## Methods

For the design of the molecular beacon, background research on other designed beacons provided information on how many base pairs the beacon should be, how long the stem of the beacon is, and what base pairs should be used in the stem for it to be stable [1]. The beacon is comprised of three regions, the 5' stem region, the unpaired loop, and the 3' stem region. On the 5' region, a modified 5' thiol is attached at the end and a 5' fluorescein is adjacent to the thiol and the beginning of the stem. The 5' stem has five base pairs that are complementary to a sequence in the viral RNA. The unpaired loop has sixteen unpaired base pairs that are complementary to a sequence in the viral RNA. Last, the 3' stem has five base pairs that are complementary to the 5' stem region and a 3' quencher at the end.
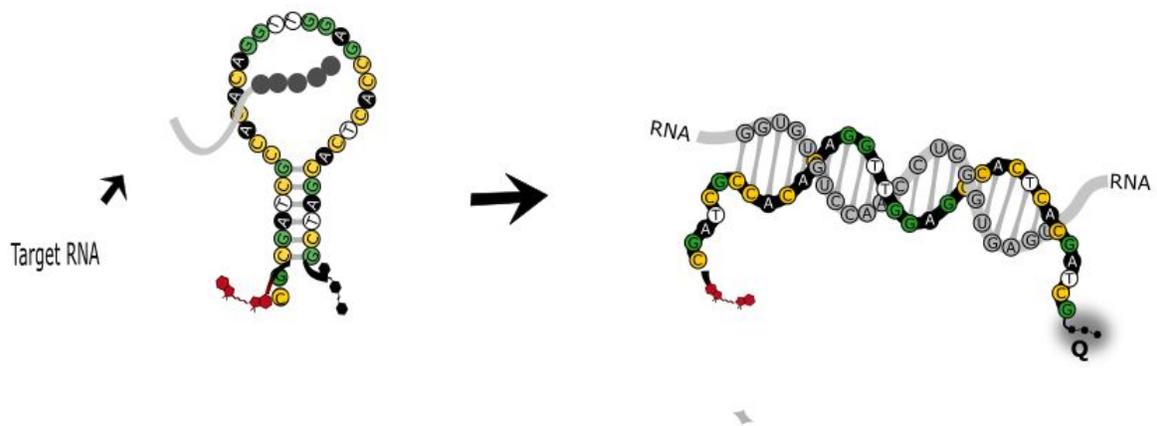

