Inhibiting the Translation of Rev Protein with Triplex Forming Oligonucleotides

Thesis Submitted in Partial Fulfillment of the Requirements of the
Jay and Jeanie Schottenstein Honors Program

Yeshiva College
Yeshiva University
September 2022

Aaron Singer

Mentor: Irina Catrina PhD, Clinical Assistant Professor of Chemistry

Table of Contents

Abstract

Despite ongoing advancements in care, patients affected by human immunodeficiency virus-1 (HIV-1) infection continue to have a shorter than average lifespan and a decreased quality of life[1]. Therefore, more research is still necessary to ensure better health outcomes for these patients.

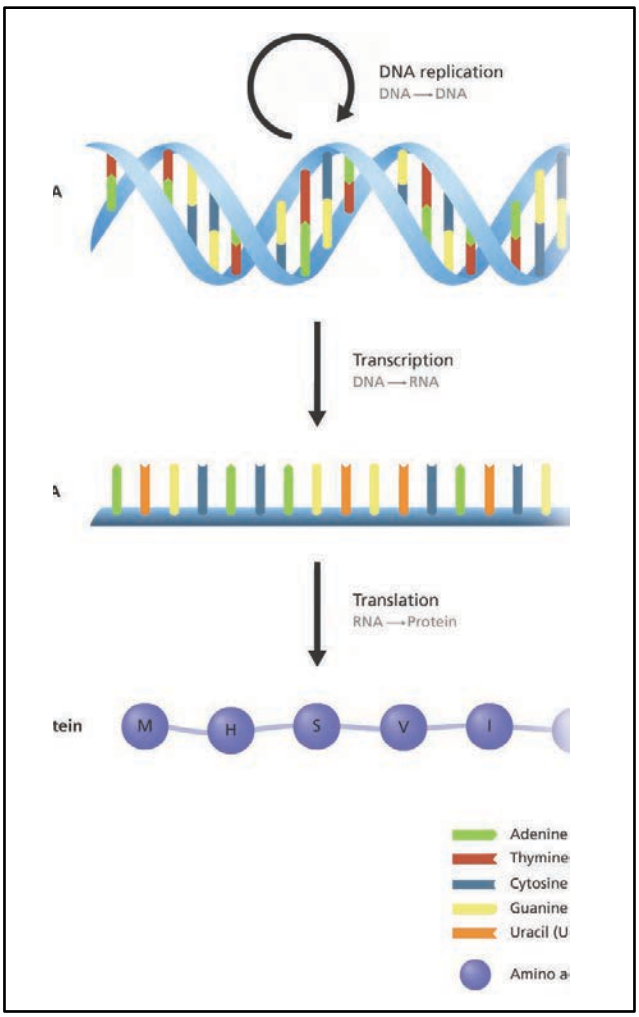
In the early step of its replication cycle, HIV-1 enters a host cell, and reverse transcribes its RNA into DNA. The virus then integrates the reverse-transcribed DNA into the host cell's genome using a viral integrase. The cell begins transcribing the integrated viral deoxyribonucleic acid (DNA) and producing fully spliced and processed messenger ribonucleic acids (mRNA)[2]. Rev protein, which itself is made from a fully spliced viral transcript, mediates the transport of intron containing viral RNAs out of the nucleus and into the cytoplasm via the CRM1 pathway. This enables the nucleocytoplasmic export of both genomic RNA (which will be packaged into virions,) and intron-containing RNA encoding for structural proteins. Rev also stabilizes viral RNAs, inhibits cellular splicing, and promotes translation of viral RNA. These functions make Rev protein essential for HIV-1's continued survival *in vivo* and make it an attractive target for anti-HIV drug design[3].

Triple helical RNA was first described in 1957[4] and triple helices have since been found to play a significant role in regulating gene expression[5]. Previous research has shown that translation of mRNA can be inhibited using triplex forming oligonucleotides (TFOs)[6]. TFOs are short sequences of nucleic acids that can form a triple helix with double stranded DNA or double stranded RNA, and a TFO can be composed of DNA, RNA, or peptide nucleic acid (PNA) backbone.

In this study, we identified a double-stranded region within the predicted Rev mRNA secondary structure(s) suitable for TFO targeting. We designed a PNA-based TFO that bound to this region. This was done using our Python program *TFOFinder*[7]. We confirmed triple helix formation using a hairpin that modeled this Rev mRNA region. Then, we used a commercial transcription/translation kit[8] to transcribe Rev mRNA and test the ability of our TFO(s) to inhibit translation of the Rev protein *in vitro*.

Introduction

Nucleic acids have multiple roles in biology. In addition to enzymatic and other functions, nucleic acids are best known for serving as the medium in which organisms store and transmit their genetic information. This information is stored within DNA in the nucleus of eukaryotes. Transcription factors are special proteins that help RNA polymerases bind to certain regions of the DNA and transcribe the DNA into pre-mRNA. This precursor mRNA is processed by 5' capping, 3' polyadenylation and splicing to yield the mature mRNA, which is exported into the cytoplasm via the Nxt1/Tap RNA export pathway and then translated by the ribosomes into proteins[9]. DNA as the genetic code, the transcription of DNA into mRNA, and the translation of mRNA into protein make up the central dogma of biology (Figure 1).



igure 1: Central dogma of biology: DNA plicates and is transcribed into RNA. RN translated into long chains of amino acic hich make up proteins[A].

There are four types of nucleotides that comprise DNA: adenine, guanine, thymine, and cytosine (Figure 2). These are often abbreviated as A, G, T, and C. Adenine and guanine are purines (denoted by the letter R), with nitrogenous bases containing two fused rings, while thymine and cytosine are pyrimidines (denoted by the letter Y), with nitrogenous bases

containing one aromatic ring. In addition to the nitrogenous ring(s), nucleotides also contain a backbone made of a (deoxy)ribose sugar moiety connected to a phosphate group. The sugar component of the backbone in DNA is deoxyribose, while the sugar component of the backbone in RNA is ribose. Additionally, thymine is replaced with another pyrimidine nucleotide, uracil, in RNA (Figure 2).

DNA is generally double stranded, with the two strands of DNA winding around each other in a double helix (Figure 2).

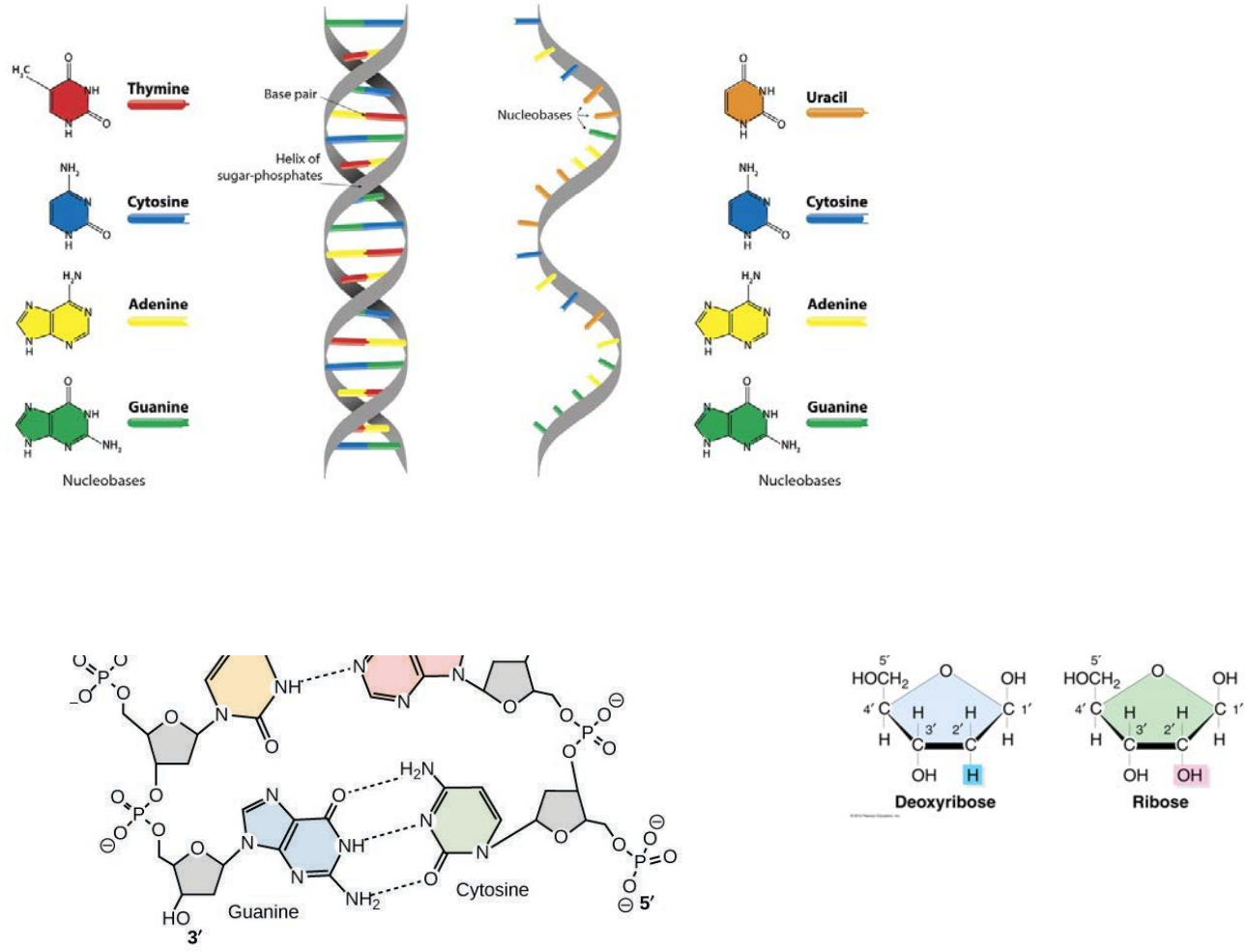
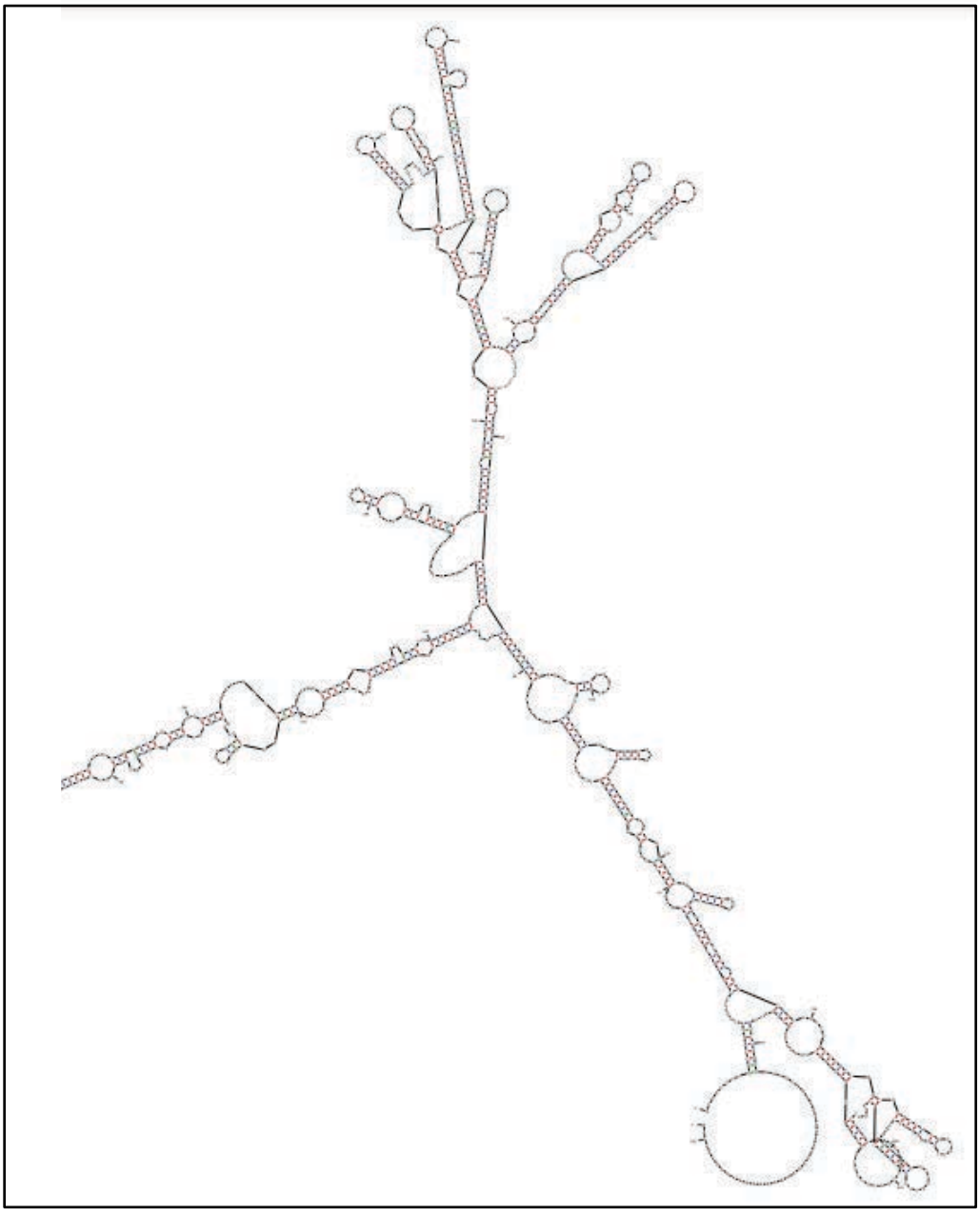


