Thermodynamic Design Principles for Double-Strand Displacement microRNA Biosensors

By: Natasha Smith

A project submitted to
Oregon State University
University Honors College,

In partial fulfillment of
the requirements for the
degree of:

Honors Baccalaureate of Science in Biochemistry and Biophysics
(Honors Associate)

Presented May 9 2016
Commencement June 2016
AN ABSTRACT OF THE THESIS OF

Natasha Smith for the degree of Honors Baccalaureate of Science in Biochemistry and Biophysics presented on May 9, 2016. Title: Thermodynamic Design Principles for Double-Strand Displacement microRNA Biosensors.

Abstract approved: _____________________________________________________

Sean M. Burrows

In recent years, microRNA’s (miRNA’s) have emerged as important in regulating gene expression. Fluctuations in their levels are associated with the up or down-regulation of tumor-suppressor genes and oncogenes. Thus, the ability to visualize changes in miRNA is an important step in understanding cancer emergence and development. The Burrows group has developed a novel double-strand displacement biosensor with high sensitivity that mitigates false signals caused by endonuclease degradation. The selectivity of the biosensor for a particular miRNA is determined by the nucleic acid sequence of the reporter. This work is focused on the development of a method to select a reporter that will yield the best selectivity. In particular, I’ve focused on selecting a reporter for miR-29b-1-5p. Variations in the expression levels of this miRNA are associated with lung cancer, cervical cancer, and lymphocytic leukemia. The principles outlined in this work can be extended to select a reporter for any miRNA utilizing the double-strand displacement biosensor design.

Key Words: Biosensor, double-strand displacement, reporter-probe complex, probe-analyte complex, hairpin conformation, quenching, micro-ribonucleic acid (miRNA), reporter selection, miR29b-1-5p

Corresponding e-mail address: smithna2@oregonstate.edu
Thermodynamic Design Principles for Double-Strand Displacement microRNA Biosensors

By: Natasha Smith

A project submitted to
Oregon State University
University Honors College,

In partial fulfillment of
the requirements for the
degree of:

Honors Baccalaureate of Science in Biochemistry and Biophysics
(Honors Associate)

Presented May 9 2016
Commencement June 2016
Honors Baccalaureate of Science in Biochemistry and Biophysics project of Natasha Smith presented on May 9th, 2016.

APPROVED:

Sean M. Burrows, Mentor, representing Department of Chemistry

David Hendrix, Committee Member, representing Department of Biochemistry and Biophysics, School of Electrical Engineering and Computer Science

Lulu Zhang, Committee Member, representing Department of Chemistry

Toni Doolen, Dean, University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, University Honors College. My signature below authorizes release of my project to any reader upon request.

Natasha Smith, Author
Acknowledgments:

First and foremost, I would like to thank the incredible Burrows team for their support and encouragement throughout this entire project. To Sean Burrows, thank you for being a wonderful mentor and for everything you have taught me. Your support means the world to me. A big thank you also goes to Corrine Brucks, without whom, I could not have done this. Thank you for teaching me and always being there to help. Thank you to my committee members, David Hendrix and Lulu Zhang for all your work and advice. A special thank you to Nicholas Larkey and Kyle Almlie for all their help and for making the Burrows lab such a wonderful place to be. I would also like to thank the DeLoach Work Scholarship for their aid in funding this research. I am truly fortunate to have had the support of so many exceptional people.
Outline:

Introduction ...........................................................................................................................................................................1

Materials and Methods ..........................................................................................................................................................3

Results and Discussion:

MATLAB Generation and Filtering ..............................................................................................................................6

Thermodynamic Analysis ..................................................................................................................................................8

Molar Fraction Analysis ...................................................................................................................................................15

Sequence Analysis ...........................................................................................................................................................16

Hairpin Probabilities .......................................................................................................................................................18

R19 Performance .............................................................................................................................................................19

Base-Pair Change Comparison ......................................................................................................................................20

Ideal Non-Complementary Length ...............................................................................................................................20

Conclusion .........................................................................................................................................................................21

Current and Future Directions .......................................................................................................................................21

References .........................................................................................................................................................................23
Introduction:

With the advancement of medicine and a corresponding increase in life expectancy there remains a persistent threat to our health: cancer. According to the CDC, cancer is responsible for the deaths of more than a half of a million Americans every year. Looked at another way, this is 1,596 people dying every day (according to 2012 statistics). At that rate, one in four American deaths are caused by cancer.

Part of the reason cancer has such a significant impact is because it is poorly understood. This is partially due to a lack of robust and proper analytical tools to examine it on a molecular level. Tools to examine proteins have been known for decades, but only within the last ten to twenty years has the importance of micro-ribonucleic acids (miRNA’s) in cancer regulation become apparent. MiRNA’s are single stranded ribonucleic acids that only contain 21-24 nucleotide bases. Consequently, few tools exist to visualize these small non-coding RNA’s.

In the cell, miRNA is synthesized in the nucleus and exported to the cytoplasm. Once in the cytoplasm it associates with a RISC complex. The miRNA-RISC complex will then hybridize with either full or partial complementarity to messenger-RNA (mRNA) and disrupt protein synthesis. Under normal cellular function these miRNA help regulate the life cycle mRNA expression.

However, disease associated changes in a cell’s microRNA expression pattern can regulate the progression of certain cancers and many other diseases. There are two categories of genes associated with cancer: oncogenes and tumor suppressor genes. Oncogenes cause cancer when activated at the wrong time, whereas tumor suppressor cause cancer when inactivated during the normal cell cycle. MiRNA’s target both oncogenes and tumor suppressor genes. If the cancer lowers miRNA’s that target oncogenes, then the cell will grow uncontrollably. On the other hand, if the cancer increases miRNA’s that target tumor suppressor genes, then the cell will not know to stop growing.

