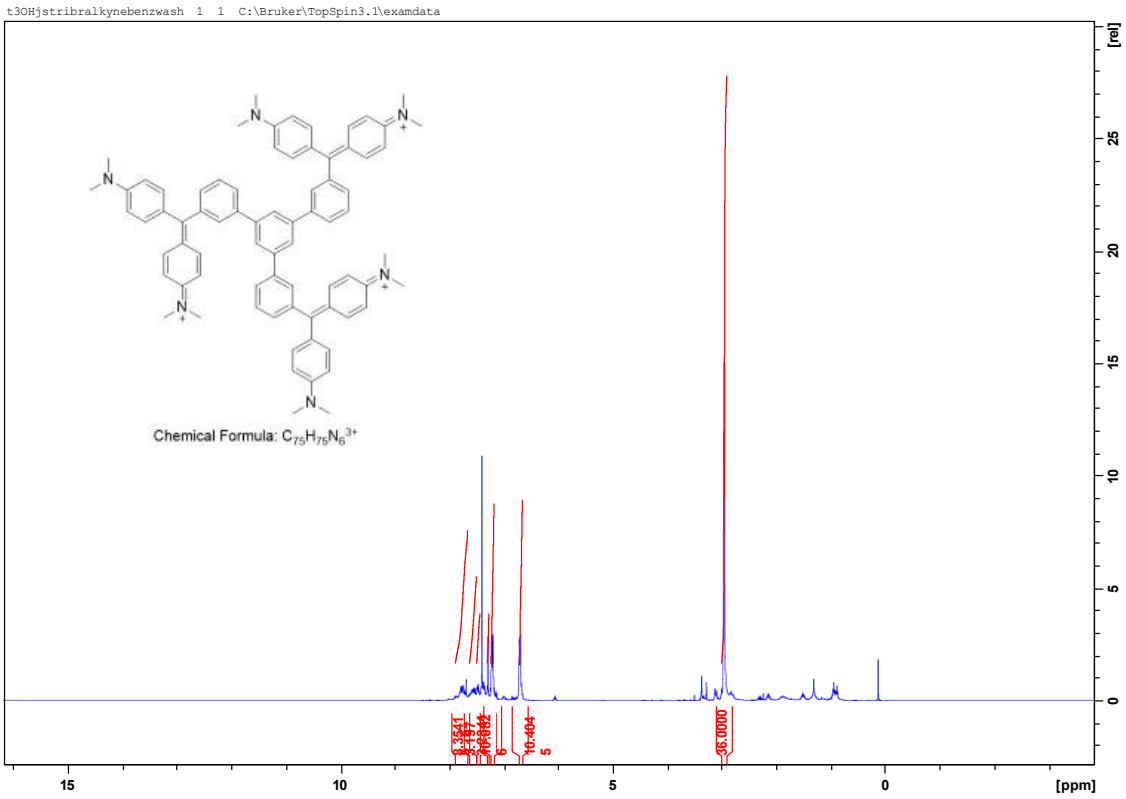
IV. Evaluating Importance of Depsie-Linkage

I. NMR data

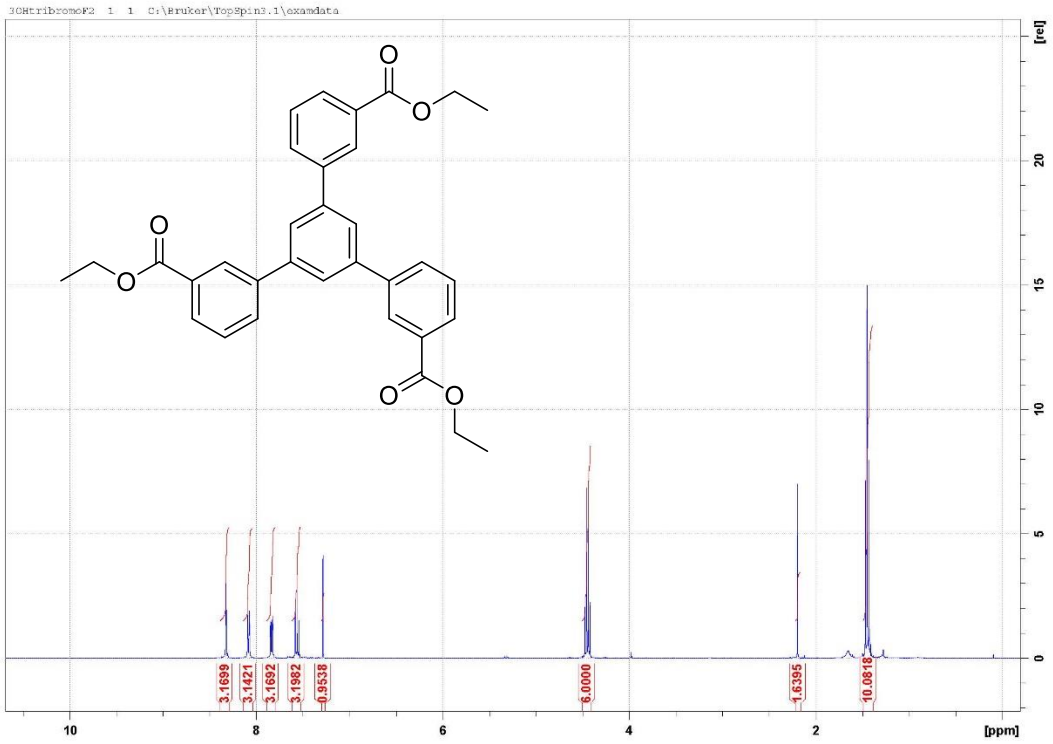
TM2 ^{1}H

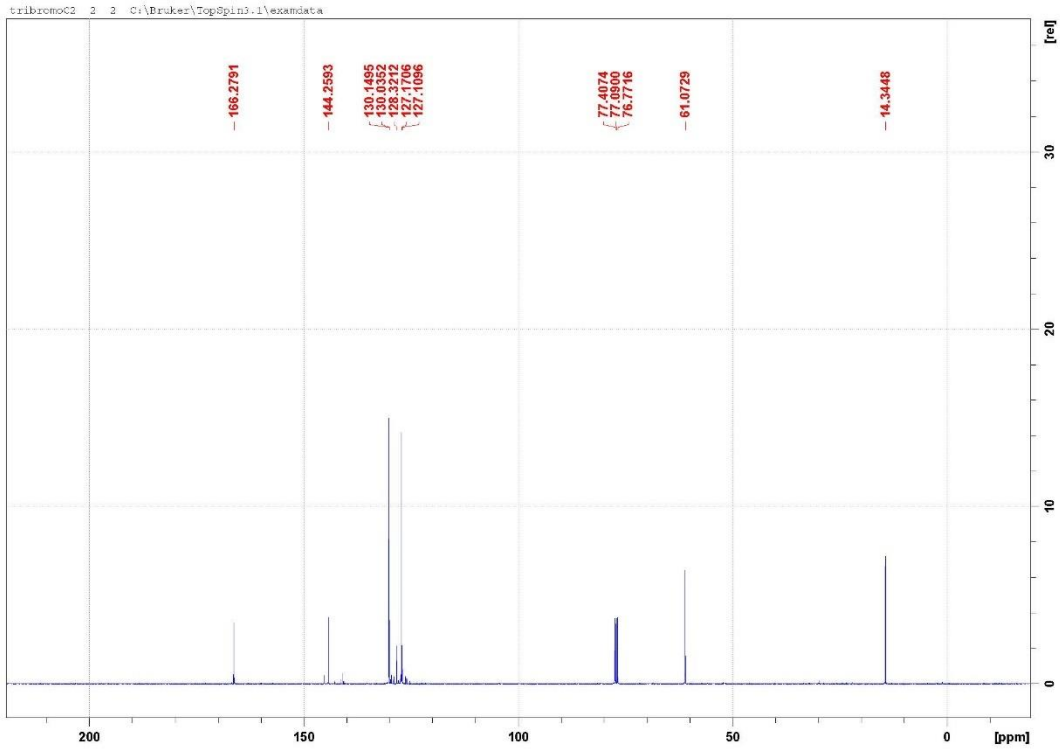


TM3 ^{1}H

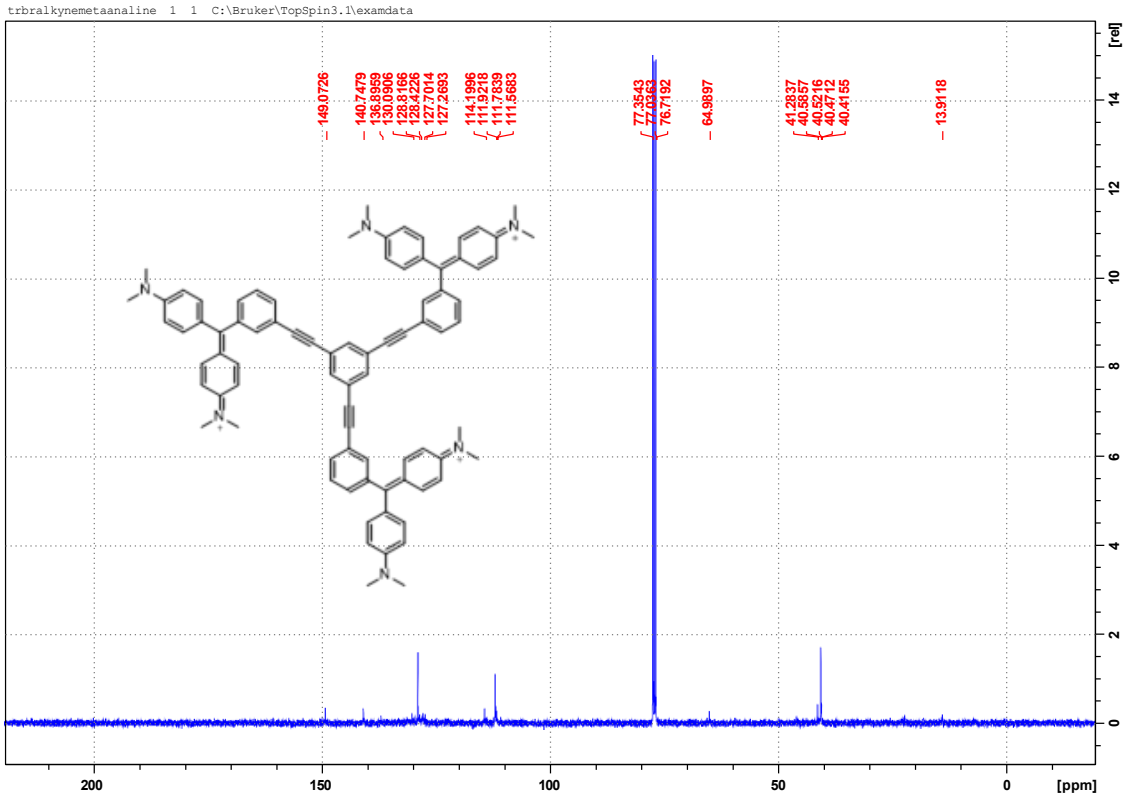


1H & 13C Meta- tri ester substituted intermediate:

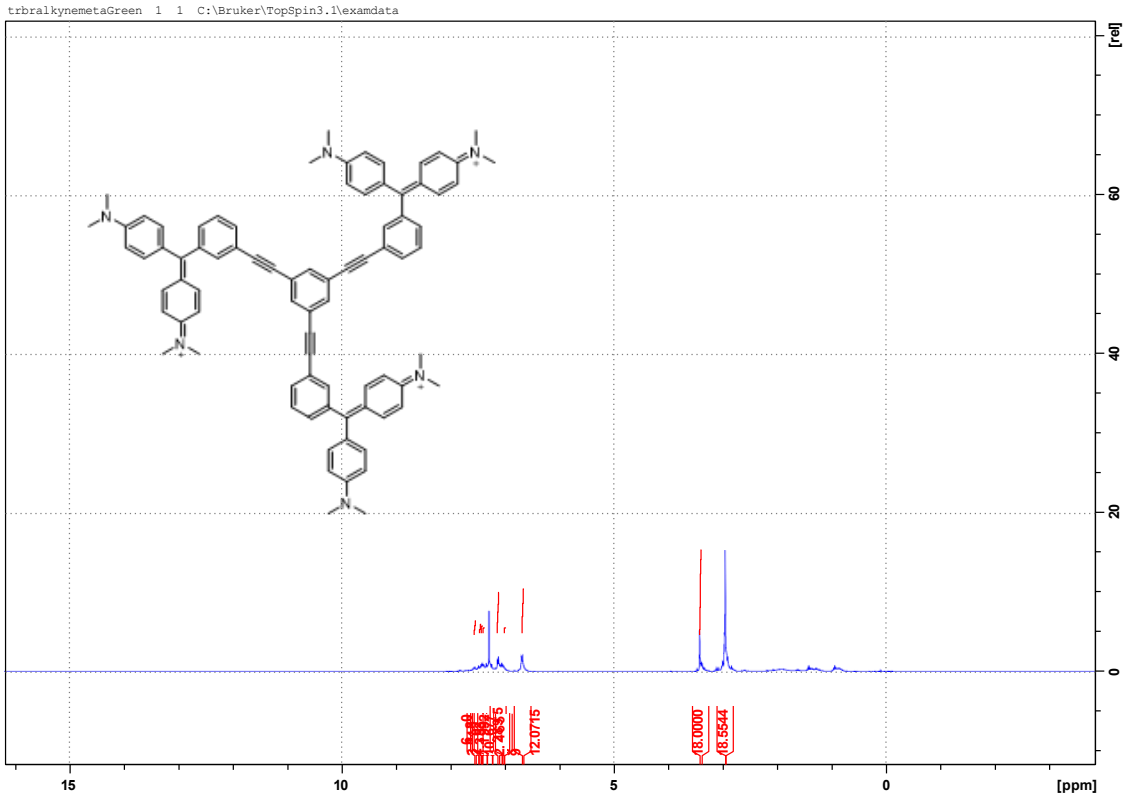

3.1699
3.1421
3.1692
3.1982
0.9538
6.0000
1.6395
10.0818
[ppm]
[rel]


166.2791
144.2593
130.1495
130.0352
128.3212
127.1706
127.1096
77.4074
77.0900
76.7716
61.0729
14.3448
[ppm]
[rel]

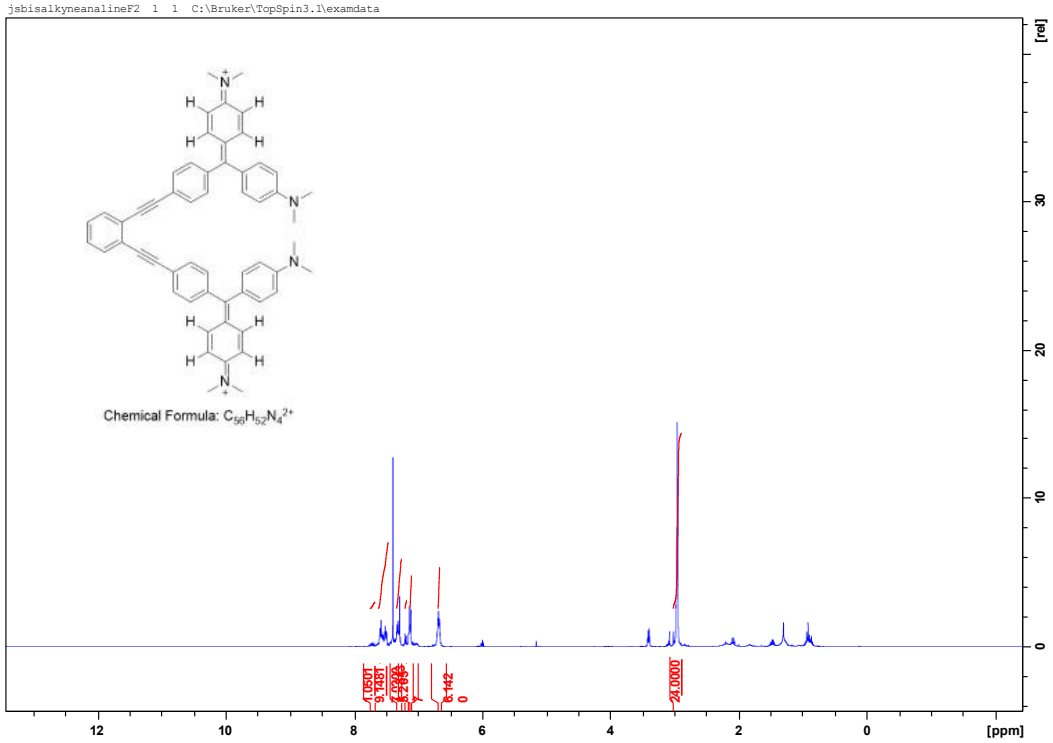
TM5 ^{13}C

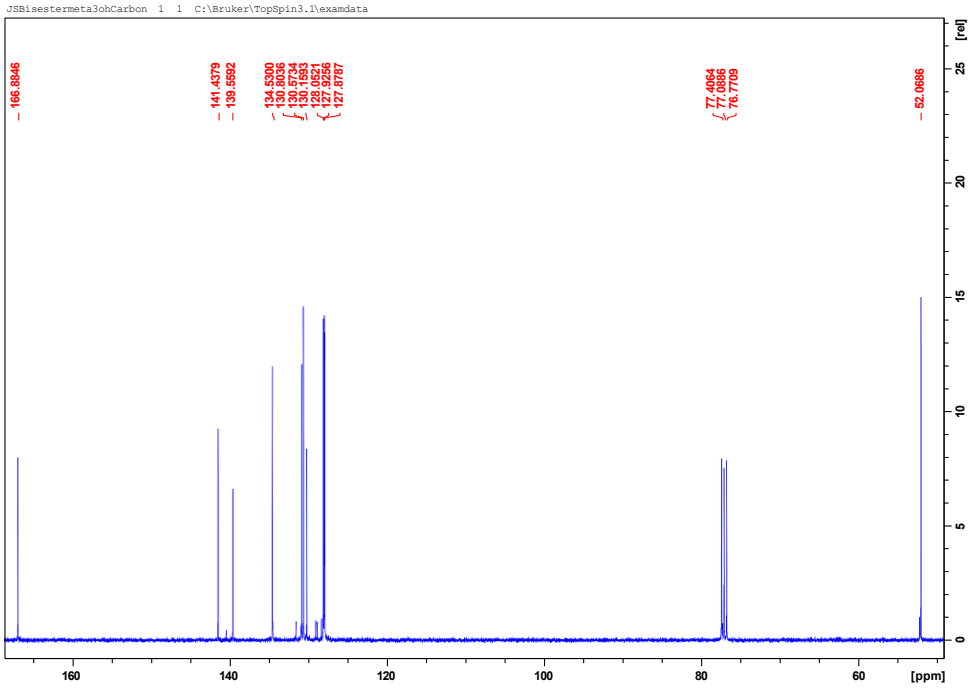

trbralkynemetaanaline 1 1 C:\Bruker\TopSpin3.1\examdata
149.0726
140.7479
136.8959
130.0966
128.8456
128.4226
127.7014
127.2693
114.1996
111.9218
111.7839
111.5683
77.3543
77.0363
76.7192
64.9897
41.2837
40.5857
40.5216
40.4712
40.4155
13.9118
[rel]
[ppm]
200
150
100
50
0

TM5 ^{1}H

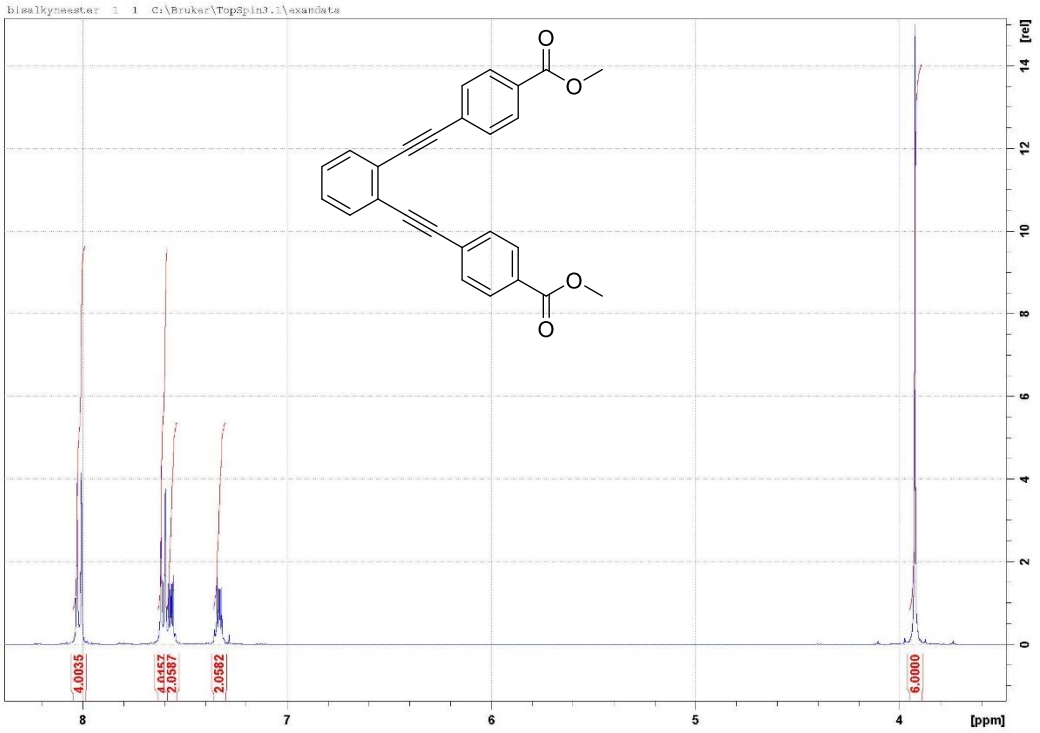

trbralkynemetaGreen 1 1 C:\Bruker\TopSpin3.1\examdata
[rel]
60
40
20
0
12.0715
8.0000
8.5544
15
10
5
0
[ppm]