Figure 2: Structure of DNA and the nucleotides that comprise it. The double helix of DNA is made of the sugar-phosphate backbone with complementary nucleobases pairing up with each other[B,C,D].

Each strand has its sugar/phosphate backbone facing outward and its nitrogenous bases facing inward. The two strands are antiparallel to each other, and bar any mutations, are fully complementary to each other. Adenine on one strand will match up with thymine on the other strand, and cytosine will match up with guanine. The two bases form hydrogen bonds with each other, adenine and thymine forming two hydrogen bonds with each other, and cytosine forming three hydrogen bonds with guanine. Due to the higher number of hydrogen bonds, the non-covalent/H-bonding of cytosine and guanine is stronger than that of adenine and thymine (Figure 2).

RNA, unlike DNA, is generally made up of only one strand, but this strand will fold upon itself under physiological conditions (Figures 2, 3). The conformation that achieves the lowest free energy when folding will be most likely to form. Because the RNA is a single strand that folds upon itself, there will often be short regions with high complementarity that form canonical base pairs, and large loops containing low complementarity (Figure 3). The nucleic acid sequence of an RNA molecule can be input into energy minimization algorithms, such as *mfold*®[10], which return the minimum free energy (MFE) and the most likely sub-optimal RNA secondary structures to form at 1M NaCl. An example of a predicted MFE mRNA structure is shown in Figure 3.



DNA and RNA both form naturally in every organism. Peptide nucleic acid (PNA), first synthesized in 1991, is an artificial nucleic acid modification with a backbone made of a modified version of glycine, an amino acid[11] (Figure 4).

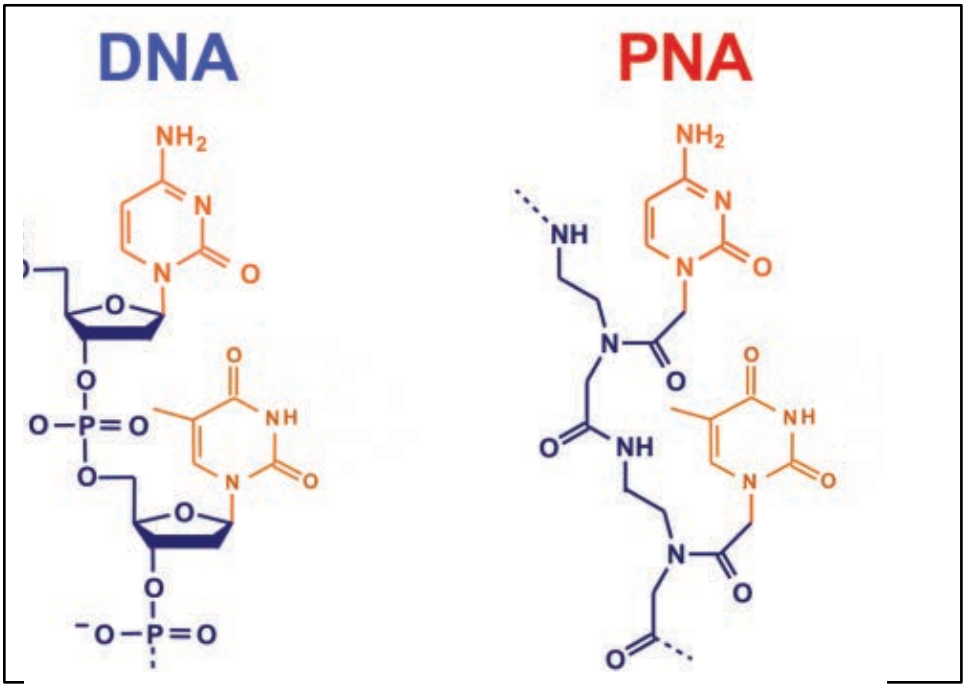


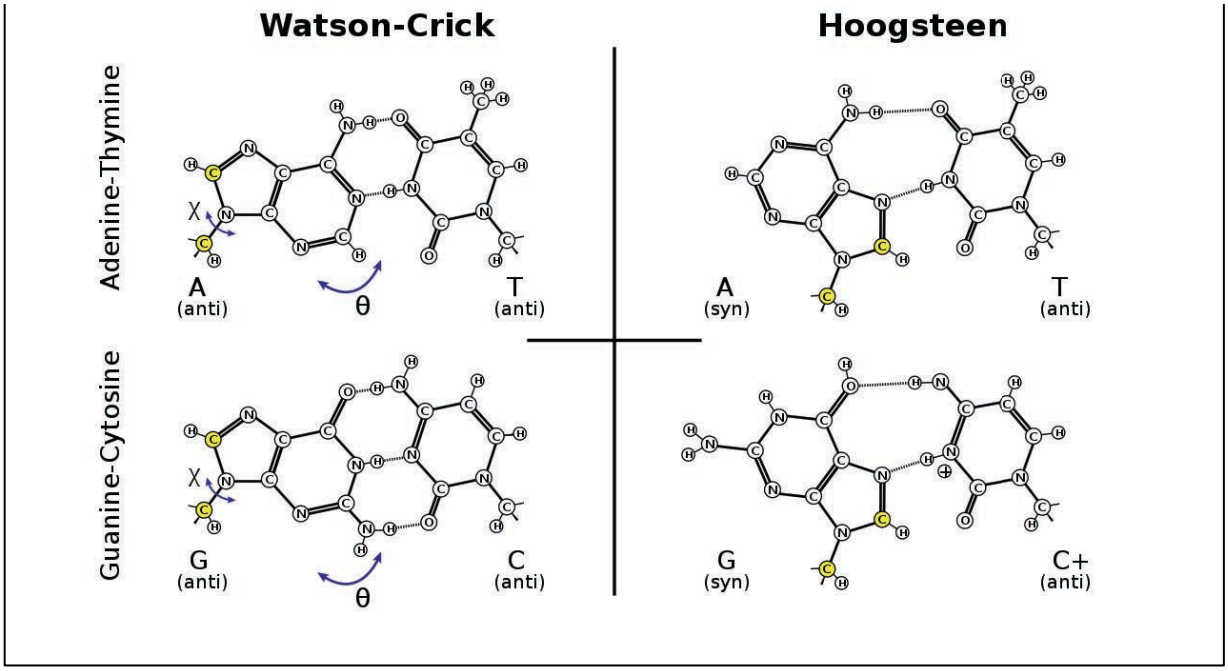
Figure 4: Structure of DNA compared to PNA. Note that the nucleosides are connected to a sugar-phosphate backbone in DNA, while they are connected to an amino acid backbone in PNA[E].