The characteristic combinations of miRNA’s exhibited by a cell are called profiles. Successful detection of these miRNA profiles will allow earlier, more precise, and more accurate diagnosis of disease. Additionally, the ability to detect miRNA profiles will advance personalized medicine and contribute to our understanding of the etiology of cancer.

There are many miRNA detection methods, but few are applicable to cells and tissue. Thus, in order to visualize miRNA’s before a cell becomes cancerous and throughout the progression of cancer there needs to be a biosensor capable of detecting miRNA in a living environment. One of the greatest challenges of this arises in detecting the low levels, often nanomolar to femtomolar concentrations, of miRNA in the cells. The biosensor must be capable of generating a signal that can be resolved from the background. Unfortunately, current biosensors bury the signal in the background because biosensor degradation by endonucleases gives rise to false
Furthermore, current in-situ analysis is slow and often fails to distinguish between similar types of miRNA. \(^5\)

*Burrows et al.* recently developed an innovative biosensor that reduced false signals by 20% - 50%. \(^5\) Using a transduction mechanism that brought two dyes together in order for a signal change to be observed mitigated false signals. Selectivity was achieved through the double strand displacement biosensor. As a result, a signal change occurs only when the reporter is displaced and folds into a hairpin as the probe binds to the analyte miRNA. See Figure 1 for a pictorial depiction of the biosensor mechanism.

![Figure 1](image)

**Figure 1**: The double-strand displacement biosensor design utilizes a reporter molecule that dissociates from the biosensor and folds into a hairpin. When this occurs, two fluorescent dyes are brought into proximity and the signal is quenched. A non-complementary region on the reporter is crucial to driving the reaction to the right in the presence of miRNA. Figure adapted from Larkey et al. \(^5\)

An important aspect of the double strand displacement biosensor is the non-complementary region of the reporter. The system is designed so that the probe is only partially complementary to the reporter, but is fully complementary to the miRNA. This difference in stability between the reporter-probe and probe-miRNA drives the reaction to the right.

Alternatively, the mechanism can be described by the following equation:

\[
\text{Reporter-Probe} + \text{miRNA} \rightarrow \text{Probe-miRNA} + \text{Reporter Hairpin} \hspace{1cm} (\text{Equation 1})
\]

\((\text{Signal On}) \hspace{1cm} \text{(Signal Off)}\)

The reaction of reporter-probe with miRNA must be highly favorable as the system moves from reporter-probe to probe-miRNA to ensure complete binding.
(Equation 1). An important constraint of the system is the reporter-probe must be more favorable than the reporter hairpin in order to form the biosensor. Equation 2 describes the formation of the reporter-probe:

\[
\text{Reporter Hairpin} + \text{Probe} \rightarrow \text{Reporter-Probe} \quad \text{(Signal Off)} \rightarrow \text{Signal On}
\]

One challenge lies in selecting the reporter that will yield the best signal change while still maintaining selectivity. The Gibbs energy, enthalpy, and entropy give an indication of the overall thermodynamic forces at work, but they don’t indicate how much reporter-probe will form or how many base pairs will occur in the reporter hairpin, reporter-probe, or probe-miRNA. There is free-ware available to investigate these metrics; thermodynamics helps narrow the pool of potential reporters down to a reasonable number. Furthermore, thermodynamic values only reveal the favorability of a reaction; they do not give any information about the activation barrier.

Therefore, it would be beneficial to develop a systematic method to design biosensors. This research focused on analysis of reporters to create and identify design metrics of thermodynamics, base pairs, molecular structure, and the molar fraction of reporter-probe formed. If successful, the model could easily be extended to select reporters for the detection of any type of miRNA using this biosensor design.

The Burrows lab is currently investigating three miRNA’s with the ultimate goal of simultaneous detection. These miRNA’s are miR-29b-1-5p, miR-26a-2-3p, and miR-27a-5p. Changes in expression levels of all of these miRNA’s are associated with cancer.\(^8\text{–}^{10}\) For the purpose of this work, we sought develop a method to create a biosensor for miR29b-1-5p. Variations in the expression levels of this miRNA are associated with lung cancer, cervical cancer, and lymphocytic leukemia\(^11,10,12\).

**Materials and Methods:**

I generated a set of potential reporters using a MATLAB reporter generation function developed by fellow Burrows group member, Corrine Brucks. First the reporter generation function takes the microRNA sequence of interest as an input to make a probe that is 100% complementary to the analyte miRNA. Next the function uses the probe sequence to generate a set of partially complementary reporters. The user specifies which regions of the probe will be complementary to the reporter. The generation function fills in the corresponding nucleic acid according to Watson-Crick base pairing to generate the complementary sequence. Watson-Crick base-pairings follow the convention that Adenine’s (A’s) are always paired with Thymine’s (T’s) and Cytosine’s (C’s) are always paired with Guanine’s (G’s). For non-complementary regions, the function generates a reporter with every permutation of the non-complementary nucleotide except that with a Watson-Crick base-pair. Thus, if the probe contains an A, the reporter may contain an A, G, or C. The result of this is that
$3^N$ reporters are generated, where $N$ is the number of nucleotides in the non-complementary region(s).

A series of MATLAB filters were applied to reduce the number of potential reporters. The filters eliminated reporters that bound significantly to the analyte as well as off-analytes. For miR29b-1-5p seven filters were applied to remove reporters that unfavorably bound to the probe, analyte, miR26a-2-3p and its probe, and miR27a-5p and its probe.