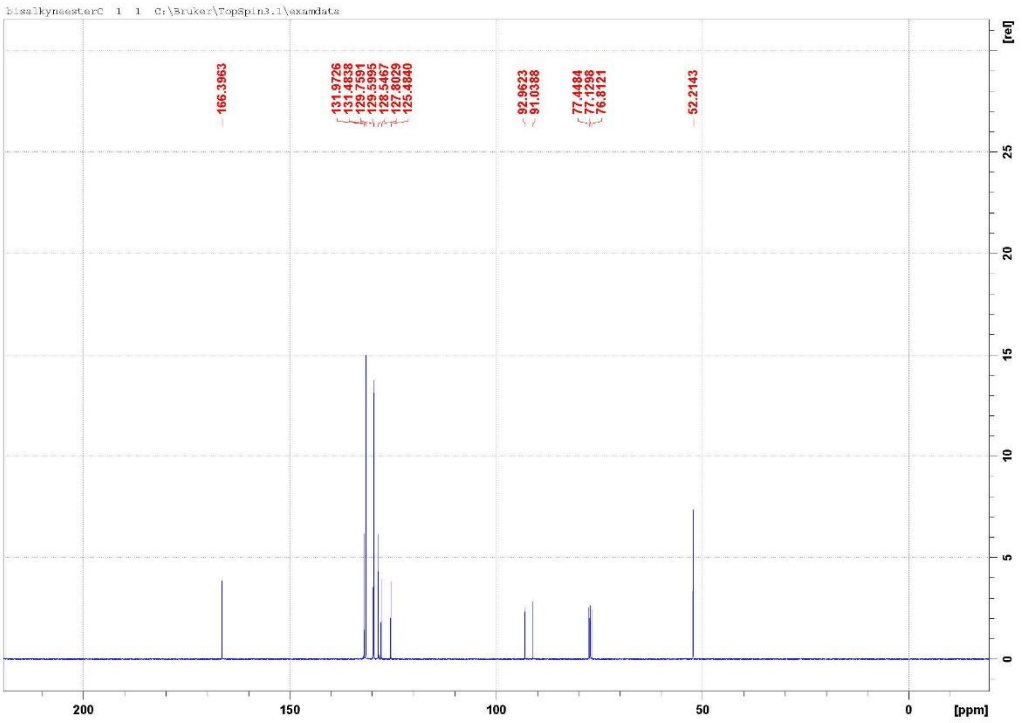
TM1:^{1}H &, ^{13}C respectively:



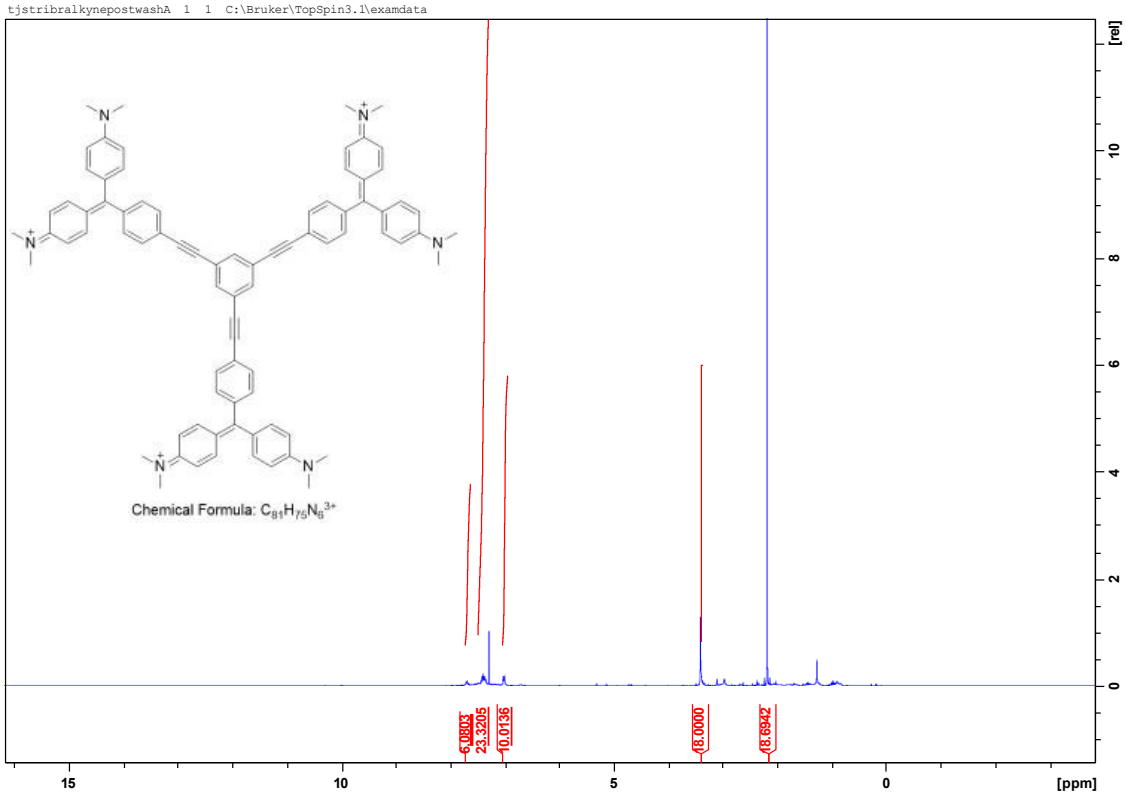


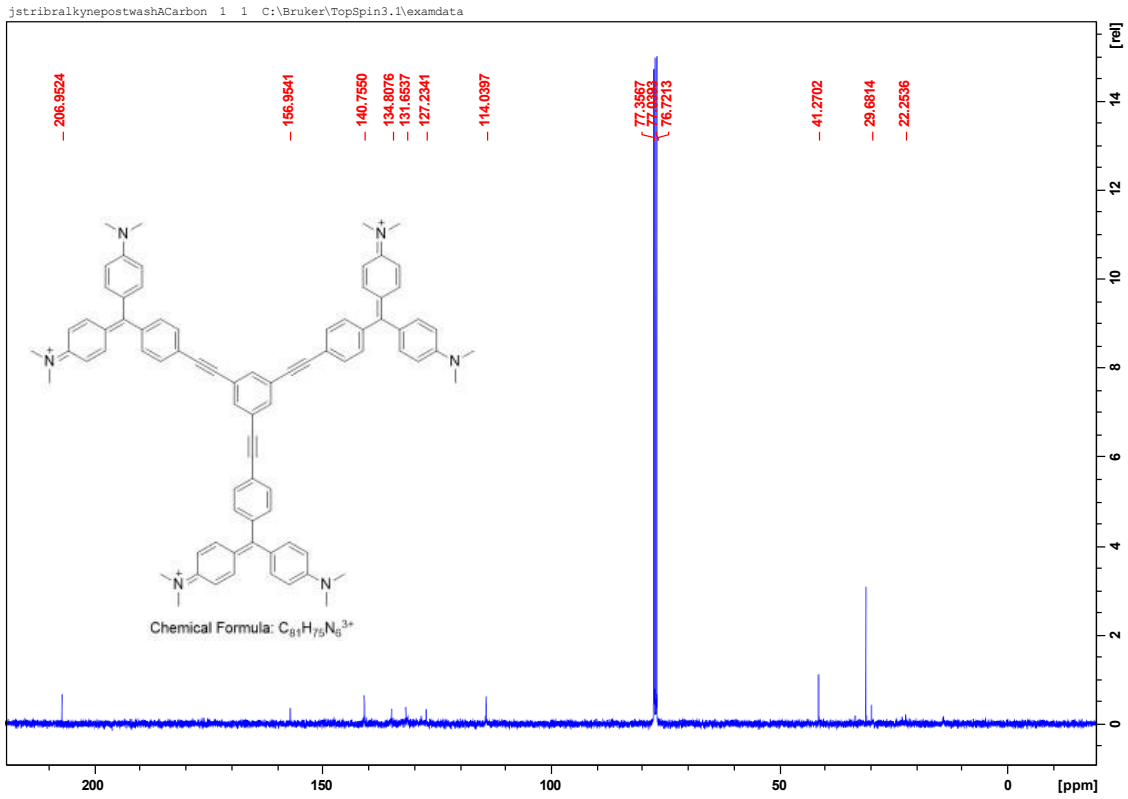
1H & 13C Para-di ester substituted intermediate:





TM4 ^{1}H & ^{13}C, respectively:





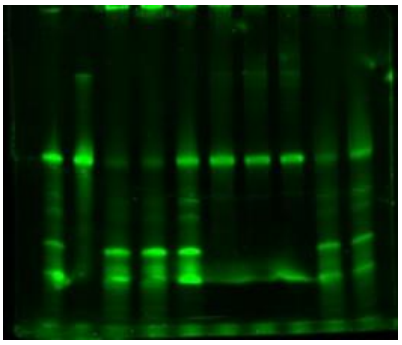
II. Walkthrough of Activity Guided Fractionation of NPEs:

Key sample Gels & cat-ELCCA that led to the isolation of active NPs:

Initial Screening of Extracts

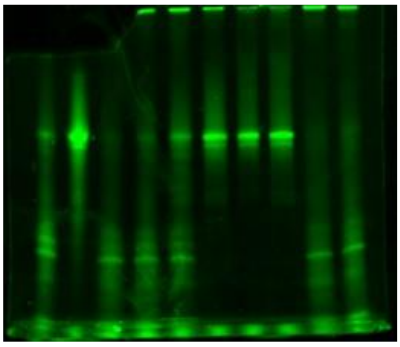
- First set of natural product extracts screened on denaturing gel

pre-miR21



miR21: +, -, 9w/1A, 3w/1A, 3w/2A, 9w/11A, 3w/7A,3w/17A, 1A, wash

pre-let7d



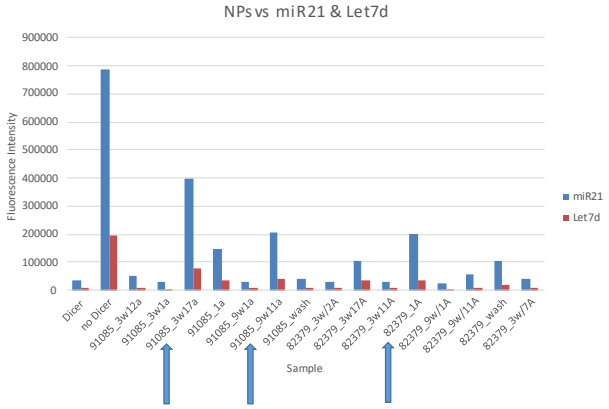
Let7: +, -, 9w/1A, 3w/1A, 3w/2A, 9w/11A, 3w/7A,3w/17A, 1A, wash

- Goal was to establish a working concentration for cat-ELCCA, 7.0 mg/ml

I followed up on several NPEs from a broader screen with a preliminary dicer digestion gel assay (above) (controls:+ dicer, - no dicer). Then I confirmed their activity via cat-ELCCA. We decided to follow up with fraction **82379_1A** since it repeated (below).