PNA is more resistant to enzymatic degradation than DNA or RNA. PNA/RNA and PNA/DNA duplexes have a higher H-bonding strength than RNA/RNA and DNA/DNA duplexes, respectively, but a mismatched pair of nucleotides within a PNA duplex is much more destabilizing than one within an RNA or DNA duplex[12].

Although DNA usually occurs as a duplex, in rare circumstances, three strands of DNA can form a triplex, with a short strand of DNA binding to the major (larger) groove of a long duplex (Figure 5), and the same phenomenon can occur with RNA.

Figure 5: DNA triple helix. Note the two larger green strands run antiparallel to each other while the shorter triplex forming oligonucleotide runs within the major groove of the duplex[F].

In duplexes, the two strands of nucleotides form classic Watson–Crick base pairing (Figure 5). In triplexes, the third strand of nucleotides forms Hoogsteen base pairs with a Watson–Crick duplex (Figure 6). For a triplex to form, the duplex needs to have a region in which one strand is comprised of purines while the other strand is comprised of pyrimidines[13]. Hoogsteen base pairs are similar to Watson–Crick pairs, but they are much rarer in nature and involve base flipping and purine rotation around the glycosidic bond[14] (Figure 6).



DNA triplexes can prevent transcription from occurring, and RNA triplexes can prevent translation from occurring[6,15]. A PNA strand can form a triplex with a DNA-DNA or RNA-RNA duplex. Pyrimidine-rich oligonucleotides bind in a parallel orientation under slightly acidic conditions (pH<6.0), with T and protonated C forming Hoogsteen hydrogen bonds with AT and GC base pairs, generating the base triplets T–AT and C+–GC, respectively[16].

Computer programs, such as *TFOFinder*, can be used to design an oligonucleotide that will form a triplex with a given DNA or RNA duplex based on energy minimization, intramolecular interactions, and accessibility of the targeted regions. *TFOFinder* finds purine rich double stranded regions in the predicted secondary structure of the RNA target of interest and excludes target regions that form bulges on the 3' strand of the duplex.

Eukaryotic genomes contain introns and exons. Introns are not transcribed, while exons are transcribed and translated (except for UTRs- untranslated regions - which are regions at the ends of the RNA that are not translated) (Figure 7). The production of Rev protein is essential to HIV-1's survival in the host organism because it allows the transportation of intron-containing viral RNA out of the nucleus and into the cytoplasm of the host cell[3].

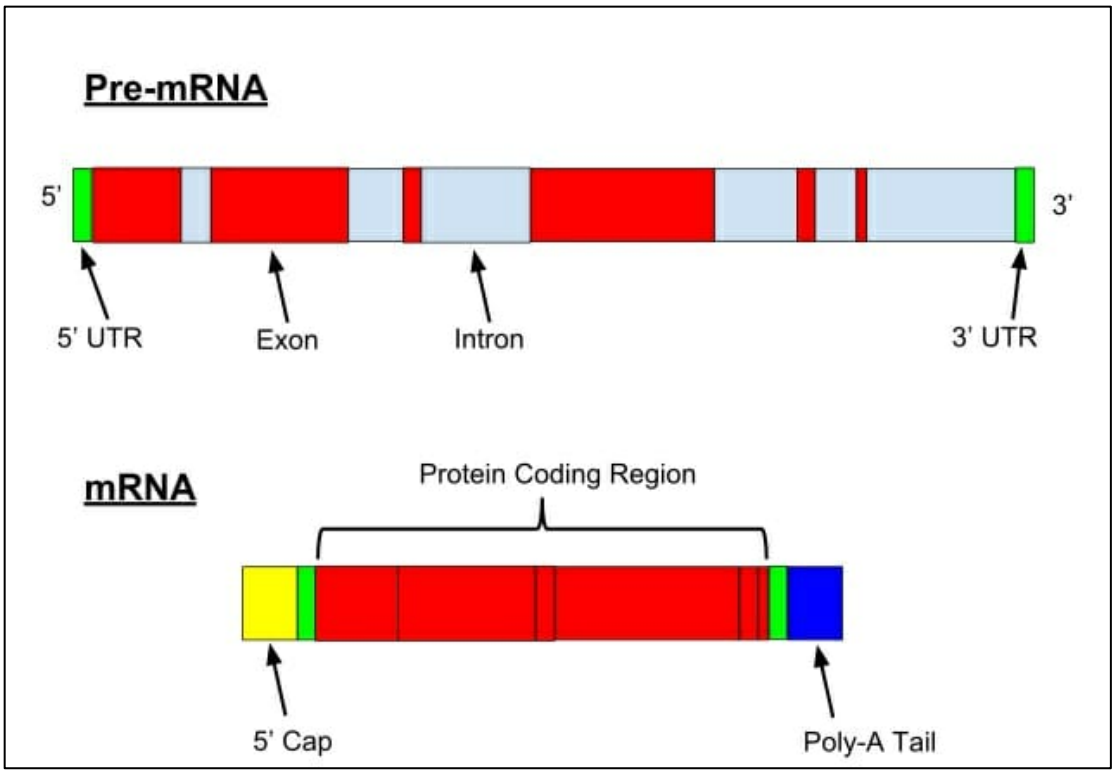


Figure 7: Immature vs. mature mRNA. The transcript is spliced, meaning the introns, are removed, and protected caps are placed on both ends of the mRNA. Introns are untranslated regions of the transcript[H].

Once a pre-mRNA has been transcribed and processed to yield the mature mRNA, the mRNA assumes a secondary (and tertiary) structure where different parts of the same strand of RNA interact with each other (Figure 8). RNA folding into the native conformation is required for performing proper cellular functions such as RNA transport and localization, translation, etc.

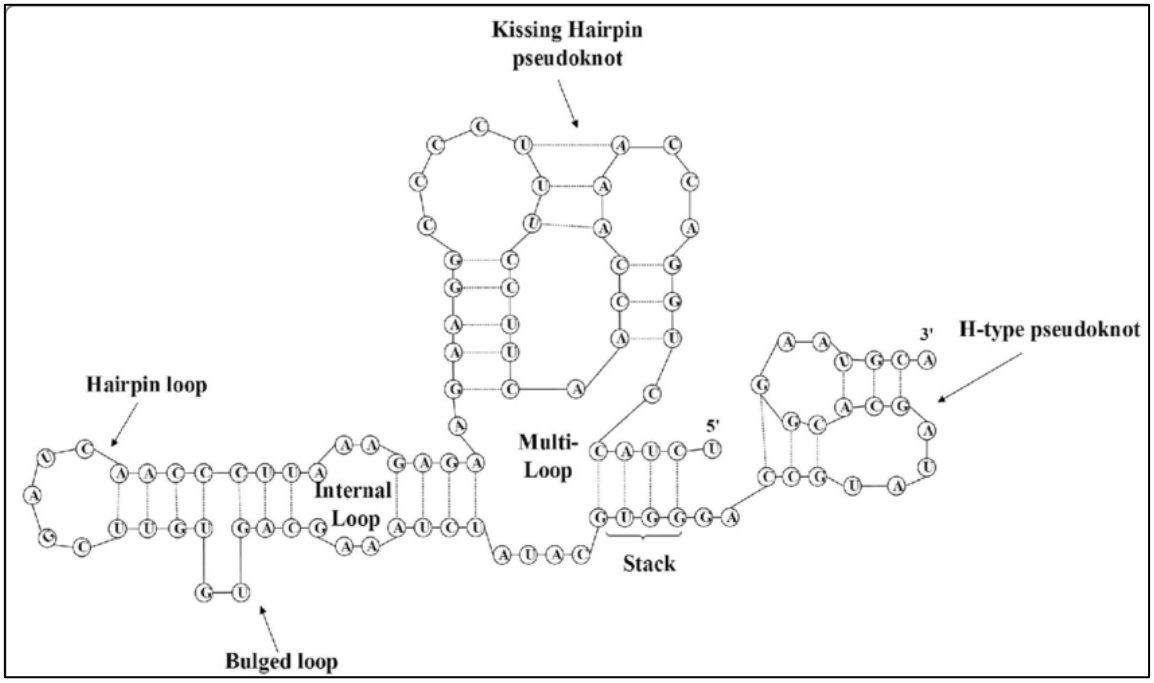


Figure 8: Different structural motifs that can occur within the secondary structure of mRNA[I].

The capped, polyadenylated, spliced and folded mRNA then gets transported out of the nucleus and into the cytoplasm where it will be translated into protein(s) by ribosomes.

Recent research by Rozners et al. has shown that specifically designed PNA molecules can inhibit translation of mRNAs both *in vitro* and in cells. This is true even for PNA molecules as short as nine nucleotides long. This effect is highly specific, and these researchers found that a mismatch of even one nucleotide eliminated any effects[6].

In this study, we explore the conformational and translational effects of the addition of specifically designed PNA molecules to Rev mRNA in solution and in *vitro* transcription/translation-coupled assays.

Materials and Methods

PNA OLIGOMER DESIGN

The secondary structure of *Rev* mRNA was predicted using *mfold* and *RNAstructure* programs and the MFE and suboptimal structures were used for oligomer design. Using TFOFinder, we designed PNA1 (NH_2-JJJJJTJJJTT-LysLysLys-$CONH_2$) and PNA2 (NH_2-TGCGTJJJJJTJJJTT-LysLysLys-$CONH_2$, which were predicted to form a triple helix with the 285-293 and 280-293 region of *Rev* mRNA, respectively.

Table 1: TFO sequences we designed. It should be noted that three lysine residues were added to the C-terminus of the TFOs in order to add a positive charge to the molecule, which improves its solubility in aqueous solutions.