Reporters that had the fewest undesirable binding interactions were run through The RNA Institute’s DINAMelt application to obtain estimated Gibbs energy, enthalpy, entropy, and melting temperature values. Reporters predicted to form dimers with melting temperatures above 18 °C were eliminated. Additionally, any reporters with off-analyte interactions that had a negative Gibbs energy ($\Delta G < -6$ kcal/mol) were eliminated. The remaining reporters were subjected to a further analysis addressing the number of bound base pairs, molecular structure, and the molar fraction of reporter-probe predicted to form.

Secondary structure prediction software from Rochester’s RNA Structure Web Server was used to predict the probability of hairpin formation. The RNA Institute’s DINAMelt application was used to perform an analysis of the number of bound base pairs in the reporter-probes and reporter hairpins. DINAMelt was also used to predict the molar fraction of reporter-probe formed at the working temperature. Ideally, all of the reporter–probe should be complexed in order to observe a signal change. All DINAMelt calculations were performed at 25C, [Na$^+$]=10 mM, [Mg$^+$]= 2.5 mM, and DNA strand concentration = 1M. For a complete description of how the DINAMelt program works see Zuker paper. This process could be adapted for a different environment simply by changing which off-analytes are being compared and the salt-temperature conditions imputed into DINA-Melt’s mfold.

The Burrows lab designed a few reporters for the miR29b-1-5p biosensor in the past (R40, R43, R55, and R99), but they did not reach the desired signal change within the expected time period. R40, R43, and R99 showed no significant signal change upon addition of miRNA. R55 did show the expected signal changes, but did not do so in the expected time frame of a few hours. These reporters will be discussed throughout the results section as a point of comparison. The naming convention stems from the order in which the reporter sequence was generated by the MATLAB program. R stands for reporter and the number is the index of the reporter in the generated list.

The reporter designed for miR29b-1-5p as a result of my contribution was R19. Two Cyanine (Cy) dyes were attached on opposite ends of the reporter. On the 3-prime end was Cy3 (Absorbance Max: 544 nm, Emission Max, 566 nm) and on the 5-prime end was Cy5 (Absorbance Max: 649 nm, Emission Max, 667 nm). The reporter was suspended in buffer containing 2.5 mM MgCl$_2$, 10 mM Tris buffer, 0.005 % Tween-20, and Phosphate Buffered Saline (PBS). The reporters were characterized by a Mai Tai laser and spectrometer equipped with an Electron-Multiplied CCD camera (EM-
Figure 1, referenced from Larkey, et al., 2014 is a visual of the experimental set up of the custom-built fluorimeter used to run experiments performed in this work. The Mai Tai laser was tuned to 740 nm to induce fluorescence from the reporters. The fluorescence signal was collected using the spectrometer/EMCCD. Here, the grating of the spectrometer was centered at 670 nm and the spectral range collected was 627.6193 to 712.2952 nm. The EMCCD camera recorded 3 frames at 500 ms per frame.

Figure 2: Experimental set up of the custom-built fluorimeter used to acquire fluorescence data. Acronyms defined: Laser= Mai Tai, M1= Mirror #1, M2= Mirror #2, BS= Beam Sampler, PD= Photo diode, LP-DM= 705 nm Long-Pass Dichroic Mirror, L= 25 mm Focal Length Lens, CH= Cuvette Holder, SPF= 720 nm Short-Pass Filter, Obj= Objective (10x, 0.25 NA), FC= Fiber Coupler, SM= Spectrometer, PM= Power Meter and CPU= Central Processing Unit.

The main component of this custom-built fluorimeter is a Titanium-Sapphire (Mai Tai, Spectra Physics, Newport) laser capable of producing 100 femtosecond (fs) long pulses at a rate of 80 MHz. The Mai Tai laser has a tunable wavelength range spanning from 690 nm to 1040 nm. The wavelength was controlled via the Mai Tai GUI program (Spectra Physics, Newport), and the power of the laser was managed by a combination of half-wave plate and Glan-polarizer. Rotation of the half-wave plate enabled precise control of the average power directed towards the sample. The average power was monitored using a beam sampler that reflected about 4% of the beam to the photodiode that was readout on the power meter (Newport, 1918-R). For these experiments the power was 3.00 mW.

Back-scattered laser light from the sample and sample holder was filtered out using long-pass and short-pass filters. A 705 nm long-pass dichroic mirror was used to allow any light above 705 nm to be removed from the fluorescence signal. The dichroic mirror reflects fluorescence signals below 705 nm and directs this light to a 720 short-pass filter. The short-pass filter further rejects back-scattered photons as the dichroic mirror is not perfect. After the filter, the light reaching the fiber-optic and detector is mostly any color less than 720 nm.
The intensity of the signal is proportional to the number of reporters in the open reporter-probe complex plus the number of reporters in the hairpin (Equation 3).

\[ I \propto [\text{Reporters Open}] + [\text{Reporters Closed}] \quad (\text{Equation 3}) \]

**Results and Discussion:**

When designing a reporter for a biosensor, there are two stages: reporter sequence generation and reporter selection. In the first stage, the MATLAB program uses the probe sequence to generate reporters with the desired non-complementary region. The second stage involves a detailed analysis of various parameters including: thermodynamics of the system, analyte and off-analyte interactions, molar fractions of reporter-probe complex present, possible hairpin structures, and the probability of reporter hairpin forming.

*MATLAB Generation and Filtering:*

To begin, the MATLAB program developed by the Burrows group generates a list of potential reporters. The program generates all possible reporter sequences that vary only in the nucleotide composition of the non-complementary region. The length of the non-complementary region(s) depends on the number of nucleotides that are not complementary to the probe. As a result, $3^N$ reporters are generated from the MATLAB program, where $N$ is the number of nucleotides in the non-complementary region(s).