Figure 1: Visual of the target sequence hybridizing to the molecular beacon

To determine the target sequence of the viral RNA, the program ViPR or Virus Pathogen Database and Analysis Resource was used to align the virus chosen to target with other similar viruses [2]. The virus used in this project was West Nile Virus, a member of the Flaviviridae family of viruses. The alignment is important, because the beacon needs to be specific to only West Nile Virus. It is necessary to find a sequence that is conserved between all West Nile Viruses, but not conserved amongst the other members of the Flaviviridae family. The picture shown below gives an idea of how ViPR was used to find a target sequence that is specific to West Nile Virus. Areas that are blank are the areas of conservation to the reference sequence, which is the West Nile Virus used for the experiment.
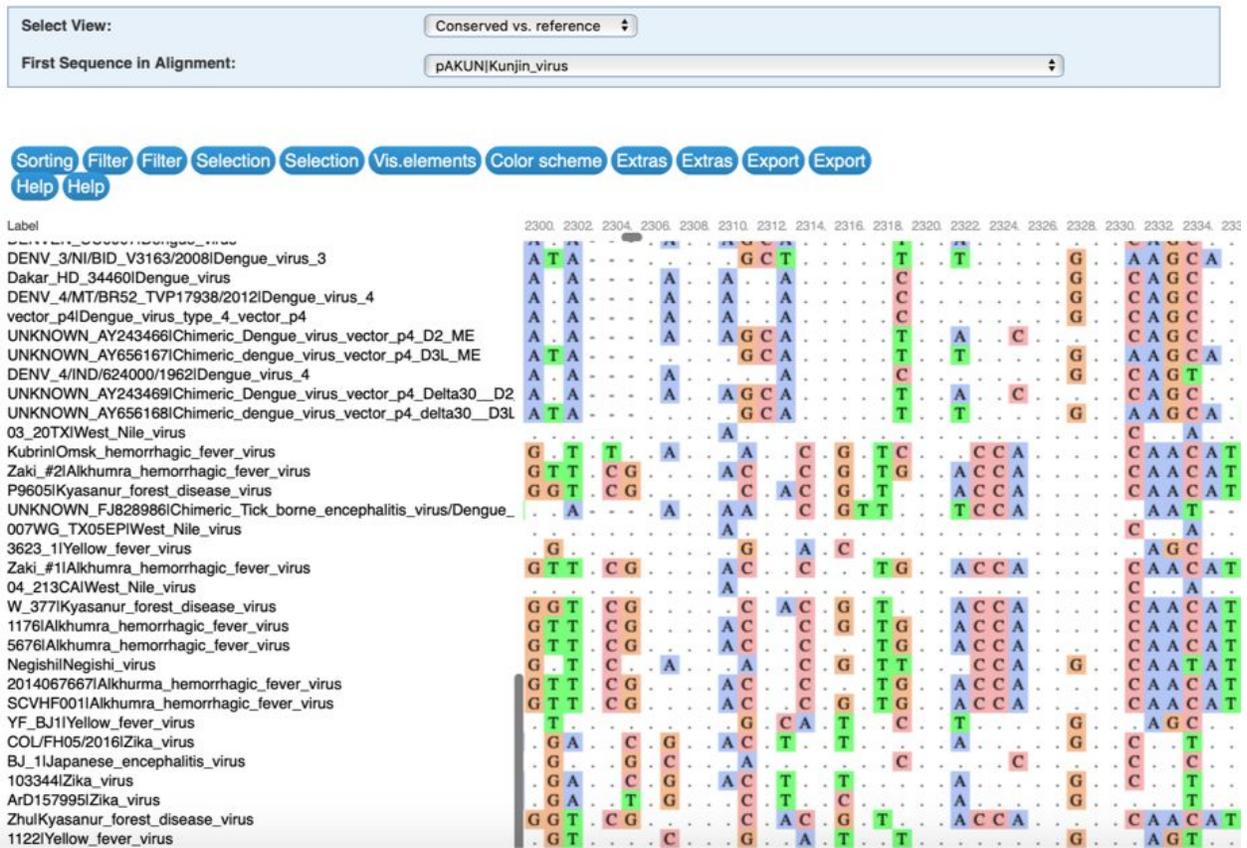
Figure 2: Screenshot of virus alignment on ViPR