TFO sequence designed for stem	N-terminus JJJJJTJJJTT-LysLysLys C-terminus
TFO sequence designed for stem and loop	N-terminus TGCGTJJJJJTJJJTT-LysLysLys C-terminus

J represents pseudoisocytosine, which replaces C and can form Hoogsteen base pairs at neutral pH[17]. These two PNA TFOs were synthesized and purified via HPLC using the services of Panagene Inc.

TRIPLE HELIX FORMATION

The RNA ideal model hairpin oligonucleotide (5' AACCCUCCCCCCAACCGACGCAGGGGGAGGG 3') was purchased from Integrated DNA Technologies Incorporated (Table 2).

Table 2: Natural and idealized hairpin sequence

Natural sequence of hairpin	5'AACCCACCUCCCAACCGACGCAGGGGGUGGG 3'
Idealized sequence of hairpin	5'AACCCUCCCCCCAACCGACGCAGGGGGAGGG 3'

0.8 µL of 50 µM RNA hairpin was added to 9.2 µL of 20 mM incubation buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 0.5 mM EDTA) in one tube, and 0.8 µL of 50 µM RNA hairpin and 8.2 µL of incubation buffer were added to another tube. The tubes were heated in boiling water for two minutes and placed on ice for 30 minutes. 1.0 µL of 100 µM PNA oligomer was added to the second tube and the tubes were incubated at room temperature for at least 1.5 hours. 1.0 µL of 80% glycerol was added to each tube and they were loaded and run through 2% agarose gel electrophoresis containing ethidium bromide (4 µL 10mg/ml in 100 mL) with 1X TBE buffer. Gel electrophoresis was run at 170V for ~30 minutes at 4°C and was analyzed under ultraviolet light and using a Bio-Rad Gel Imaging System.

A similar set-up was used to examine samples with varying concentrations of PNA1, PNA2, and a negative control RNA hairpin (5' CCTTTTTTTTTTTAAAATCCAAAAAAAAAAGG 3'), as presented in Table 3.

Table 3: Setup for the PNA binding assay with the Rev model RNA hairpin, analyzed using horizontal agarose gel electrophoresis.

	1	2	3	4	5	6	7	8	9
Rev HP	1µL of 20µM	1µL of 20µM	1µL of 20µM	1µL of 20µM	1µL of 20µM	1µL of 20µM	1µL of 20µM	1µL of 20µM	1µL of 20µM
PNA 1	0	1 µL of 20µM	2 µL of 20µM	0	0	0	0	1 µL of 20µM	2 µL of 20µM
PNA 2	0	0	0	0	1 µL of 20µM	2 µL of 20µM	0	0	0
Negative control HP	0	0	0	0	0	0	1 µL of 20µM	1 µL of 20µM	1 µL of 20µM
1X Incubation buffer	1µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL	1 µL
Water	8 µL	7 µL	6 µL	8 µL	7 µL	6 µL	8 µL	7 µL	6 µL

In order to increase the resolution of the bands within the gel, similar experiments were conducted using 12% polyacrylamide gel. The gel was made from 12 mL of water, 2 mL of 10X TBE buffer, 6 mL of 19:1 40% Accugel, 175 µL of 10% APS, and 15 µL of TEMED. The samples (as can be seen in Table 4) were loaded into the gel and ran at 125 Volts at 4°C for ~30 minutes.

Table 4: Setup for the PNA binding assay of PNAs 1 and 2 with the Rev hairpin, analyzed using vertical native polyacrylamide gel electrophoresis.

	1	2	3	4	5	6	7	8	9
RNA HP	2µL of 20µM	2µL of 20µM	2µL of 20µM	2µL of 20µM	2µL of 20µM	2µL of 20µM	2µL of 20µM	2µL of 20µM	2µL of 20µM
65% glycerol	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL	2 µL
1X incubation buffer	2 µL	1 µL	0	0	1 µL	2 µL	1 µL	0	0
PNA 1	0	1µL of 10µM	2µL of 10µM	2µL of 20µM	1µL of 100µM	0	0	0	0
PNA 2	0	0	0	0	0	0	1µL of 10µM	2µL of 10µM	2µL of 20µM

To get quantitative data on the bands, the gels were analyzed with a software called ImageJ which was used to measure the brightness of each band in the gel. The relative brightness was calculated by comparing all values to the brightness of the hairpin alone. Three gels were run and analyzed.

AMPLIFICATION OF Rev AND Rev-EGFP DNA USING PCR

A pUASp-Rev-EGFP plasmid was used to amplify Rev and Rev-EGFP coding DNA region by Polymerase Chain Reaction (PCR) to be used as template for the *in vitro* transcription/translation assay. One forward primer with the T7 promoter sequence (underlined

sequence) added at the 5' end (5' AATAATTAATACGACTCACTATAGGGAGAATGGCAGGAAGAAGCGGAG 3') was used but two different reverse primers were used (5' CTATTCTTTAGTTCCTGACTCCAATACTG 3' and 5' GTTAACGTTCGAGGTCGACTCTAG 3'), which amplified the Rev, and Rev-EGFP fragment, respectively. The reactions were prepared as per Table 5.

Table 5: Reaction preparation for PCR. An aliquot of each PCR reaction was used to confirm the samples were run through agarose gel with ethidium

Solution	**Volume (µM)**
Water	32.5
5X Phusion HF Buffer	10
10 mM dNTPs	1
10 µM Primer (forward)	2.5
10 µM Primer (reverse)	2.5
DNA template	1
Phusion enzyme	0.5
Total	50

The PCR reaction was performed with an initialization temperature of 98°C for 30 seconds. Each cycle involved a denaturation stage of 98°C for 10 seconds, an annealing stage of 64°C for 15 seconds, and an elongation stage of 72°C for 15 seconds. 35 cycles were run, followed by a final elongation stage at 72°C for 10 minutes, and a final hold temperature of 4°C. The samples were run through 0.5X TAE 1% agarose gel electrophoresis to confirm the presence of DNA fragments of the expected size (374 bp for Rev and 1091 bp for Rev-EGFP).

IN VITRO T7 TRANSCRIPTION/TRANSLATION-COUPLED ASSAY

In vitro T7 transcription/translation was performed using Rev DNA with and without PNA present using the TnT® T7 Quick for PCR DNA System (Promega, catalog number

L5900). A master mix was prepared by adding the following: 52µL TNT T7 Quick for PCR DNA master mix, 1.3µL 1mM methionine, 5.85µL PCR product, and 2.6µL Fluorotect (lysine with BODIPY dye). The master mix was split among three tubes, prepared as described in Table 6. The samples were incubated at 30°C for 60 minutes and then placed on ice. They were then combined with 5X Laemmli buffer and run in an agarose gel as previously described.

Table 6: Setup for T7 transcription/translation assay with and without the presence of PNA TFO.

	Sample 1	Sample 2	Sample 3
Master mix	19 µL	19 µL	19 µL
Water	1 µL	0	0
PNA1	0	1 µL of 20 µM	0
PNA2	0	0	1 µL of 20 µM

Because we could not detect the protein product for the above transcription/translation reactions, we conducted another test to determine if a positive control protein supplied with the kit could be produced. Two samples were prepared, one with 9µL TnT® T7 Quick Coupled Transcription/Translation System master mix, 1µL 1mM methionine, 1µL EGFP PCR DNA, and 1µL Fluorotect, and the other sample with 1µL of Fluorotect. These samples were run with glycerol in 1X TBE agarose gel at 80 Volts.

GEL ANALYSIS AND BAND QUANTIFICATION

Gel images acquired with a BioRad gel imager instrument were imported into ImageJ and analyzed using the gel analysis plugin and further analyzed with a custom R-script (Appendix). The experimental data was fitted and the equilibrium constants of PNA1 and PNA2 dissociation from the RNA hairpin were determined using a sigmoidal dose response model, the log-logistic model with four parameters (two fixed) from the *drc* R-package[18].