Since the relationship between the number of non-complementary nucleotides and the number of possible reporters is exponential, a large number of possible reporters are often generated. In order to reduce the list of reporters, a series of filter functions are applied. The purpose of a filter is to eliminate reporters whose non-complementary regions have Watson-Crick base pairing with other known oligonucleotides in the sample. These oligonucleotides are termed off-analytes. Each filter operates by comparing the reporter sequence to off-analyte sequences imputed by the user. These off-analytes include other miRNA’s and their respective probes commonly expressed with the miRNA of interest. This helps to choose a reporter with selectivity for the appropriate miRNA. By comparing to other miRNA and their probes, it also allows us to insert multiple reporter-probe biosensors into the same sample without worrying about cross-reactivity, or unintended binding, between reporter-probe biosensors. Any reporters with potential nucleation sites (three or more adjacent Watson-Crick base pairs) are eliminated. The number of filters is dependent on the number of off-analytes the user decides to compare.
A typical set of filters is shown in Figure 3. The resulting list after seven filters is typically 25% of the original list given by the generation function. Thus for a non-complementary region of desired length N, the entire process will take only a few minutes and will give:

$$\text{Final \# reporters produced} \approx 0.25*(3^N) \quad (\text{Equation 4})$$

**Figure 3**: MATLAB Filter Flow Chart. Red boxes represent reporters predicted to have non-complementary regions bind to the comparative strand and are removed from the list of possible reporters. Green boxes represent those reporters that have passed through a filter and do not bind to the comparative strand. The reporters are initially compared against the probe to ensure the non-complementary region does not bind to any region of the probe. Next, the reporters are run against the miRNA of interest to ensure the reporters will not bind to the analyte. Lastly, the reporters are run through a series of filters evaluating their binding with off-analytes. Off-analytes include other miRNA’s and their respective probes typically co-expressed with the miRNA of interest.
A major advantage of using this computational approach to generate and initially filter reporters is it allows thousands of possibilities to be examined within minutes to hours. Before the development of this work, reporter generation was done by hand-picking a random sequence for the non-complementary region. This allowed only a few reporters to be analyzed at a time. The MATLAB program allows us to generate a much larger pool of reporters. Non-computational methods would require a much smaller starting set and might unintentionally overlook a good reporter design.

The above computational method was applied to generate and filter a set of reporters for miR29b-1-5p, a miRNA associated with lung cancer, cervical cancer, and lymphocytic leukemia. An important factor to consider when generating reporters is how long to make the non-complementary section. This is not a simple question as the length of the non-complementary section influences how favorably the system moves from reporter-probe to reporter hairpin and vice versa. In the case of miR29b, three rounds of the MATLAB program were run with N=5, N=7, and N=9. The results are compared and discussed later in this work.

Now that a list of all possible reporters had been generated I proceeded with further refinement and selection of the most ideal reporter for the biosensor. The next stage of refinement involved thermodynamic analysis of all components in the system as it moved from reporter hairpin to reporter-probe and then to probe-analyte (Equations 1 and 2).

**Thermodynamic Analysis:**

There are three types of molecular interactions that need to be analyzed. The first are hairpins where a single strand of DNA folds onto itself. The second are heterodimers that form between two different strands of DNA. The third are homodimers that form between two identical strands of DNA. These types of interactions can occur with reporters, probes, and miRNA. For a given miRNA biosensor, the only heterodimers that should form are the reporter-probe and probe-analyte. The reporter is the only hairpin.

In order for the system in Figure 1 to work properly, several thermodynamic parameters must be met. First, the reporter must bind to the probe for miR-29b-1-5p more favorably than any other possible binding interactions. These interactions include: reporter hairpin formation, the binding between the probe and any off-analytes, reporter homodimers, and the reporter and any other off-analytes. For example, if the reporter hairpin is more stable than the reporter-probe complex, then the reporter-probe complex would never form. Secondly, in order to drive the reaction of reporter-probe plus miRNA to completion, the binding between the probe and the miRNA must be the most favorable reaction of all.
Order of necessary stability of various constituents in solution (most stable to least):
1. Probe—miR29b-1-5p
2. Reporter—Probe       (dG < -15)
3. Reporter Hairpin

Gibbs Free Energy, Enthalpy, Entropy, and melting temperature must be assessed when analyzing stability of the various constituents in the sample matrix. According to the definition of Gibbs Free Energy, a reaction will be spontaneous (i.e. favorable) only if $\Delta G$ is negative. If the Gibbs Energy is negative, but the melting temperature is 2-10 degrees below the working temperature (room temperature or 37°C), then the desired interactions will not likely occur. Conversely, when the melting temperature is above the working temperature, but the Gibbs Energy is positive, the interaction will not likely occur. Further consideration will only be perused when a reporter hairpin or reporter-probe complex is predicted to have a negative Gibbs energy and a melting temperature well above the working temperature.

In addition to ensuring the correct order of stability for the analyte, probe, and reporter interactions, there are a number of off-analyte interactions to take into account during thermodynamic analysis. In the Burrows lab, we are interested in simultaneous detection of miR-29b-1-5p, miR-26a-2-3p, and miR-27a-5p. Thus, when designing a reporter for miR-29b-1-5p we analyzed the thermodynamic interactions between it and potential off-analyte miRNA’s: miR-26-2-3p and miR27a-5p. It was also necessary to consider interactions between the miR29b-1-5p reporter and the other miRNA probes. In addition, homodimers of reporters and probes needed to be evaluated.