The area of the viral RNA that was searched for conservation was the envelope region of the virus. This area is specific to the West Nile Virus, which will allow a beacon to be designed specific to West Nile Virus and nothing else. Also, there is little secondary structure that forms in the envelope region. After searching the envelope region for areas that are unique to West Nile Virus, eight sequences were identified as potential targets. The program mFold was used to analyze the secondary structure of the envelope region and the potential targets were highlighted to make it easy to see if secondary structure would create issues with the binding of the beacon [3]. Below shows examples of mFold predicting secondary structure with target sequences highlighted.
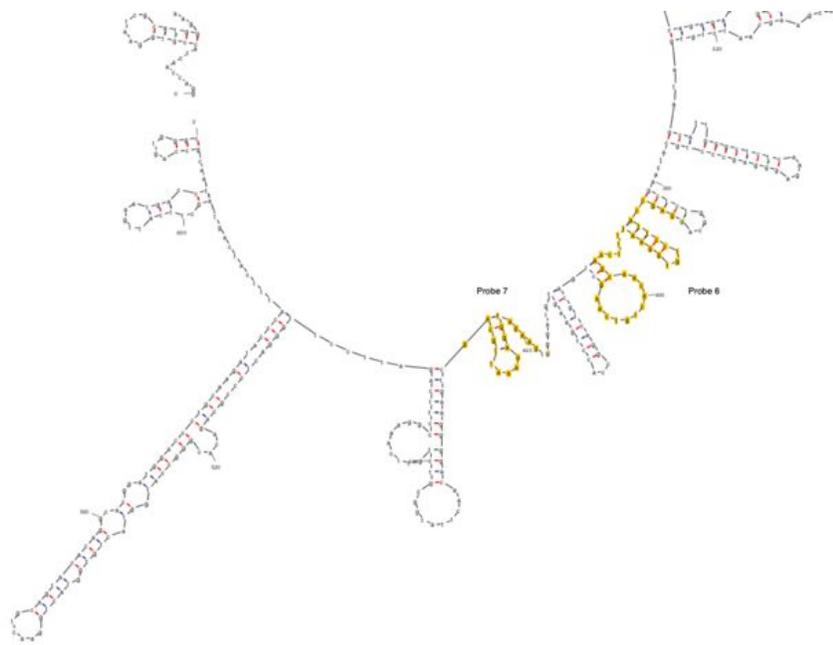
Figure 3: *mFold showing secondary structure for target sequences 6 and 7 (yellow)*
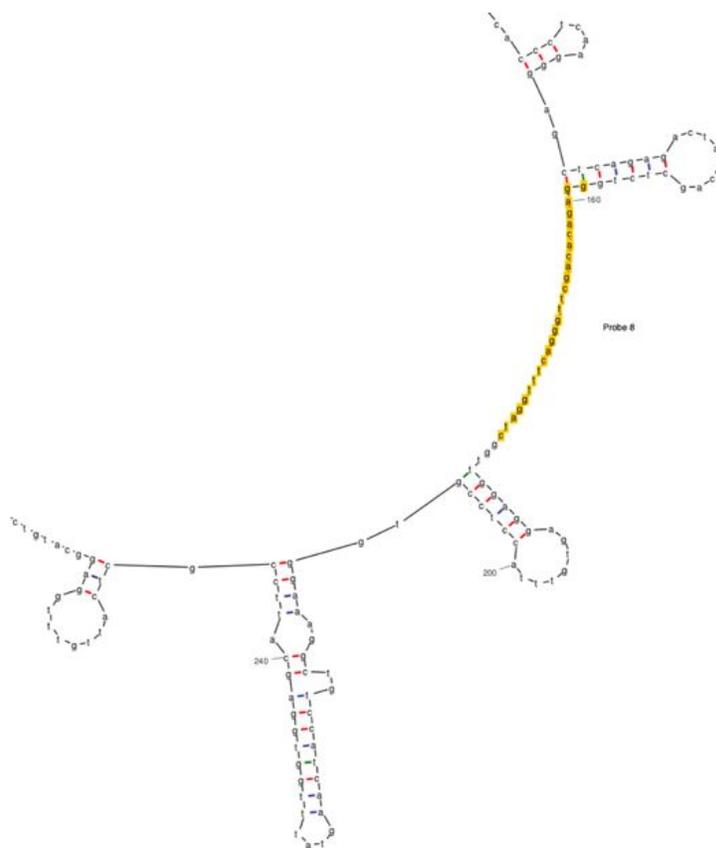
Figure 4: *mFold showing secondary structure for target sequences 8 (yellow)*

NUPACK was the last software tool used for analysis of the molecular beacon. First, NUPACK was used to design the beacon by inputting code designing a beacon that has a complementary sequence to the target area and designing how the bases should pair. NUPACK then will output a preview design based on the input information. An analysis is then ran on the design by looking for other secondary structures that can form and the probability of finding each base in that position [4]. NUPACK will also calculate the minimum free energy and the melting profile for the design [4]. This was the most important tool for deciding which design to choose. Below shows an example of the script used for a candidate design on NUPACK and the output structure.

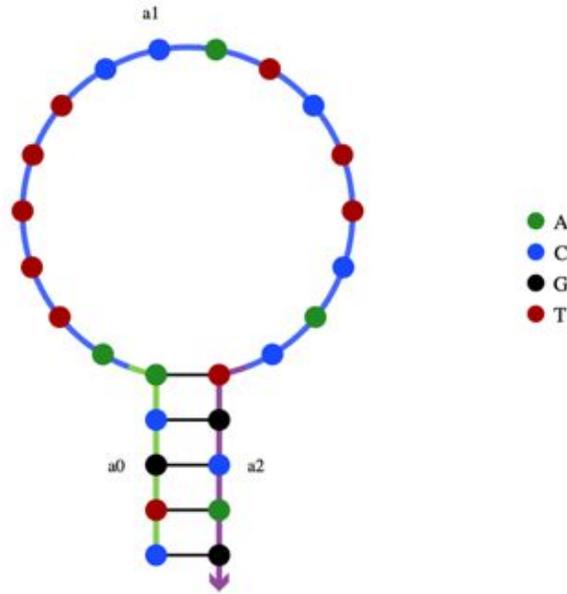*Figure 5: Example script in NUPACK used to design and analyse possible sequences*