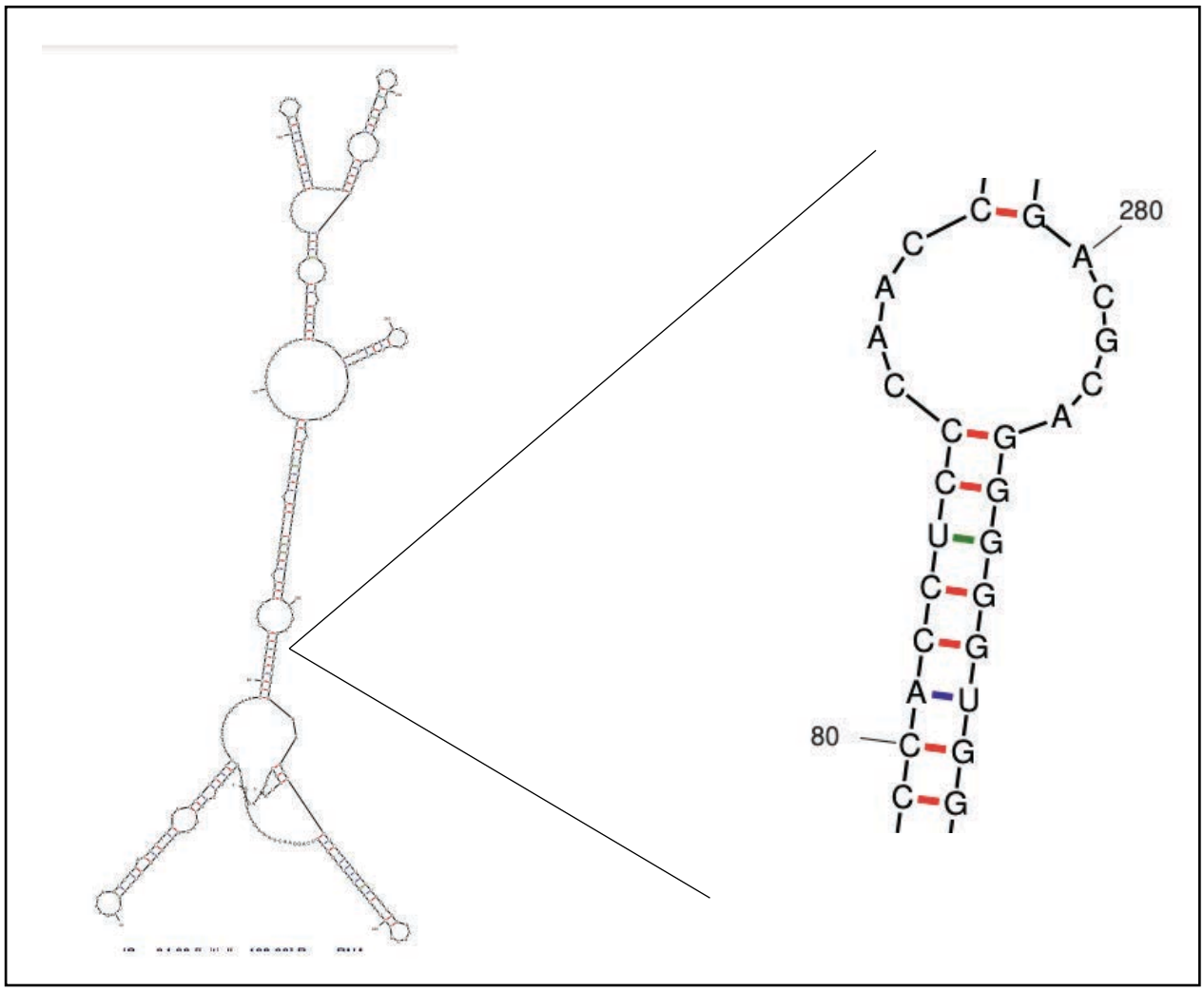
Results and Discussion

SELECTION OF THE Rev-mRNA TARGET REGION

Specific TFOs, which were predicted to form a triplex with Rev mRNA, were designed using *TFOfinder*. *TFOfinder* outputs many possible locations for TFO binding, so the candidate with the highest GA composition and lowest single strand count fraction (1 for fully single stranded and 0 for fully double stranded) was chosen.

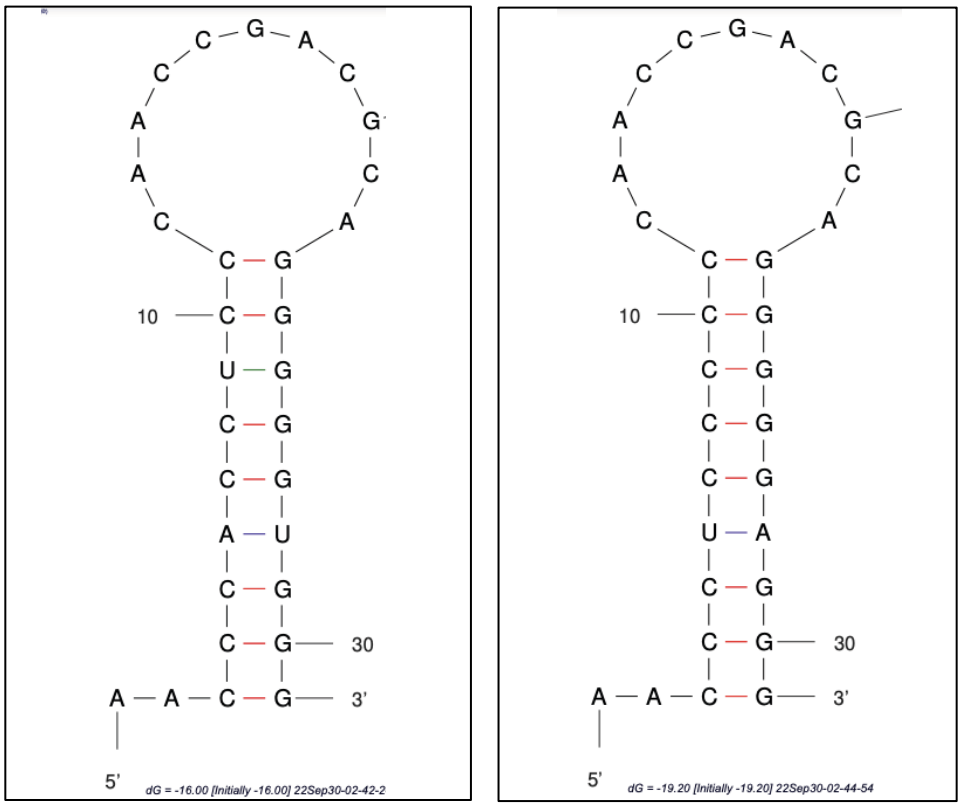
Multiple purine-rich duplexes are predicted to form within Rev-mRNA. The candidates chosen were TFOs that would theoretically bond to a part of the Rev mRNA that formed an internal loop (Figure 9).

Hairpins are similar to internal loops and occur when single-stranded RNA folds upon itself to form a double stranded stem with a loop.



igure 9: Predicted secondary structure of Rev mRNA given by *mfold*. Note the hairpin stem nd loop magnified in the figure. This region was targeted with PNA.

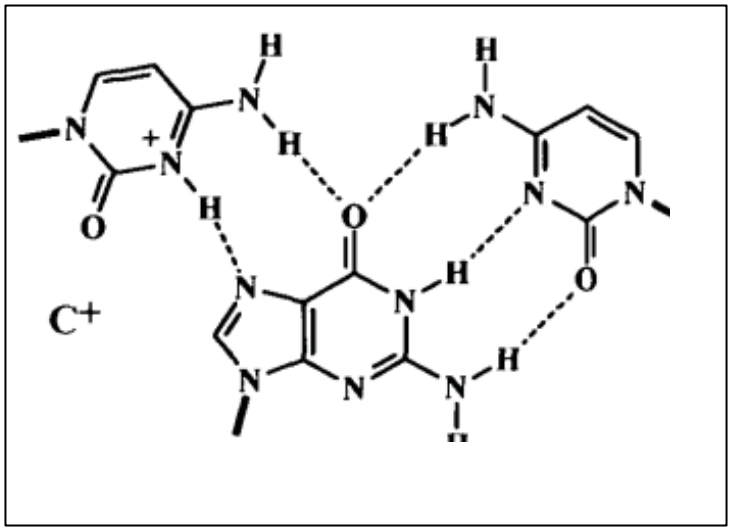
The part of the RNA that formed the duplex that the TFO was hypothesized to bind to was isolated, and an idealized version was created (Figure 10). Isolating this part of the RNA gave a hairpin, comprised of a 5' pyrimidine double stranded stem and a single stranded loop. This idealized version also had two changes. First, the uracil at position 84 was swapped with a cytosine. This turned a G-U wobble pair into a Watson-Crick pair and was done to stabilize the stem of the model hairpin. The second change was that the AU pair connecting positions 81 with 290 was changed to a UA pair. This was done so that 3' strand of the stem of the hairpin would be entirely comprised of purines, a requirement for efficient triplex formation. However, it was reported that triplexes can form with up to two pyrimidine inversions[6]



igure 10. 2D secondary structure of natural and idealized hairpin iven by *mfold*.

DESIGNING THE TFOs

A challenge that came with designing the TFOs was that in order to form a triplex with a GC pair, a protonated cytosine is needed (see Figure 11).



The p*K*a of cytosine is around 4.5, much lower than the pH of the body, and it needs to be protonated to form Hoogsteen base pairs. In order to overcome this issue, cytosine was replaced with pseudoisocytosine, a modified form of cytosine (Figure 12).

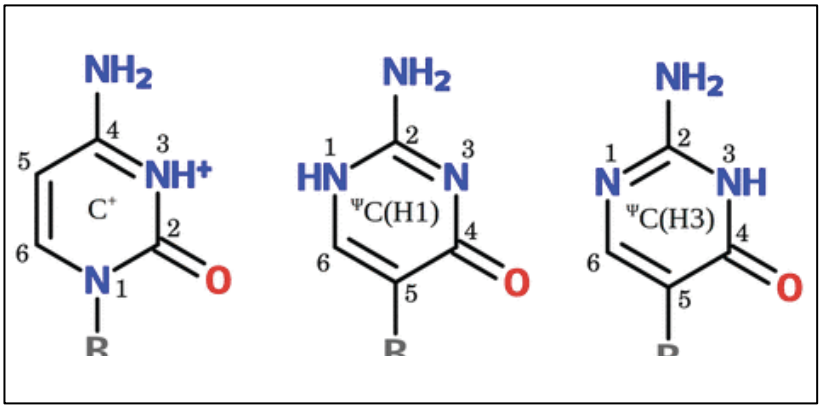
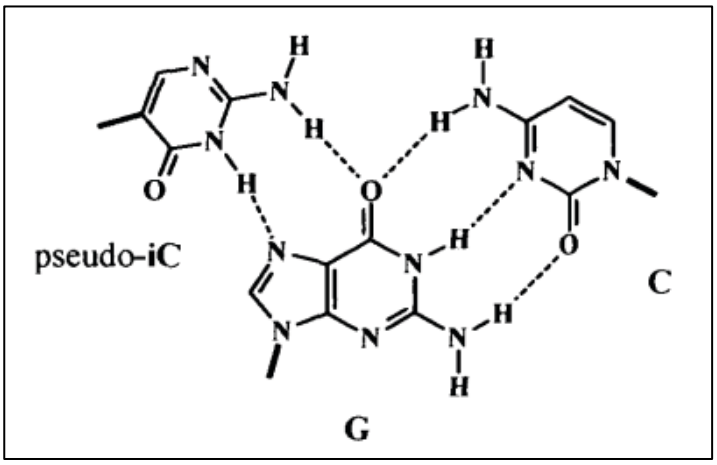


Figure 12: Cytidine, along with two tautomers of pseudoisocytidine. Note that unlike cytosine, a nitrogen is present opposite the oxygen on the ring, allowing the other nitrogen to carry a hydrogen which can participate in hydrogen bonding[K].