The process outlined in this work will present the thermodynamic analysis in the order that results in the most efficient process. This process is depicted in Figure 4. When the number of nucleic acids, $N$, is less than 6, the order in which the thermodynamic analysis is done makes little impact on the efficiency of the process. However, when more nucleic acids comprise the non-complementary region, the large number of reporters generated makes it advantageous to perform the analysis in a particular order, such that the list is reduced as early as possible.
The first step, analysis of reporter dimerization, is a simple way to quickly reduce the list of possible reporters. To do this, find the thermodynamics of each reporter dimer by entering their sequences in DINAMelt’s Two-State Hybridization software. This application can process around 600 reporters at a time and will give changes in Gibbs energy, enthalpy, and entropy, as well as melting temperature values. Any dimers with a melting temperature greater than 18 °C can be eliminated. This temperature was chosen because it is a few degrees below the working temperature of 20-25°C and any reporters below this threshold will not have stable dimers. It is important to eliminate any reporters that have self-binding affinity, because they could significantly distort the analytical performance of the sensor. Ideally, the melting temperature of the reporter homodimer should be at least 10 °C lower than the working temperature, but with longer non-complementary regions (N>8), the reporter homodimers exhibit higher melting temperatures and it is necessary to include these reporters in order to have options remaining at the end of the selection process.

Figure 4: Reporter selection flow chart for N ≥ 6. Steps earlier in the process eliminate a greater number of reporters allowing the large list to be reduced quickly. The temperatures displayed are based on a 25 °C working temperature design. This could be adapted to 37 °C by simply adding 12 °C to each T_m listed.
The reason reporters predicted to form homodimers are removed early is because the binding of two reporters could cause the dyes on different strands to approach and quench or FRET with each other. The effects of this would be dependent on the relative stability of the homodimer complex in comparison to the reporter hairpin and reporter-probe complex and would be difficult to predict. Thus removing potential homodimers is an important step in designing a functional biosensor. Typically there are a large number of reporter homodimers with melting temperatures above 18°C, so this step significantly reduces the number of reporter candidates. If the number of remaining reporters is too large, the results can be further pared down by eliminating homodimers with negative ΔG values and T_m’s between 15 - 18°C. This is beneficial when the length of the non-complementary region is greater than or equal to 9 (N ≥ 9).

The next step is to evaluate the formation of all other reporter-off-analyte complexes through the Two-State Hybridization application. Any reporters that have ΔG values < -6 kcal/mol and a T_m > 15 °C are eliminated. These parameters are set to mitigate off-analyte binding at the working temperature.

Once reporters with off-analyte and dimer interactions have been eliminated, the predicted structures of the hairpins are analyzed. The formation of the proper hairpin structure is essential for the performance of the sensor since the signal change is dependent on the two dyes coming together in the hairpin (Figure 1). A reporter may have multiple hairpins that fall into one of two categories which we have called ideal hairpins (Figure 5a) and non-ideal hairpins (Figure 5b). Typically a reporter sequence has one to five possible hairpins. Ideal hairpins have a secondary structure where the 5’ and 3’ ends lie next to each other. This brings the Cy5 and Cy3 dyes within the desired proximity of one another for a signal change. In contrast, non-ideal hairpins do not

![Figure 5: The structure of an ideal hairpin (a) and non-ideal hairpin (b). Ideal hairpins have the 5’ and 3’ ends next to each other to facilitate signal change. The 3’ and 5’ ends of a non-ideal hairpin are separated and show poor signal change.](image-url)
have the 5’ and 3’ ends next to each other and often contain one or more internal hairpins (Figure 5). When this occurs the dyes are not in proximity to one another and the signal change observed is poor.

DINAMelt’s Quickfold application was used to predict the number and structure of potential reporter hairpins. After running each reporter through Quickfold, there are two important pieces of information to examine. The first is the number of hairpins. When dealing with a large number of possible reporters, it is simplest to eliminate all reporters with more than one predicted hairpin. In our studies, this eliminated about 60% of the reporters. It is important to have one predicted hairpin because this ensures the reporter will close up on itself, successfully quenching the dyes. The second piece of critical information is whether the predicted structure is an ideal hairpin or a non-ideal hairpin. This can be determined by analyzing the secondary structure generated by Quickfold. All reporters with a non-ideal hairpin should be eliminated. However, sometimes there is a shortage of reporters with a single hairpin. In this case it is good to include some reporters with multiple hairpins as long as the hairpins are ideal hairpins. Reporters with non-ideal hairpins should not be considered because the dyes will not be brought into proximity upon hairpin closing. Based on my experiences, I found that it is best to have at least ten reporter possibilities remaining at this point or there is a chance that subsequent steps will eliminate all the possible reporters.

Once the number of reporters have been narrowed down, it is time to analyze the reporters hairpin melting temperatures relative to the reporter-probe complex. Recall the reporter-probe complex should be more stable than the reporter hairpin so that the reaction moves easily to the right (Equation 2). To reflect this relative stability the melting temperature of the hairpin should be lower than the melting temperature of the reporter-probe complex. However, it cannot be too low, as it still needs to be stable in solution so the signal change operates correctly. Based on experimental findings of previous biosensor designs the Tm of the reporter hairpin should be 10 ± 5 °C < Tm of the reporter-probe complex for the best results. For most reporter-probe complexes the melting temperature is around 40 - 60°C depending on the length of the non-complementary region. A shorter non-complementary region gives a higher melting temperature due to the increased number of paired complementary bases.