**Figure 6:** *Output design from example script (figure 5)*

Results

Thirty-two design combinations were tested in NUPACK after finding eight potential target sequences for West Nile Virus: the beacon with the target sequence being a part of the stem (a), the beacon with a random combination of five base pairs in the stem (b), and these two combinations paired with the target sequence. The table below shows the data collected to help determine the best designed beacon to test.

Target Sequence 5'-3'	Reverse Compliment 5'-3'	# of base pairs for beacon	Delta G of beacon	Tm of beacon (C)	Tm of Pair (C)
1a. cttggaatgagcaacagaga (976-995)	tctctgttgctcattccaag	25	-5.16	64	62
1b. " "	" "	30	-6.05	75	71
2a. ggagcaacatgggtgga (1015-1031)	tccacccatgttgctcc	22	-3.1	62	73
2b. " "	" "	27	-5.27	73	73
3a. accatcgatgtgaagatgatgaa (1084-1106)	ttcatcatcttcacatcgatggt	28	-3.09	46	74
3b. " "	" "	33	-8.26	75	74
4a. gacacatgcgccaaatttgcctg (1306-1328)	caggcaaatttggcgcatgtgtc	28	-4.88	75	83
4b. " "	" "	33	-6.57	70	83
5a. acagtggactgtgaaccacg (1525-1544)	cgtggttcacagtccactgt	25	-7.12	64	77
5b. " "	" "	30	-9.42	69	75
6a. ggagccattcctgtggaattttcaagcaacactgtcaag (1768-1806)	cttgacagtgttgcttgaaaattccacaggaatggctcc	44	-4.18	31	88
6b. " "	" "	49	-6.79	55	85
7a. gtgaagatggaaaaattgcag (1834-1854)	ctgcaatttttccatcttcac	26	-2.89	58	69
7b. " "	" "	31	-5.01	70	70
8a. ggagacacagcttgggactttggatc (2230-2255)	gatccaaagtcccaagctgtgtctcc	31	-2.9	42	82
8b. " "	" "	36	-8.01	75	81

Table 1: Target sequences for designs and data from NUPACK

The ideal candidate beacon would have a higher minimum free energy and a gap between the melting temperature of the beacon and the melting temperature of when its paired to the target sequence. A trend showed when using five random base pairs as the stem, it increased the amount of secondary structure forming and lowered the minimum free energy due to the increase in stability from

hybridization. Most designs had a lot of secondary structure forming in the unpaired loop and the target sequence also formed a lot of secondary structure not making ideal designs. From table 1 and looking at secondary structure of the target in mFold (Figures 3 and 4) , sequences 7a and 8a were chosen as the top two designs. This was due to both designs having higher gibbs free energy while still maintaining and good melting profile. mFold also show less secondary structure forming in the target areas, making it more accessible. Below are images from NUPACK of the probability for structure formation of the top two designs.
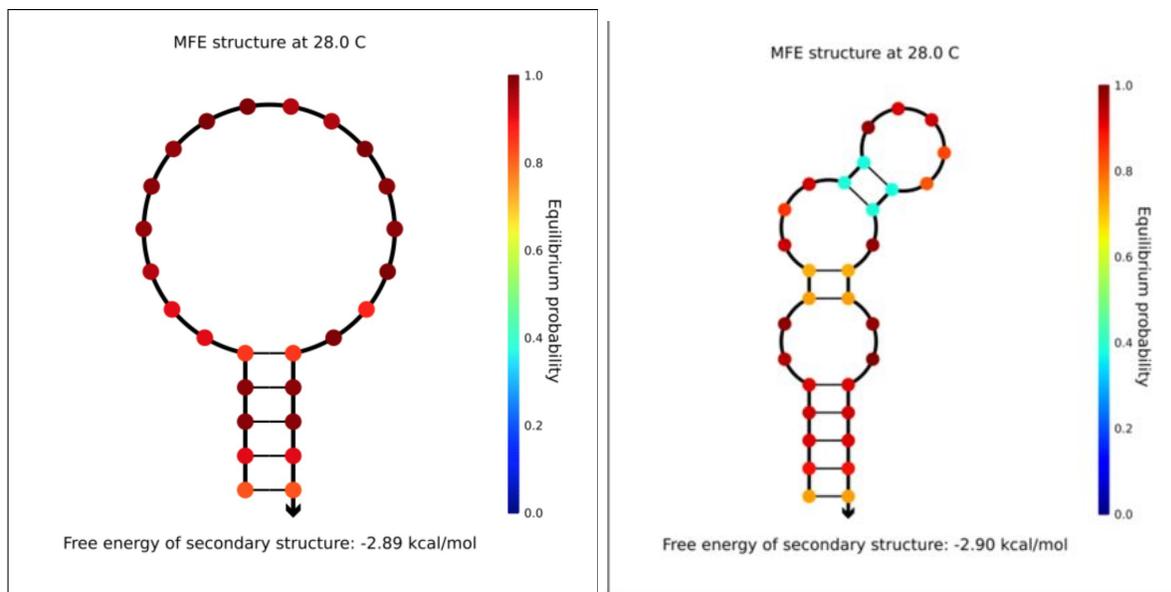