Pseudoisocytosine forms a triplex with a GC pair in a similar way to how a protonated cytosine would (Figure 13).



In ac[illegible]ner was ordered. This TFO was designed to not only bind to the stem of the hairpin to form a parallel triple helix,

but to form a duplex with a 5-nt 5' segment from the loop of the hairpin. We hypothesized that the additional length would help the TFO bind more tightly to the hairpin. Both TFOs also had an extra two T bases that would remain unbound near the base of the hairpin to add stability (Figure 14).

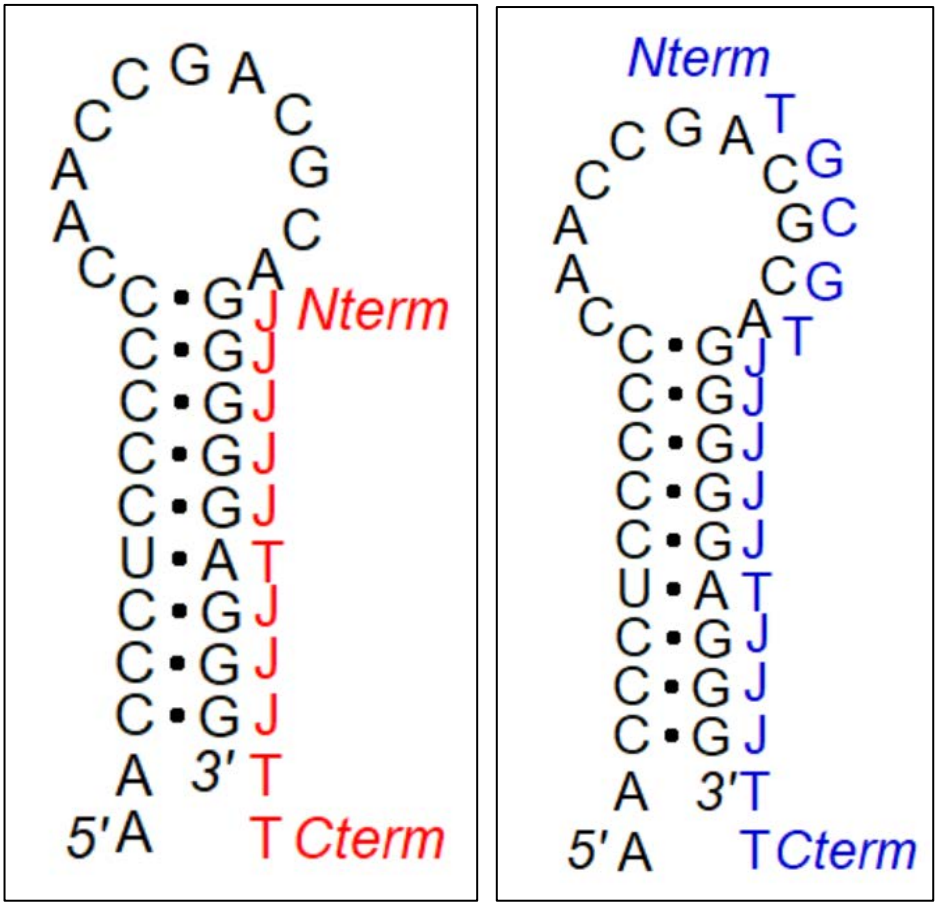


Figure 14: Structure of both PNAs and their predicted bonding arrangement to the idealized hairpin.

TRIPLE HELIX FORMATION

The model Rev RNA hairpin was incubated with each of the two PNAs at varying concentrations, and native/non-denaturing gel electrophoresis was used to evaluate and analyze PNA binding to the hairpin. Simultaneously, the model Rev RNA hairpin alone was run through the same agarose gel in an adjacent well. Ethidium bromide fluoresces under ultraviolet light

when complexed with double-stranded DNA or RNA, it does not bind to PNA/DNA or PNA/RNA hybrids, and thus, was used to detect the RNA hairpin.

GEL ANALYSIS

As presented in Figures 15 and 16, the RNA model hairpin/PNA samples migrated as two bands within the gel, while the model hairpin alone formed one band. The higher band in the RNA/PNA samples (top of the gel) indicated that the PNA formed a triplex with the RNA hairpin. The triplex complex was larger in size than the duplex and therefore moved more slowly through the gel. The second band ran the same distance as the band of the RNA hairpin alone sample and was certainly not the PNA alone, as the single stranded PNA does not bind ethidium bromide.

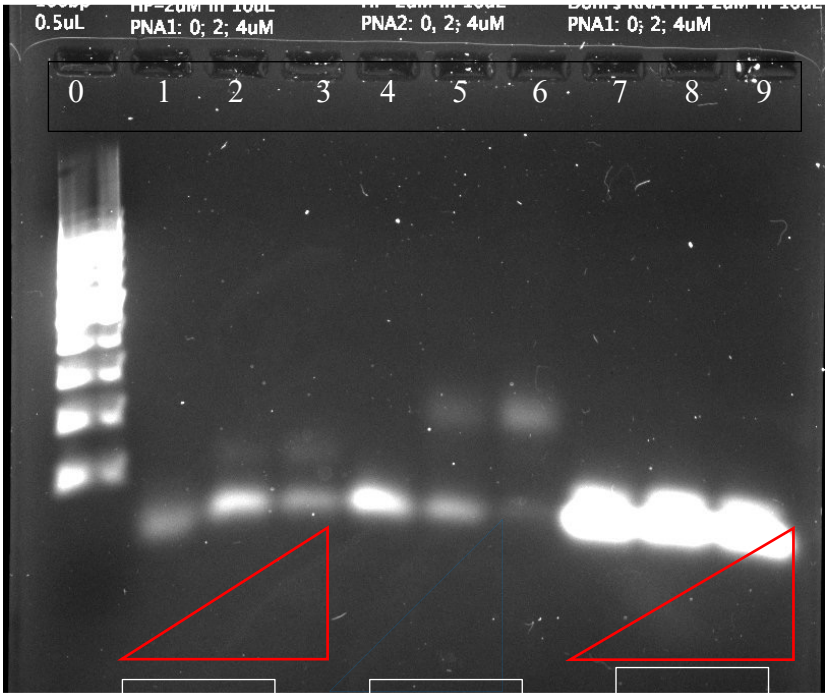


Figure 15: RNA/PNA binding assay on agarose gel. Lane 0 is a DNA ladder. Lanes 1-6 contain the model hairpin and varying amounts of PNA. Lanes 1-3 contain increasing amounts of PNA1, lanes 4-6 contain increasing amounts of PNA2, and lanes 7-9 contain increasing amounts of PNA1 with a negative control hairpin. Note that there is only one band in lanes 7-9, indicating that our TFO is specific.

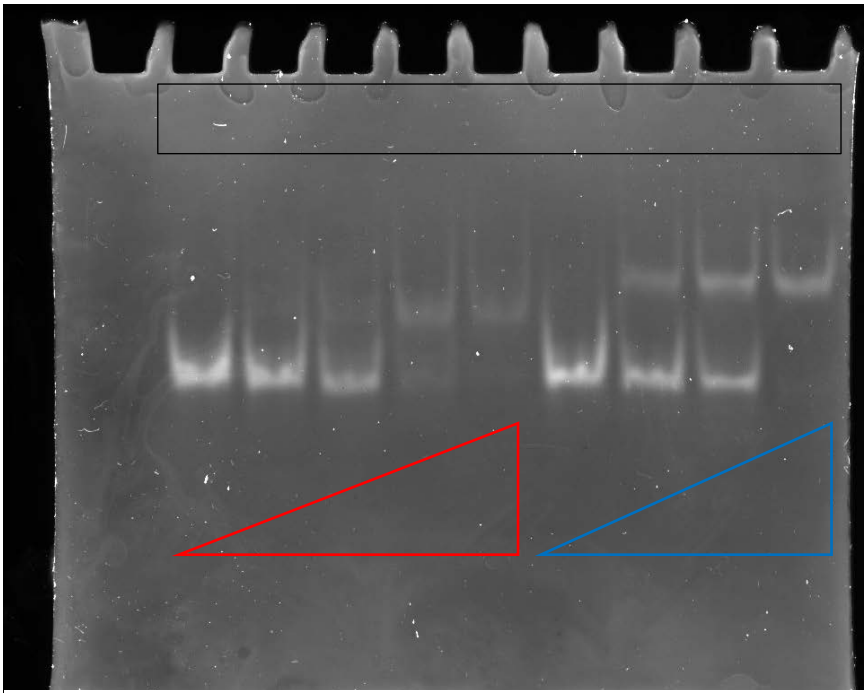


Figure 16: RNA/PNA binding assay on polyacrylamide gel. All lanes contain the model hairpin. Lanes 1-5 contain increasing amounts of PNA1 and lanes 6-9 contain increasing amounts of PNA2. Note the upper bands indicating PNA bound to the hairpin. Also note location of the upper bands in lanes 7-9 and how they are higher up in the gel than the upper bands in lanes 2 5. This is due to the PNA2 being longer and therefore larger thar PNA1.

Surprisingly, at high concentrations of PNA, less total RNA (sum of the intensity of the bands) was visible in the gel as compared to the RNA control sample. This effect was probably due to the PNA invading the stem of hairpin and forming an RNA-PNA duplex instead of an RNA-RNA/PNA triplex. RNA-PNA duplexes also do not bind ethidium bromide, explaining the "disappearance" of RNA in the gel.

This "disappearing" effect that PNA had on RNA by unfolding it was even more pronounced with the second PNA that we designed, which extended past the double-stranded

region of the RNA hairpin and was also complementary to five nucleic acids that were part of the single-stranded loop region. The effect with this PNA was so extreme that at a concentration of 4 μM, absolutely no RNA was visible in the gel, indicating that it had all been unraveled.

The quantitative data from ImageJ can be seen in Table 7 and Figure 17:

Table 7: Fractions of duplex and triplex after PNA complexing with RNA, as determined using ImageJ to measure the relative brightness of the bands within polyacrylamide gel.

	Trial 1		Trial 2		Trial 3		Average	
sample	**duplex**	**triplex**	**duplex**	**triplex**	**duplex**	**triplex**	**duplex**	**triplex**
1	1.01	0.00	1.23	0.00	1.12	0.00	1.12	0.00
2	0.69	0.00	1.54	0.00	0.81	0.00	1.01	0.00
3	0.38	0.02	0.49	0.00	0.52	0.02	0.46	0.01
4	0.04	0.11	0.01	0.11	0.06	0.09	0.04	0.10
5	0.00	0.10	0.00	0.13	0.03	0.09	0.01	0.11
6	0.99	0.00	0.77	0.00	0.88	0.00	0.88	0.00
7	0.56	0.12	0.55	0.07	0.46	0.14	0.52	0.11
8	0.38	0.20	0.48	0.16	0.30	0.13	0.39	0.17
9	0.01	0.30	0.00	0.14	0.03	0.18	0.01	0.21

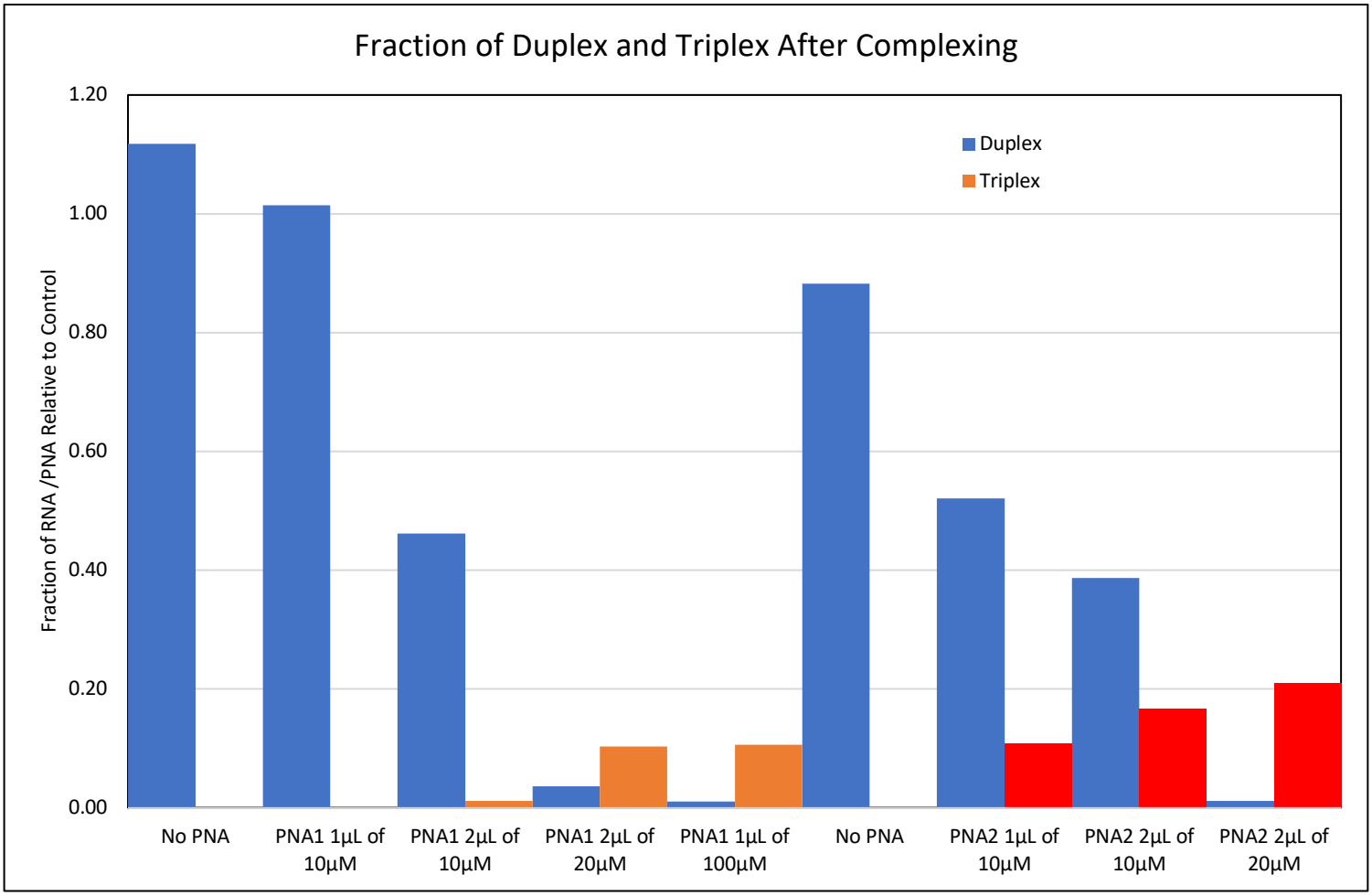
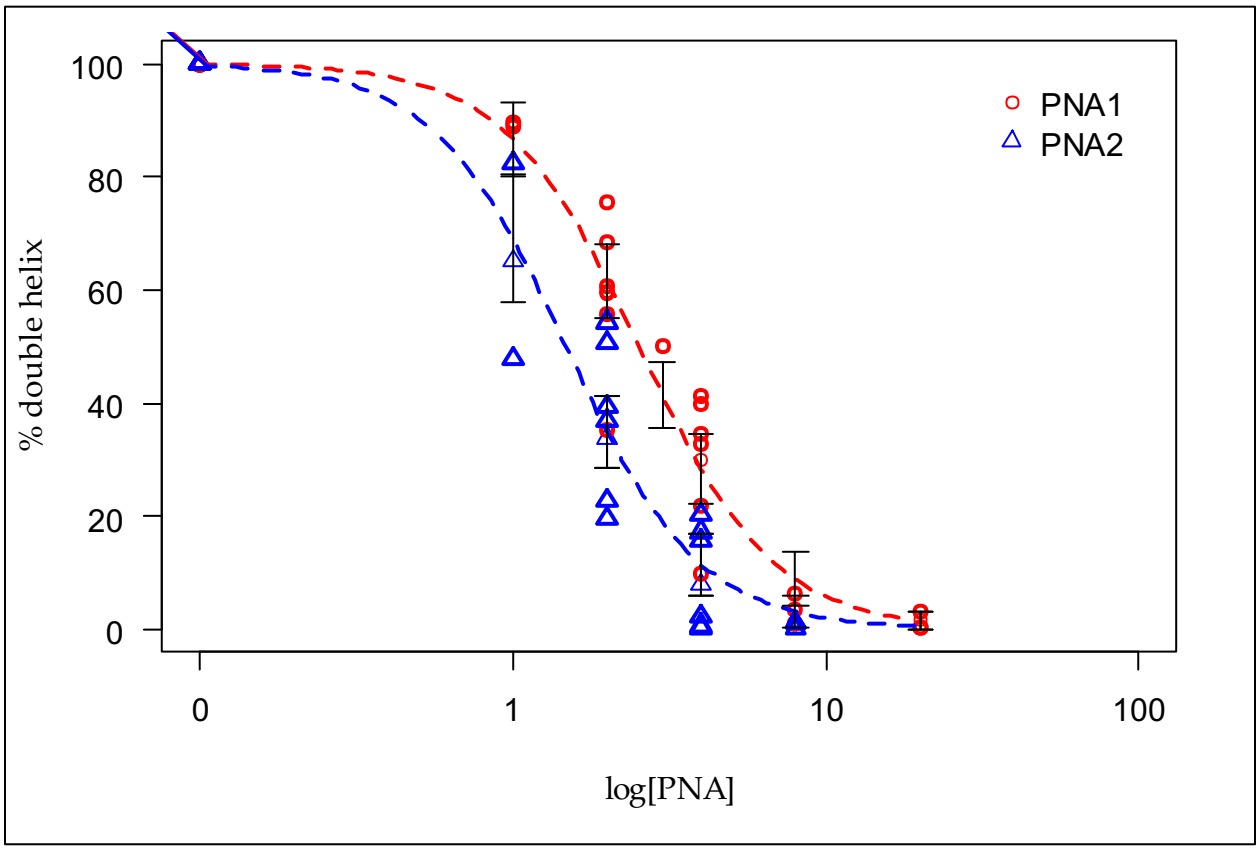


Figure 17: Fraction of duplex and triplex after TFOs complex with RNA. Visual representation of data within Table 7. Note the increase in triplex fraction with increasing PNA concentrations and the almost total disappearance of the duplex at the highest concentration of each PNA.

These data were also input into a sigmoidal dose-response model, as mentioned previously. This can be seen in Figure 18.



As a negative control, the experiment included an RNA hairpin that cannot for a triple helix with our PNAs. As expected, no effect was observed on the RNA hairpin migration on a native gel electrophoresis (Figure 15: lanes 7-9), with the RNA presenting as one band both, alone and with PNA added, meaning that no triple helix formed, and no invasion of the double helix occurred.

TRANSCRIPTION/TRANSLATION ASSAY

Once we had determined that our TFO was forming a triplex with the hairpin, we attempted to determine whether this meant the TFO was capable of preventing translation of Rev

mRNA from occurring. A transcription/translation kit was used on T7-Rev DNA both in the presence and absence of PNA. Unfortunately, despite repeated attempts, including we were not able to detect any translated protein. We tried to perform a control transcription/translation assay using Luciferase DNA, but this was also unsuccessful.

There are many possible reasons that translation was not successful. It could be that our PCR fragments are incomplete or somehow contain an error which prevents transcription from occurring. There could also be a problem with the transcription/translation kit that we used, either with the enzymes that transcribe the RNA or that translate the protein. It is possible that the detection sensitivity is also problem, because even though we were able to detect the fluorescently labeled lysine, Rev contains only 3 Lys residues. In addition, the recommended lower limit for template size is 300 bp for in vitro transcription assays and the Rev DNA template is close to this limit at 374 bp long.

FUTURE EXPERIMENTS

As mentioned previously, the control transcription/translation of our PCR fragments was not successful. To determine which part of the process is not working properly, Dr. Irina Catrina and Jonah Rocheeld will attempt to transcribe the PCR fragments without translating them. To monitor RNA duplex invasion by TFO, a fluorescently labeled RNA hairpin will be designed.

Conclusions

The appearance of a second band in the gels at a level slightly higher than the RNA, and the positive correlation between the disappearance of RNA and the addition of larger amounts of our TFOs indicates that both TFOs successfully bound to the RNA model hairpin, forming a

triplex, and invading the duplex. In addition to being a successful result for this project's purpose, this also strengthens the efficacy of *TFOFinder,* showing that it can generate TFOs reliably.

Despite repeated attempts, we were not able to analyze the effect of TFO on RNA translation into protein, however alternative detection methods such as Western blotting and silver staining may help answer this question.

References

Sources

1. Powell MK, Benková K, Selinger P, Dogoši M, Kinkorová Luňáčková I, et al. (2016) Opportunistic Infections in HIV-Infected Patients Differ Strongly in Frequencies and Spectra between Patients with Low CD4+ Cell Counts Examined Postmortem and Compensated Patients Examined Antemortem Irrespective of the HAART Era. PLOS ONE 11(9): e0162704. https://doi.org/10.1371/journal.pone.0162704

2. Delelis, O., Carayon, K., Saïb, A. et al. Integrase and integration: biochemical activities of HIV-1 integrase. Retrovirology 5, 114 (2008). https://doi.org/10.1186/1742-4690-5-114

3. Truman CT, Järvelin A, Davis I, Castello A. HIV Rev-isited. Open Biol. 2020 Dec;10(12):200320. doi: 10.1098/rsob.200320. Epub 2020 Dec 23. PMID: 33352061; PMCID: PMC7776567.

4. G. Felsenfeld, David R. Davies, and Alexander Rich *Journal of the American Chemical Society* **1957** *79* (8), 2023-2024 DOI: 10.1021/ja01565a074

5. Jain A, Wang G, Vasquez KM. DNA triple helices: biological consequences and therapeutic potential. Biochimie. 2008 Aug;90(8):1117-30. doi: 10.1016/j.biochi.2008.02.011. Epub 2008 Feb 21. Erratum in: Biochimie. 2018 Mar 23;: PMID: 18331847; PMCID: PMC2586808.

6. T. Endoh, D. Hnedzko, E. Rozners, N. Sugimoto, Angew. Chem. Int. Ed. 2016, 55, 899.

7. Tfo_rna/tfo_finder_win.Py at Master · Icatrina/tfo_rna - Github.com. https://github.com/icatrina/tfo_rna/blob/master/tfo_finder_win.py.

8. TNT T7 Quick for PCR DNA - Fred Hutch. https://research.fredhutch.org/content/dam/stripe/hahn/methods/mol_biol/tnt_quick_PCR_manual.pdf.

9. RNA Export Mediated by Tap Involves NXT1-dependent Interactions with the Nuclear Pore Complex * Lévesque, Lyne et al. Journal of Biological Chemistry, Volume 276, Issue 48, 44953 – 44962

10. RNA Folding Form, http://www.unafold.org/mfold/applications/rna-folding-form.php.

11. Nielsen P.E., Egholm M., Berg R.H., Buchardt O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. Science. 1991;254:1497–1500

12. Gupta A, Mishra A, Puri N. Peptide nucleic acids: Advanced tools for biomedical applications. J Biotechnol. 2017 Oct 10;259:148-159. doi: 10.1016/j.jbiotec.2017.07.026. Epub 2017 Jul 29. PMID: 28764969; PMCID: PMC7114329.

13. Maldonado R, Filarsky M, Grummt I, Längst G. Purine- and pyrimidine-triple-helix-forming oligonucleotides recognize qualitatively different target sites at the ribosomal DNA locus. RNA. 2018 Mar;24(3):371-380. doi: 10.1261/rna.063800.117. Epub 2017 Dec 8. PMID: 29222118; PMCID: PMC5824356.

14. Aishima J, Gitti RK, Noah JE, Gan HH, Schlick T, Wolberger C. A Hoogsteen base pair embedded in undistorted B-DNA. Nucleic Acids Res. 2002 Dec 1;30(23):5244-52. doi: 10.1093/nar/gkf661. PMID: 12466549; PMCID: PMC137974.

15. Goñi JR, de la Cruz X, Orozco M. Triplex-forming oligonucleotide target sequences in the human genome. Nucleic Acids Res. 2004 Jan 15;32(1):354-60. doi: 10.1093/nar/gkh188. PMID: 14726484; PMCID: PMC373298.

16. Rodrigo Maldonado, Michael Filarsky, Ingrid Grummt, and Gernot Längst RNA March 2018 24: 371-380; Published in Advance December 8, 2017, doi:10.1261/rna.063800.117

17. Role of Pseudoisocytidine Tautomerization in Triplex-Forming Oligonucleotides: In Silico and in Vitro Studies. Yossa Dwi Hartono, Y. Vladimir Pabon-Martinez, Arzu Uyar, Jesper Wengel, Karin E. Lundin, Rula Zain, C. I. Edvard Smith, Lennart Nilsson, and Alessandra Villa *ACS Omega* **2017** *2* (5), 2165-2177 DOI: 10.1021/acsomega.7b00347

18. Ritz, Christian, et al. “Dose-Response Analysis Using R.” *PLOS ONE*, Public Library of Science, https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0146021.

Images

A. Sanger. “What Is the 'Central Dogma'?” @Yourgenome · Science Website, 21 July 2021, https://www.yourgenome.org/facts/what-is-the-central-dogma/.
B. “DNA vs. RNA – 5 Key Differences and Comparison.” *Genomics Research from Technology Networks*, https://www.technologynetworks.com/genomics/lists/what-are-the-key-differences-between-dna-and-rna-296719.
C. “Discovery of the Structure of DNA (Article).” *Khan Academy*, Khan Academy, https://www.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-discovery-and-structure/a/discovery-of-the-structure-of-dna.
D. Carr, Dr Steven M. Ribose & Deoxyribose Sugars, https://www.mun.ca/biology/scarr/iGen3_02-07.html.
E. “Peptide Nucleic Acid " Laboratory of Applied Biomolecular Research: Boston University.” *Laboratory of Applied Biomolecular Research RSS*, https://sites.bu.edu/biomol/projects/pna/.
F. Bank, RCSB Protein Data. “1BWG: DNA Triplex with 5' and 3' Junctions, NMR, 10 Structures.” RCSB PDB, https://www.rcsb.org/structure/1BWG.
G. Nikolova, Evgenia N, et al. “Transient Hoogsteen Base Pairs in Canonical Duplex DNA.” *Nature*, U.S. National Library of Medicine, 24 Feb. 2011, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3074620/.
H. Editors, BD. “Intron - Definition, Function and Structure.” *Biology Dictionary*, 6 Aug. 2017, https://biologydictionary.net/intron/.
I. Kai, Zhang & Yuting, Wang & Yulin, Lv & Jun, Liu & He, Juanjuan. (2019). An efficient simulated annealing algorithm for the RNA secondary structure prediction with Pseudoknots. BMC Genomics. 20. 10.1186/s12864-019-6300-2, https://www.researchgate.net/figure/RNA-Secondary-Structure-and-Substructures_fig1_338191138
J. Egholm, M, et al. “Efficient PH-Independent Sequence-Specific DNA Binding by Pseudoisocytosine-Containing Bis-PNA.” *Nucleic Acids Research*, U.S. National Library of Medicine, 25 Jan. 1995, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC306657/?page=2.
K. *ACS Publications: Chemistry Journals, Books, and References Published ...* https://pubs.acs.org/doi/10.1021/acsomega.7b00647.

Acknowledgements

This entire project would not have been possible without the immense amount of time and effort put in by Dr. Catrina. Her guidance in the lab and with the writing of the paper was truly a blessing. I would also like to thank the Honors Department, specifically Dr. Eliezer Schnall for approving the funding and Dr. Benjamin Stein for placing the IDT order. Finally, I

would like to thank the Biology Department for allowing us use their equipment to conduct our research.

Appendix

```
# Irina E. Catrina
# July 6th, 2022
# This script is for the analysis of duplex disappearance for Rev mRNA hairpin model RNA
# make these packages and their associated functions
# available to use in this script
library(dplyr)
library(ggplot2)
library(tidyverse)
library(drc)
#theme_set(
#  theme_bw() +
#    theme(legend.position = "top")
#)
#clear R's brain
rm(list = ls())
#import data
myfile2  <- read.csv("C:/Users/User/OneDrive - Yeshiva
University/Desktop/AAA_YU/Research/AAA_R/AS_DH_2.csv")
View(myfile2)
myfile2 <- na.omit(myfile2)
glimpse(myfile2)
tbl_df
model.LL3 <- drm(DH ~ Conc, data = myfile2, PNA, fct = LL.3(fixed=c(NA, 100, NA), names
= c("Slope", "Upper Limit", "Kd")))
summary(model.LL3)
#mselect(model.LL3, fctList = list(W1.3(fixed=c(NA, 100, NA)),W1.4(), W2.3(fixed=c(NA,
100, NA)), W2.4(),  LL.4()),linreg=TRUE)
#par(mfrow = c(1, 2))
#plot(DH~Conc, data = myfile2, log = "x")
windowsFonts(A = windowsFont("Palatino Linotype"))
plot(model.LL3, log = "x", xlab="log[PNA]", ylab = "% double helix", lty=0, lwd = 1, xlim =
c(0, 100), col = c("red", "blue"), family = "A")
plot(model.LL3, add = TRUE, type = "all", legend = (""), lty=2, lwd = 2, col = c("red", "blue"),
family = "A")
plot(model.LL3, add = TRUE, col = c("red", "blue"), type = "bars", legend = (""), lty = 0, family
= "A")
#main="LL.3(fixed=c(NA, 100, NA))"
```