Thus, the melting temperature of the reporter hairpin should be in the range of 25-55 °C. The temperature parameters will depend on whether the application is intended to be at room temperature or 37 °C. For optimal performance at body temperature it is best to keep the reporter hairpin Tm above 40 °C and the reporter-probe Tm above 50°C.

At this point, the potential reporters have been cross-checked with all off-analytes for potential unfavorable interactions, analyzed for potential to form homodimers, optimized for one ideal hairpin, and fall in the appropriate melting temperature range. The next step is to analyze the thermodynamics of the entire system as it moves from reporter-probe to probe-miRNA. See Equation 5:
\[ \Delta G_{\text{Probe-Analyte}} - \Delta G_{\text{Reporter-Probe}} = \Delta G_{\text{system}} \]  

(Equation 5)

The same equation applies for \( \Delta H_{\text{system}}, \Delta S_{\text{system}}, \) and \( \Delta T_m\text{ system}. \) The next goal is to determine what thermodynamic values are optimal. In order for the reaction to be spontaneous as it moves from reporter-probe to probe-miRNA, the Gibbs free energy must be negative. Since the probe-miRNA complex should be more stable than the reporter-probe complex, a positive \( \Delta T_m\text{ system} \) would be expected. Additionally, the change in enthalpy and entropy will influence the dynamics of the system according to the following equation.

\[ \Delta G = \Delta H - T\Delta S \]  

(Equation 6)

The \( \Delta H_{\text{system}} \) is expected to be negative as the system moves to a more stable state and energy is released. As the system moves from a partially complementary reporter-probe complex to a fully complementary, and thus more ordered, probe-miRNA complex we can also expect a negative \( \Delta S_{\text{system}}. \) Based on Equation 6 it is clear the temperature of the system is an important factor to consider. At high temperatures, the entropic factors will outweigh the enthalpic factors and the reaction will be unfavorable with a positive \( \Delta G_{\text{system}}. \)

To pinpoint the appropriate ranges for each thermodynamic parameter, I analyzed the thermodynamic values of other reporters the Burrows lab had already investigated. These reporters were designed prior to the design process described here. The design process outlined here was formulated based on observations from successful and unsuccessful reporters. A summary of the thermodynamic properties of each reporter can be seen in Figure 6.

R4 for Let7a and R25 for miR-26a-2-3p were used as a point of reference. These reporters were selected because they performed well and gave a significant signal change within a few hours of miRNA analyte addition. R40, R43, and R99 for miR-29b-1-5p performed poorly with little change in signal upon binding their respective analytes. R55 showed signal reduction upon addition of the miRNA, but proceeded too slowly.
Figure 6: A thermodynamic comparison of reporters previously tested by the Burrows group as the system moves from reporter-probe to probe-miRNA.

Close examination of each figure indicates that a more negative $\Delta H$ and $\Delta G$ does not indicate better performance (R40, R43, and R99). However, the $\Delta H$ and $\Delta G$ must be sufficiently negative or the reaction proceeds as seen in R55. Likewise, a more positive $-T\Delta S$ and $\Delta T_m$ are not indicators of reporter success. We also see from R55 that too low of a $-T\Delta S$ and $\Delta T_m$’s results in poor performance. In fact, the two successful reporters, R4 and R25 have thermodynamic values that lie between those of R40, R43, R99 and R55. This data indicates there is an optimum zone for thermodynamic values that lies somewhere between the following ranges:

- $-83.9 < \Delta H < -24.9$ (kcal/mol)
- $19.8 < -T\Delta S < 69.2$ (kcal/mol)
- $-14.6 < \Delta G < -5.1$ (kcal/mol)
- $6.1 < \Delta T_m < 20.7$ (°C)

Based on this data, reporters with RP to PA values that fell between these ranges were selected for, but reporters that did not fulfill these criteria were eliminated. In the case of miR29b-1-5p this left three reporters of particular interest: R8, R19, and R244. R8 and R19 have non-complementary regions of length 7, whereas R244 has a non-complementary region of length 9.
Figure 7: Thermodynamic comparison of miR-29b-1-5p reporters as the system moves from reporter-probe to probe-analyte. In order to ensure the reaction proceeds from reporter-probe to probe-analyte, the magnitude of each thermodynamic parameter should be larger than those of R55. Analysis of R244, R8, and R19 thermodynamics shows this is the case. R4-Let-7a and R25-miR-26a are reporters designed previously by the Burrows lab that worked well. The thermodynamic values of R244, R8, and R19 are near those of R4-Let7a and R25-miR26a, indicating they have potential to work well.

Molar Fraction Analysis:

Once the number of possible reporters has been narrowed down a molar fraction analysis can be performed. Each reporter must be run individually and so it is advantageous to have only 3-5 reporters to analyze at this point. The DINAMelt Hybridization of Two Different Strands application is used to predict the molar fraction of reporter-probe complexed at operation temperature. Ideally, 100% of the reporter-probe should be complexed in order to yield the greatest signal change. In the case of miR-29b-1-5p three reporters were submitted to molar fraction analysis: R8, R19, and R244.
Figure 8: The molar fraction as a function of temperature for (a) R244, (b) R8, and (c) R19. R244 shows only 0.84 of the reporter and probe are complexed at 25 °C. At body temperature (37 °C) only about 0.54 of R244 binds its probe. R8 and R19 show near 1.00 association at room temperature and around 0.97 association at body temperature. This indicates R8 and R19 would both be good reporters.

As can be seen by the plots above in Figure 8, the fraction of reporter and probe bound together at 25 °C for R244 was only 0.84. At body temperature, this fraction drops even further to 0.54. Ideally, the molar fraction should be near 1.00 at the working temperature. The reason for the rapid drop in molar fraction of reporter-probe as a function of temperature for R244 is unclear. It may have to do with the fact that the non-complementary region is 9 nucleotides, making the reporter-probe complex less stable. Based on this information, R244 can be eliminated as a potential reporter.