NP extracts on cat -ELCCA

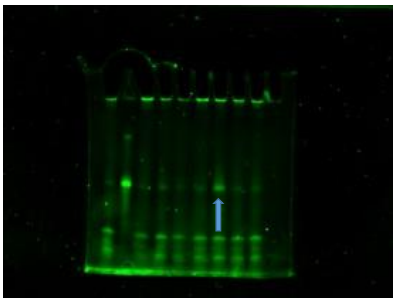
- cat-ELCCA of new NPs, 91085 & 82379, @ 0.7 mg/ml



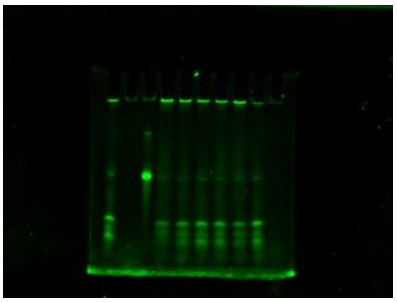
- From this screen, we were interested in following up on three extracts.

Follow Up On 82379 samples

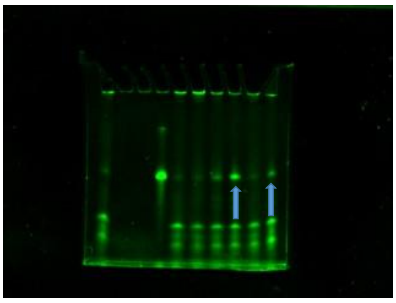
- Followed up on refined samples of 82379 samples



+,-, F1-F7



+,-, F8- F13



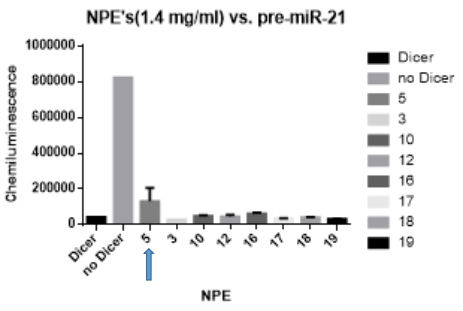
+,-, F14-F19

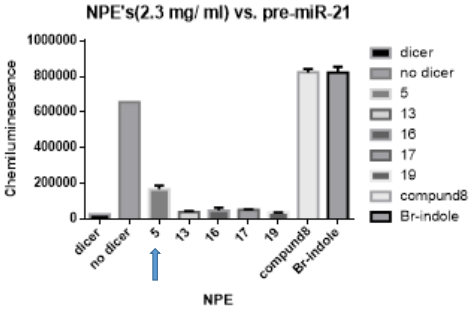
- Again, attempting to establish a working concentration. (1.4 mg/ml)

(above) Fractionated 82379 samples were tested via dicer digestion gel assay and (below) cat-ELCCA. **Fraction 5** repeated, further fractionated this sample.

Cat-ELCCA of 382379 Fractions

- Followed up new extracts in cat-ELCCA



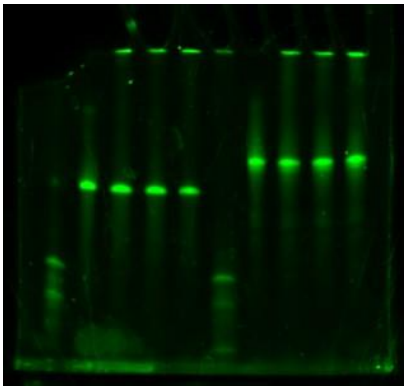


- Fractions 17 and 19 did not repeat, but fraction 5 was consistently active

(below) Follow up on select isolated NPs from **fraction 5,** derived from **82379_1A,** against premiR21 (left) and pre-let7d (right) in dicer digestion assay and confirmed via cat-ELCCA vs. premiR21.

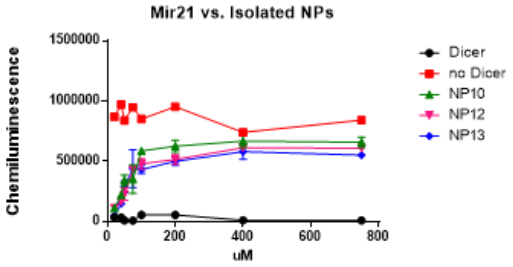
Isolated NPs

NPs: @ 750uM



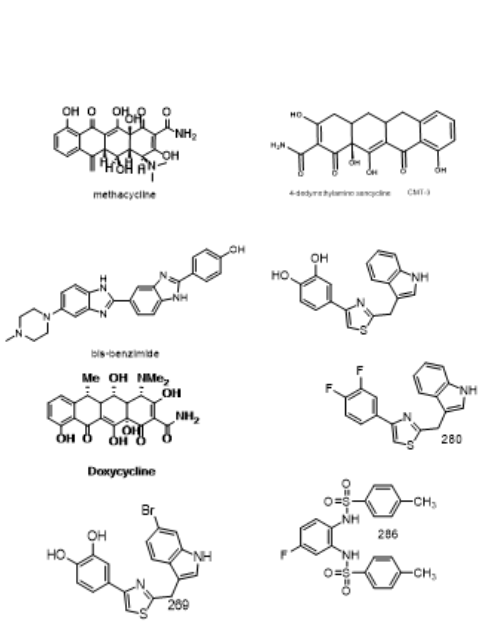
miR21: +, -,mir-10,mir-12, mir-13 let7d: +, -, mir-10,mir-12, mir-13

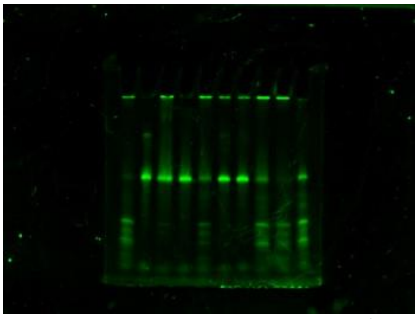
- Received and tested Isolated NP's
- Wanted to establish a titration curve for new sample



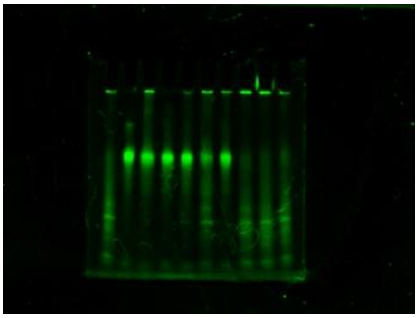
16 point, lower 8 points were all level though

III. Dicer- digestion gel assay of initial activity confirmation





miR21: +, -, methacycline, bis- benz, doxycycline, br-8, CMT, 8, di-fluoro, sulfonamide(286)



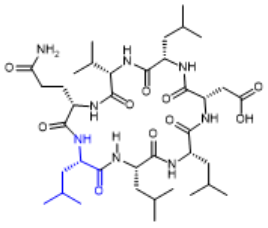
Let7: +, -, methacycline, bis- benz, doxycycline, br-8, CMT, 8, di-fluoro, sulfonamide

(above) preliminary dicer digestion gel assay of small molecules (catechols, sulfonamides, tetracyclines, Hoechst moieties) from small molecule screen, before association with chelating behavior.

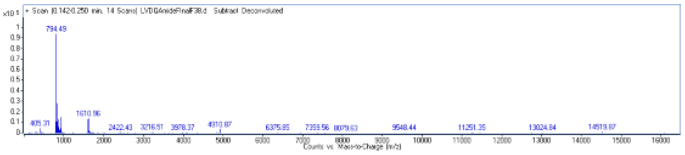
IV. Evaluating importance of depsie-linkage

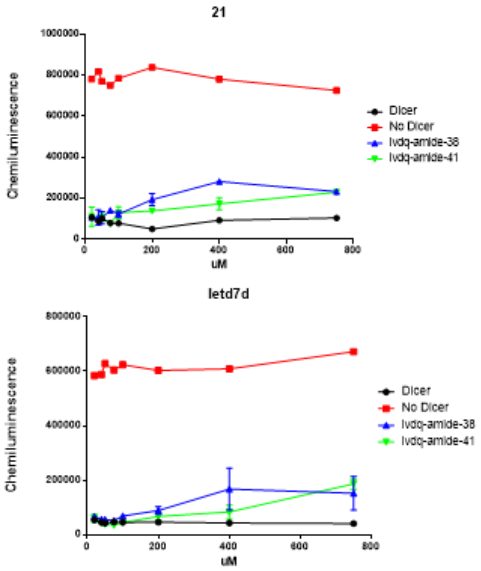
Evaluating Importance of Depsi- Portion

- I wanted to determine if Depsi bond was necessary for activity
- Turns out it may be necessary



Molecular Weight: 794.9920





(above) macrocycle without the depsi-linkage MS data and cat-ELCCA vs. premiR21 and pre-let7d.

References

1. Kulkarni, J. A. *et al.* The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol. 2021 166* **16**, 630–643 (2021).

2. Anderson, E. J. *et al.* Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *N. Engl. J. Med.* **383**, 2427–2438 (2020).

3. Rossi, J. J. & Rossi, D. Oligonucleotides and the COVID-19 Pandemic: A Perspective. *Nucleic Acid Ther.* **30**, 129 (2020).

4. Jackson, M., Marks, L., May, G. H. W. & Wilson, J. B. The genetic basis of disease. *Essays Biochem.* **62**, 643 (2018).

5. Hofker, M. H., Fu, J. & Wijmenga, C. The genome revolution and its role in understanding complex diseases. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842**, 1889–1895 (2014).

6. Fonseca Guerra, C., Bickelhaupt, F. M., Snijders, J. G. & Baerends, E. J. Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment. *J. Am. Chem. Soc.* **122**, 4117–4128 (2000).