Figure 7: NUPACK probability analysis for design 7a. (left) and 8a. (right)

Design 7a (left) has the highest probability of forming the beacon structure without secondary structure forming in the unpaired region. When looking at mFold, target 7 forms some secondary structure (Figure 3), but target 8 does not form secondary structure making it more accessible for the beacon (Figure 4). Since sequence 7a had the highest probability for folding correctly, this design was chosen to test experimentally.

Once the design sequence was chosen, IDT or Integrated DNA Technologies was used to order the molecular beacon. IDT allowed for easy modification and ordering of the molecular beacon [5]. A 5' thiol and 5' fluorescein was added to the 5' end of the design sequence and a 3' Dabcyl was added to the 3' end of the sequence.

While experiments were being done on the designed beacon, molecular dynamic simulations were run using Python scripts (oxDNA) developed by prior Colorado State University research. The MD simulations were used to test the folding of the designed beacon into the predicted configuration and to show the hybridization of the molecular beacon with the target sequence. The simulations were able to show the beacon folded into the correct structure and the target strand was able to hybridize to the designed beacon (Figure 9 and 10).
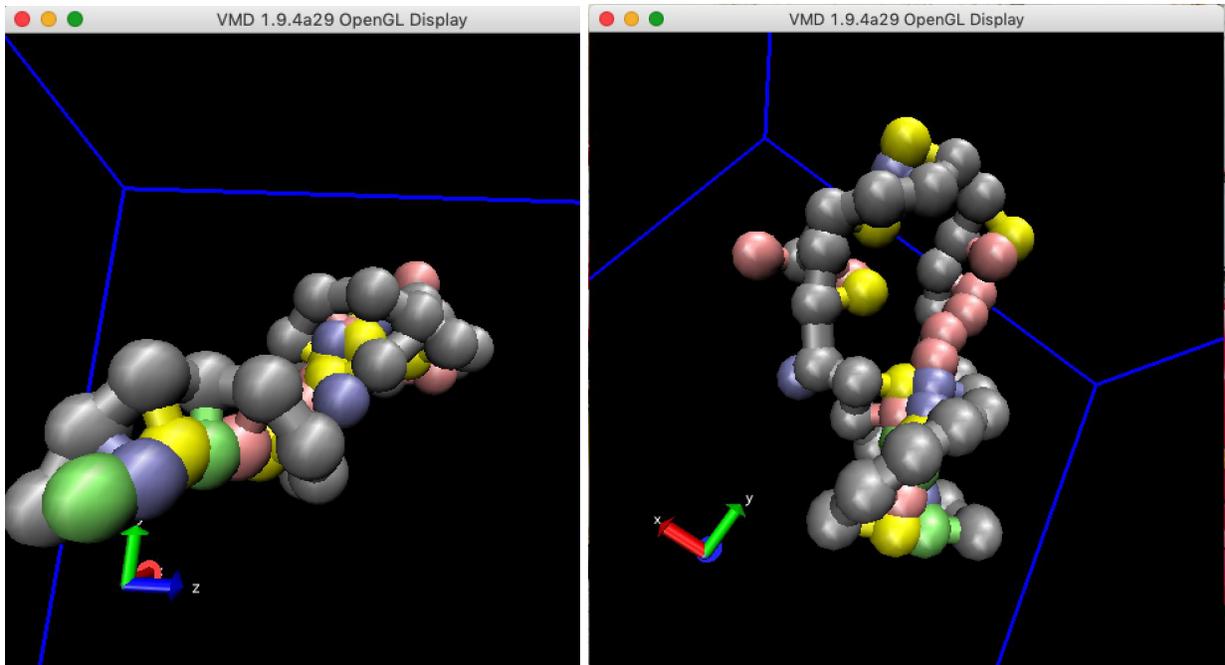
Figure 9: *Molecular Dynamic simulation of molecular beacon folding into target structure*
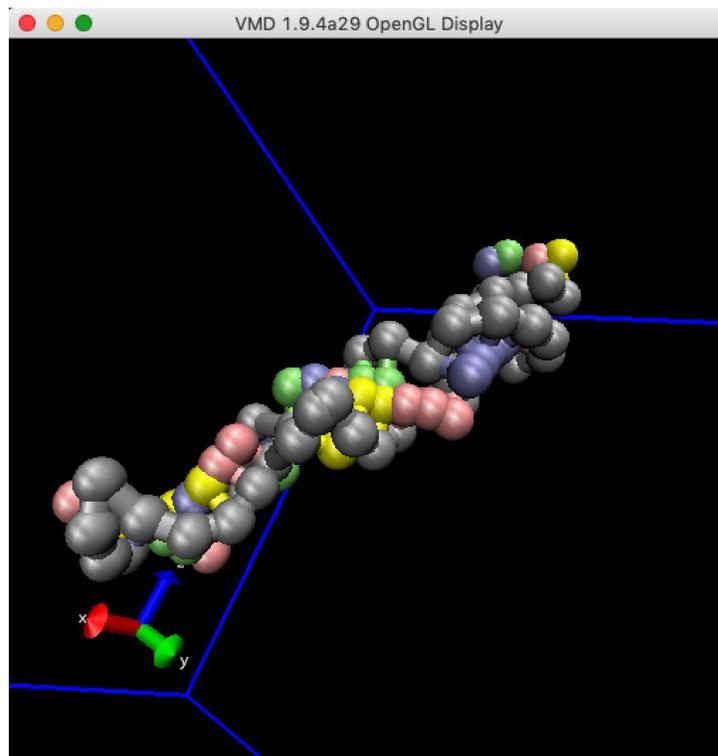
Figure 10: *Molecular Dynamic simulation of molecular beacon hybridizing with target sequence*

Preliminary lab tests show that the designed beacon was able to load and concentrate into the protein crystal. Data also shows the fluorescence signal intensifies upon the addition of the target sequence. See experimental page for an overview of experiments done and results.

Conclusion

Protein crystals with a fluorescently tagged beacon that intensifies in the presence of a specific virus can allow for the development of a noninvasive biosensor for detection of different viruses. Small protein crystals can be concentrated in a small area, giving an amplified signal once the complementary target is present. ViPR, mFold, and NUPACK all became helpful tools for the design of a molecular beacon. With these methods along with the experimental testing, the BIOMOD research team at Colorado State University can continue to develop new biosensors for a wide variety of common viruses.

Future Direction

To continue this project, further experiments and molecular design simulation need to be done for our first design to better show design feasibility. After, more beacon designs can be made for other viruses following the methods presented here. Once more beacons are designed and tested, the team wants to research into ways to concentrate the loaded crystals into a small area for testing of sample viruses.

References

1. **Molecular Beacons:** http://www.molecular-beacons.org
2. **ViPR:** https://www.viprbrc.org/brc/home.spg?decorator=vipr
3. **mFold:** or UNAFold: http://unafold.rna.albany.edu/?q=mfold
4. **NUPACK**: nucleic acid package**:** http://nupack.org/partition/new
5. **IDT:** https://www.idtdna.com/pages/products/qpcr-and-pcr/custom-probes/molecular-beacons