On the other hand, R8 and R19 have nearly optimal molar fractions (0.99) at room temperature (25 °C) and reporter-probe molar fractions of 0.97 at body temperature (37 °C). R8 and R19 appear to be fairly similar up to body temperature, but R8 falls off slightly faster as temperature increases further. However, the differences in slopes was slight (0.001) and not reason enough to choose one reporter over the other. Further analysis was needed to choose the best reporter.

Sequence Analysis:

To further investigate the differences between R244, R8, and R19 I examined the sequences themselves (Figure 9). The region complementary to the probe can be seen in the conserved, blue region, whereas the non-complementary region is variable. Recall that R55 is a reporter for miR-29b-1-5p that did show signal change, but was too slow. Here it is used as a comparison point. One interesting thing to note is the new reporters have a higher A and T content than R55. Since G/C’s form three hydrogen bonds, in comparison to A/T’s two, the high G/C content of R55 could be contributing to the slow reaction by making it harder for the miRNA to displace the reporter. If this is the case we would expect a higher A/T content in the non-complementary region (as in R8, R19, and R244) will make it easier for the miRNA to displace the reporter. The other thing to note is that R244 has a longer non-
complementary region. This may be contributing to the instability of the reporter-probe complex relative to R8 and R19 as seen in Figure 8.

<table>
<thead>
<tr>
<th>Stem</th>
<th>Conserved Region</th>
<th>Variable Region</th>
<th>Stem</th>
</tr>
</thead>
<tbody>
<tr>
<td>R55</td>
<td>5’- CGATG</td>
<td>GCTGGTTTCATAT G G T G T C G C G</td>
<td>CATCG - 3’</td>
</tr>
<tr>
<td>R8</td>
<td>5’- CGATG</td>
<td>GCTGGTTTCATAT G G G A T A A A T</td>
<td>CATCG - 3’</td>
</tr>
<tr>
<td>R19</td>
<td>5’- CGATG</td>
<td>GCTGGTTTCATAT G G G T T A A A T</td>
<td>CATCG - 3’</td>
</tr>
<tr>
<td>R244</td>
<td>5’- CGATG</td>
<td>GCTGGTTTCATAT C T G A A G A A T</td>
<td>CATCG - 3’</td>
</tr>
</tbody>
</table>

**Figure 9**: Sequence alignment for possible reporters for the miR29-b-1-5p biosensor. R55 which was tested previously and proceeded forward slowly is compared to the three new possibilities: R8, R19, and R244. The green regions are the reporter stems that bind upon hairpin formation. The light blue indicates that all four reporters share that nucleotide. Yellow indicates that the three new reporters share a nucleotide different than that of R55. Orange indicates two of the new reporters share a nucleotide. Red nucleotides are unique to that reporter.
Hairpin Probabilities:

The next factor to consider is the probability of each individual hairpin forming. Below are the secondary structure outputs from Rochester’s RNA Structure Web Server.

![Diagram of hairpin structures](image)

**Figure 10:** The hairpin probability maps of (a) R55 (b) R8, and (c) R19 generated by the RNA Structure Web Server. The probability (p) of each nucleotide as bound or free is indicated by the following colors: red (p ≥ 99%), orange (99% > p ≥ 95%), yellow (99% > p ≥ 95%), dark green (90% > p ≥ 80%), bright green (80% > p ≥ 70%), light blue (70% > p ≥ 60%), dark blue (60% > p ≥ 50%), and purple (p ≥ 50%). Warm colors (reds) indicate a high probability, whereas cool colors (purples) indicate a low probability.

The probability of a given nucleotide being paired or free is depicted by color. Warm colors (reds) indicate high probability and cool colors (purples) indicate low probability. There is a clear difference between the binding probabilities of the stem on R55 as compared to the new reporters: R8 and R19. R55’s stem has a lower probability of binding. The tight binding of this region is crucial in order for the Cy3 and Cy5 dyes on the 3’ and 5’ ends to come into proximity and initiate a signal change. This indicates that R8 and R19 will have a higher probability of bringing the dyes together.

The other major difference between R55 and the newer reporters is the second binding region. In R55 there are only two nucleotides predicted to bind with only a > 60% probability. In contrast, R8 and R19 have 3 bound nucleotides with a higher probability of binding. This secondary binding region could play a role in stabilizing the stem region further.

R8 and R19 are very similar and differ by only 1 nucleotide (red box in Figures 10 (b) and (c)), thus their secondary structures are very similar. There are two regions in which they differ: the stem and the loop. In truth, because they are so similar it is...
unlikely one is significantly better than the other. We decided to order and test only R19 because lab resources required testing one reporter at a time, but we expect R8 to work just as well and it will be tested in future studies.

**R19 Performance:**

To determine the functionality of R19, it was bound to probe and the intensity measured in the absence and presence of miR-29b-1-5p. The results of this and how it compares to previous reporters can be seen in Figure 11.

![Figure 11: The reporter intensity as a hairpin, as a reporter-probe complex, and in the presence of miRNA. The newest reporter design, R19, is compared to previously tested reporters. R100 showed little signal change upon addition of the probe and miRNA. R55 did respond to probe and miRNA addition, but took a long time for the signal to decrease. R19 showed an excellent response, returning to near original levels within a few hours.](image)

For an optimally functioning reporter, the signal should increase as the probe is added due to opening of the reporter hairpin and separation of the dyes. When the miRNA is added the reporter should return to its hairpin conformation and the signal should return to its original levels (given that the miRNA to probe ratio is 1:1). Analysis of previous reporters (R100, R99, R40, and R43) show what to expect if the reporter
does not bind well to the probe. R55 shows the reporter binding well to the probe, but a slow decrease in signal upon addition of the miRNA. R19 demonstrated good signal contrast for both the probe and the miRNA. The signal rose upon addition of probe, and returned almost to starting levels within a few hours after adding probe.