7. Travers, A. & Muskhelishvili, G. DNA structure and function. *FEBS J.* **282**, 2279–2295 (2015).

8. Marr, C., Geertz, M., Hütt, M. T. & Muskhelishvili, G. Dissecting the logical types of network control in gene expression profiles. *BMC Syst. Biol.* **2**, 1–9 (2008).

9. Travers, A. A., Muskhelishvili, G. & Thompson, J. M. T. DNA information: from digital code to analogue structure. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **370**, 2960–2986 (2012).

10. Travers, A. A. & Klug, A. The bending of DNA in nucleosomes and its wider implications. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **317**, 537–561 (1987).

11. Gartenberg, M. R. & Crothers, D. M. DNA sequence determinants of CAP-induced bending and protein binding affinity. *Nat. 1988 3336176* **333**, 824–829 (1988).

12. Hud, N. V. & Plavec, J. A unified model for the origin of DNA sequence-directed curvature. *Biopolymers* **69**, 144–158 (2003).

13. McFail-Isom, L., Sines, C. C. & Williams, L. D. DNA structure: cations in charge? *Curr. Opin. Struct. Biol.* **9**, 298–304 (1999).

14. Hud, N. V., Sklenář, V. & Feigon, J. Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J. Mol. Biol.* **286**, 651–660 (1999).

15. Shui, X., Sines, C. C., McFail-Isom, L., VanDerveer, D. & Williams, L. D. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* **37**, 16877–16887 (1998).

16. Hud, N. V. & Feigon, J. Localization of divalent metal ions in the minor groove of DNA A-tracts. *J. Am. Chem. Soc.* **119**, 5756–5757 (1997).

17. Rouzina, L. & Bloomfield, V. A. DNA bending by small, mobile multivalent cations. *Biophys. J.* **74**, 3152–3164 (1998).

18. Cheatham, T. E. & Kollman, P. A. Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of d[ACCCGCGGGT]2 in the presence of hexaamminecobalt(III). *Structure* **5**, 1297–1311 (1997).

19. Young, M. A., Jayaram, B. & Beveridge, D. L. Intrusion of Counterions into the Spine of Hydration in the Minor Groove of B-DNA: Fractional Occupancy of Electronegative Pockets. *J. Am. Chem. Soc.* **119**, 59–69 (1997).

20. Tereshko, V., Minasov, G. & Egli, M. The Dickerson-Drew B-DNA dodecamer revisited at atomic resolution. *J. Am. Chem. Soc.* **121**, 470–471 (1999).

21. Sines, C. C., McFail-Isom, L., Howerton, S. B., VanDerveer, D. & Williams, L. D. Cations mediate B-DNA conformational heterogeneity. *J. Am. Chem. Soc.* **122**, 11048–11056 (2000).

22. Hud, N. V. & Polak, M. DNA-cation interactions: The major and minor grooves are flexible ionophores. *Curr. Opin. Struct. Biol.* **11**, 293–301 (2001).

23. Hud, N. V. & Feigon, J. Characterization of divalent cation localization in the minor groove of the AnTn and TnAn DNA sequence elements by 1H NMR spectroscopy and manganese(II). *Biochemistry* **41**, 9900–9910 (2002).

24. Peticolas, W. L., Wang, Y. & Thomas, G. A. Some rules for predicting the base-sequence dependence of DNA conformation. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2579–2583 (1988).

25. Minchin, S. & Lodge, J. Understanding biochemistry: structure and function of nucleic acids. *Essays Biochem.* **63**, 433 (2019).

26. Kielkopf, C. L., Ding, S., Kuhn, P. & Rees, D. C. Conformational flexibility of B-DNA at 0.74 Å resolution: d(CCAGTACTGG)2. *J. Mol. Biol.* **296**, 787–801 (2000).

27. Brukner, I. *et al.* Evidence for opposite groove-directed curvature of GGGCCC and AAAAA sequence elements. *Nucleic Acids Res.* **21**, 1025–1029 (1993).

28. Brukner, I., Susic, S., Dlakic, M., Savic, A. & Pongor, S. Physiological concentration of magnesium ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *J. Mol. Biol.* **236**, 26–32 (1994).

29. Lu, X. J., Shakked, Z. & Olson, W. K. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **300**, 819–840 (2000).

30. Mohanty, C. S. & D. Sequence- and structure-based analysis of proteins involved in miRNA biogenesis. *J. Biomol. Struct. Dyn. ISSN* 1–14 (2017) doi:10.1080/07391102.2016.1269687.

31. Hud, N. V. & Plavec, J. A unified model for the origin of DNA sequence-directed curvature. *Biopolymers* **69**, 144–158 (2003).

32. Homsy, J. *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262–1266 (2015).

33. Lane, P. A. Sickle cell disease. *Pediatr. Clin. North Am.* **43**, 639–664 (1996).

34. Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* **21**, 210–219 (2014).

35. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet. 2015 471* **47**, 8–12 (2014).

36. Abraham, B. J. *et al.* Small genomic insertions form enhancers that misregulate oncogenes. *Nat. Commun. 2017 81* **8**, 1–13 (2017).

37. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

38. Wilhelm, S. M. *et al.* BAY 43-9006 Exhibits Broad Spectrum Oral Antitumor Activity and Targets the RAF/MEK/ERK Pathway and Receptor Tyrosine Kinases Involved in Tumor Progression and Angiogenesis. *Cancer Res.* **64**, 7099–7109 (2004).

39. Musella, A. *et al.* PARP inhibition: A promising therapeutic target in ovarian cancer. *cellmolbiol.org* **61**, 44–61 (2015).

40. Moore, K. *et al.* Maintenance Olaparib in Patients with Newly Diagnosed Advanced Ovarian Cancer. *N. Engl. J. Med.* **379**, 2495–2505 (2018).

41. Godzieba, M. & Ciesielski, S. Natural DNA Intercalators as Promising Therapeutics for Cancer and Infectious Diseases. *Curr. Cancer Drug Targets* **20**, 19–32 (2020).

42. Wartell, R. M., Larson, J. E. & Wells, R. D. Netropsin. *J. Biol. Chem.* **249**, 6719–6731 (1974).

43. Wade, W. S., Mrksich, M. & Dervan, P. B. Design of Peptides That Bind in the Minor Groove of DNA at 5′-(A,T)G(A,T)C(A,T)-3′ Sequences by a Dimeric Side-by-Side Motif. *J. Am. Chem. Soc.* **114**, 8783–8794 (1992).

44. Characterization of a Duocarmycin-DNA Adduct-recognizing Protein in Cancer Cells | Cancer Research | American Association for Cancer Research. https://aacrjournals.org/cancerres/article/59/21/5417/505608/Characterization-of-a-Duocarmycin-DNA-Adduct.

45. Hawkins, C. A. *et al.* Controlling Binding Orientation in Hairpin Polyamide DNA

Complexes. *J. Am. Chem. Soc.* **122**, 5235–5243 (2000).

46. Baird, E. E. & Dervan, P. B. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.* **118**, 6141–6146 (1996).

47. Mrksich, M., Parks, M. E. & Dervan, P. B. Hairpin Peptide Motif. A New Class of Oligopeptides for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA. *J. Am. Chem. Soc.* **116**, 7983–7988 (1994).

48. Mrksich, M. & Dervan, P. B. Design of a Covalent Peptide Heterodimer for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA. *J. Am. Chem. Soc.* **116**, 3663–3664 (1994).

49. Wade, W. S., Mrksich, M. & Dervan, P. B. Binding affinities of synthetic peptides, pyridine-2-carboxamidonetropsin and 1-methylimidazole-2-carboxamidonetropsin, that form 2:1 complexes in the minor groove of double-helical DNA. *Biochemistry* **32**, 11385–11389 (1993).

50. Mrksich, M. *et al.* Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Biochemistry* **89**, 7586–7590 (1992).

51. Dose, C., Farkas, M. E., Chenoweth, D. M. & Dervan, P. B. Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit. *J. Am. Chem. Soc.* **130**, 6859–6866 (2008).

52. Poulin-Kerstein, A. T. & Dervan, P. B. DNA-Templated Dimerization of Hairpin Polyamides. *J. Am. Chem. Soc.* **125**, 15811–15821 (2003).

53. Pilch, D. S. *et al.* Binding of a hairpin polyamide in the minor groove of DNA: Sequence-specific enthalpic discrimination (polyamide-DNA binding affinity/isothermal calorimetry/2:1 pyrrole-imidazole DNA binding motif/polyamide-base hydrogen bonds/imidazole-guanine interaction). *Biochemistry* **93**, 8306–8311 (1996).

54. Renneberg, D. & Dervan, P. B. Imidazopyridine/Pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.* **125**, 5707–5716 (2003).

55. Ellervik, U., Wang, C. C. C. & Dervan, P. B. Hydroxybenzamide/pyrrole pair distinguishes T·A from A·T base pairs in the minor groove of DNA. *J. Am. Chem. Soc.* **122**, 9354–9360 (2000).

56. Reinhold, W. C., Thomas, A. & Pommier, Y. DNA-Targeted Precision Medicine; Have We Been Caught Sleeping? *Trends in Cancer* **3**, 2–6 (2017).

57. Khan, G. S., Shah, A., Zia-Ur-Rehman & Barker, D. Chemistry of DNA minor groove binding agents. *J. Photochem. Photobiol. B Biol.* **115**, 105–118 (2012).

58. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).

59. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, 855–862 (1993).

60. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68–D73 (2014).

61. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).

62. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).

63. Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231–235 (2004).

64. Gregory, R. I. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235–240 (2004).

65. Han, J. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).

66. Lee, Y. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).

67. Svobodova, E., Kubikova, J. & Svoboda, P. Production of small RNAs by mammalian Dicer. *Pflugers Arch. Eur. J. Physiol.* **468**, 1089–1102 (2016).

68. Zeng, Y., Yi, R. & Cullen, B. R. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J.* **24**, 138–148 (2005).

69. Bohnsack, M. T. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).

70. Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science (80-.).* **303**, 95–98 (2004).

71. Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).

72. Grishok, A. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**, 23–34 (2001).

73. Hutvagner, G. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science (80-.).* **293**, 834–838 (2001).

74. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science (80-.).* **293**, 1146–1150 (2001).

75. Mourelatos, Z. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720–728 (2002).

76. Tabara, H. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**, 123–132 (1999).

77. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science (80-.).* **305**, 1434–1437 (2004).

78. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**, 663–666 (2005).

79. Catalanotto, C., Cogoni, C. & Zardo, G. MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. *Int. J. Mol. Sci.* doi:10.3390/ijms17101712.

80. Lau, P.-W. *et al.* The molecular architecture of human Dicer. *Nat. Struct. Mol. Biol.* **19**, 436–440 (2012).

81. Ke, X. *et al.* Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate Article Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate. *Cell* 1–13 (2018) doi:10.1016/j.cell.2018.03.080.

82. Macrae, I. J. Structural basis for double-stranded RNA processing by Dicer. *Science (80-.).* **311**, 195–198 (2006).

83. Koh, H. R., Ghanbariniaki, A. & Myong, S. RNA stem structure governs coupling of dicing and gene silencing in RNA interference. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10349–E10358 (2017).

84. Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H. & Tomari, Y. Recognition of the pre-miRNA structure by Drosophila Dicer-1. *Nat. Struct. Mol. Biol.* **18**, 1153–1158 (2011).

85. Chakravarthy, S., Sternberg, S. H., Kellenberger, C. A. & Doudna, J. A. Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J. Mol. Biol.* **404**, 392–402 (2010).

86. Noland, C. L., Ma, E. & Doudna, J. A. siRNA repositioning for guide strand selection by human Dicer complexes. *Mol. Cell* **43**, 110–121 (2011).

87. Im, H. I. & Kenny, P. J. MicroRNAs in neuronal function and dysfunction. *Trends Neurosci.* **35**, 325–334 (2012).

88. Li, Z. & Rana, T. M. Therapeutic targeting of microRNAs: current status and future challenges. *Nat. Rev. Drug Discov.* **13**, 622–638 (2014).

89. Lorenz, D. A. & Garner, A. L. Approaches for the Discovery of Small Molecule Ligands Targeting microRNAs. in *Top Med Chem* vol. 27 79–110 (2017).

90. Ha, M. & Narry Kim, V. Regulation of microRNA biogenesis. *Nat. Publ. Gr.* **15**, (2014).

91. Park, C. Y., Choi, Y. S. & McManus, M. T. Analysis of microRNA knockouts in mice. *Hum. Mol. Genet.* **19**, (2010).

92. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).

93. Southwell, A. L., Skotte, N. H., Bennett, C. F. & Hayden, M. R. Antisense oligonucleotide therapeutics for inherited neurodegenerative diseases. *Trends Mol. Med.* **18**, 634–643 (2012).

94. Liang, X. H. *et al.* Antisense oligonucleotides targeting translation inhibitory elements in 5' UTRs can selectively increase protein levels. *Nucleic Acids Res.* **45**, 9528–9546 (2017).

95. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293–296 (2000).

96. Jackson, A. L. *et al.* Widespread siRNA 'off-target' transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179–1187 (2006).

97. Liang, X. H., Vickers, T. A., Guo, S. & Crooke, S. T. Efficient and specific knockdown of small non-coding RNAs in mammalian cells and in mice. *Nucleic Acids Res.* **39**, (2011).

98. Khvorova, A. & Watts, J. K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **35**, 238–248 (2017).

99. Crooke, S. T., Wang, S., Vickers, T. A., Shen, W. & Liang, X. H. Cellular uptake and trafficking of antisense oligonucleotides. *Nat. Biotechnol.* **35**, 230–237 (2017).

100. Goodchild, J., Kim, B. & Zamecnik, P. C. The Clearance and Degradation of Oligodeoxynucleotides Following Intravenous Injection into Rabbits. *https://home.liebertpub.com/ard* **1**, 153–160 (2009).

101. Beltinger, C. *et al.* Binding, uptake, and intracellular trafficking of phosphorothioate-modified oligodeoxynucleotides. *J. Clin. Invest.* **95**, 1814 (1995).

102. Eckstein, F. Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic Acid Ther.* **24**, 374–387 (2014).

103. Manoharan, M. 2'-carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation. *Biochim. Biophys. Acta* **1489**, 117–130 (1999).

104. Mangos, M. & Damha, M. Flexible and frozen sugar-modified nucleic acids--modulation of biological activity through furanose ring dynamics in the antisense strand. *Curr. Top. Med. Chem.* **2**, 1147–1171 (2002).

105. Prakash, T. P. An overview of sugar-modified oligonucleotides for antisense therapeutics. *Chem. Biodivers.* **8**, 1616–1641 (2011).

106. Egli, M. *et al.* Probing the influence of stereoelectronic effects on the biophysical properties of oligonucleotides: comprehensive analysis of the RNA affinity, nuclease resistance, and crystal structure of ten 2'-O-ribonucleic acid modifications. *Biochemistry* **44**, 9045–9057 (2005).

107. Bijsterbosch, M. K. *et al.* In vivo fate of phosphorothioate antisense oligodeoxynucleotides: predominant uptake by scavenger receptors on endothelial liver cells. *Nucleic Acids Res.* **25**, 3290–3296 (1997).

108. Miller, C. M. *et al.* Stabilin-1 and Stabilin-2 are specific receptors for the cellular internalization of phosphorothioate-modified antisense oligonucleotides (ASOs) in the liver. *Nucleic Acids Res.* **44**, 2782–2794 (2016).

109. Loke, S. L. *et al.* Characterization of oligonucleotide transport into living cells. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 3474–3478 (1989).

110. Yakubov, L. A. *et al.* Mechanism of oligonucleotide uptake by cells: involvement of specific receptors? *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6454–6458 (1989).

111. Bushby, K. *et al.* Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *Lancet. Neurol.* **9**, 77–93 (2010).

112. Tanner, G. *et al.* Therapeutic strategy to rescue mutation-induced exon skipping in rhodopsin by adaptation of U1 snRNA. *Hum. Mutat.* **30**, 255–263 (2009).

113. Eagle, M. *et al.* Survival in Duchenne muscular dystrophy: Improvements in life expectancy since 1967 and the impact of home nocturnal ventilation. *Neuromuscul. Disord.* **12**, 926–929 (2002).

114. Hoffman, E. P. The discovery of dystrophin, the protein product of the Duchenne muscular dystrophy gene. *Febs J.* **287**, 3879 (2020).

115. Aartsma-Rus, A. & Van Ommen, G. J. B. Antisense-mediated exon skipping: A versatile tool with therapeutic and research applications. *RNA* **13**, 1609 (2007).

116. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nat. 1998 3916669* **391**, 806–811 (1998).

117. Stein, C. A. & Castanotto, D. FDA-Approved Oligonucleotide Therapies in 2017. *Mol. Ther.* **25**, 1069–1075 (2017).

118. Brad Wan, W. & Seth, P. P. The Medicinal Chemistry of Therapeutic Oligonucleotides. *J. Med. Chem.* **59**, 9645–9667 (2016).

119. Cheloufi, S., Dos Santos, C. O., Chong, M. M. & Hannon, G. J. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* **465**, 584–589 (2010).

120. Usmani, S. S. *et al.* THPdb: Database of FDA-approved peptide and protein therapeutics. *PLoS One* **12**, (2017).

121. Landthaler, M., Yalcin, A. & Tuschl, T. The Human DiGeorge Syndrome Critical

Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis. *Curr. Biol.* **14**, 2162–2167 (2004).

122. Takakura, K., Abe, H., Tanaka, R., … K. K.-J. of & 1986, undefined. Effects of ACNU and radiotherapy on malignant glioma. *thejns.org*.

123. Agarwal, S., Jangir, D. K., Mehrotra, R., Lohani, N. & Rajeswari, M. R. A Structural Insight into Major Groove Directed Binding of Nitrosourea Derivative Nimustine with DNA: A Spectroscopic Study. *PLoS One* **9**, e104115 (2014).

124. Aich, P. & Dasgupta, D. Role of Magnesium Ion in Mithramycin-DNA Interaction: Binding of Mithramycin-Mg2+ Complexes with DNA. *Biochemistry* **34**, 1376–1385 (1995).

125. Nelson, S. M., Ferguson, L. R. & Denny, W. A. Non-covalent ligand/DNA interactions: Minor groove binding agents. *Mutat. Res. Mol. Mech. Mutagen.* **623**, 24–40 (2007).

126. Arcamone, F. *et al.* Adriamycin, 14-hydroxydaimomycin, a new antitumor antibiotic from S. Peucetius var. caesius. *Biotechnol. Bioeng.* **11**, 1101–1110 (1969).

127. Cortés-Funes, H. & Coronado, C. Role of anthracyclines in the era of targeted therapy. *Cardiovasc. Toxicol.* **7**, 56–60 (2007).

128. oncology, R. W.-S. in & 1992, undefined. The anthracyclines: will we ever find a better doxorubicin? *europepmc.org*.

129. Fourmy, D., Recht, M. I., Blanchard, S. C. & Puglisi, J. D. Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274**, 1367–1371 (1996).

130. Kumana, C. R. & Yuen, K. Y. Parenteral aminoglycoside therapy. Selection, administration and monitoring. *Drugs* **47**, 902–913 (1994).

131. Andreana, I. *et al.* Nanomedicine for Gene Delivery and Drug Repurposing in the Treatment of Muscular Dystrophies. *Pharm. 2021, Vol. 13, Page 278* **13**, 278 (2021).

132. White, P. J., Anastasopoulos, F., Pouton, C. W. & Boyd, B. J. Overcoming biological barriers to in vivo efficacy of antisense oligonucleotides. *Expert Rev. Mol. Med.* **11**, e10 (2009).

133. Relizani, K. *et al.* Efficacy and Safety Profile of Tricyclo-DNA Antisense Oligonucleotides in Duchenne Muscular Dystrophy Mouse Model. *Mol. Ther. - Nucleic Acids* **8**, 144–157 (2017).

134. Fleischmann, W. A., Greenwood-Quaintance, K. E. & Patel, R. In Vitro Activity of Plazomicin Compared to Amikacin, Gentamicin, and Tobramycin against Multidrug-Resistant Aerobic Gram-Negative Bacilli. *Antimicrob. Agents Chemother.* **64**, (2020).

135. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov. 2018 178* **17**, 547–558 (2018).

136. Costales, M. G. *et al.* Small Molecule Inhibition of microRNA-210 Reprograms an Oncogenic Hypoxic Circuit. *J. Am. Chem. Soc.* **139**, 3446–3455 (2017).

137. Velagapudi, S. P. *et al.* Design of a small molecule against an oncogenic noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5898–5903 (2016).

138. Lorenz, D. A., Song, J. M. & Garner, A. L. High-Throughput Platform Assay Technology for the Discovery of pre-microRNA-Selective Small Molecule Probes. *Bioconjug. Chem.* **26**, 19–23 (2015).

139. Lorenz, D. A. & Garner, A. L. A click chemistry-based microRNA maturation assay optimized for high-throughput screening. *Chem. Commun. Chem. Commun* **52**, 8267–8270 (2016).

140. Daniel A. Lorenz, Steve Vander Roest, Martha J. Larsen & and Amanda L. Garner. Development and Implementation of an HTS-Compatible Assay for the Discovery of Selective Small-Molecule Ligands for Pre-microRNAs. *SLAS Discov.* 1–8 (2017) doi:10.1177/2472555217717944.

141. Lorenz, D. A., Song, J. M. & Garner, A. L. High-Throughput Platform Assay Technology for the Discovery of pre-microRNA-Selective Small Molecule Probes. *Bioconjug. Chem.* **26**, 19–23 (2015).

142. Disney, M. D. *et al.* Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem. Biol.* **11**, 1720–1728 (2016).

143. Chen, Y. *et al.* Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nat. Chem. Biol.* **12**, 717–723 (2016).

144. Nguyen, L. *et al.* Rationally Designed Small Molecules That Target Both the DNA and RNA Causing Myotonic Dystrophy Type 1. *J. Am. Chem. Soc.* **137**, 14180–14189 (2015).

145. Haraguchi, T. *et al.* A potent 2 0 -O-methylated RNA-based microRNA inhibitor with unique secondary structures. doi:10.1093/nar/gkr1317.

146. Luu, L. M. *et al.* A Potent Inhibitor of Protein Sequestration by Expanded Triplet (CUG) Repeats that Shows Phenotypic Improvements in a Drosophila Model of Myotonic Dystrophy. *ChemMedChem* **11**, 1428–1435 (2016).

147. Robertson, A. W. *et al.* Discovery of Surfactins as Inhibitors of MicroRNA Processing Using Cat-ELCCA. *ACS Med. Chem. Lett.* **12**, 878–886 (2021).

148. Nuñez, O., Chavez, B., Shaktah, R., Garcia, P. P. & Minehan, T. Synthesis and DNA binding profile of monomeric, dimeric, and trimeric derivatives of crystal violet. *Bioorg. Chem.* **83**, 297–302 (2019).

149. Yu, A. M., Choi, Y. H. & Tu, M. J. RNA Drugs and RNA Targets for Small

Molecules: Principles, Progress, and Challenges. *Pharmacol. Rev.* **72**, 862–898 (2020).

150. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).

151. Jain, S., Venkataraman, A., Wechsler, M. E. & Peppas, N. A. Messenger RNA-based vaccines: Past, present, and future directions in the context of the COVID-19 pandemic. *Adv. Drug Deliv. Rev.* **179**, (2021).

152. Verbeke, R., Lentacker, I., De Smedt, S. C. & Dewitte, H. The dawn of mRNA vaccines: The COVID-19 case. *J. Control. Release* **333**, 511–520 (2021).

153. Lin, M. & Guo, J. T. New insights into protein–DNA binding specificity from hydrogen bond based comparative study. *Nucleic Acids Res.* **47**, 11103 (2019).

154. Rohs, R. *et al.* Origins of Specificity in Protein-DNA Recognition. *https://doi.org/10.1146/annurev-biochem-060408-091030* **79**, 233–269 (2010).

155. Dršata, T. *et al.* Structure, stiffness and substates of the dickerson-drew dodecamer. *J. Chem. Theory Comput.* **9**, 707–721 (2013).

156. Chen, L., Glovert, J. N. M., Hogan, P. G., Rao, A. & Harrison, S. C. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **392**, 42–48 (1998).

157. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci.* **78**, 2179–2183 (1981).

158. Oguey, C., Foloppe, N. & Hartmann, B. Understanding the Sequence-Dependence of DNA Groove Dimensions: Implications for DNA Interactions. *PLoS One* **5**, e15931 (2010).

159. Li, J. *et al.* Structural basis for DNA recognition by STAT6. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 13015–13020 (2016).

160. Lambert, M., Jambon, S., Depauw, S. & David-Cordonnier, M. H. Targeting Transcription Factors for Cancer Treatment. *Mol. 2018, Vol. 23, Page 1479* **23**, 1479 (2018).

161. Carabet, L. A., Rennie, P. S. & Cherkasov, A. Therapeutic Inhibition of Myc in Cancer. Structural Bases and Computer-Aided Drug Discovery Approaches. *Int. J. Mol. Sci. 2019, Vol. 20, Page 120* **20**, 120 (2018).

162. Nair, S. K. & Burley, S. K. X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA: Molecular Bases of Regulation by Proto-Oncogenic Transcription Factors. *Cell* **112**, 193–205 (2003).

163. da Rosa, G. *et al.* Sequence-dependent structural properties of B-DNA: what have we learned in 40 years? **1**, 3.

164. Derose, E. F., Perera, L., Murray, M. S., Kunkel, T. A. & London, R. E. Solution Structure of the Dickerson DNA Dodecamer Containing a Single Ribonucleotide.

Biochemistry **51**, 2407 (2012).

165. Dickerson, U. S. A. *et al.* The Dickerson-Drew B-DNA Dodecamer Revisited at Atomic Resolution. *Proc. Natl. Acad. Sci. U.S.A* **78**, 5756–5757 (1981).

166. Fishert, D. E., Parentt, L. A. & Sharpt, P. A. Myc/Max and other helix-loop-helix/leucine zipper proteins bend DNA toward the minor groove (DNA bending/transcription factors/immun buln enhancer). *Biochemistry* **89**, 11779–11783 (1992).

167. Dans, P. D. *et al.* Long-timescale dynamics of the Drew–Dickerson dodecamer. *Nucleic Acids Res.* **44**, 4052–4066 (2016).

168. Doyle, L. A., Vivero, M., Fletcher, C. D. M., Mertens, F. & Hornick, J. L. Nuclear expression of STAT6 distinguishes solitary fibrous tumor from histologic mimics. *Mod. Pathol. 2014 273* **27**, 390–395 (2013).

169. STAT6 signal transducer and activator of transcription 6 [Homo sapiens (human)] - Gene - NCBI. https://www.ncbi.nlm.nih.gov/gene/6778.

170. Cascoń, A. & Robledo, M. MAX and MYC: A heritable breakup. *Cancer Res.* **72**, 3119–3124 (2012).

171. Amati, B. & Land, H. Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr. Opin. Genet. Dev.* **4**, 102–108 (1994).

172. Teixeira, L. K. *et al.* NFAT1 transcription factor regulates cell cycle progression and cyclin E expression in B lymphocytes. *Cell Cycle* **15**, 2346 (2016).

173. Werneck, M. B. F., Vieira-De-Abreu, A., Chammas, R. & Viola, J. P. B. NFAT1 transcription factor is central in the regulation of tissue microenvironment for tumor metastasis. *Cancer Immunol. Immunother.* **60**, 537–546 (2011).

174. Chu, Y., Hoffman, D. W. & Iverson, B. L. A pseudocatenane structure formed between DNA and A cyclic bisintercalator. *J. Am. Chem. Soc.* **131**, 3499–3508 (2009).

175. Kumar, S., Xue, L. & Arya, D. P. Neomycin-neomycin dimer: An all-carbohydrate scaffold with high affinity for AT-rich DNA duplexes. *J. Am. Chem. Soc.* **133**, 7361–7375 (2011).

176. MÜLLER, W. & GAUTIER, F. Interactions of heteroaromatic compounds with nucleic acids. A - T-specific non-intercalating DNA ligands. *Eur. J. Biochem.* **54**, 385–394 (1975).

177. Kim, S. K. & Nordén, B. Methyl green. A DNA major-groove binding drug. *FEBS Lett.* **315**, 61–64 (1993).

178. Wakelin, L. P. G., Adams, A., Hunter, C. & Waring, M. J. Interaction of Crystal Violet with Nucleic Acids. *Biochemistry* **20**, 5779–5787 (1981).

179. Schuster, I. I., Colter, A. K. & Kurland, R. J. Magnetic Resonance Studies of

Triphenylcarbonium Ions. I. Fluorine-19 Nuclear Magnetic Resonance Studies of Conformational Equilibria and Interconversion. *J. Am. Chem. Soc.* **90**, 4679–4687 (1968).

180. Balazy, M. *et al.* Dimeric and trimeric derivatives of the azinomycin B chromophore show enhanced DNA binding. *Org. Biomol. Chem.* **15**, 4522–4526 (2017).

181. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, NA-NA (2010).

182. Boger, D. L., Fink, B. E., Brunette, S. R., Tse, W. C. & Hedrick, M. P. A simple, high-resolution method for establishing DNA binding affinity and sequence selectivity. *J. Am. Chem. Soc.* **123**, 5878–5891 (2001).

183. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. doi:10.1038/nprot.2006.202.

184. Greenfield, N. J. Analysis of Circular Dichroism Data. *Methods Enzymol.* **383**, 282–317 (2004).

185. Intramolecular DNA Coiling Mediated by a Metallo-Supramolecular Cylinder. https://onlinelibrary.wiley.com/doi/epdf/10.1002/1521-3773%2820010302%2940%3A5%3C879%3A%3AAID-ANIE879%3E3.0.CO%3B2-X.

186. Triantafillidi, K., Karidi, K., Malina, J. & Garoufis, A. Oligopyridine-ruthenium(II)-amino acid conjugates: synthesis, characterization, DNA binding properties and interactions with the oligonucleotide duplex d(5¢-CGCGCG-3¢) 2. (2009) doi:10.1039/b904951g.

187. Garbett, N. C., Ragazzon, P. A. & Chaires, J. B. Circular dichroism to determine binding mode and affinity of ligand-DNA interactions. (2007) doi:10.1038/nprot.2007.475.

188. Peng, C. K. *et al.* Long-range correlations in nucleotide sequences. *Nat. 1992 3566365* **356**, 168–170 (1992).

189. Chatzidimitriou-Dreismann, C. A. & Larhammar, D. Long-range correlations in DNA. *Nat. 1993 3616409* **361**, 212–213 (1993).

190. DeFina, S. C. & Dieckmann, T. Synthesis of selectively 15N- or 13C-labelled malachite green. *J. Label. Compd. Radiopharm.* **45**, 241–248 (2002).

191. Jenkins, T. C. Optical absorbance and fluorescence techniques for measuring DNA-drug interactions. *Methods Mol. Biol.* **90**, 195–218 (1997).

192. Nielsen, P. E. & Appella Editors, D. H. Peptide Nucleic Acids Methods and Protocols Second Edition Methods in Molecular Biology 1050.

193. Pasquinelli, A. E. Conservation of the sequence and temporal expression of let-7

heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).

194. Johnson, S. M. *et al.* RAS Is Regulated by the let-7 MicroRNA Family. *Cell* **120**, 635–647 (2005).

195. Mayr, C., Hemann, M. T. & Bartel, D. P. Disrupting the Pairing Between let-7 and Hmga2 Enhances Oncogenic Transformation. *Science* **315**, 1576 (2007).

196. Sampson, V. B. *et al.* MicroRNA Let-7a Down-regulates MYC and Reverts MYC-Induced Growth in Burkitt Lymphoma Cells. *Cancer Res.* **67**, 9762–9770 (2007).

197. Takamizawa, J. *et al.* Reduced Expression of the let-7 MicroRNAs in Human Lung Cancers in Association with Shortened Postoperative Survival. *Cancer Res.* **64**, 3753–3756 (2004).

198. Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin-28. *Science* **320**, 97 (2008).

199. Chin, L. J. *et al.* A SNP in a let-7 microRNA Complementary Site in the KRAS 3′ Untranslated Region Increases Non–Small Cell Lung Cancer Risk. *Cancer Res.* **68**, 8535–8540 (2008).

200. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science (80-.).* **294**, 853–858 (2001).

201. Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet. 2011 1212* **12**, 861–874 (2011).

202. Rupaimoole, R. & Slack, F. J. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discov. 2017 163* **16**, 203–222 (2017).

203. Click Chemistry: Diverse Chemical Function from a Few Good Reactions - Kolb - 2001 - Angewandte Chemie International Edition - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/1521-3773(20010601)40:11%3C2004::AID-ANIE2004%3E3.0.CO;2-5.

204. Peypoux, F., Bonmatin, J. M. & Wallach, J. MINI-REVIEW Recent trends in the biochemistry of surfactin.

205. Golub, L. M., McNamara, T. F., D'angelo, G., Greenwald, R. A. & Ramamurthy, N. S. A Non-antibacterial Chemically-modified Tetracycline Inhibits Mammalian Collagenase Activity. *http://dx.doi.org/10.1177/00220345870660080401* **66**, 1310–1314 (2016).

206. Garner, A. L. *et al.* Tetracyclines as Inhibitors of Pre-microRNA Maturation: A Disconnection between RNA Binding and Inhibition. *ACS Med. Chem. Lett.* **10**, 816–821 (2019).

207. Guerra, W., Silva-Caldeira, P. P., Terenzi, H. & Pereira-Maia, E. C. Impact of metal coordination on the antibiotic and non-antibiotic activities of tetracycline-based drugs. *Coord. Chem. Rev.* **C**, 188–199 (2016).

208. Kim, Y. J. o., Wu, W., Chun, S. E., Whitacre, J. F. & Bettinger, C. J. Catechol-Mediated Reversible Binding of Multivalent Cations in Eumelanin Half-Cells. *Adv. Mater.* **26**, 6572–6579 (2014).

209. Bhushan, R. & Brückner, H. Marfey's reagent for chiral amino acid analysis: A review. *Amin. Acids 2004 273* **27**, 231–247 (2004).

210. Athanassiou, Z. *et al.* Structural mimicry of retroviral Tat proteins by constrained β-hairpin peptidomimetics: Ligands with high affinity and selectivity for viral TAR RNA regulatory elements. *J. Am. Chem. Soc.* **126**, 6906–6913 (2004).

211. Pai, J. *et al.* Screening of Pre-miRNA-155 Binding Peptides for Apoptosis Inducing Activity Using Peptide Microarrays. *J. Am. Chem. Soc.* **138**, 857–867 (2016).

212. Bose, D. *et al.* Selective inhibition of miR-21 by phage display screened peptide. *Nucleic Acids Res.* **43**, 4342–4352 (2015).

213. Nahar, S., Bose, D., Pal, S., Chakraborty, T. K. & Maiti, S. Cyclic Cationic Peptides Containing Sugar Amino Acids Selectively Distinguishes and Inhibits Maturation of Pre-miRNAs of the Same Family. *Nucleic Acid Ther.* **25**, 323–329 (2015).

214. Pai, J. *et al.* High-throughput profiling of peptide-RNA interactions using peptide microarrays. *J. Am. Chem. Soc.* **134**, 19287–19296 (2012).

215. Robertson, A. W. *et al.* Discovery of Surfactins as Inhibitors of MicroRNA Processing Using Cat-ELCCA. *ACS Med. Chem. Lett.* **12**, 878–886 (2021).

216. Chemistry and Biology of Cyclic Depsipeptides of Medicinal and Biological Interest | Bentham Science. https://www.eurekaselect.com/article/6972.

217. Zhongmin Liu, A. *et al.* Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate Data resources 5ZAK 5ZAL 5ZAM GSE110516. *Cell* **173**, (2018).

218. Diaz, J. P. *et al.* Association of a peptoid ligand with the apical loop of pri-miR-21 inhibits cleavage by Drosha. *RNA* **20**, 528 (2014).

219. Shortridge, M. D. *et al.* A Macrocyclic Peptide Ligand Binds the Oncogenic MicroRNA-21 Precursor and Suppresses Dicer Processing. *ACS Chem. Biol.* **12**, 1611–1620 (2017).

220. Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin-28. *Science* **320**, 97 (2008).

221. Michlewski, G., Guil, S., Semple, C. A. & Cáceres, J. F. Posttranscriptional Regulation of miRNAs Harboring Conserved Terminal Loops. *Mol. Cell* **32**, 383 (2008).

222. Thomas, J. R., Liu, X. & Hergenrother, P. J. Biochemical and thermodynamic characterization of compounds that bind to RNA hairpin loops: Toward an

understanding of selectivity. *Biochemistry* **45**, 10928–10938 (2006).

223. Thomas, J. R., Liu, X. & Hergenrother, P. J. Size-specific ligands for RNA hairpin loops. *J. Am. Chem. Soc.* **127**, 12434–12435 (2005).

224. Liu, X., Thomas, J. R. & Hergenrother, P. J. Deoxystreptamine dimers bind to RNA hairpin loops. *J. Am. Chem. Soc.* **126**, 9196–9197 (2004).

225. Connelly, C. M., Thomas, M. & Deiters, A. High-Throughput Luciferase Reporter Assay for Small-Molecule Inhibitors of MicroRNA Function. *J. Biomol. Screen.* **17**, 822 (2012).

226. Connelly, C. M. & Deiters, A. Cellular microRNA sensors based on luciferase reporters. *Methods Mol. Biol.* **1095**, 135–146 (2014).

227. Gumireddy, K. *et al.* Small-molecule inhibitors of microRNA miR-21 function. *Angew. Chemie* **47**, 7482–7484 (2008).

228. Sarabia, F., Chammaa, S. & García-Castro, M. Synthetic studies on stevastelins. 2. Synthesis of lipidic- and peptidic-modified analogues. *J. Org. Chem.* **70**, 7858–7865 (2005).

229. Bisek, N., Wetzel, S., Arndt, H. D. & Waldmann, H. Synthesis and Conformational Analysis of Stevastelin C3 Analogues and Their Activity Against the Dual-Specific Vaccina H1-Related Phosphatase. *Chem. – A Eur. J.* **14**, 8847–8860 (2008).

230. Sarabia, F., Chammaa, S. & García-Ruiz, C. Solid phase synthesis of globomycin and SF-1902 A5. *J. Org. Chem.* **76**, 2132–2144 (2011).

231. Kogen, H. *et al.* Crystal structure and total synthesis of globomycin: Establishment of relative and absolute configurations [5]. *J. Am. Chem. Soc.* **122**, 10214–10215 (2000).

232. Alcaro, M. C. *et al.* On-resin head-to-tail cyclization of cyclotetrapeptides: optimization of crucial parameters. *J. Pept. Sci.* **10**, 218–228 (2004).

233. Batiste, S. M. & Johnston, J. N. Evidence for Ion-Templation During Macrocyclooligomerization of Depsipeptides. *J. Am. Chem. Soc.* **140**, 4560–4568 (2018).

234. Pagadoy, M. Der & Wallach, J. Solid-Phase Synthesis of Surfactin , a Powerful Biosurfactant Produced by Bacillus subtilis , and of Four Analogues. **11**, 195–202 (2005).

235. Bollhagen, R., Schmiedberger, M., Barlos, K. & Grell, E. A new reagent for the cleavage of fully protected peptides synthesised on 2-chlorotrityl chloride resin. *J. Chem. Soc. Chem. Commun.* 2559–2560 (1994) doi:10.1039/C39940002559.

236. Wester, A., Hansen, A. M., Hansen, P. R. & Franzyk, H. Perfluoro-tert-butanol for selective on-resin detritylation: a mild alternative to traditionally used methods. *Amino Acids* **53**, 1455–1466 (2021).

237. Ferrer-Gago, F. J. & Koh, L. Q. Methods and Approaches for the Solid-Phase

Synthesis of Peptide Alcohols. *Chempluschem* **85**, 641–652 (2020).

238. Batiste, S. M. & Johnston, J. N. Rapid synthesis of cyclic oligomeric depsipeptides with positional, stereochemical, and macrocycle size distribution control. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14893–14897 (2016).

239. Van Dongen, S. F. M. *et al.* Single-step azide introduction in proteins via an aqueous diazo transfer. *Bioconjug. Chem.* **20**, 20–23 (2009).

240. MacRae, I. J., Ma, E., Zhou, M., Robinson, C. V. & Doudna, J. A. In vitro reconstitution of the human RISC-loading complex. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 512–517 (2008).

241. Lorenz, D. A. *et al.* Expansion of cat-ELCCA for the Discovery of Small Molecule Inhibitors of the Pre-let-7−Lin28 RNA−Protein Interaction. (2018) doi:10.1021/acsmedchemlett.8b00126.

242. White, C. J. & Yudin, A. K. Contemporary strategies for peptide macrocyclization. *Nat. Chem. 2011 37* **3**, 509–524 (2011).

243. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed. Engl.* **56**, 3770–3821 (2017).

244. Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215–218 (2000).

245. Xie, G. *et al.* Substrate spectrum of tyrocidine thioesterase probed with randomized peptide N-acetylcysteamine thioesters. *Bioorg. Med. Chem. Lett.* **12**, 989–992 (2002).

246. Tao, J., Hu, S., Pacholec, M. & Walsh, C. T. Synthesis of proposed oxidation-cyclization-methylation intermediates of the coumarin antibiotic biosynthetic pathway. *Org. Lett.* **5**, 3233–3236 (2003).

247. Tripathi, A., Choi, S. S., Sherman, D. H. & Kim, E. S. Thioesterase Domain Swapping of a Linear Polyketide Tautomycetin with a Macrocyclic Polyketide Pikromycin in Streptomyces sp. CK4412. *J. Ind. Microbiol. Biotechnol.* **43**, 1189 (2016).

248. Koch, A. A. *et al.* A Single Active Site Mutation in the Pikromycin Thioesterase Generates a More Effective Macrocyclization Catalyst. *J. Am. Chem. Soc.* **139**, 13456–13465 (2017).

249. Kittendorf, J. D. & Sherman, D. H. The Methymycin/Pikromycin Biosynthetic Pathway: A Model for Metabolic Diversity in Natural Product Biosynthesis. *Bioorg. Med. Chem.* **17**, 2137 (2009).