Base Pair Change Comparison:

Now that we have a set of tested reporters it is useful to perform an analysis of how each system changes in regards to the number of bound base pairs to see if any patterns emerge. A summary of the bound base pairs for R244, R8, R19, R55, R4, and R25 in complexes is depicted in Table 1.

Table 1: A summary of the number of bound base pairs in the reporter stem, reporter-probe complex (RP), and probe-analyte complex (PA) for R244, R8, R19, R55, R4, and R25.

<table>
<thead>
<tr>
<th>Reporter</th>
<th>Analyte</th>
<th>Reporter Stem</th>
<th>RP</th>
<th>PA</th>
<th>R→RP</th>
<th>RP→PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>R244</td>
<td>miR29b</td>
<td>9</td>
<td>13</td>
<td>21</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>R8</td>
<td>miR29b</td>
<td>5</td>
<td>15</td>
<td>21</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>R19</td>
<td>miR29b</td>
<td>5</td>
<td>15</td>
<td>21</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>R55</td>
<td>miR29b</td>
<td>7</td>
<td>17</td>
<td>21</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>R4</td>
<td>Let7a</td>
<td>9</td>
<td>13</td>
<td>20</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>R25</td>
<td>miR26a</td>
<td>7</td>
<td>17</td>
<td>22</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

The ideal RP-PA change appears to be between 5-7 nucleotides. If the change exceeds this number, the reporter-probe complex’s long non-complementary region increases the risk that the molar fraction present at working temperature will be too low, as in the case of R244. If the RP-PA change is less than 5 nucleotides, as in the case of R55, the reporter-probe complex is more stable and the reaction may proceed slowly because the miRNA will have a harder time displacing the reporter.

Ideal Non-Complementary Length:

Another parameter to examine is the length of the non-complementary region. Reporters with a non-complementary region of length nine appear to affect the stability of the reporter probe complex such that the molar fraction is lower than 100 % at working temperatures. This was demonstrated by R244. On the other end of the spectrum, non-complementary regions of length 5 yielded no ideal reporters for miR-29b-1-5p. The thermodynamic changes between reporter-probe and probe-analyte were too small (less than those of R55). The ideal non-complementary length appears to be around seven nucleotides based on the success of R19.
Conclusion:

When selecting a reporter a variety of factors must be evaluated including: analyte and off-analyte interactions, thermodynamic values, hairpin structure, molar fractions of biosensor formed, and the probability of base pair binding in the hairpin. A net change of 5-7 base pairs moving from reporter-probe to probe-analyte yields the best results. In addition, a non-complementary region of length seven appears to provide a stable reporter-probe complex that proceeds to the probe-analyte complex in the desired period of time.

This reporter design process produced R19. This biosensor was shown to be sensitive to the presence of miR-29b-1-5p within the desired time period of a few hours. This supports the validity of this design method for reporters of double-strand displacement biosensors. This work also made significant improvements over previous design methods by increasing the pool of possible reporters, streamlining the selection process, and contributing to the understanding of the important parameters behind the double-strand displacement biosensor design.

Current and Future Directions:

Thermodynamic parameters give an indication of how favorable a reaction is, but they do not give any information about the activation barrier. To our knowledge, no model currently exists for predicting the activation energy for this type of system. Therefore, my current and future research is focusing on development of a model to predict reporter success that takes into account the activation energy, changes in melting temperature, Gibbs free energy, enthalpy, and entropy. If successful, the model could quickly and easily be extended to select reporters for probes that will subsequently detect any type of miRNA using this biosensor design.

In order to develop the model, I plan to perform a series of kinetic binding experiments to find the forward rate constant of reporter-probe to probe-miRNA. The

![Figure 12](image-url)

**Figure 12**: The method for determining the activation energy of the biosensor as it moves from reporter-probe to probe-analyte. The activation energy can be obtained by multiplying the slope by $-\frac{E_a}{R}$, where $R$ is the universal gas constant.
kinetic studies will measure the time required for a signal change of various reporters after probe addition as a function of temperature. Next, similar experiments will be performed except miRNA analyte will be added the reporter-probe biosensor.

A spectrometer will be used to measure the change in intensity over time. A small concentration of probe or miR analyte will be added to a large amount of the reporter or reporter-probe biosensor, respectively. I will use the Arrhenius equation to relate the rate constant to the activation energy, see Equation 1. By plotting ln(k) vs. 1/T, the activation energy can be found from the slope (–Ea/R), where k is the rate constant and 1/T is the inverse of the temperature.

\[
\ln (k) = -\frac{E_a}{RT} + \ln (A) \quad \text{(Equation 7)}
\]

Before temperature studies can be performed, the order of the reaction must be determined. Currently, I have performed a set of experiments with varying probe concentrations and constant reporter concentrations in order to determine this. After data workup, I will be ready to proceed to temperature studies and be able to determine the activation energy using the Arrhenius equation.

If successful, this research will significantly enhance and streamline the biosensor design. These sensors can then be utilized for the early detection and identification of specific cancer cells. Additionally, by monitoring fluctuations in miRNA expression levels, this work also has the potential to aid in understanding the pathophysiology of many diseases.
References:


