2015-07-30

Design of Novel Bioluminescent Nucleic Acid Sensing Systems for Clinical Analysis

Eric Anthony Hunt
University of Miami, hunt.eric@gmail.com

UNIVERSITY OF MIAMI

DESIGN OF NOVEL BIOLUMINESCENT NUCLEIC ACID SENSING SYSTEMS
FOR CLINICAL ANALYSIS

By

Eric A. Hunt

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2015

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

DESIGN OF NOVEL BIOLUMINESCENT NUCLEIC ACID
SENSING SYSTEMS FOR CLINICAL ANALYSIS

Eric A. Hunt

Approved:

<div style="display:flex">

Sapna K. Deo, Ph.D.
Associate Professor of Biochemistry
& Molecular Biology

Leonidas Bachas, Ph.D.
Professor of Chemistry
Dean of the College of Arts &
Sciences

</div>

Angel Kaifer, Ph.D.
Professor of Chemistry

Dean of the Graduate School

Pirouz Daftarian, Ph.D.
Research Assistant Professor of
Biochemistry & Molecular Biology

HUNT, ERIC A.   (Ph.D., Chemistry)

Design of Novel Bioluminescent Nucleic Acid   (August 2015)
Sensing Systems for Clinical Analysis

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Sapna K. Deo.
No. of pages in text. (133)

Nucleic acids can be pointed biomarkers for a variety of diseases, medical disorders, and even injuries. Their importance is ever growing in the modern clinical setting, which is seeing rapid development in technologies, such as high-throughput sequencing of genomic content, utilizing large data sets to make new correlations, ultimately leading to the discovery of more novel biomarker targets. In the wake of these rapidly accumulating discoveries, there is a need for the creation of biosensor technologies capable of distilling the data-heavy discoveries into clinically applicable diagnostic tests. Herein, work pertaining to the development of nucleic acid biosensors, and the continued improvement of the bioluminescent proteins upon which they are based, is discussed. In particular, bioluminescent stem-loop probes utilizing *Renilla* luciferase as a reporter are described for the detection of nucleic acid targets, and demonstrated in detecting a miR-21 microRNA target from human serum samples. To improve the sensitivity and utility of these bioluminescent probes, the genetic alteration of *Gaussia* luciferase to produce truncated variants exhibiting unique spectral and kinetic characteristics is discussed, as well as the bacterial expression of a truncated *Vargula* luciferase for high-throughput detection platforms. Finally, the prospective research pertaining to the overlap of both of these areas is highlighted, as new bioconjugation techniques are developed for the chemical synthesis of these bioluminescent sensing systems.

for Angie

"To finish the moment, to find the journey's end in every step of the road, to live the greatest number of good hours, is wisdom."

Ralph Waldo Emerson 1803–1882

Thank you for adventure, for beauty and meaning in every moment.
I love you.

in memoriam Tyler

"Life is not an end result, it's a journey. Treasure every moment that you have. Value the people who chose to spend time with you."

Tyler Joseph Wilson, M.D. May 16, 1986–July 18, 2015

Our brother, you will be sorrowfully missed, and joyfully remembered. Be at peace.
1 Thessalonians 4:13–14

to Sapna

I would like to extend special thanks to my advisor, Sapna. I consider myself blessed to have a mentor who has given me guidance and encouragement in all aspects of my life. These years we have spent together have made me a better scientist, a better teacher, and a better person. I admire you and the quiet yet strong leadership you have demonstrated with your students.

There Is No Leaving

Winter is not coming
Even when you leave—
The mangroves will keep stretching
Fingers of Miami past
And present,
Spreading and then growing
Until they cover you like the humidity.

Quinn Smith (Written August 31, 2014)

Contents

3   Truncated *Gaussia* Luciferase Variants   54

4   Expression of a Truncated *Vargula* Luciferase   81

List of Figures

List of Tables

Chapter 1

Introduction

MicroRNA Detection: Current Technology and Research Strategies[1]

Overview

The relatively new field of microRNA (miR) has experienced rapid growth in methodology associated with its detection and bioanalysis as well as with its role in -omics research, clinical diagnostics, and new therapeutic strategies. The breadth of this area of research and the seemingly exponential increase in number of publications on the subject can present scientists new to the field with a daunting amount of information to evaluate. This review aims to provide a collective overview of miR detection methods by relating conventional, established techniques [such as quantitative reverse transcription polymerase chain reaction (RT-qPCR), microarray, and Northern blotting (NB)] and relatively recent advancements [such as next-generation sequencing (NGS), highly sensitive biosensors, and computational prediction of miR targets] to common miR research strategies. This should guide interested readers toward a more focused study of miR research and the surrounding technology.

---

1 [1]   Eric A Hunt et al. "MicroRNA Detection: Current Technology and Research Strategies". In: *Annual Review of Analytical Chemistry* 8.3 (2015), pp. 3.1–3.21, reproduced with permission of Annual Reviews, from Annual Review of Analytical Chemistry.

1.1 Perspective

It has been slightly more than two decades since the first microRNA (miR) was discovered [2] and slightly more than one decade since the widespread regulatory power of miR was realized [3]. In this short amount of time, the young field has experienced rapid growth in methodology associated with miR detection and bioanalysis as well as the role of miR in -omics research, clinical diagnostics, and new therapeutic strategies. The breadth of this area of research and the seemingly exponential increase in number of publications on the subject can present scientists new to the field with a daunting amount of information to review; indeed, even for scientists working in the field it can be a formidable task just to remain well informed. This review aims to provide a collective overview of miR detection methods by relating conventional, established techniques and relatively recent advancements (within the past few years) to common miR research strategies.

1.1.1 What Is MicroRNA?

MiR is a class of small, noncoding RNAs (ncRNAs) that post-transcription-ally regulate gene expression by acting on messenger RNA (mRNA) through association with an RNA-induced silencing complex (RISC) to suppress its translation or effect its degradation. This review does not go into great detail on the biogenesis of miR and assumes a previous knowledge on the topic. Readers interested in a more in-depth review of the biogenesis and cellular function of miR are encouraged to consult some of the excellent reviews already published on these topics [4–9]. The biogenesis of miR presents particular challenges, which this review references as they pertain to different detection methods. In particular, the presence of primary (pri-) and precursor (pre-) miRs and the potential for different sequence isoforms (isomiRs) derived from slight differences in upstream processing can be problematic for many

detection techniques. Additionally, miRs of the same family may differ by only one base, making specificity a critical challenge in methods attempting to distinguish individual members. Most miRs are approximately 22 nucleotides in length, which is approximately the length of a standard polymerase chain reaction (PCR) primer, and comprise only approximately 0.01% of the total ribonucleic acid (RNA) typically extracted from a sample [10]. This means that miR detection techniques not only need to be specific, but also sensitive. Finally, because of their small size, miRs exhibit a wide range of melting temperature (T_m)s, making it difficult to optimize parallel reactions involved in their detection.

1.1.2 MicroRNA in Disease

As the function of miR was elucidated, it became clear that these small RNAs were capable of great things on the cellular level. To date, there is a large (and constantly growing) number of publications highlighting different mRNA targets under the control of miR. The expansive set of published miR sequences is housed in a searchable repository known as miRBase[2], which is available for free online. It has been shown that a single miR can regulate hundreds of mRNAs and thereby control an entire expression network [11]. Due to its immense regulatory power, aberrant miR expression levels have been implicated as a biomarker in several different forms of disease including neurodegenerative diseases (and CNS injury), diabetes, cardiovascular disease, kidney disease, liver disease, and even immune dysfunction [12–15]. This biomarker trait is not only tissue specific, as miRs have also been isolated from extracellular, less invasive samples such as serum, saliva, and urine [16, 17]. Of special interest is the involvement of miR in cancer, where it has been shown to be a biomarker for metastasis, chemoresistance, diagnosis, and prognosis

---

[2] `http://mirbase.org/`

and implicated in potential oncogenic moieties [17–20]. The utilization of miR is also being explored as a potential form of therapy or treatment [18, 21]. Using miR as a biomarker for disease is one of the many prominent trends driving the field of personalized medicine into fruition [22].

1.1.3 Common MicroRNA Research Strategies

The field of miR is at a crucial turning point. Although there is still much that is unknown about the mechanisms underlying miR function on the cellular level, researchers have answered many questions pertaining to how miR exerts its regulatory control over post-transcriptional processes, heralding a new era of miR research. In terms of miR detection, the nature of the research goal is the driving motive force behind selection of the appropriate method(s) of detection. Each set of methods has its own strengths and weaknesses and as such, these requirements should be carefully weighed to develop a successful miR research strategy.

1.2 Current MicroRNA Detection Techniques

1.2.1 Conventional Techniques

Northern blotting. Northern blotting (NB) has been the method used in miR research since the initial discovery of lin-4 as a negative regulator of lin-14 in 1993 [2]. The technique combines an electrophoretic separation—typically by denaturing urea-polyacrylamide gel—followed by transfer to a membrane, usually a positively charged nylon membrane by semidry capillary transfer. The miR is then hybridized with labeled probes and imaged. Traditionally ^{32}P-labeled deoxyribonucleic acid (DNA) probes are used to visualize the blotted RNA. NB is the only technique that allows for the quantitative visualization

of miR. The size separation step of NB enables the technique to be used for quantitative expression analysis of mature and pri-/pre-miR as well as for analysis of size variation of isomiRs from imprecision of Drosha and Dicer cleavage in upstream biogenesis of the mature miR [23–25]. In comparison to other conventional techniques, NB suffers from low sensitivity (nM–pM) [26], low throughput, and high input RNA requirements (typically on the order of 5–50 μg total RNA per sample) [27–31].

Radioisotopes (^{32}P) are the most commonly used labeling system for NB detection, but the use of radioisotopes poses several safety concerns for the researchers using them and the environment where they are disposed. Strict constraints pertaining to the use of radioisotopes often makes their use impractical or impossible, especially when an institution prohibits their usage [29, 31]. In addition to these safety concerns, the use of radioisotopes greatly increases the amount of time required to perform the NB technique, and in some cases ^{32}P labels must be exposed for days to detect weak signals. To improve safety for researchers and reduce impact on the environment, the introduction of hapten-labeled probes coupled with enzymatic detection methods has been established [29]. By labeling DNA probes with a 3′-digoxigenin (DIG) hapten, the authors were able to reduce exposure time to minutes or hours using a DIG-antibody conjugated alkaline phosphatase and a chemiluminescent substrate. Although this approach reduces exposure time, the use of hapten labels is not as sensitive as radioisotopic labeling. To compensate for this drop in sensitivity, other modifications have been made to the traditional NB method.

The use of locked nucleic acid (LNA) in NB probes yielded a tenfold increase in sensitivity as compared to traditional DNA probes and improved mismatch specificity [30]. Another method reported an approximately 20-fold increase in sensitivity by cross-linking the RNA to the membrane using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) [28]. Typically, RNA

is cross-linked to the membrane by UV irradiation, which likely proceeds through the uridine of the RNA. It is possible that this mechanism of linkage disrupts the subsequent hybridization of labeled probes and thereby reduces sensitivity. The EDC linkages should occur through the 5′-phosphate of the RNA, thereby linking the RNA to the membrane in a fashion more amenable to the subsequent hybridization step. A more recent method of NB for miR analysis utilizes these improvements in conjunction with DIG-labeling in a technique termed LED (for LNA-modified probes, EDC cross-linking, and DIG-labeled). Kim et al. [27] reported an approximately 1,000-fold decrease in exposure time and could detect as low as 0.05 fmol of target RNA from approximately 3 μg of total RNA. One of the major drawbacks of this method, however, is that LNA is proprietary technology and can increase the cost of hybridization probes by approximately 50-fold over traditional DNA probes. To help alleviate cost, another method termed DSLE (for DIG-labeled, splinted-ligation, and EDC cross-linking) was introduced [32]. In this method, the miR target is attached to a universal DIG-labeled probe by splinted ligation—using an unlabeled bridge oligonucleotide to bring the two into proximity and T4 DNA ligase to join the 3′-OH of the miR with the 5′-phosphate of the universal DIG-labeled probe [33] (Figure 1.1).

The method could be completed in 6–8 h and could detect 2 fmol of the RNA target using as little as 4 μg of total RNA per sample. Although
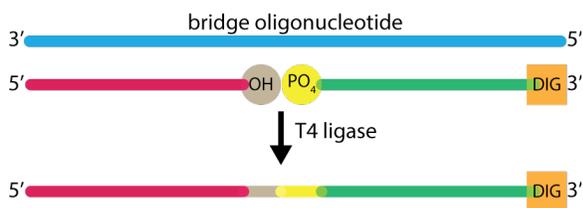
Figure 1.1. DSLE [digoxigenin (DIG)-labeled, splinted-ligation, and EDC cross-linking] method for Northern blot analysis of miR. 5′-Phosphate of DIG-labeled probe (green) and 3′-hydroxyl of miR target (purple) are brought into proximity by adjacent hybridization to the bridge oligonucleotide (blue). T4 ligase repairs the nick, thereby generating labeled miR.

the authors acknowledge that their method was 200 fold less sensitive than another method utilizing the same splinted-ligation procedure incorporating ^{32}P-labeled probes—presenting the same aforementioned safety concerns—DSLE at least demonstrated comparable results to methods utilizing LNA-modified probes [34]. Although radioisotopic labeling is still the most sensitive method and LNA probes provide superior hybridization of labeled probes, these additional modifications to the traditional NB procedure provide important advancements that make it a viable method for miR detection in any laboratory setting.

Microarray. Similar to NB, microarray analysis of miR relies on the sensitive, specific hybridization of the target miR to a complementary DNA probe, the major difference being that it is the miR that is labeled in the microarray technique (note, however, some methods discussed below that do not require labeling of the miR target). Microarrays depend on the spatial organization of complementary capture probes on a solid phase. This limits hybridization of a specific target to a localized spot that is easily visualized with fluorescence/imaging instrumentation. It is this principle of design that has allowed microarrays to be among the first technologies capable of massively parallel analysis of hundreds of miRs simultaneously from one sample [10, 11]. There are some inherent drawbacks to the microarray method. First, it is only semiquantitative and is most readily suited to compare relative expression levels of miR between different cellular states (e.g., diseased versus healthy). Therefore, the method requires some other form of validation, such as (), to quantify expression. Second, microarrays have a smaller dynamic range than other methods of detection such as RT-qPCR or next-generation sequencing (NGS), often causing a fold change compression that underestimates relative changes in miR abundance [10, 11, 35]. Finally, as microarray is a hybridization-based method, specificity can be an issue between closely

related sequences. Several advancements and modifications have been made to address these issues related to microarrays.

Because all target sequences are analyzed in parallel on a microarray, it follows that the entire chip must undergo the same set of hybridization conditions. Because miRs exhibit a wide range of T_ms (in the range of 45–74 °C) [36] it is has been historically problematic to obtain T_m-matched probe sets for full expression profiling using DNA oligonucleotides. The introduction of LNA into the microarray probe sets allowed for T_m normalization and simultaneously improved specificity and mismatch discrimination using 2.5–5 μg of total RNA [37]. An additional benefit to the use of this platform is that it required no pre-amplification or fractionation steps to enrich the miR content. In a similar fashion, oligonucleotides modified with 2′-O-(2-methoxyethyl) nucleic acid analogs—which exhibit enhanced hybridization affinity/specificity to RNA—have been employed to normalize the T_m of microarray capture probe sets [38].

The most common method of miR labeling is by enzymatic attachment of the label. There are two main approaches used. In one approach, T4 ligase is used to attach a fluorescent-labeled nucleotide or short oligonucleotide to the 3′-OH of the RNA. However, because miRs have a 5′-phosphate, there is the possibility of circularization by intramolecular ligation instead of labeling [10, 11, 39]. This complication can be avoided by adding another enzymatic step to dephosphorylate the miR before the labeling step. In the second approach, a 3′-polyA tail is added to the miR using polyadenylate polymerase (PAP). Once the tail is in place, a bridge oligonucleotide with complementary polyT region aligns the 3′-OH of the miR with the 5′-phosphate of another labeled oligonucleotide for splinted ligation (as previously outlined for NB). Although this method avoids potential circularization, the polyadenylation is not controlled and a variable number of adenosine ribonucleotides may be added to the tail of the miR, thereby possibly affecting subsequent hybridiza-

tions [10, 11]. Enzymatic labeling methods are convenient for miR analysis; however, the enzymes used often exhibit a bias toward certain substrates or sequences and introduce artifacts into the apparent abundance of affected miRs. Additionally, the presence of pri-/pre-miR can also introduce artifacts into the microarray results. The addition of 5′ hairpins to the capture probes has been used to distinguish the mature miR from its pri-/pre-miR forms [40] (part *a* of Figure 1.2). Li and Ruan [39] highlight a more expansive overview of miR labeling methods.

Although technical replicates for microarrays typically show good reproducibility, several studies have indicated a lack of interplatform agreement in expression profile data [35, 41–44]. Much of this inherent variability arises from the labeling step and selection of controls. For commercially available
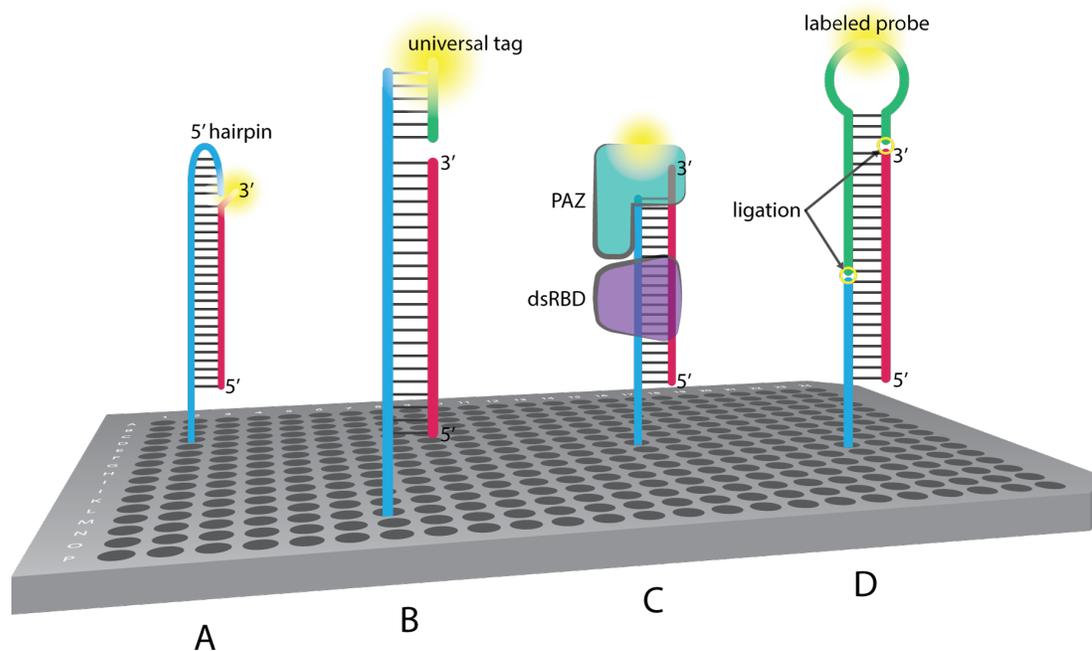
Figure 1.2. Various methods for microarray-based miR detection. (a) 5′ hairpins help select for mature miR form only. (b) SHUT assay: universal tag hybridization dependent on base-stacking stabilization provided by adjacently hybridized miR. (c) Label-free method utilizes PAZ-dsRBD fusion which recognizes 3′ overhang and dsRNA structure of hybridized miR. (d) LASH assay: label-free method uses adjacent binding of capture probe and labeled hairpin probe to facilitate ligation. Abbreviations: LASH; ligase-assisted sandwich hybridization; PAZ, Piwi/Argonaute/Zwille; SHUT, stacking-hybridized universal tag.

microarrays, each company goes about treatment of the RNA sample differently and thereby introduces different artifacts and biases into the data as a result of the imperfections inherent in each labeling method [36, 45]. Because of the bias introduced by labeling, some methods aim to replace or remove the traditional miR labeling step with alternative approaches—as Lee et al. [46] describe using a biotin-labeled structure-specific RNA binding protein (PAZ-dsRBD derived from Argonaute proteins) to recognize array captured miR targets (part c of Figure 1.2 on the preceding page). A hybridization mechanism dependent on base stacking (termed stacking-hybridized universal tag or SHUT) allowed for the use of a universal reporter probe for all miR sequences on the array [47, 48]. The capture probes are designed such that the universal reporter probe requires base-stacking stabilization provided by the target miR to remain bound (part b of Figure 1.2 on the previous page). Without the presence of the target miR, the universal reporter probe dissociates and is washed away, resulting in no signal generation. Another method that replaced the miR labeling step is the RNA-primed array-based Klenow enzyme assay, which performs a post-hybridization labeling of the capture probe using the annealed miR target as a primer [49]. The capture probes are covalently attached via their 5′ end to the microarray solid phase. Moving out from the solid phase, the probes consist of a common spacer, three thymidine residues, and an antisense sequence to the miR of interest. Following hybridization of the target miR, remaining single-stranded probes are digested away from the solid-phase with exonuclease I, and the Klenow fragment of DNA polymerase I is used to add biotinylated adenosine residues, priming from the bound miR. These biotin are then used to generate a signal using fluorophore-conjugated streptavidin. A recent publication outlined a method termed ligase-assisted sandwich hybridization, which utilized both base-stacking stabilized hybridization and a process similar to the previously discussed splinted-ligation method. This method was able to detect a syn-

thetic miR target down to 30 fM (10 amol) and closely matched results from RT-qPCR using 1 μg of total RNA extracted from blood [50]. This was also considered a label-free method, as a labeled, miR-specific hairpin probe was ligated to the target miR and capture probe to yield a signal (part d of Figure 1.2 on page 9). Currently, most available, established commercial microarrays are label based and require on the order of tens to hundreds of nanograms of total RNA, exhibiting nanomolar to picomolar detection limits [10, 26].

Quantitative reverse transcription polymerase chain reaction. If any method could be considered a single gold standard among miR detection techniques, it would be RT-qPCR as it offers a good balance between cost, precision, and sample size along with a large functional dynamic range. It is used to validate results from whole-genome screening methods, such as microarrays and NGS, and in the screening of clinically relevant subsets of miRs [51]. The goal of many research strategies is to obtain a full miR profile of a specific tissue or patient-derived sample, and in that regard, as a trusted method of miR detection, RT-qPCR has been advanced to achieve these goals. However, given the breadth of miR research strategies, application of RT-qPCR to all of these realms presents significant challenges, especially when considering the wide range of miR sources and the host of methods used to extract RNA from them (see Sidebar on the next page, *The Importance Of Sample Preparation and RNA Extraction*). Although it stands true for every method, it is especially true for RT-qPCR that the quality and reliability of the result is dependent on the quality of the input RNA, as any degradation can introduce errors that will be amplified further during the reverse transcription (RT) and quantitative polymerase chain reaction (qPCR) steps. It is important to have good quality control and to check the integrity of the RNA before beginning the actual detection portion of the protocol.

The Importance of Sample Preparation and RNA Extraction

Given the evidence that miR profiles could potentially become predictive biomarkers for disease states, researchers are endeavoring to develop powerful bioinformatics approaches by analyzing routinely preserved samples such as formalin-fixed paraffin-embedded (FFPE) and laser-capture microdissected samples [51–53]. Although mRNA is not stable in such sample preparations, miR has been demonstrated to be quite resilient. The performance of RT-qPCR in profiling such samples has been shown to be largely dependent on the quality, i.e., purity and integrity, of input RNA. Purity is typically measured by spectrophotometric methods, and most modern instruments can monitor integrity using automated capillary electrophoresis to determine the 28S/18S ribosomal RNA (rRNA) ratio. Low integrity RNA contains small RNA fragmentation that can cause overestimation of the miR contribution [54]. Additionally, large RNAs usually exhibit a carrier effect and therefore the amount of miR present is dependent on the amount of total RNA [51].

The first step in RT-qPCR is to convert the extracted RNA from the sample to complementary DNA (cDNA) by RT. There are two main approaches typically used to do this: (a) using miR-specific RT primers and (b) extending all miRs with a common sequence so that RT may be performed using a universal primer (Figure 1.3 on page 16). Each method has its benefits and disadvantages. Some laboratories may prefer to perform a fractionation of the RNA extract to remove the pri-miR sequences and to lower unwanted background in the qPCR; the pre-miR sequences are smaller and not always removed in this step. However, newer methods of cDNA synthesis, specifically the use of miR-specific stem-loop primers, avoid the

need for precedent fractionation and can distinguish between different miR forms [55]. The miR-specific primers may also possess a linear segment in place of a stem-loop structure at their 5′-end, although both of these variants contain an miR antisense portion of approximately 6–8 nucleotides on their 3′-end. Independent of structure, the 5′-end contains a common sequence for priming qPCR [10, 11, 51]. Although the linear primers are easier to design, the stem-loop primers are better at specifically targeting the mature miR form, and thus provide enhanced specificity and reduced qPCR background. Although this is beneficial for getting more accurate quantitation of mature miR, it does hinder the RT of isomiR sequences [10]. In both cases, RT primers should be designed such that annealing can be done at lower temperatures (usually 16 °C), as this helps preserve pri-/pre-miR secondary structures (thus keeping the template sequence sequestered) and thereby reduce RT of those longer, potentially unwanted miR forms [51]. In contrast to using miR-specific primers, the second method uses a universal RT primer to synthesize cDNA. This is done in one of two ways. In the first, PAP is used to extend the miR population with a polyA tail. An RT primer containing a universal qPCR primer sequence followed by a polyT segment then binds to the extended miR for cDNA synthesis—the number of thymine residues incorporated in the cDNA is selected for by aligning the RT primer to the 3′-end of the miR using degenerate bases (usually three) that can base pair with more than one base. A downside to this method is that PAP extends all RNAs in the sample, including pri-/pre-miR sequences, and it does not allow for optimized primers for miR sets with widely varying T_ms [51]. In the second universal method, T4 ligase is used to ligate a common sequence to all miRs in the sample and then cDNA is synthesized with a universal RT primer. The use of these common sequence RT methods shifts the selection process for mature forms of miR to the qPCR step, which we discuss below. Another novel method of cDNA synthesis worth mentioning is the RT method used in miR-ID. In this method,

miR in the sample is circularized through intermolecular ligation and reverse transcribed to produce tandem repeats of the cDNA sequence. This approach allows for great control over qPCR primer placement to distinguish between miR forms [56].

Once the cDNA is synthesized, qPCR can begin. This proceeds in a straightforward fashion using a miR-specific and universal qPCR primer set. The two fluorescent systems used for monitoring qPCR of miR are SYBR Green and TaqMan probes (Figure 1.3 on page 16). The major challenge in the qPCR step arises again from the broad range of T_ms exhibited by miR, which makes running an array of simultaneous reactions difficult. This problem can be alleviated by using LNA-containing primers to tune all reactions to an optimal set of conditions. Indeed, the added specificity of hybridization brought about by using LNA-modified primers for qPCR (such as the miRCURY LNA qPCR platform offered by Exiqon) enables discrimination between mature and pri-/pre-miR forms independent of the cDNA synthesis method used [10, 11, 45, 51]. It is important to perform internal tests such as a ten-fold dilution (which should correlate to 3.32 cycles) to determine amplification efficiency and to perform a melting curve experiment to determine amplification specificity. The latter is especially important for the SYBR Green method in which the dye intercalates into any double-stranded DNA (dsDNA) product [13]. Perhaps the most difficult part of RT-qPCR is analyzing the data after it is generated. As miR represents only ~0.01% of total RNA extracted from a sample, the effective amount present can experience wide variability depending on the sample preparation and extraction yield and integrity (see Sidebar on page 12). As such, the normalization of miR profile data is of utmost importance, not only in RT-qPCR, but in all profiling techniques [57]. The variety of sample types profiled—i.e., cells, tissues, and extracellular fluids such as blood, urine, and saliva—make the use of common values in other analyses such as cell count invalid as normalization factors. Similarly, miR enrichment procedures

often remove other RNAs that may serve as normalization factors, such as rRNA. Most groups currently use a so-called housekeeping or reference gene such as small nuclear RNA (snRNA) U6, or a geometric mean of multiple reference genes to normalize their data [13, 57–59]. In addition to these types of normalization, it is also important to include the proper internal and external controls to account for plate or slide variation and potential human error such as pipetting error [36, 57].

1.2.2 Biosensor Techniques

Often simple advances in methodology can be applied to several different systems with drastic improvement to sensitivity and/or selectivity (e.g., two temperature hybridization protocols for solid-phase and surface-modified capture systems) [60, 61]. This has been increasingly evident as traditional detection systems have been gradually adapted to conductive surfaces for electrochemical detection schemes. Generally, the most significant advances for uniplex or small-scale multiplex platforms have been those that increase assay sensitivity for a specific subset of miR targets, whereas array-type profiling systems have seen the greatest benefit from more capable discrimination of isomiRs and familial variation for enhanced interassay (or even intra-assay) agreement. Although these characteristics are in no way divergent, the end-user of a particular platform or methodology has often been required to balance sensitivity and selectivity to maximize assay utility toward their unique application. This becomes most apparent as the number of miR targets increases, when T_m diversity and expression variants begin to exert more influence on overall signal quality [11]. Thus, many recent attempts at platform development (both uniplex and multiplex) have capitalized on adapting high-sensitivity transducers to novel, selectivity-enhanced recognition elements. As many excellent reviews [26, 62–66] have provided in-depth
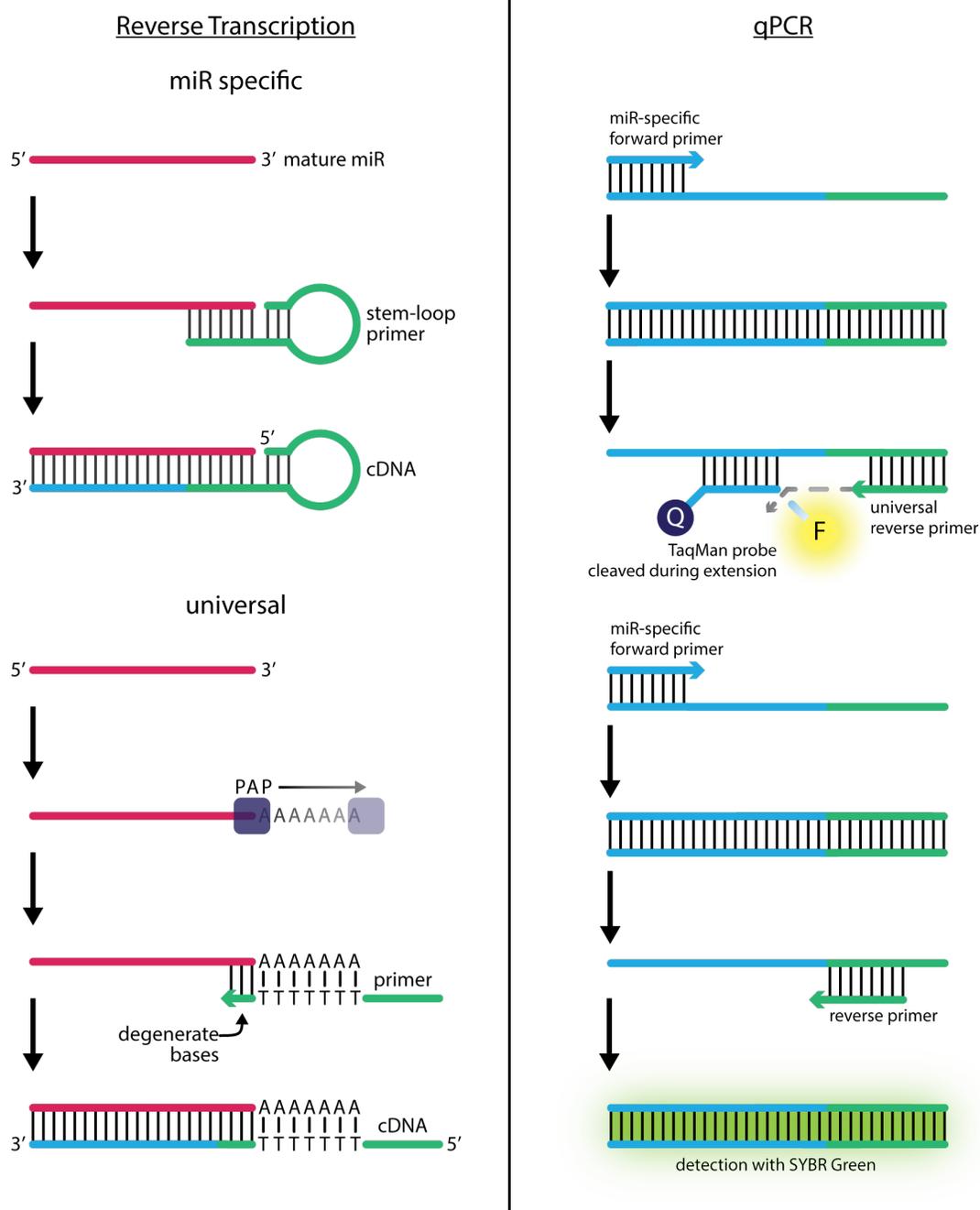
Figure 1.3. RT-qPCR: methods for cDNA synthesis by RT using miR-specific (a) or universal (b) primers. Two fluorescent methods are used for monitoring miR qPCR: TaqMan (c) polymerase exonuclease activity releases fluorophore during PCR extension; SYBR Green (d) fluorescent dye intercalates into dsDNA produced by PCR amplification. Abbreviations: PAP, polyadenylate polymerase; PCR, polymerase chain reaction; qPCR, quantitative polymerase chain reaction; RT-qPCR, quantitative reverse transcription polymerase chain reaction; RT, reverse transcription.

background on the various biosensor platforms and techniques, we limit our discussion to recent examples of this trend that utilize direct (unlabeled) miR detection methods.

Electrochemical-based detection. A recent development in hybridization-based miR assays is the incorporation of electrochemical impedance spectroscopy (EIS), a technique frequently associated with electrode surface characterization during sensor construction, as a means of detecting specific hybridization events [67]. This rather complex technique has been adapted to detect impedance changes associated with variable charge transfer at the surface of an electrode, with measurements typically obtained using coordinated metal ion redox probes. One strategy involves the selective deposition of a charge-transfer resistant (insulating) layer [68, 69] that accumulates on a capture probe-functionalized electrode surface in the presence of surface-targeted anionic species (miR). Immobilization of this layer can be performed via catalytic polymerization using horseradish peroxidase or, in the case of [68] , G-quadruplex hemin in conjunction with H_2O_2. To minimize background, these systems can also incorporate a charge-neutral capture sequence composed of peptide nucleic acids or Morpholinos [70], although Ren et al. [71] managed approximately the same detection limit using standard DNA anti-miR capture probes. Additionally, they found that the capture probe monolayer could double as the insulating layer and were able to leverage signal amplification by combining detection and target recycling via removal of target-hybridized capture probes using duplex-specific nuclease.

Selective removal of capture probes can also be applied to other electrochemical techniques, as clearing the electrode surface often enhances charge-transfer events. Gao and Peng [72] applied this concept to the amperometric detection of three miRs differing in length by up to 10 nucleotides. By combining Surveyor® mismatch-selective nuclease and exonuclease I single-

strand-specific nuclease with an electrode-bound monolayer of miR-specific capture probes, only perfectly bound target miR could prevent capture probe digestion. When combined with glucose oxidase-functionalized reporter probes, this platform provided a detection limit of approximately 10 fM with extremely high mismatch selectivity (Table 1.1 on the next page).

For many electrochemical techniques, however, inaccessibility to or loss of the selective capture region on the electrode surface is detrimental to overall assay sensitivity. In these cases, it often becomes a sensitivity compromise between the density of surface-immobilized capture probes and the diffusion of electroactive species. A recent advancement involves the use of DNA-based scaffolds [73–75] that form a tetrahedral platform for capture probe immobilization. In addition to providing reproducible capture probe orientation, this scaffold provides a rigid, isolated pedestal that abrogates nonspecific interaction of the capture moiety with the surface while providing reduced surface density for better mass transport characteristics. These enhancements serve to increase the signal-to-noise ratio and allow detection limits as low as 10 aM [73, 75].

Increasingly, electrochemistry has proven highly amenable to incorporation of the Carnation Italian ringspot virus (CIRV) protein p19 as a selectivity agent for mature miR transcripts [76]. This molecular caliper can sequester double-stranded RNAs (dsRNAs) ranging from approximately 21 to 23 nucleotides with no interference from single-stranded RNAs (ssRNAs) or heteroduplexes [77] and has shown utility for discriminating miR/probe duplexes using various techniques including differential guanine oxidation [78], resistance changes [77], and square-wave voltammetry/EIS [79]. Given the difficulty of applying electrochemical techniques to complex matrices, p19 also offers an attractive secondary benefit of purifying miR from total RNA extracts [80] or even from high-background matrices such as serum [81] for a broader repertoire of direct analyses.

Table 1.1. Abbreviations: EIS, electrochemical impedance spectroscopy; EXPAR, exponential amplification reaction; LNA, locked nucleic acid; LOD, limit of detection; PEG, poly(ethylene glycol); PNA, peptide nucleic acid; TMB, tetramethylbenzidine.

Principle	Reporter	Detection	LOD	Sample	Ref
PNA-decorated Au nanobead	G-quadruplex hemin, H_2O_2-initiated polyaniline deposition	EIS	0.50 fM	Total RNA extract	[68]
Carbon nanotube-bridged field-effect transistor decorated with p19	None	Resistance	1 aM	Total yeast RNA	[77]
Capillary electrophoresis, p19	Fluorescent dye	Laser-induced fluorescence	0.5 fM	Serum	[81]
DNA capture probe-functionalized Au electrode	Glucose oxidase, $Os(bpy)_2(API)Cl$	Amperometry	10 fM	Total RNA extract	[72]
Target-assisted exponential amplification reaction on Au electrode	Streptavidin-alkaline phosphatase, α-naphthol	Differential pulse voltammetry	99 fM	None	[82]
Graphene/Au nanoparticle-decorated Au electrode with LNA stemloop and bio-barcode	Streptavidin-Horseradish peroxidase, H_2O_2, hydroquinone	Chrono-amperometry	6 fM	Total RNA extract	[85]
DNA tetrahedral scaffold on Au electrode	Avidin-Horseradish peroxidase, H_2O_2, TMB	Amperometry	10 aM	Total RNA extract	[75]
Hybridization chain reaction on grapheme/Au nanoparticle substrate	Methylene blue	Differential pulse voltammetry	3.3 fM	Serum	[84]
PNA-decorated Si nanowire	None	Resistance	1 fM	Total RNA extract	[86]
Nanopore diffusion using PEG barcode	None	Ionic current	100 fM	None	[87]

Other methods to increase analytical sensitivity have focused on alternative signal amplification techniques in an attempt to provide direct detection without requiring external enrichment of endogenous targets. One strategy relies on internal target enrichment via the isothermal exponential amplification reaction (EXPAR). Although specific methodologies are diverse [82, 83], the amplification of a dual-domain probe containing a central, antisense recognition site for a nicking enzyme is the basic requirement for target enrichment. The 5′ domain will always correspond to the target sequence, whereas the 3′ domain can recognize either target or primer. Capture at the 3′ domain allows polymerase to extend this sequence along the entire capture probe for duplication of the target and incorporation of the nicking site. A nicking endonuclease subsequently regenerates the intact target sequence for detection or entry into the amplification cycle as either primer or primer generator. Alternatively, the signaling moiety can be amplified, and authors have explored the use of hybridization chain reaction [73, 84] in which a hybridization cascade is initiated upon target capture; triple signal amplification [85] that allows a barcoded, branched structure containing 3 distinct signaling moieties to be captured by target in a bridged format; and poly-HRP80 [75] containing hundreds of individual HRP molecules.

Optical-based detection. A significant benefit to optical detection lies in platform flexibility, as surface immobilization is not required for signal transduction. Additionally, optical methods are often less susceptible to interference from matrix components, as signal attenuation becomes most apparent only at high concentrations of absorbing species. As a result, optical methods are ideally suited to direct detection with minimal purification requirements. A diverse set of optical techniques are available, and as these are thoroughly reviewed elsewhere [26, 88], our discussion is limited to recent

enhancements that offer potential for increased sensitivity and selectivity using fairly standard techniques.

The microarray has been the workhorse of miR profiling for more than a decade, but a significant drawback to this technology is the requirement for end-labeling miR sequences prior to analysis. As previously mentioned, the SHUT assay [47] makes use of the energy-minimizing arrangement of bases that occurs upon sequence annealing to utilize a universal signaling probe that will reliably hybridize to an immobilized capture probe only when a target sequence completely complementary to the capture probe is present. This arrangement allows mismatch selectivity at either terminus and reasonable sensitivity (low femtomolar limit of detection (LOD)) on an array-style platform that has been exploited in traditional fluorescence-based detection schemes [47, 89] as well as a unique (although preliminary) electrochemiluminescence (ECL) platform [90]. As with electrochemistry, ECL benefits from signal amplification techniques arising from variations of hybridization chain reaction. Enhancement results from the dramatic increase in double-stranded reporter construct as the hybridization cascade proceeds, given that various coreactants in the ECL reaction can be intercalated within the reporter [91, 92]. Although the potential for false positives in a format already designed around competitive target displacement seems rather substantial, sensitivities fell within the femtomolar range, indicating that target recycling was likely increasing the signal-to-noise ratio. This effect was demonstrated in a solution-phase molecular beacon assay that relied on signal enhancement from double-stranded nuclease-induced target recycling [93]. Upon commencement of the target recycling phase, the detection limit decreased from 0.16 nM to 0.6 fM, whereas the dynamic range increased from 2 to 5 orders of magnitude. A similar technique described previously for electrochemical-based biosensors (EXPAR) can also be adapted for optical systems [94]. Although this approach yields detection limits moderately higher than those demonstrated on

electrochemical systems, the method itself provides important benefits, such as occurring in the solution-phase, being self-contained, and possibly being adaptable to in situ or in vivo techniques using a tethered double-stranded nuclease. For methods that normally are label-free such as surface plasmon resonance and silicon photonic microring resonators, the addition of semiselective labels that recognize only the duplex formed upon target capture allows signal amplification analogous to target recycling. This approach can take the form of antibodies that recognize heteroduplexes [95] or the target-bridged capture of streptavidin functionalized with biotinylated antisense probes [96]. In both cases, the labels enhance the target signal by a reproducible value that permits quantification. In many instances, time-to-result (TTR) is a critical aspect of biosensor application, and there is significant diversity of platform design that enables rapid detection. One advancement in this area involved the CIRV protein p19, mentioned previously for electrochemical applications. In this work, p19 has been paired with fluorescence polarization to provide selective detection of miR-122 at low picomolar levels within 3 min [97]. Likewise, Arata et al. [98] designed a microfluidic device based on laminar flow-assisted dendritic target amplification that generated a 0.5 pM detection limit within 20 min without a requirement for external power. A familiar point-of-care packaging was used to house a lateral-flow assay for the presence of miR-215, and a 75 pM aliquot could be visually observed and quantitatively detected within 20 min due to the effects of Au nanoparticle aggregation [99]. As these techniques continue to evolve and advance, increased detection efficiency concurrent with decreased TTR will remain important benchmarks for biosensor quality assessment.

Application of nanotechnology. The merging of nanotechnology with commonly implemented assay design has sought to improve the efficiency of charge transfer for increased sensitivity detection as well as to increase

the surface area available for probe immobilization. At the forefront of both goals, carbon nanotubes have practically revolutionized electrode design. High susceptibility to changes in conductivity coupled with a three-dimensional structure have allowed 1 aM detection limits in the previously described p19/field-effect transistor assay (Figure 1.4 on the following page), and Tran et al. [100] were still able to maintain a dynamic range of 10 fM-1 nM using multi-wall carbon nanotubes decorated simply with DNA capture probes. Carbon nanotubes also possess the capability to integrate photonic and electrochemical techniques, as demonstrated by a signal-off assay incorporating a quantum dot-labeled capture probe that is noncovalently associated with single-wall carbon nanotubes (SWCNTs) immobilized on an indium tin oxide chip [101].

In the absence of the target, the capture probe remained associated with the SWCNTs and converted photonic quantum dot emission to a high photovoltaic current. However, upon target hybridization, dissociation of the capture probe from the SWCNT surface resulted in signal quenching, whereas DNase I digested the DNA portion of the heteroduplex for amplification via target recycling. A similar system of signal amplification designed for solution-phase detection relied on the formation of a three-way junction consisting of the target sequence, an assistant probe, and a Hg^{2+}-intercalated molecular beacon [93]. Upon complex formation, the intercalated Hg^{2+} is liberated, being made available to quench Ag nanocluster reporters, while a nicking endonuclease cleaves the molecular beacon probe and liberates the target for recycling.

Operating on essentially the same principle as carbon nanotubes, nanowires have the potential to reduce systemic complexity by replacing the traditional electrode while still providing greatly enhanced surface area. Zhang et al. [86] realized a detection limit of 1 fM with excellent single-mismatch selectivity using a network of silicon nanowires decorated with peptide nucleic acid capture probes, whereas capture-probe-decorated Au nanowires operating in
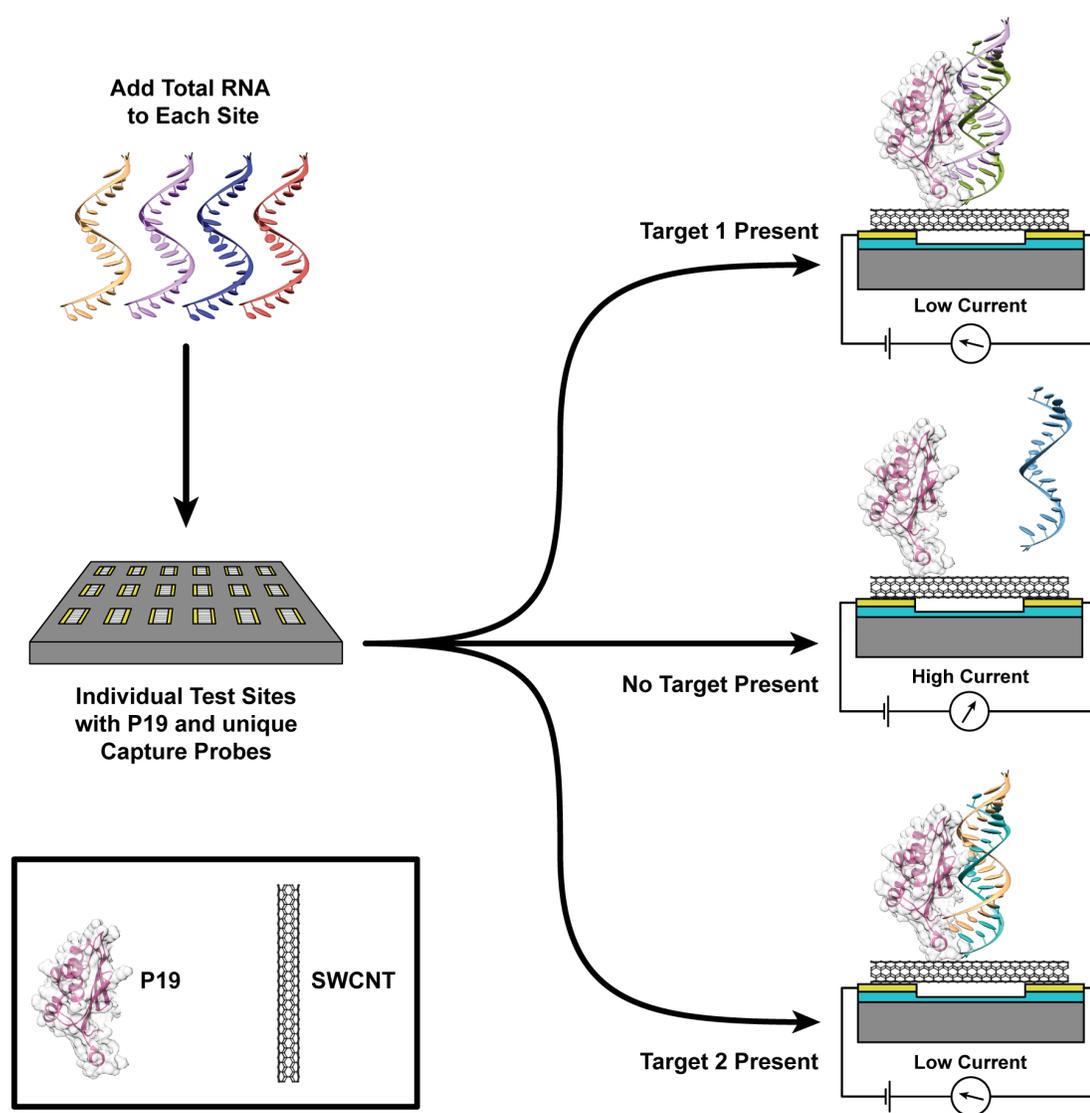
Figure 1.4. Example of an ideal electrochemical miR detection platform [77]. Total RNA (ideally in a bodily fluid matrix) is transferred to the wells of an electrochemical array. Each well contains Carnation Italian ringspot virus (CIRV) p19 covalently immobilized to single-wall carbon nanotubes (SWCNTs) that bridge a gap between two Au electrodes. Unique capture probes are then added to each well, and any mature, antisense target miR will be captured selectively by the combined actions of the capture probe and the dsRNA-specific p19. Target immobilization will be evident in a dose-dependent increase in resistance across the SWCNT.

a target-bridged reporter format were able to achieve a 100 aM LOD using surface-enhanced Raman scattering [60] (Figure 1.5 on the next page).

In addition to the moderately well-established miR detection techniques utilizing nanotubes and nanowires, progress has been made recently through the application of nanopores for miR detection. Although not nearly as well developed as these other nanotechnologies, nanopores have the opportunity to rival the capabilities of this diverse set of platforms owing to their extreme selectivity, label-free operation, and (in the case of synthetic nanopores) flexibility in design for extreme analyte specificity in complex matrices [102]. A drawback to this approach is that the regularity of miR structure (especially size) makes it currently impossible to distinguish individual miRs based on nanopore transit [87]. A fairly recent report demonstrated divergent signals for similarly sized duplex DNA and RNA as well as for transfer RNA (tRNA) on a silicon nitride membrane containing synthetic nanopores [80], but no attempt was made to discriminate divergent miR sequences. Zhang et al. [87] addressed this limitation by generating a "clicked" polyethylene glycol barcode of variable length that could be appended to a selective capture strand. The size of the barcode determined the blocking efficiency of ionic current through the nanopore and served as a specific marker for positive capture of a unique miR sequence. This design allowed them to discriminate four unique miR signatures, and while the estimated detection limit was acceptable ($\sim$100 fM), no data concerning mismatch selectivity was presented.

1.2.3 Other Techniques

Another method that may interest readers is size-coded ligation-mediated PCR in which miR sequences act as a guide for the adjacent hybridization of two size-coded DNA probes that are subsequently ligated together. Probes of different, unique lengths are designed to use a target miR to guide ligation, and
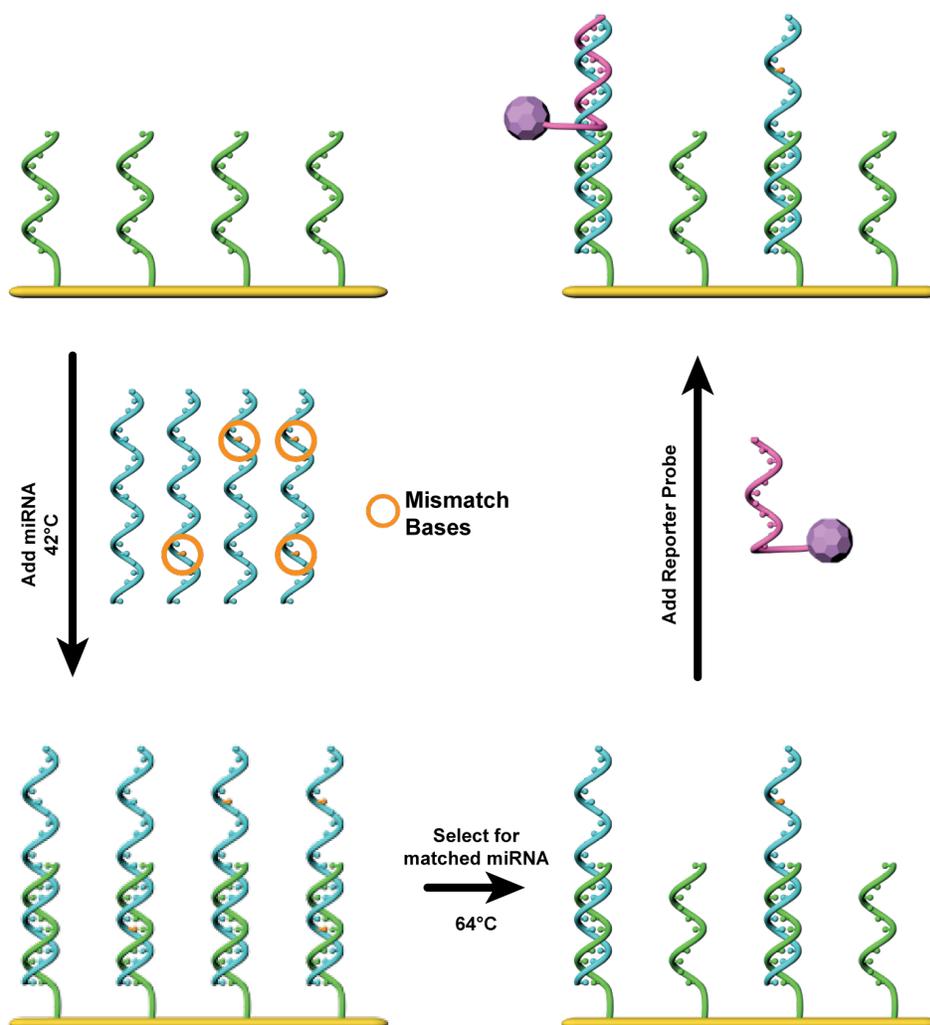
Figure 1.5. Example of a recent optical-based miR detection scheme [60]. Peptide nucleic acid capture probes are covalently immobilized on Au nanowires. Upon introduction of target at 42 °C, some tolerance for mismatches exists, but only single mismatches will be allowed. After initial hybridization, the temperature is ramped to 64 °C in the presence of reporter-functionalized probes, and only perfectly matched target sequences retain the reporter strand and remain immobilized on the nanowire. Detection results from surface-enhanced Raman scattering, as only probes that remain proximal to the nanowire are detected.

each probe contains a universal primer sequence for PCR amplification. Upon successful amplification, the products can be separated by gel electrophoresis and identified by their size, which directly correlates to the miR used to initiate ligation [103]. Another hybridization-based platform with similar application to microarrays or RT-qPCR is the fully automated, digital-count profiling NanoString nCounter® system [104]. In this method, miR sequences act as a guide for adjacent hybridization of a sequence-specific capture probe labeled with biotin and a reporter probe labeled with a unique four-color, seven-position barcode. The hybridized constructs are purified and bound to a streptavidin-coated slide where a voltage is applied to elongate the molecules. The elongation allows for the digital imaging and counting of the uniquely barcoded miR targets.

Next-generation sequencing. Although each of the aforementioned techniques and platforms present beneficial approaches to miR detection with promising avenues for advancement, we anticipate NGS technology becoming the leading methodology in miR research. This does not mean other methodologies such as microarrays will disappear; on the contrary, validation of results will always be a requirement in miR research [45]. Some of the primary reasons that NGS has not been considered leading in miR bioanalysis are its cost and the relative complexity involved in analyzing the large amounts of data it produces. These issues are being resolved, however, as the technology itself matures and bioinformatics infrastructures capable of handling and processing large amounts of data are becoming more commonplace [11, 13, 105]. NGS of RNA (also called deep sequencing or RNA-seq) is perhaps the only technique that exposes the immense variation inherent in miR processing. The heterogeneity of miR (e.g., isomiR or single-nucleotide changes) can be problematic for other techniques that lack the ability to identify these other miR forms, which could potentially exert similar regulatory pressures as the

related known miR targets. However, one should keep in mind that the NGS method is not constrained to previously reported miR sequences logged in miRBase (as is the case with microarrays or RT-qPCR), and although this means novel miR sequences may be discovered, not every small RNA read obtained will be a functional miR. This is part of the issue keeping NGS technology from being widely adapted in the clinical setting—it takes considerable computational and validation effort to distinguish meaningful data from the noise.

Currently, the two leading NGS technologies are Illumina HiSeq 2500 and Sequencing by Oligonucleotide Ligation and Detection (SOLiD), especially in the realm of miR analysis. 454 is also a leading NGS technology but by 2016 will no longer be supported by Roche [106]; it is therefore not discussed here. For an excellent overview of NGS technology, readers should consult the previous review [107] on this topic. Although the relatively short maximum read length of SOLiD could be considered a disadvantage in other applications, this characteristic in miR applications is not a problem. In fact, it is well suited for miR analysis as it is the second highest throughput system (second to the Illumina HiSeq 2500 system) and offers the lowest error rate in reads as each sequence is read both forward and backward [106]. A recent publication demonstrated the viability of NGS in miR profiling by comparing it to microarray and NanoString nCounter® methods, along with validating results with RT-qPCR, analyzing cell line and xenograft samples, as well as flash-frozen and FFPE samples [35]. Results demonstrated that, as with RT-qPCR, NGS exhibits a wide dynamic range and does not suffer from the same fold-change compression as the microarray or NanoString nCounter® methods. It was also demonstrated that NGS shows good correlation between flash-frozen and FFPE samples in global miR profiling and could therefore be applicable for the analysis of clinical samples. Although NGS certainly still has a long way to go before it is as readily adopted as conventional

techniques such as microarray or RT-qPCR, it definitely shows promise in being a powerful tool for miR detection and discovery.

1.3   Bioinformatics and MicroRNA Target Prediction

The difficulty in correlating miR expression with mRNA targets for clinical applications is multifaceted but stems most directly from the inherent uncertainty with which we identify miR target sequences. As opposed to siRNA, miR does not require a perfect antisense match against a potential mRNA target, nor is the target region as clearly defined as it was when the local search area was confined to the 3′ untranslated region (UTR). Currently, bioinformatics approaches still rely on 3′ UTR matches to the seed region at the 5′ end of putative miR, although newer approaches are integrating the coding region into search parameters [108, 109]. These dual-region searches can also take advantage of the higher conservation of coding regions [109] to increase search accuracy across species. For in-depth descriptions of current identification algorithms, several excellent reviews have recently become available [22, 110–112].

Target validation presents an additional challenge in that disruption of a pathway is rarely dependent on a single miR, so it becomes difficult to conclude with certainty that an observed in vivo effect can be correlated to manipulation of one pathway interactor. Moreover, the lack of in vivo validation means that consensus-derived target prediction suffers from an incomplete set of identification parameters; as such, prediction becomes more of an educated guess, having to sort through massive databases for a short list of potentially therapeutic miRs. It is, therefore, imperative that collaborative inquiry bridges the gap that develops between the clinical and molecular methods with a strong biostatistical foundation to provide only the most relevant information

for integration into translational therapies. As the technologies discussed in this review continue to be developed and applied, this foundation will continue to expand as the function and mechanism of miR are elucidated, ultimately narrowing this gap.

List of Acronyms

T_m melting temperature – Temperature at which a double-stranded nucleic acid molecule is 50% denatured (i.e. single-stranded). 3, 8, 13–15

DNA deoxyribonucleic acid 4–8, 17–19, 23, 25

    dsDNA double-stranded deoxyribonucleic acid 14

    cDNA complementary deoxyribonucleic acid – DNA synthesized from mRNA by reverse transcription. 12–14, *see* mRNA & RT

LNA locked nucleic acid – An RNA analog modified with a 2′-O 4′-C methylene bridge that "locks" the ribose sugar into the 3′-endo conformation, which greatly improves hybridization to other nucleic acids (DNA/RNA).
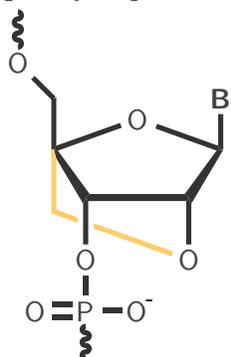
5–8, 14, 19

LOD limit of detection 19, 21, 25

miR microRNA – Small non-coding RNAs capable of altering cellular processes through post-transcriptional regulation of mRNA by initiating degradation of the mRNA or interfering with the translation process. 1–15, 17, 18, 21, 25, 27–30

    pri-miR primary microRNA – In the biogenesis of mature miR: in animals, a 55–70 nucleotide RNA exhibiting hairpin structure directly transcribed from the genome and processed by Drosha in the nucleus to yield pre-miR. 2, 5, 9, 12–14

    pre-miR precursor microRNA – In the biogenesis of mature miR: an ∼30 base-pair RNA with a terminal loop which is excised by Dicer in the

cytoplasm to yield an ∼22 nucleotide mature miR to be incorporated into the RISC. 2, 5, 9, 12–14

NB Northern blotting 1, 4–8

NGS next-generation sequencing 1, 7, 11, 27, 28

PCR polymerase chain reaction 3, 25, 27

    qPCR quantitative OR "real-time" polymerase chain reaction – In this method, DNA is amplified by the polymerase chain reaction while simultaneously being detected and quantified by fluorescence. 11–14

    RT-qPCR quantitative reverse transcription polymerase chain reaction – In this method, RNA is converted to complementary DNA (cDNA) by an enzymatic process using a reverse transcriptase. The cDNA is then amplified by the polymerase chain reaction while simultaneously being detected and quantified by fluorescence. 1, 7, 11, 12, 14, 27–29

RISC RNA-induced silencing complex – A complex comprised of multiple proteins and a ssRNA (e.g. mature miR) which acts as a guide for binding to a specific mRNA. see 2

RNA ribonucleic acid 3–6, 8, 10–15, 18, 19, 25, 27, 28

    ssRNA single-stranded ribonucleic acid 18

    dsRNA double-stranded ribonucleic acid 18

    mRNA messenger ribonucleic acid – A class of RNA responsible for relaying genetic information from DNA to ribosome where it is translated to amino acids. 2, 3, 12, 29

    ncRNA noncoding ribonucleic acid – A class of RNA demonstrating cellular functions apart from mRNA (i.e. not involved in translation). 2

    rRNA ribosomal ribonucleic acid – RNA component of the ribosome involved in protein production. 12, 15

    snRNA small nuclear ribonucleic acid – Involved in splicing of introns/exons in the nucleus. 15

tRNA transfer ribonucleic acid – RNA molecule with high secondary structure bound to an amino acid corresponding to its anticodon, which aligns with the triplet codons of mRNA at the ribosome in the translation process. 25

RT reverse transcription – In this method, RNA is converted to complementary DNA (cDNA) by an enzymatic process using a reverse transcriptase. see 11–13

SOLiD Sequencing by Oligonucleotide Ligation and Detection 28

UTR untranslated region 29

List of Terms

Argonaute Class of proteins which bind small ncRNAs (e.g. miR) and are an important part of the RISC. 10, *see* ncRNA, &

Biogenesis also biosynthesis, The process by which something is created through a metabolic or biosynthetic pathway. 2, 5

Biotin Also known as vitamin H and B_7, D-biotin is a small molecule which functions as a cofactor in carboxylase catalyzed reactions. It exhibits exceptionally strong binding to avidin proteins, a property which is heavily utilized in biotechnology and molecular biology applications. see 10, 27

    Biotinylated Being chemically modified with a biotin. see 10, 22

Drosha and Dicer Ribonuclease III family dsRNA-specific endoribonucleases involved in the biogenesis of mature miR from pri- and pre-miR. see 5

Hybridization The process by which two ssDNA/ssRNA molecules anneal through hydrogen bonding interactions between the nucleobases adenine & thymine/uracil and guanine & cytosine to form dsDNA/dsRNA. 6–8, 10, 14, 15, 17, 19–21, 23, 25, 27

isomiR MiRs containing slight sequence variations with regard to the known reference miR due to alternate pathways in miR biogenesis, such as deletion, addition, or substitution of bases. 2, 5, 13, 15, 27

Ligase specifically DNA ligase, An enzyme capable of making phosphodiester bonds to join DNA molecules or otherwise repair breaks in the nucleic acid backbone. 6, 8, 10, 13

Microarray A solid phase platform used to interrogate several known target nucleic acid sequences at once via direct hybridization. The sequences are distinguished from each other by spatial arrangement on the solid phase. 1, 7–11, 21, 27–29

Nuclease An enzyme capable of breaking phosphodiester bonds as found between nucleotides in nucleic acid molecules. 17, 18, 21, 22

Endonuclease An enzyme capable of breaking phosphodiester bonds within a nucleic acid molecule, typically at some specific sequence of nucleotides, as is the case with restriction endonucleases. 20, 23

Exonuclease An enzyme capable of breaking phosphodiester bonds at the extremities of a nucleic acid molecule, specifically by removing one nucleotide at a time from either end (e.g. $5'$ to $3'$ or $3'$ to $5'$ exonucleases). 10, 17

Polymerase An enzyme capable of synthesizing nucleic acid polymers (i.e. ssDNA/ssRNA) using a ssDNA/ssRNA molecule as a template. Some polymerases exhibit $3'$ to $5'$ exonuclease activity as part of a "proofreading" error correction mechanism which allows for the re-synthesis of mistakes. 8, 10, 20

Streptavidin A tetrameric protein isolated from *Streptomyces avidinii* exhibiting an extremely high affinity for the small molecule biotin ($K_d \sim 10^{-15}$) see 10, 19, 22, 27

Chapter 2

Bioluminescent Stem-Loop Probes for Highly Sensitive Nucleic Acid Detection[3,4]

Overview

Here, we report the first bioluminescent stem-loop probe, which is 50-fold more sensitive and able to achieve a LOD 25-fold lower than fluorescent stem-loop probes. Chemical generation of a signal from *Renilla* luciferase reduces background noise for improved quantitative utility in nucleic acid biomarker detection.

2.1 Perspective

Fluorescent stem-loop probes (FSLPs), termed "molecular beacons," were first developed by Tyagi and Kramer in 1996 [114]. A typical FSLP design consists of a 15–30 nt loop flanked by two short 5–7 nt self-complementary regions (stem) modified with a fluorophore and a quencher. The probe switches between closed and open conformations upon hybridization of a complementary target to the loop. Fluorescence is quenched by direct energy transfer or fluorescence resonance energy transfer (FRET) [115] in the closed conformation, and in the open conformation fluorescence may be observed by a suitable spectroscopic method. The stem-loop design offers superior specificity and reduces procedural complexity by eliminating the need for removal of excess/unhybridized probes prior to an instrumental reading. Since their first reported

---

3   [113]   E. A. Hunt and S. K. Deo. "Bioluminescent stem-loop probes for highly sensitive nucleic acid detection". In: *Chemical communications* 47.33 (2011), pp. 9393–5, reproduced by permission of The Royal Society of Chemistry.
4   The original *Supplementary Information* to the article has been included as *Methods* on page 44.

use, FSLPs have become a widely used tool in chemical, biological, and medical sciences, including more specific applications in qPCR, SNP and genetic variation screening, in vitro and in vivo detection of DNA and RNA, and small molecule and protein detection [116–118]. FSLPs have also been adapted for applications requiring immobilization to a solid phase, such as biochips and fiber optic biosensors [116, 119].

Background fluorescence from the excitation source is one limitation on FSLP sensitivity and is often accentuated when immobilizing the probes to a solid phase due to autofluorescence of the substrate or steric interactions which prevent the stem from closing properly [116]. Perhaps the greatest limitation to FSLP sensitivity is that the effective number of probes able to generate a signal directly corresponds to the number of complementary target molecules present in the assayed sample. In the past decade, a number of advancements have been made to address this issue [116]. Locked nucleic acid (LNA) and peptide nucleic acid (PNA) have been used to improve target specificity by stabilizing the stem duplex in the closed conformation, thereby reducing the inherent background noise arising from inadequate quenching [120–122]. Less conventional approaches include implementing poly-thymidine·Hg^{2+} duplex and poly-thymidine·PNA-poly-adenine triplex stems for more stable closure [123, 124]. The problem with these improvements is that the features employed to increase target specificity are typically also detrimental to sensitivity [125]. An additional approach to decrease background noise has been through the use of alternative quenchers including graphene oxide, gold nanoparticles, and "superquencher" assemblies [126–128]. Other approaches have attempted to enhance the probe signal by swapping the traditional organic dye fluorophore for a more robust reporter, such as inorganic quantum dots [129] or phosphorescent Eu^{3+} complexes [130]. Apart from the Eu^{3+} complex, however, the problem of background fluorescence caused by external excitation still exists.

We have approached this issue from a new angle and developed the first reported bioluminescent stem-loop probe (BSLP). Utilizing a mutant version of the photoprotein *Renilla* luciferase containing eight mutations (Rluc8) [131] as the reporter

(em. 485 nm) and 4-(4'-dimethylaminophenylazo) benzoic acid (DABCYL) as the quencher (ex. 475 nm), our design (Figure 2.1 on the following page) exhibits a unique advantage over traditional FSLPs.

2.2   Discussion

Rluc8 generates light through the oxidative decarboxylation of its native substrate coelenterazine to coelenteramide, which eliminates nearly all background noise as external excitation is not necessary. Bioluminescent reporters have been shown to yield superior sensitivity and lower detection limits in sensing system development [132, 133]. A large molecule such as Rluc8 (39.5 kDa) is much bulkier than the typical organic dyes used in FSLP design. In initial designs for the BSLP, the steric effects of having a large molecule attached to the probe resulted in a very low T_m for the stem. This prevented the probe from closing tightly and properly quenching the bioluminescent signal. For this reason, the BSLP used in this work was designed to detect a 22 nt target sequence and contained a 6 bp, 83% GC stem to ensure adequate closure. The Rluc8 was expressed and purified from *E. coli* [133] (see Section 2.4.1 on page 44) and chemically conjugated to the DABCYL-modified oligonucleotide sequence (Eurofins MWG Operon, see Section 2.4.2 on page 44 for the full sequence) using succinimidyl 4-formylbenzoate (SFB) and succinimidyl 4-hydrazinonicotinate acetone hydrazone (SANH) (see Section 2.4.2 on page 44 and Figure 2.4 on page 46 for a synthetic scheme). Analysis by SDS-PAGE (Figure 2.5 on page 47) showed that polyconjugates (i.e. multiple DNA probes attached to one Rluc8) are formed; we believe that this helps achieve a higher level of quenching and is not detrimental to the proper functioning of the BSLP. The BSLP was purified by affinity chromatography using a biotin-modified thymidine incorporated into the stem as an affinity handle (left side of Figure 2.1 on the following page, see Section 2.4.3 on page 47). The concentration of purified BSLP was determined by bicinchoninic acid (BCA) protein assay and by UV-Vis absorbance of the DABCYL quencher. The concentration
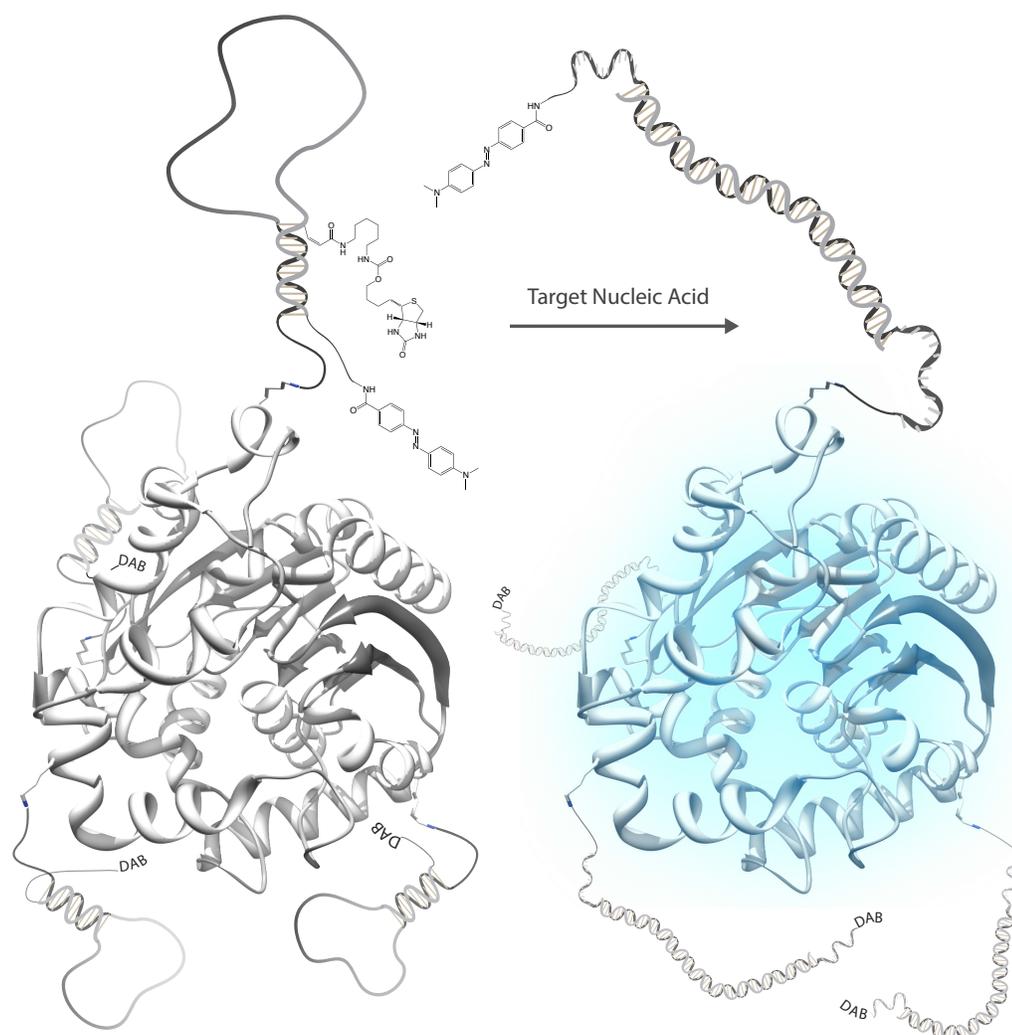
Figure 2.1. The structure of the BSLP (not to scale). Rluc8 is attached via exposed lysine residues. Polyconjugates of multiply-modified Rluc8 are formed (see Section 2.4 on page 44). The biotin-modified thymidine is only shown in the top SLP of the closed conformation (left). Upon binding to target, the probe changes conformation and bioluminescence can be observed. Abbreviations: DAB, DABCYL quencher.

obtained by UV-Vis absorbance of the DABCYL was used for assay development because it was more accurately representative of the number of DNA stem-loops available for target binding. For direct comparison, a traditional FSLP was used in parallel with the BSLP in all experiments. The FSLP was synthesized (Eurofins MWG Operon) with the same sequence as the BSLP, utilizing 6-carboxyfluorescein (6FAM) as the fluorescent reporter (ex. 495 nm, em. 524 nm) and Black Hole Quencher 1 (BHQ1) as the quencher (ex. 534 nm). After initial calibrations were carried out per previously established methods [114], S/N was optimized by holding the target oligonucleotide concentration constant while varying the concentration of each probe (see Section 2.4.5 on page 50). Using 10 pmol of FSLP gave an optimal S/N ratio of 1.6, whereas 0.5 pmol of BSLP gave an optimal S/N ratio of 9 (see part *a* of Figure 2.2 on the next page). At 0.5 pmol, the FSLP signal was too weak to be distinguished from background noise (see Figure 2.10 on page 53 for the calibration curves used to create part *a* of Figure 2.2 on the next page).

This drastic improvement in signal amplitude is characteristic of bioluminescent reporters and the inherently low background signal associated with the absence of external excitation. This increased signal amplitude allows the BSLP to generate a reliable quantitative response even when added at a low concentration. The exceptional sensitivity afforded by the bioluminescent signal was evidenced in the calibrations performed at the optimal probe concentrations of ~3 nM BSLP and ~67 nM FSLP (see part *b* of Figure 2.2 on the following page). By slope comparison, the BSLP signal is 50-fold more sensitive than the FSLP signal; consequently, the BSLP was able to achieve a LOD of 0.4 nM, whereas the FSLP exhibited a LOD of 10 nM (see part *b* of Figure 2.2 on the next page). This 25-fold lower LOD is an exceptional improvement over the traditional FSLP as well as many of the advancements made to improve the traditional fluorescence method (Table 2.1 on page 42).

In addition to these characterization studies, we have demonstrated the viability of the BSLP in directly detecting nucleic acid levels in a human serum matrix. MicroRNAs (miRNA) are small (approximately 22 nt) regulatory RNAs, which control
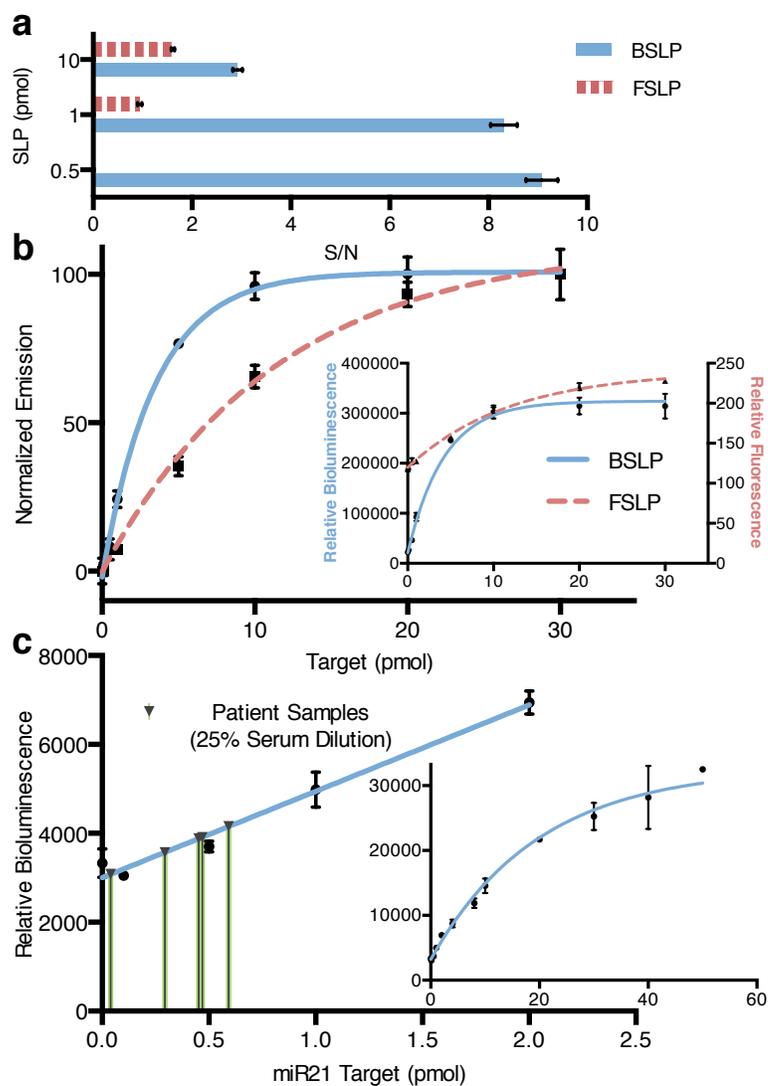
Figure 2.2. Enhanced sensitivity using a bioluminescent reporter versus a traditional fluorophore. (a) The BSLP was varied from 0.5 to 20 pmol with target held constant at 4 pmol. The FSLP was varied from 1 to 50 pmol with target held constant at 10 pmol. (b) A calibration was performed using the optimal conditions determined for each probe. (c) Serum samples from patients diagnosed with varying stages of breast cancer were directly tested for miR21 levels (green triangles) in reference to a calibration done in a normal mouse serum matrix spiked with synthetic DNA miR21 target. Abbreviations: RLU, relative luminescence units.

Table 2.1. LOD comparison for several FSLP modifications to BSLP. Abbreviations: LOD, limit of detection; BSLP, bioluminescent stem-loop probe; FSLP, fluorescent stem-loop probe; LNA, locked nucleic acid; PNA, peptide nucleic acid; T, thymidine.

Probe type	LOD (nM)	Reference
BSLP	0.4	this study
FSLP	10	this study
LNA	7.4	[121]
Graphene oxide	2	[126]
PNA	1.6	[122]
Fiber optic	1.1	[119]
$Hg^{2+}.T_7$	0.5	[123]
Eu^{3+} complex	0.5	[130]

a wide variety of cellular functions though the degradation and/or translational repression of mRNA [134]. Recent discoveries indicate that miRNA expression patterns are drastically altered in cancerous cell states [135, 136] and in some cases are considered oncogenic [137, 138]. As such, miRNAs can be used as biomarkers for the onset of cellular disease states [139, 140]. We designed a BSLP in which the loop region was complementary to the mature sequence of miRNA miR21

<div align="center">

TAGCTTATCAGACTGATGTTGA

</div>

which has been shown to be up-regulated in breast [139] and several other forms of cancer [141, 142]. Of six serum samples drawn from patients in varying stages of breast cancer, four demonstrated detectable levels of miR21 without any extraction step (part *c* of Figure 2.2 on the previous page, also see Section 2.4.6 on page 51). The variance between samples is credited to the disparity in type and prognosis of breast cancer between patients. While further studies will be done on more specific subsets of patient samples, this initial data shows that direct detection of miRNA in human serum is possible using our BSLP. The stem-loop hybridization format also affords a high degree of specificity and is able to distinguish target sequences containing a single nucleotide mismatch (Figure 2.3 on the next page, see Section 2.4.6 on page 51 for detail on mismatch target design).
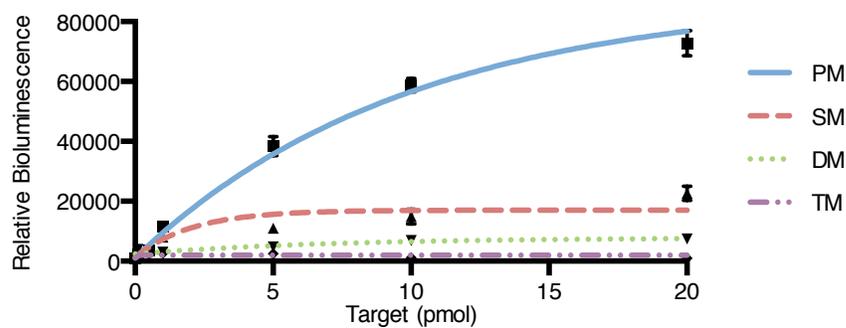
Figure 2.3. SLPs exhibit a high degree of specificity due to the kinetic energy barrier associated with the probe stem duplex. This allows the BSLP to distinguish target containing a single nucleotide mismatch. Abbreviations: PM, perfect match; SM, single mismatch; DM, double mismatch; TM, triple mismatch; RLU, relative luminescence units.

2.3 Conclusion

To conclude, we have demonstrated as a proof of concept that a bioluminescent photoprotein such as Rluc8 can be used as the reporter in a stem-loop probe for the detection of nucleic acids. The use of Rluc8 as a bioluminescent reporter eliminates background noise from external excitation. It is also an extremely cost effective reporter as it can be expressed and purified easily from genetically modified *E. coli*. This first-generation bioluminescent platform has shown promising sensitivity and future modifications will expand its application in detection and imaging. The glow-type kinetics of Rluc8 bioluminescence allows for a stronger signal to be obtained by increasing integration time, but will require further modification to lower background such as replacing the single quencher moiety with a molecular assembly of multiple quenchers to further improve sensitivity [128]. We are currently developing a BSLP using *Gaussia* luciferase, which we have demonstrated to withstand heating up to 90 °C with minimal loss in activity (data not shown), for qPCR applications. Like the FSLP, the BSLP can be modified to accommodate any number of adaptations that may improve the sensitivity through background noise reduction, and given the number of significant improvements these modifications have had on traditional

FSLPs, it is reasonable to consider that the extension of these modifications to the BSLP could also greatly improve its detection capabilities.

2.4 Methods[5]

2.4.1 Expression and Purification of *Renilla* Luciferase

The gene for the Rluc8 mutant [131] of the wild-type *Renilla* luciferase was cloned into the pBAD/Myc-HisA vector (Invitrogen) and transformed into *E. coli* LMG194 for expression. 200 mL cultures were grown at $37\,^{\circ}$C to an OD_{595nm} of $\sim$0.6, at which point expression was induced with 0.2% L-arabinose. Expression was allowed to continue for 5 h at $37\,^{\circ}$C. Cells were collected by centrifugation and lysed by sonicating (2 min, 5 s on/off pulse) in 20 mM sodium phosphate, 500 mM sodium chloride, pH 7.2 buffer. Crude extract was purified by affinity chromatography using a Ni·NTA agarose solid phase. The protein was allowed to bind for 2 h at ambient temperatures before eluting with 500 mM imidazole. The purified Rluc8 was then concentrated using 10 kDa MWCO Millipore filters; buffer-exchanged into a 100 mM sodium phosphate, 150 mM sodium chloride, pH 7.2 buffer; and stored at $4\,^{\circ}$C. The purified Rluc8 is stable for several months.

2.4.2 Conjugation of the Stem-Loop Probe

The custom oligonucleotide sequence:

$5'$[AminoC6]

GGGGGA TCAACATCAGTCTGATAAGCTA [Bio-dT]CCCCC

[Dabcyl-Q]$3'$

was ordered from Eurofins MWG Operon. The $5'$ amino modification was chemically converted to a benzaldehyde moiety by succinimidyl 4-formylbenzoate (SFB, Pierce – Thermo Fisher Scientific Inc.) in 50× excess in a 50 mM sodium phosphate, 75 mM

---

5 Originally published as *Supplementary Information* to the article.

sodium chloride, pH 7.2 buffer. Amine residues of purified Rluc8 were modified with succinimidyl 4-hydrazinonicotinate acetone hydrazone (SANH, Pierce – Thermo Fisher Scientific Inc.) in $10\times$ excess in a 100 mM sodium phosphate, 150 mM sodium chloride, pH 7.2 buffer. Once a stable amide bond is formed between 4-hydrazinonicotinate and Rluc8, the alkyl hydrazone is hydrolyzed by buffer exchanging into a 100 mM sodium citrate, 150 mM sodium chloride, pH 6.0 buffer, which yields the aldehyde-reactive hydrazine moiety. SFB-modified oligonucleotide was added in $1.5\times$ excess to form a stable hydrazone conjugate (Figure 2.4 on the next page).

A large molecule such as Rluc8 (39.5 kDa) is much bulkier than the typical organic dyes used in FSLP design. In initial designs for the BSLP, the steric effects of having a large molecule attached to the probe resulted in a very low T_m for the stem. This prevented the probe from closing tightly and properly quenching the bioluminescent signal, therefore a six base pair stem containing 83% GC pairs was used to increase the T_m to an adequate level. To remove the excess Rluc8 remaining in solution after conjugation (Figure 2.5 on page 47), we incorporated a biotin affinity-handle on the probe stem.

It is also evident from the SDS-PAGE analysis that while the stoichiometry of the conjugation reaction is controlled to promote a 1:1 ratio of probe to Rluc8, the formation of polyconjugates is inevitable and quite prominent. An estimation of the average number of probe molecules attached to each Rluc8 was performed by quantifying the purified conjugate by BCA protein assay, and by the molar extinction coefficient for the DABCYL modification. By these two methods, there was an approximately 4:1 ratio of probe to Rluc8. While the BCA assay is not an extremely accurate or precise method of protein quantification, this does provide further evidence that polyconjugates are readily formed in the conjugation reaction. Due to the size of Rluc8, however, the formation of polyconjugates should not inhibit normal functionality of the BSLP. It may, however, cause higher order polyconjugates to be more effectively quenched than those containing only one or two DABCYL functionalities.
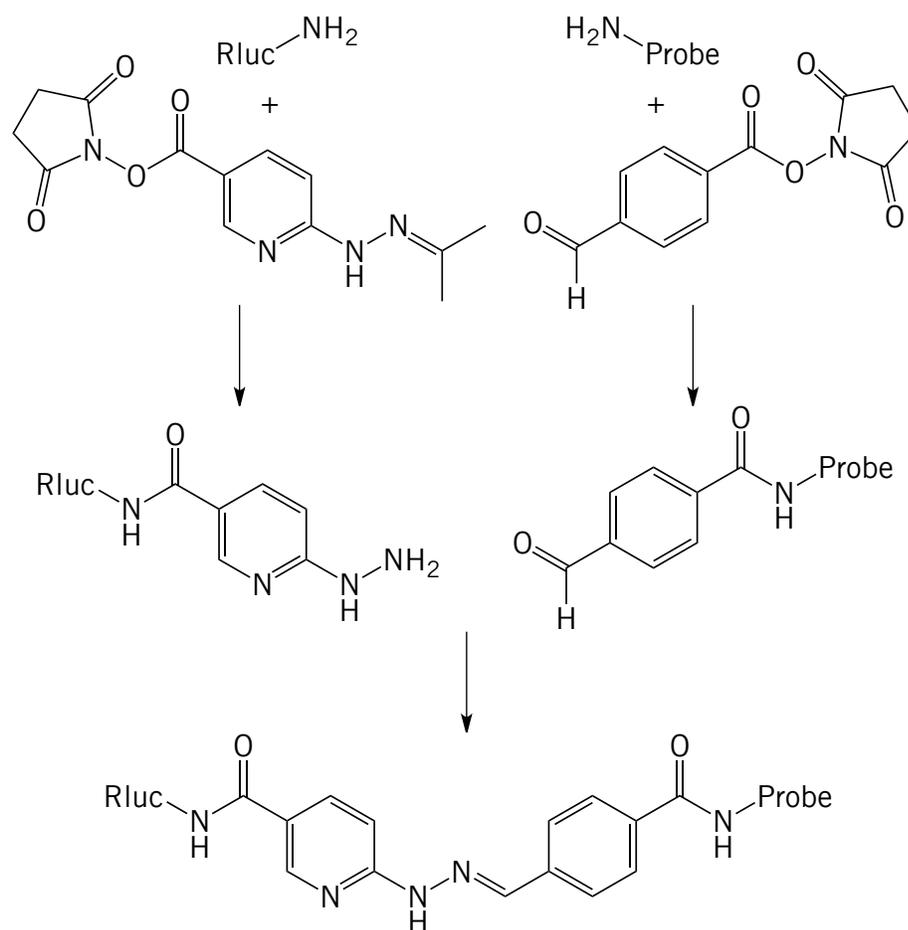
Figure 2.4. Schematic of chemical reaction for attaching the stem-loop probe to Rluc8. The SANH succinimidyl ester reacts with amine side chains on Lys and/or the N-terminus of the protein, and the SFB succinimidyl ester reacts with the 5′ amino-modification on the stem-loop probe. Transfer of the SANH-Rluc8 intermediate to pH 6.0 citrate buffer hydrolyzes the alkyl hydrazone of SANH, yielding the hydrazine functionality which readily reacts with the benzaldehyde of the SFB-Probe intermediate. The final product is a stable covalent linkage between the two biomolecules.
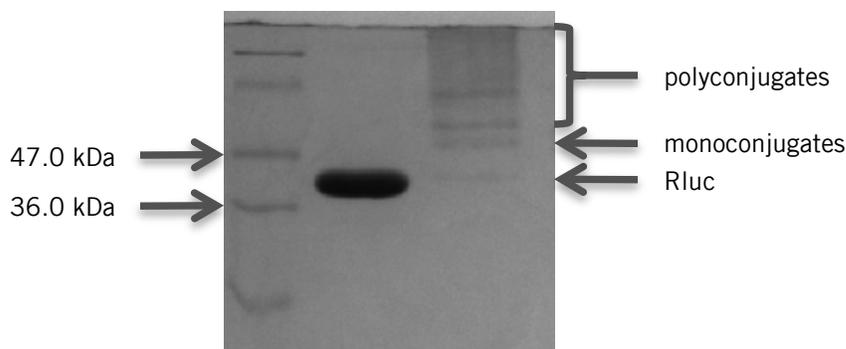
Figure 2.5. A 12.5% SDS-PAGE gel of the non-purified BSLP following Rluc8 conjugation, stained with Coomassie Blue. Lane 1: Fisher BioReagents EZ-Run Prestained Protein Marker 0-118 kDa; Lane 2: purified Rluc8 (control); Lane 3: BSLP non-purified conjugate. Note that lanes 2 and 3 are not concentration matched. The discernable bands above that of the monoconjugate in lane 3 represent polyconjugate BSLPs.

2.4.3 Purification of the Bioluminescent Stem-Loop Probe

A purification column was prepared using 1 mL of monomeric avidin-immobilized agarose. The packed resin was washed with 8 mL of wash buffer (100 mM sodium phosphate, 150 mM sodium chloride, pH 7.0). Next, the column was washed with 6 mL of elution buffer (wash buffer containing 2 mM D-biotin) to block any irreversible biotin binding sites on the resin. The column was then regenerated with 12 mL of regeneration buffer (100 mM glycine, pH 2.8) followed by 8 mL of wash buffer. A 5 μL aliquot of the non-purified stem-loop probe was removed for comparison, and the remainder was added to the column and forced into the bed with 1 mL of wash buffer. The column was then sealed and incubated at ambient temperature for 1 h. To wash the probe, 1 mL aliquots of wash buffer were collected until the bioluminescent emission ($\lambda_{\text{max em.}} = 485\,\text{nm}$, Tecan GENios) resulting from the addition of the native substrate coelenterazine returned to baseline. The probe was then eluted with 200 μL aliquots of elution buffer until the absorbance of the 3′ DABCYL modification ($\lambda_{\text{max abs.}} = 475\,\text{nm}$, $\varepsilon_{475}\,\text{nm} \approx 30,000$) returned to baseline (Figure 2.6 on the next page). Elution fractions were then collected based on their 475 nm absorption and subsequently pooled and concentrated using Millipore – Microcon YM-30 Centrifugal

Filter Units. The column was washed with 4 mL of regeneration buffer and stored upright at $4\,^{\circ}\text{C}$ after washing with 5 mL of wash buffer containing 0.01% sodium azide.
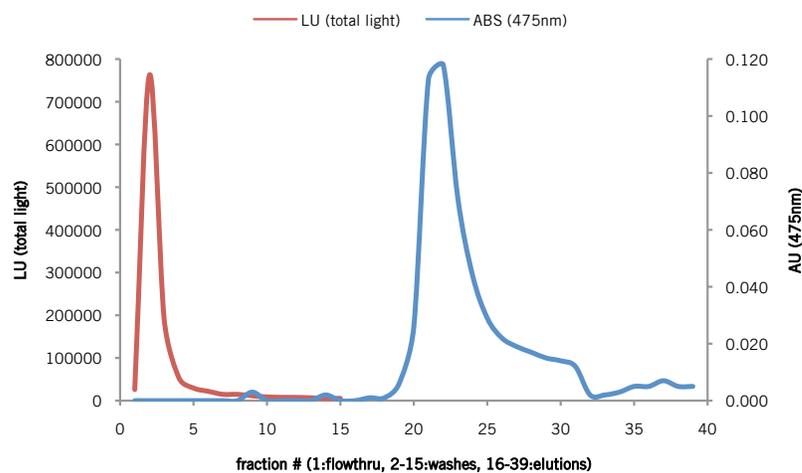
Figure 2.6. Purification of the stem-loop probe conjugate. Rluc8 elution was measured by total light bioluminescence in the presence of coelenterazine. Probe elution was monitored by the presence of the $3'$ DABCYL modification absorbing at 475 nm.

Purification of the BSLP was necessary due to background noise caused by any unconjugated Rluc8. To address this issue, the thymidine in the stem of the probe was replaced with a biotin-modified thymidine (bio-T). This afforded a convenient affinity handle for purification without detriment to the structural integrity of the SLP sequence needed for proper stem-loop formation. The BSLP was then conjugated to Rluc8 using SANH/SFB chemistry and purified by affinity chromatography using an agarose solid phase modified with monomeric avidin. The monomeric avidin-biotin binding affinity is considerably lower than the typical avidin-biotin interaction ($K_d = 10^{-8}\,\text{M}$ versus $10^{-15}\,\text{M}$) and biotinylated molecules may be displaced with biotin containing buffer. Purification effectively removed free Rluc8, lowering the residual bioluminescent background and allowing for an approximately 6-fold signal change to be seen between the fully open and closed conformations (Figure 2.7 on the following page).
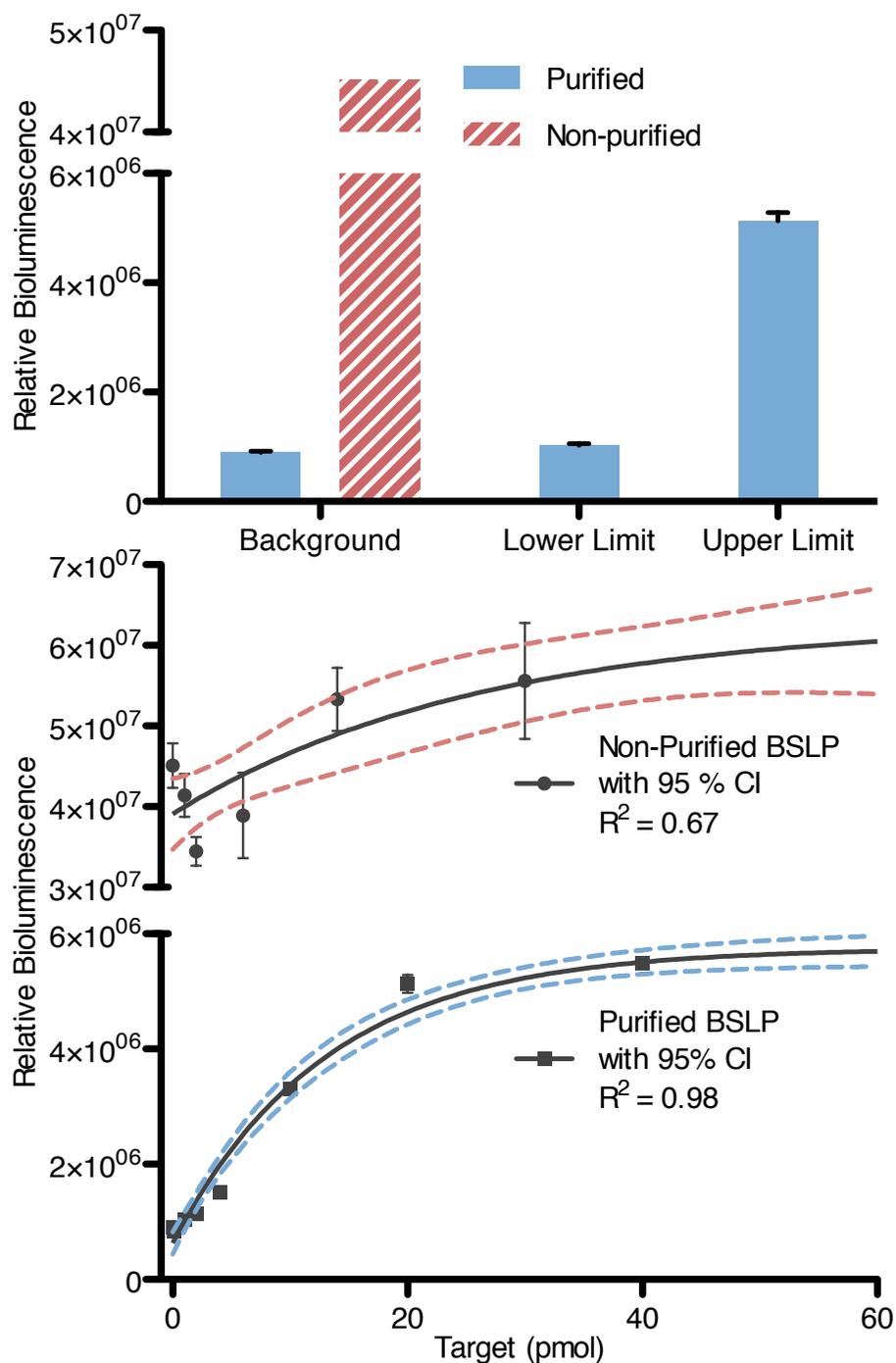
Figure 2.7. The background noise caused by free Rluc8 is drastically reduced after purification. Without purification, signal at the upper and lower limits of the calibration is masked completely and no target can be detected.

The fully purified BSLP is stored in 100 mM sodium phosphate, 1 mM EDTA, pH 7.4 at 4 °C, and is stable with only minor losses in activity for over 1 year. We have noticed no aggregation of the purified BSLP, but it was centrifuged for 1 min at $16,000 \times$g before every use to ensure the removal of any insoluble matter. It should be noted that aggregation is unlikely due to the high charge density on the surface of the Rluc8 once functionalized with DNA.

2.4.4 Assay Setup

A standard white polypropylene 96-well plate was washed three times for 30 minutes at ambient temperature with 0.5% BSA in 100 mM sodium phosphate, 1 mM EDTA, pH 7.4 buffer. The plate was then rinsed and dried. Calibrations were performed using varying amounts of target mixed with 0.5 pmol of BSLP in a 100 mM sodium phosphate, 1 mM EDTA, pH 7.4 buffer and allowed to hybridize at ambient temperature for 3 h (hybridization for 1 h will also provide sufficient quantitative signal). Total-light bioluminescent readings were taken using a PerkinElmer Victor X Light bioluminometer equipped with a syringe auto-injector. The bioluminescent reaction was initiated by injecting 50 μL of a 2.5:1000 dilution ($\sim$6 μM) of 1 mg/mL (in acidified methanol) native coelenterazine (NanoLight Technology) prepared in 100 mM sodium phosphate, 1 mM EDTA, pH 7.4 buffer immediately before use (1 mg/mL stock made in degassed, acidified ethanol and stored under argon at -80 °C) and shaking for 5 s. Total well volume including injected coelenterazine was held at 150 μL. The amount of coelenterazine used does not have a noticeable effect on the background signal; therefore, the amount of coelenterazine injected was in excess of what the Rluc8 was able to process.

2.4.5 Stem-Loop Probe Characterization

In order to compare the BSLP to a traditional FSLP, a FSLP using a 6-carboxy-fluorescein (6FAM)/Black Hole Quencher 1 (BHQ1) fluorophore/quencher pair was
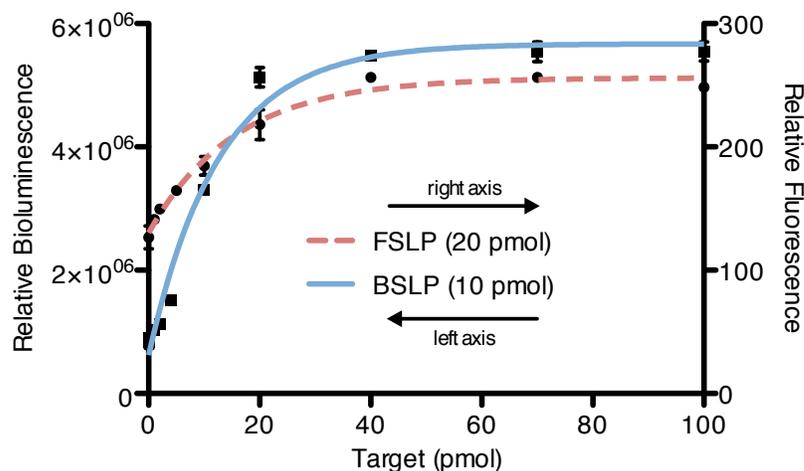
Figure 2.8. Initial calibrations run according to previous methods [114].

ordered with the same sequence (minus the bio-T modification). The FSLP was characterized by fixing the target concentration at a point in the middle of the linear portion of the initial calibration curve (Figure 2.8, run according to previously published methods) [114] and varying the concentration of FSLP used to achieve the best S/N ratio (Figure 2.9 on the following page and Figure 2.10 on page 53). The same procedure was also done for the BSLP (Figure 2.10 on page 53). The optimal concentration for each SLP was chosen by varying the amount of SLP while holding the target concentration constant. The concentration at which the best S/N ratio was achieved was used for all following assays. The decreasing trend following the point of maximum S/N seen in Figure 2.10 on page 53 for both SLPs (less prominent in the FSLP) occurs due to the accumulation of background noise, a byproduct of incomplete quenching in the closed state which diminishes the sensitivity of the response. The trend presents itself differently in the BSLP because there is no background noise from an external excitation source, therefore the S/N is much higher, especially at lower concentrations.

2.4.6 MicroRNA Detection in a Human Serum Matrix

Normal mouse serum spiked with miRNA target miR-21:

TAGCTTATCAGACTGATGTTGA
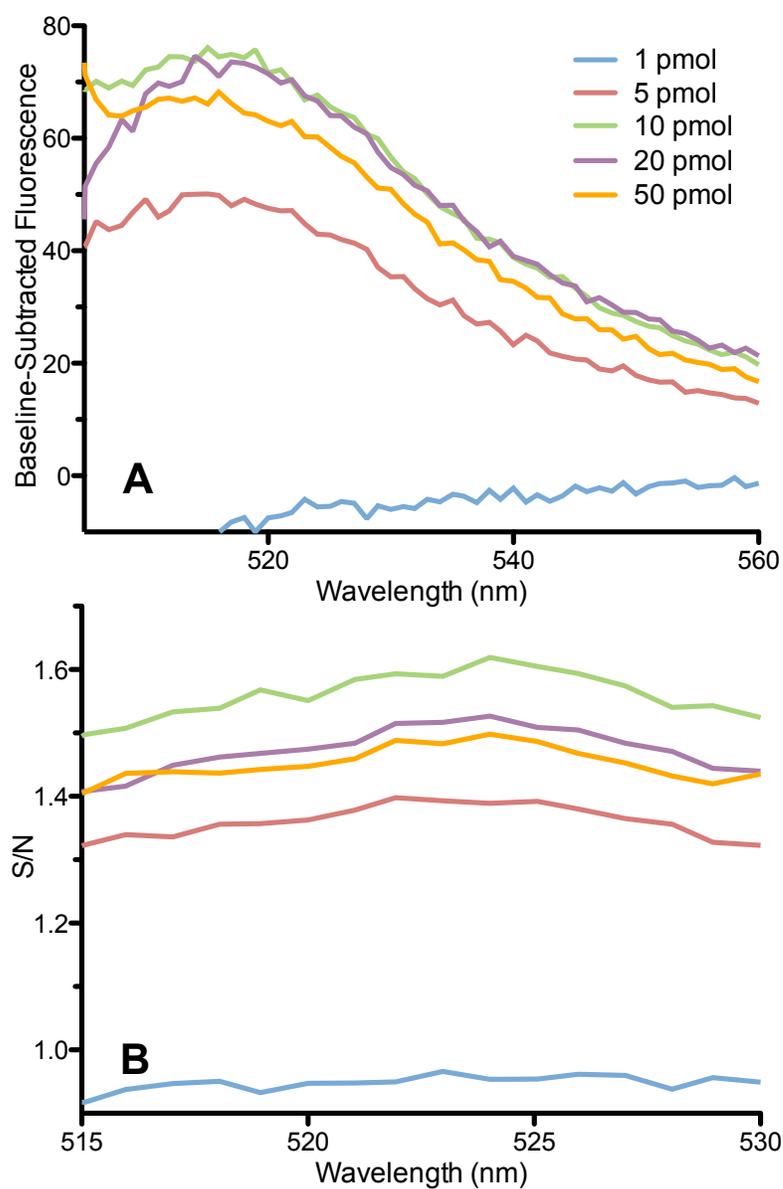
Figure 2.9. 10 pmol of target was used to characterize the FSLP, and 4 pmol for the BSLP. These points were chosen because they both fell in the middle of the initial calibrations run according to previous methods [114]. (A) The fluorescent emission of varying amounts of FSLP in the presence of 10 pmol of target. (B) The background normalized S/N response of varying amounts of FSLP in the presence of 10 pmol of target.
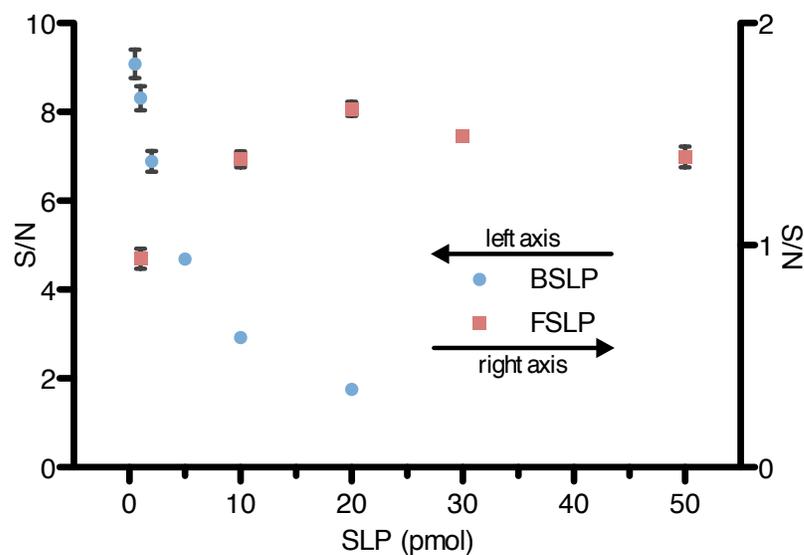
Figure 2.10. To determine the optimal concentration of SLP to use in the assay design, the amount of SLP was varied while holding the concentration of target constant.

was used to prepare a calibration curve. All serum samples were diluted to 25% initial concentration with 100 mM sodium phosphate, 1 mM EDTA, pH 7.4 buffer and boiled for 5 min. Once cooled, 50 μL of the serum was added to each well before the addition of synthetic target or BSLP. The total well volume before coelenterazine injection was still held at 100 μL.

The mismatch probes contained the following mutations to the wild-type miR21 sequence: single mismatch (SM) – T6A; double mismatch (DM) – A12T, G18C; triple mismatch (TM) – T5A, A12T, G18C.

Chapter 3

Truncated High-Activity Variants of *Gaussia* Luciferase Expressed in *Escherichia coli*[6]

Overview

Gaussia luciferase (Gluc)—with its many favorable traits such as small size, bright emission, and exceptional stability—has become a prominent reporter protein for a wide range of bioluminescence-based detection applications. The five internal cystine residues crucial to functional structure formation, however, make expression of high quantities of soluble protein in bacterial systems somewhat difficult. In addition to this challenge, the current lack of structural data further complicates the use of Gluc for in vitro applications, such as biosensors, or cellular delivery, both of which rely heavily on robust and reproducible bioconjugation techniques. While Gluc is already appreciably small for a luciferase, a reduction in size that still retains significant bioluminescent activity would be very beneficial in biosensor design and cellular transport studies. We have developed truncated variants of Gluc, which maintain attractive bioluminescent features, and have characterized their spectral and kinetic properties. Additionally, a C-terminal linker has been incorporated into these variants that can be used for reliable, specific modification through tyrosine-based bioconjugation techniques.

---

[6]   *Supplementary Information* has been included as *Appendix I* on page 111.

3.1 Perspective

In recent years, *Gaussia* luciferase (Gluc)—from the mesopelagic copepod *Gaussia princeps*—has emerged as a prominent reporter protein for bioluminescent detection applications. Gluc exhibits several favorable traits, including high luminescent output, small size ($\sim$20 kDa), and excellent stability (including thermostability) in part due to the high number of cystine moieties that help form its tertiary structure. The wild-type (WT) sequence for Gluc possesses a 17 amino acid N-terminal excretion tag, and is readily expressed in highly active form in mammalian cell systems [143, 144]. As such, Gluc has become a popular reporter for monitoring in vivo biological events.

While there is no doubt that Gluc is an excellent bioluminescent reporter for bioanalytical and clinical research applications, there are still some characteristics of Gluc that make it difficult to work with, and as such there has been a number of publications disclosing mutations made to Gluc in order to improve upon or add to its bioluminescent properties. In 2008, Inouye and Sahara identified two homologous domains present in Gluc, and upon isolating them both, characterized the effects of pH and halogen ions on their bioluminescent properties [145], thus setting a precedent to guide further work on optimizing the bioluminescent reaction of Gluc for specific applications.

In 2009, Maguire et al. created a variant of Gluc containing the single point mutation M43I, which provided glow-type kinetics in the presence of 0.1% Triton X-100 [146]. This enhanced stability of the bioluminescent emission was favorable for streamlining high-throughput assays utilizing Gluc as a reporter. In 2011, Kim et al. developed several variants of Gluc demonstrating "superluminescent" properties through a semi-rational site-directed approach targeting the hydrophilic region near the end of the second homologous domain in Gluc. One particular variant, F89W/I90L/H95E/Y97W (Monsta), demonstrated an $\sim$5-fold increase in intensity compared to WT Gluc as well as a red-shift in emission to $\sim$503 nm

[147]. More recently in 2013, Degeling et al. created a variant, L30S/L40P/M43V (Gluc4), which demonstrated exceptional glow-type kinetics both in the presence and absence of Triton X-100 [148].

In addition to improvements made to the bioluminescent properties of Gluc, there have also been several studies performed to improve the ease by which Gluc is able to be produced. Gluc has traditionally been expressed in mammalian systems as the number of necessary cystine residues make folding in prokaryotic systems difficult. Some solutions to this issue involve expression of Gluc as a fusion with a solubilizing partner such as a synthetic IgG-binding domain [149], or production of Gluc by cell-free protein synthesis [150]. In 2010, Rathnayaka et al. demonstrated that the use of the pCold expression system enhanced the amount of Gluc present in the soluble fraction when expressed in an LMG194 *E. coli* host [151].

The difficulty of production of Gluc in high concentration and pure form and the lack of structural data, limits its use for in vitro applications such as biosensors or cellular delivery where reliable bioconjugation is a necessity. Inouye et al. introduced a cysteine residue at the end of a linker in an effort to make bioconjugation to Gluc more straightforward [152]. While the use of cysteine residues with traditional maleamide chemistries is well known, it also introduces some extra complications when considering the number of cysteine residues in Gluc involved in disulfide bonds and their sensitivity to reducing conditions, which are often necessary to prevent dimerization before conjugation.

Even considering this previous work, there are still improvements that can be made to Gluc that will make it easier to work with in a variety of applications and easier to produce in a more affordable and facile expression system, *E. coli*. Expanding on previous work, truncated Gluc variants were created that are characterized by enhanced luminescence features, convenient size, and accessible and flexible chemical handles for specific, reliable bioconjugation through a tyrosine residue. The use of a cold-shock expression system in conjunction with a less

reducing environment allowed for the expression and purification of large amounts of soluble, properly folded Gluc from an *E. coli* host without issue.

3.2   Results

3.2.1   Variants of *Gaussia* Luciferase

All variants of Gluc created in this study are listed in Table 3.1 (see *Mutation and Truncation*, Section 3.5.2 on page 73 of the *Methods* for more detail).

Table 3.1. Gluc variants created in this study. The numbering of mutations is from the beginning of Gluc sequence used in this study and therefore may not match other publications. The M60I mutation differs from the M43V mutation found in the Gluc4 variant produced by Degeling et al. [148] and instead utilizes the M43I mutation from the variant produced by Maguire et al. [146]. As such, the variant was renamed 4luc for clarity. The plasmids and variants containing the C-terminal tyrosine linker are denoted by a Y following their name. The truncated variants and plasmids containing them are denoted by a t preceding the variant name.

Plasmid	Variant	Mutations	Ref.
pGluc(Y)/ptG(Y)	Gluc(Y)/tG(Y)	*WT (i.e. none)*	–
pMonsta(Y)/ptMon(Y)	Monsta(Y)/tMon(Y)	F89W, I90L, H95E, Y97W	[147]
p4luc(Y)/pt4(Y)	4luc(Y)/t4(Y)	L47S, L57P, M60I	[146, 148]

3.2.2   Expression and Purification of *Gaussia* Luciferase Variants from *Escherichia coli*

Gluc was purified in high concentration from *E. coli*– $\sim$15-20 mg per 1 L culture was recovered from the soluble fraction by immobilized-metal affinity chromatography (IMAC) following cell lysis. Figure 3.1 on the following page shows the full-size and truncated variants of Gluc produced in this study. The calculated molecular weight for the full-size Gluc variants with C-terminal tyrosine linker is $\sim$22.1 kDa and $\sim$14.4 kDa for the truncated variants. However, the purified proteins appear 4–5

kDa larger when analyzed by SDS-PAGE. This was also observed by Rathnayaka et al. [151].
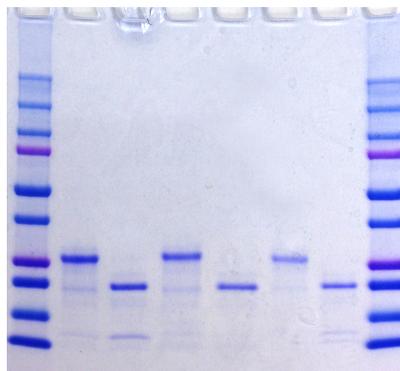
Figure 3.1. SDS-PAGE analysis of purified Gluc variants with C-terminal tyrosine linker. The first and last lanes contain Precision Plus Protein$^{\mathrm{TM}}$ Dual Color Standard (Bio-Rad Laboratories Inc., Hercules, California). Samples were loaded in the following order: GlucY, tGY, MonstaY, tMonY, 4lucY, t4Y. These samples were not treated with protease inhibitor and some degradation is visible.

3.2.3   Bioluminescence Spectra of *Gaussia* Luciferase Variants

Spectra were obtained for each Gluc variant using native coelenterazine as the substrate for the bioluminescent reaction. Each full-size variant containing a C-terminal tyrosine linker was compared to the corresponding "native" sequence without the C-terminal tyrosine linker. It was determined that the addition of a C-terminal tyrosine linker had no effect on the bioluminescent emission (see Figure 3.2a on page 60). The truncation of Gluc, however, did introduce a 10–15 nm red-shift (see Figure 3.2b on page 60) in the bioluminescent emission. The emission maxima for each variant (with tyrosine linker) are listed in Table 3.2 on the next page.

Table 3.2. Approximate emission maxima (em$_{\text{max}}$) for GlucY variants.

	em$_{\text{max}}$ (nm)
GlucY	480
MonstaY	493
4lucY	490
tGY	495
tMonY	491
t4Y	488

3.2.4   Kinetic Parameters of *Gaussia* Luciferase Variants

Michaelis–Menten kinetic parameters for each of the GlucY variants are listed in Table 3.3. The Michaelis–Menten plots used to derive these parameters are shown in Figure A1.3 on page 113. The kinetic data exhibit two distinct features requiring a slightly modified Michaelis–Menten model in order to derive typical kinetic parameters.

Table 3.3. Kinetic parameters for each GlucY variant.

	V_{max} (RLU/s)	$K_{0.5}^{h}$	h	[luc] (M)	$k_{\text{cat}}^{\text{relative}*}$	$(k_{\text{cat}}/K_{0.5}^{h})^{\text{relative}\dagger}$
GlucY	6.87×10^5	32.2	2.54	5×10^{-11}	1.0	1.0
MonstaY	4.53×10^5	10.6	2.10	5×10^{-11}	0.66	2.0
4lucY	1.86×10^4	1524	2.84	5×10^{-11}	2.7×10^{-2}	5.7×10^{-4}
tGY	6.20×10^3	4.33	1.44	5×10^{-9}	9.0×10^{-5}	6.7×10^{-4}
tMonY	1.15×10^3	6.3	1.70	5×10^{-9}	1.7×10^{-5}	8.6×10^{-5}
t4Y	4.79×10^2	11.4	1.98	5×10^{-9}	7.0×10^{-6}	2.0×10^{-5}

* Typical comparison of k_{cat} parameters requires V_{max} to have units of mol/s. For the purpose of comparing variants in this study, k_{cat} is expressed relative to that of GlucY using RLU.

† The parameter $(k_{\text{cat}}/K_{0.5}^{h})$ is considered an acceptable measure of specificity for enzymes not obeying Michaelis–Menten kinetics [153], however, as the units are not straightforward due to the inclusion of the Hill coefficient exponent, this value is expressed relative to GlucY.

(a) effect of C-terminal tyrosine linker on spectra

(b) effect of truncation on spectra
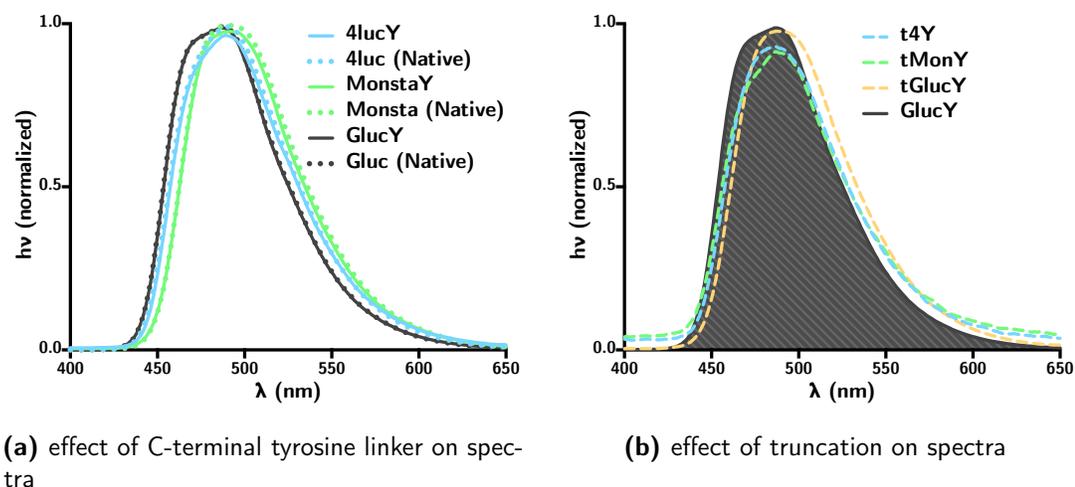
Figure 3.2. The addition of a C-terminal tyrosine linker did not alter the bioluminescence emission of Gluc variants. Truncation, however, introduced a 10–15 nm red-shift in the spectrum of each variant.

(a) WT

(b) Monsta

(c) 4luc

(d) Gluc and Rluc
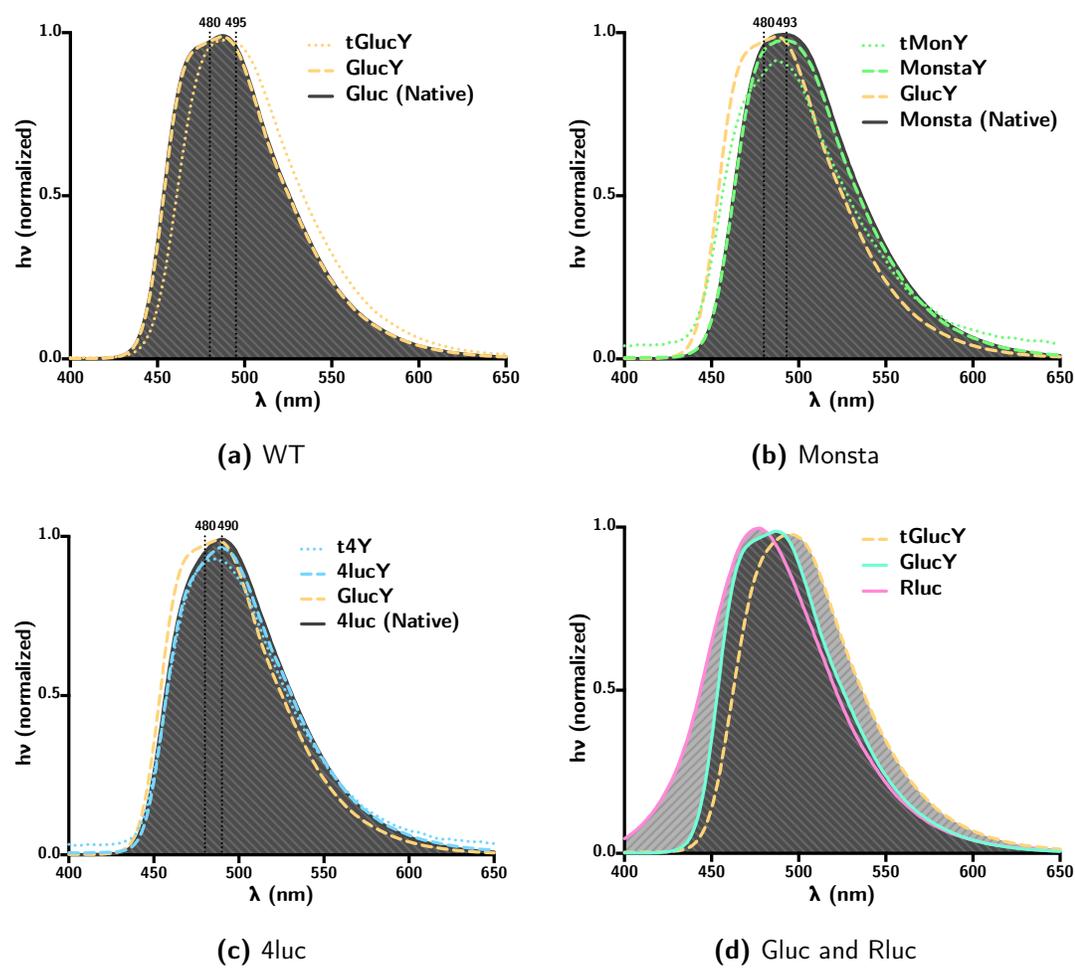
Figure 3.3. Spectra comparisons for each variant.

3.2.5 Decay Kinetics of *Gaussia* Luciferase Variants

The decay kinetics of the bioluminescent emission of each Gluc variant were determined (see Table 3.4 and Figure 3.4) by adding the native substrate coelenterazine and recording the relative luminescence until the signal returned to baseline (see *Kinetic studies*, Section 3.5.4 on page 77 of the *Methods* for more detail).

Table 3.4. Half-life parameters for GlucY variants.

	Half-Life (s)
GlucY	7
MonstaY	4
*4lucY	—
tGY	67
tMonY	51
t4Y	101

* 4lucY half-life was too long to be determined
by regression analysis of the data in Figure 3.4a

(a) decay kinetics of full-size GlucY variants **(b)** decay kinetics of truncated GlucY variants
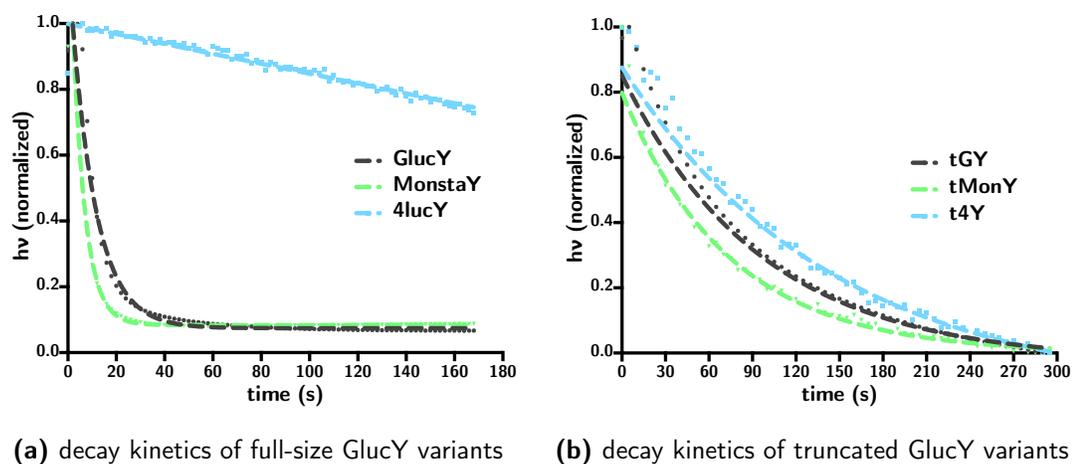
Figure 3.4. Decay kinetics for GlucY variants.

3.2.6 Secondary Structure Analysis of *Gaussia* Luciferase Variants by Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra were obtained for all GlucY variants (see Figure 3.5) as well as the full-size native variants (see Figure A1.4 on page 114). Secondary structure assignments were made using CDSSTR (see *Computational Analysis of Structure*, Section 3.5.6 on page 79 of the *Methods* for more detail) to analyze the CD spectra (see Figure 3.6 on the following page).

(a) CD spectra of full-size Gluc variants

(b) CD spectra of truncated Gluc variants
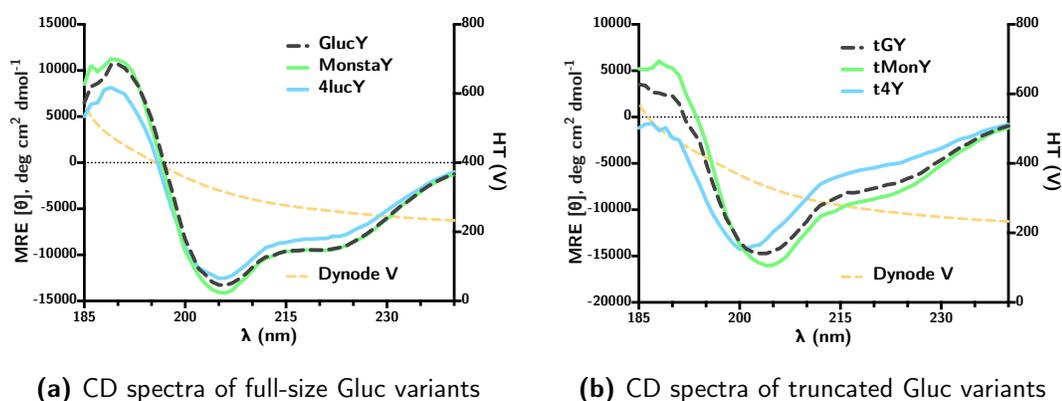
Figure 3.5. Baseline-corrected CD spectra for GlucY variants.
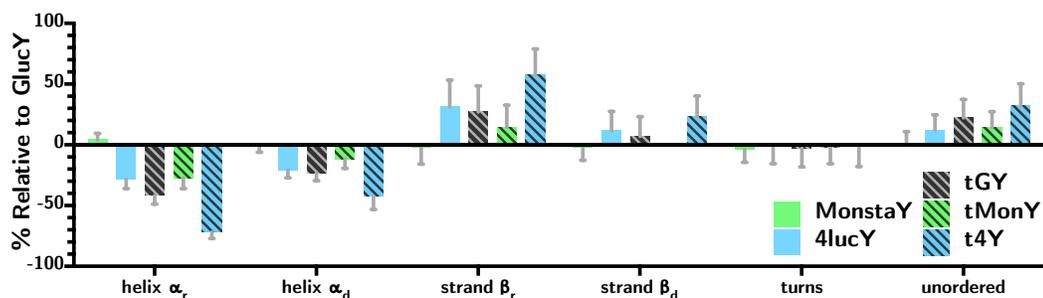
(a) change in secondary structure relative to GlucY
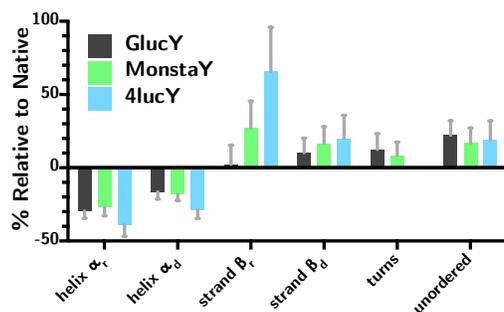
(b) change in secondary structure introduced by C-terminal tyrosine linker relative to native
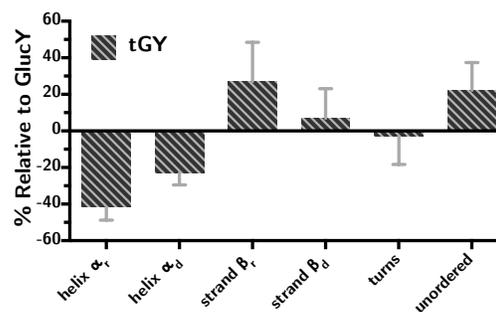
(c) change in secondary structure introduced by truncation relative to GlucY
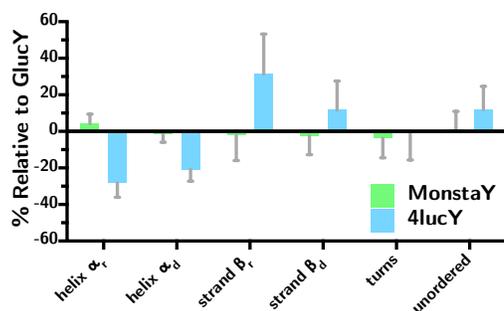
(d) change in secondary structure for full-size GlucY variants relative to GlucY
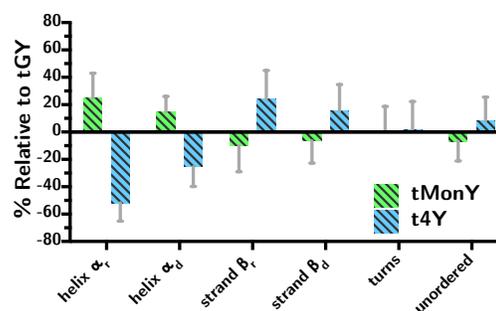
(e) change in secondary structure for truncated GlucY variants relative to tGY

Figure 3.6. Secondary structure assignment for Gluc variants using CDSSTR.

3.3 Discussion

Currently, there is no crystal structure available for Gluc. Therefore, computational techniques and mutagenesis studies must be used to develop conjectures about the regions of the protein that play a significant role in function and substrate interaction. The sequence for Gluc contains two structural domains exhibiting a high degree of homology (see Figure 3.7 on the next page) [145]. This is a common and somewhat unique trait of marine luciferases, and is observed in the *Metridia* luciferase isoforms [154], which share ∼60–70% sequence homology with Gluc (see Figure A1.5 on page 116). This observed intramolecular homology is believed to be the result of intragenic duplications and recombinations acting as a mechanism of structural and functional evolution [155, 156].

The two homologous domains present in Gluc were identified using the *Rapid Automatic Detection and Alignment of Repeats* (RADAR) tool[7] hosted by the European Bioinformatics Institute – European Molecular Biology Laboratory (EBI-BMBL) (see Table A1.1 on page 112) [155]. Each homologous domain contains five cysteine residues which are intermolecularly conserved among marine luciferases (see Figure A1.5 on page 116). Of these five cysteine residues, four are intramolecularly conserved between the homologous domains (see conserved cysteine residues marked with (∗) in Figure 3.7 on the next page).

These highly conserved cysteine residues are hypothesized to be involved in the formation of cystine disulfides necessary for the formation of correct tertiary structure crucial to functional bioluminescent activity [145, 147, 150, 151]. It is not believed, however, that these cystine residues are directly involved in the catalytic process [145].

The cytoplasmic environment in *E. coli* is reducing in large part due to the antioxidant activity of thioredoxin reductase (*trxB*) and glutathione oxidoreductase (*gor*), which poses a significant challenge for proper folding of proteins having

---

[7] `http://www.ebi.ac.uk/Tools/pfa/radar/`
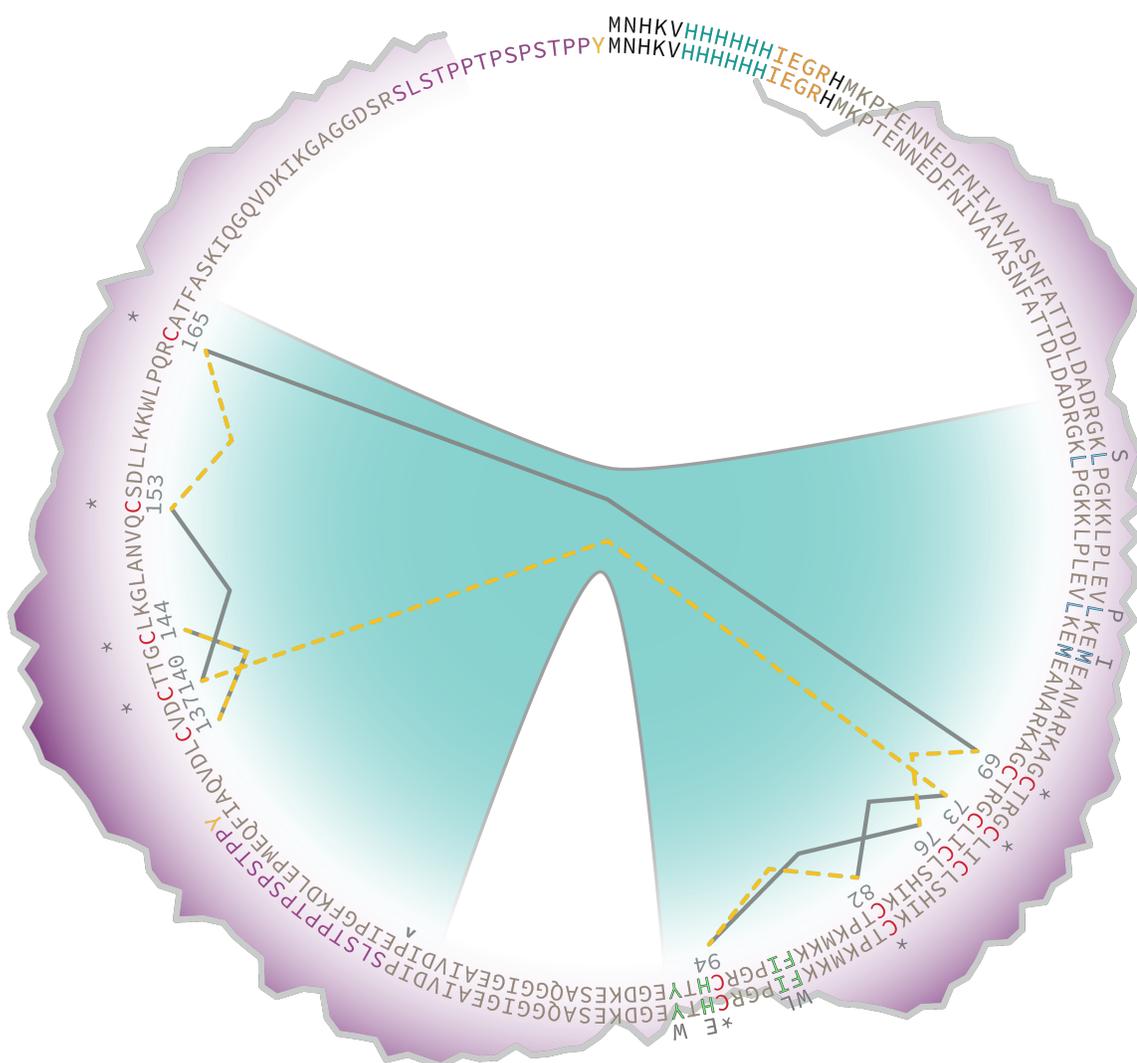
Figure 3.7. Structural map for GlucY and tGlucY. The two homologous domains of Gluc are indicated by the teal background in the interior of the ring; cysteine residues are colored in red and those conserved between the homologous domains are marked with asterisks (∗). Hydrophobicity is plotted around the ring according to the Kyte & Doolittle scale with darker purple being more hydrophobic; the hydrophilic pocket hypothesized to be involved in substrate recruitment is at the bottom of the ring. Cystine predictions are marked with solid grey lines (DiANNA 1.1) and dashed yellow lines (DISULFIND). Though the two algorithms differ slightly in their predictions, both identified two disulfides within each homologous domain. The 4luc mutation points (L47S/L57P/M60I) are shown in outlined blue text and the Monsta mutation points (F89W/I90L/H95E/Y97W) in outlined green text. The mutant residues are listed above each point. The location in the full-length sequence chosen for truncation (E117) is marked by a caret (∧).

high cystine content critical to function. An initial approach to help alleviate this problem is lowering the temperature at which the protein is expressed—this slows down the rate of protein production, thereby encouraging proper folding and reducing aggregation [157].

Indeed, Gluc is one such protein requiring special attention to expression conditions for proper folding, and while it has been expressed in mammalian systems [143] it has proven difficult to express in high concentration from bacterial systems, requiring alternative approaches such as expression with a solubilizing fusion partner [149] or cell-free expression systems where redox potentials can be more accurately controlled [150] . Recently, Rathnayaka et al. demonstrated soluble expression of Gluc using the pCold Cold Shock Expression System (Takara Bio. Inc., Japan) and carrying out protein expression at $15\,^{\circ}$C in a JM109(DE3) *E. coli* host [151].

For this reason, the pCold-I (see Figure 3.8 on the following page) expression vector was utilized in these studies, albeit with a different host strain of *E. coli*. Another approach to improve proper disulfide formation involves the use of a host strain of *E. coli* with a more favorable (i.e. less reducing) cytoplasmic environment. We chose to use Origami$^{\text{TM}}$ 2 (Novagen - EMD Millipore, Billerica, Massachusetts), a *trxB*$^{-}$/*gor*$^{-}$ strain with K12 background. Using this described setup for expression, Gluc was purified in high concentration ($\sim$15-20 mg per 1 L culture) from the soluble fraction using a genetically fused N-terminal $6\times$ histidine tag and Ni $\cdot$ NTA IMAC (see Figure 3.1 on page 58 for SDS-PAGE analysis of each GlucY variant diluted for better molecular weight determination and Figure A1.2 on page 112 of *Appendix I* for SDS-PAGE analysis of GlucY elutions directly following IMAC purification).

The location of cystine disulfides (i.e. which cysteine residues are connected via disulfides) has not yet been determined. As these cystine residues are crucial to the formation of a functional structure, computational techniques [158, 159] were used to predict the location of cystine disulfides in Gluc (see Figure 3.7 on the previous page). Two internal disulfides were predicted within each homologous domain, with
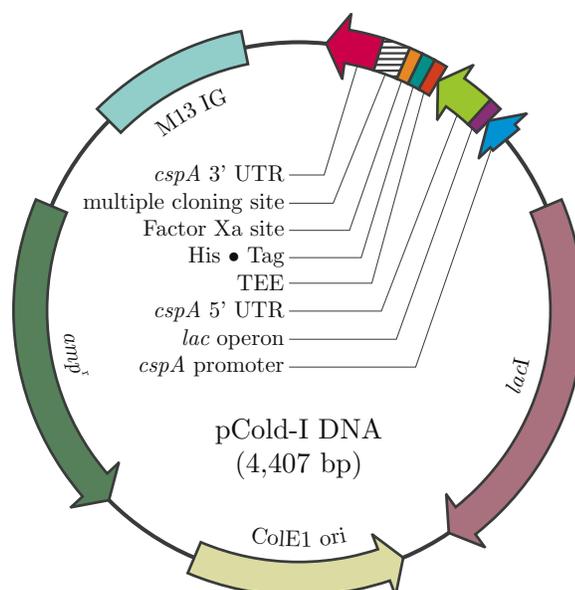
Figure 3.8. The pCold-I Cold Shock Expression System introduces an N-terminal $6\times$ histidine tag followed by a factor Xa cleavage site for tag removal following purification. The gene of interest inserted into the multiple cloning site is under the control of the *cspA* promoter and *lac* operon for expression control. The vector imparts ampicillin resistance through the *amp*^{r} gene encoding β-lactamase.

one disulfide bridging the two domains. Curiously, the conserved cysteine residues were not always the ones involved in forming the internal disulfides.

After identifying similarities between the chemical structures of the GFP chromophore (formed from S64, Y65, and G66) and the native luciferin substrate for Gluc (coelenterazine) (see Figure 3.9 on the following page), Kim et al. proposed that rather than embedding a chromophore within the protein structure, marine luciferases such as Gluc have evolved to recruit luciferin as an exogenous chromophore [147]. Indeed, an apparent hydrophilic region is observed near the boundary between the homologous domains present in Gluc (see Figure 3.7 on page 65). This same hydrophilic region between two homologous domains is also observed in other marine luciferases (e.g. *Metridia* luciferase) and is thought to play a role in substrate recruitment [147].

Two primary variants were selected to be further explored in this study (see Table 3.1 on page 57). The Monsta variant (F89W/I90L/H95E/Y97W) from Kim et al. was chosen for its high bioluminescent output and red-shifted emission [147].
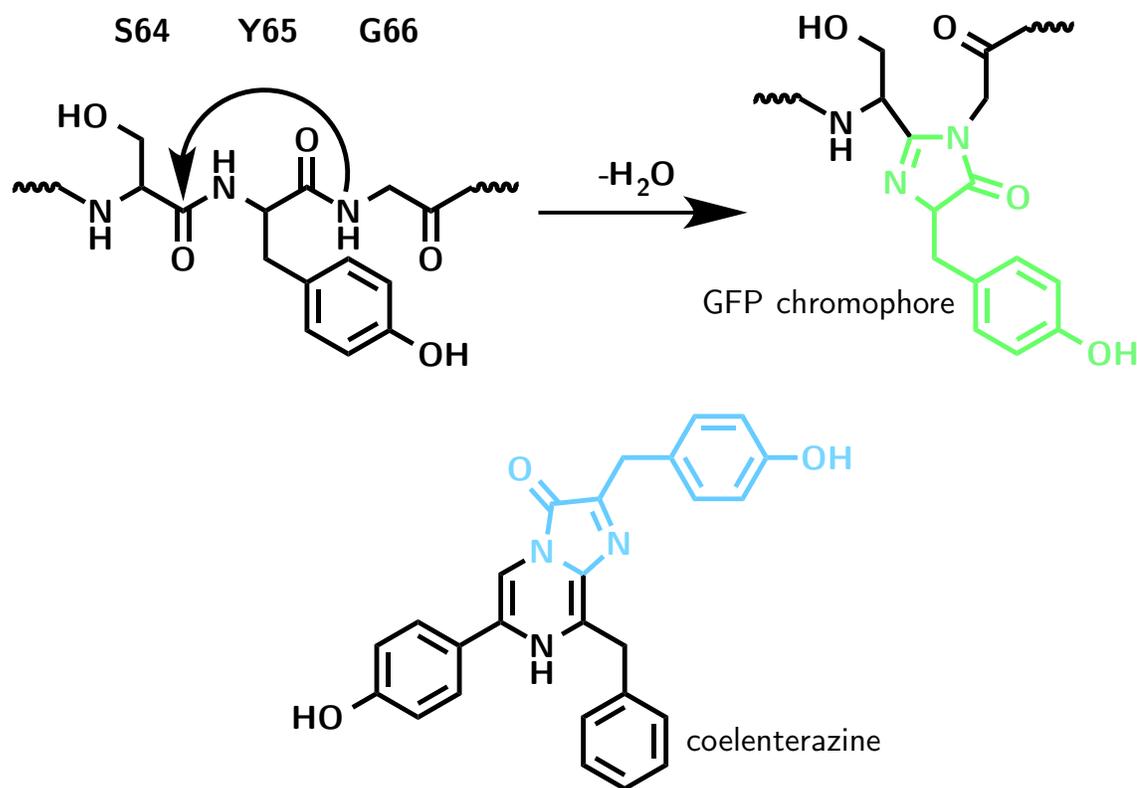
Figure 3.9. The chromophore matured from S64/Y65/G66 of GFP demonstrates some structural similarity with the chemiluminescent substrate coelenterazine.

The 4luc variant (L47S/L57P/M60I) adapted from Maguire et al., Degeling et al. was chosen for its exceptional signal stability over time (i.e. long bioluminescence half-life) [146, 148]. Interestingly, both variants harbor all mutations in the N-terminal homologous domain. As such, the N-terminal domain was chosen for creating truncated variants, in the hopes that these previous modifications to the Gluc sequence would impart improved bioluminescent activity to the truncated variants as well. In attempting to create truncated variants of Gluc that retain considerable bioluminescent activity, the location for truncation was chosen with the knowledge that proper cystine disulfide formation is critical to functional structure, and that the hydrophilic region lying between the two homologous domains is important for substrate recruitment.

One of the driving motivations behind the desire to create high-activity truncated variants of Gluc, is that the large size of bioluminescent proteins presents specific

challenges in trying to incorporate them as reporters for in vitro applications such as biosensors [113, 133] or for in vivo applications such as cellular delivery for imaging. When conjugating the luciferase to another biomolecule or recognition element, a reduction in size could ultimately equate to a reduction in the likelihood of interference to binding or diffusion of assay components. Most of these applications, especially the production of biosensors, requires accurate and reliable bioconjugation techniques to ensure well characterized product that will function in a reproducible manner. For this reason, in addition to creating truncated variants, a linker peptide containing a reactive tyrosine residue was added to the C-terminus of each variant (see Table 3.1 on page 57 and Table A2.1 on page 119 of *Appendix I*). The tyrosine residue can be easily modified to suit a variety of downstream reactions using diazonium salts [160, 161] and will be discussed further in Chapter 5 Section 5.2 on page 97.

It was determined that the addition of a C-terminal tyrosine linker did not have any adverse effect on the bioluminescent emission of the variants as confirmed by comparing spectra with the native form of each variant (see Figure 3.2a on page 60). Truncation, however, did seem to introduce an ∼10–15 nm red-shift in emission (see Figure 3.2b on page 60), therefore it can be assumed that truncation drastically changes the local environment of the substrate binding domain, but does not render it inactive.

The MonstaY variant showed a significant red-shift of ∼15–20 nm (see Figure 3.3b on page 60) and the 4lucY variant showed a slight red-shift in emission of ∼5–10 nm (see Figure 3.3c on page 60). The truncated versions of these variants, tMonY and t4Y, had similar bioluminescent spectra to their full-size relatives (see Figure 3.3b and Figure 3.3c on page 60), with the minor exception of tMonY, which was blue-shifted by ∼5 nm from MonstaY. Interestingly, tGY had a bioluminescent spectra almost identical to MonstaY with a red-shift in emission of ∼15–20 nm (see Figure 3.3a on page 60).

Considering the data obtained from CD analysis (see Figure 3.6 on page 63), the addition of a C-terminal tyrosine linker introduced an ∼20% increase in unordered structure and corresponding ∼20-30% reduction in α-helical character (see Figure 3.6b on page 63). This is somewhat expected as the linker is designed to be a soluble extension for bioconjugation. The sharp increase in β-strand character observed for 4lucY is accompanied by a large standard deviation, which in this case can be thought of as a measure of agreement between the different reference sets used for CDSSTR analysis of the CD data.

It appears that the mutations introduced for the MonstaY variant do not cause significant alterations to the secondary structure (see Figure 3.6d on page 63). Kinetically, the MonstaY variant has a $K_{0.5}^h$—which is related to K_m (see *Kinetic studies*, Section 3.5.4 on page 77 for more detail)—three fold lower than that of GlucY and a relative specificity constant ($k_{\text{cat}}/K_{0.5}^h$) twice that of GlucY (see Table 3.3 on page 59). It is possible that this increase in specificity for coelenterazine is brought about by the introduction of mutant residues allowing for hydrogen bonding to the substrate. This could also potentially explain the red-shift observed for this mutant by increasing electronic interaction between the protein and the substrate.

The mutations introduced for the 4luc variant caused more prominent changes to the secondary structure (see Figure 3.6d on page 63). Kinetically, 4lucY behaves very differently from the other variants, and while its bioluminescent output is somewhat lower, the stability of its emission is exceptional (see Figure 3.4a on page 61). Indeed, the large change in secondary structure must have drastic effects on its ability to bind the substrate coelenterazine. This is further supported by its $K_{0.5}^h$ being two orders of magnitude higher than that of GlucY or MonstaY and its relative specificity being almost four orders of magnitude lower (see Table 3.3 on page 59).

The truncated variants tGY, tMonY, and t4Y have specificity constants similar to that of 4lucY. Truncation alone does drastically alter the secondary structure

(see Figure 3.6c on page 63), but this should be expected as almost half of the protein sequence contributing to the secondary structure makeup has been removed. Among the truncated variants, the tMonY mutations tend to increase α-helical character and decrease β-strand character, while the t4Y mutations have the inverse effect to a higher degree (see Figure 3.6e on page 63). However, these mutations do not seem to change the degree of disorder. The Hill coefficient for the truncated variants is still slightly greater than one, indicating some level of cooperativity still exists even without the second homologous domain present (see Table 3.3 on page 59). It is likely that this cooperativity is a function of luciferase concentration, and truncated species are interacting in solution to form loosely-bound multimers. Indeed, when the concentration of luciferase was reduced, smaller Hill coefficients were observed (data not shown).

3.4   Conclusion

In conclusion, the successful truncation of the WT sequence and multiple variants of Gluc was carried out. Additionally, a C-terminal tyrosine linker was introduced through molecular cloning using the polymerase chain reaction (PCR). All variants created in this study were characterized by CD spectroscopy and traditional kinetic parameters were derived in classical style. These Gluc variants have wide applicability as bioluminescent reporters given the breadth of their kinetic characteristics and the smaller size of the truncated forms. The addition of a tyrosine handle for bioconjugation should prove especially useful in the development of biosensors and targeted cellular delivery systems as well as other classical analytical techniques requiring the labeling of other biomolecules or recognition elements with a bioluminescent reporter.

3.5   Methods

3.5.1   Molecular Cloning

The 185 amino acid sequence for WT Gluc from the NCBI GenBank (accession number `AAG54095`, ''`luciferase [Gaussia princeps (T. Scott, 1894)]`'') was used to create a codon-optimized, synthetic sequence for expression in an *E. coli* host. The signal peptide, which is unnecessary for bacterial expression, was identified using the online prediction tool (see Figure A1.1 on page 112) provided by the Center for Biological Sequencing Analysis at the Technical University of Denmark[8] [162] and removed from the synthetic sequence. The signal peptide region identified has been corroborated by others in the literature [145].

   With the signal peptide removed, the primary sequence for Gluc was codon-optimized for *E. coli* and synthesized commercially (GenScript USA Inc., Piscataway, New Jersey). Two endonuclease restriction sites were included in the synthetic sequence for subsequent cloning steps—a 5′ NdeI restriction site (`CA*TATG`), and a 3′ XbaI restriction site (`T*CTAGA`). The commercially synthesized sequence was received cloned into the pUC57 plasmid at the EcoRV restriction site (`GAT*ATC`). The pUC57::Gluc and pCold-I (Takara Bio. Inc., Japan) plasmids was transformed into cloning strain NEB5-α (New England Biolabs, Ipswich, Massachusetts), propagated by growing at 37 °C, and purified using a QIAprep Spin Miniprep Kit (Qiagen, Valencia, California). The purified pUC57::Gluc plasmid was treated with NdeI and XbaI restriction endonuclease in a two-step digestion (each held at 37 °C for 1 h before inactivation at 65 °C for 20 min) and purified from a 1% agarose TAE gel. The restricted synthetic Gluc sequence was inserted into the pCold-I cold shock expression vector (see Figure 3.8 on page 67) by treating pCold with NdeI and XbaI in a double digest supplemented with BSA (37 °C for 1 hr) followed by purification on a 1% agarose TAE gel and ligation with T4 DNA ligase (New England Biolabs,

---

[8]   `http://www.cbs.dtu.dk/`

Ipswich, Massachusetts) overnight at 16 °C. The newly constructed pCold-I::Gluc plasmid (pGluc, see Figure 3.10a on the next page) was transformed into NEB5-α for propagation and storage. The final plasmid was verified by Sanger sequencing. Sequencing primers used for the pCold-I vector may be found in Table A1.2 on page 115 of *Appendix I*.

3.5.2 Mutation and Truncation

Following purification from NEB5-α, site-directed mutagenesis was performed on pGluc by PCR using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, California) and the Q5® Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, Massachusetts) in order to create the variants of Gluc listed in Table 3.1 on page 57. A list of primers used for these mutations may be found in Table A1.2 on page 115 of *Appendix I*. PCR primers and methods were developed using the guidelines provided by the online tools PrimerX[9] and NEBaseChanger[10]. All mutations were confirmed by Sanger sequencing.

To assess whether Gluc could be expressed in a truncated form, a stop codon was introduced at the chosen location for truncation (E117) by mutating the GAA codon to TAA using the QuikChange Site-Directed Mutagenesis Kit. This allowed for a simple method of screening for soluble expression of the truncated form of Gluc. Once the truncated form was confirmed, a C-terminal tyrosine at the end of a linker peptide sequence was inserted into both the full-size and truncated forms of Gluc using the Q5® Site-Directed Mutagenesis Kit (see Table A1.2 on page 115 for primer sequences). For the truncated Gluc, the PCR reaction used to insert the C-terminal tyrosine linker simultaneously deleted the second homologous domain from pGluc (see truncation site marked with ($\wedge$) in Figure 3.7 on page 65), thereby drawing the transcription terminator region directly downstream of the truncation point to ensure total truncation during translation (i.e. remove the potential for

[9] http://www.bioinformatics.org/primerx/
[10] http://nebasechanger.neb.com/

stop codon read-through). In total, twelve unique plasmids were created for this study (see Table 3.1 on page 57), six containing a C-terminal tyrosine linker (see Figure 3.10b).

(a) pGluc

(b) pGlucY
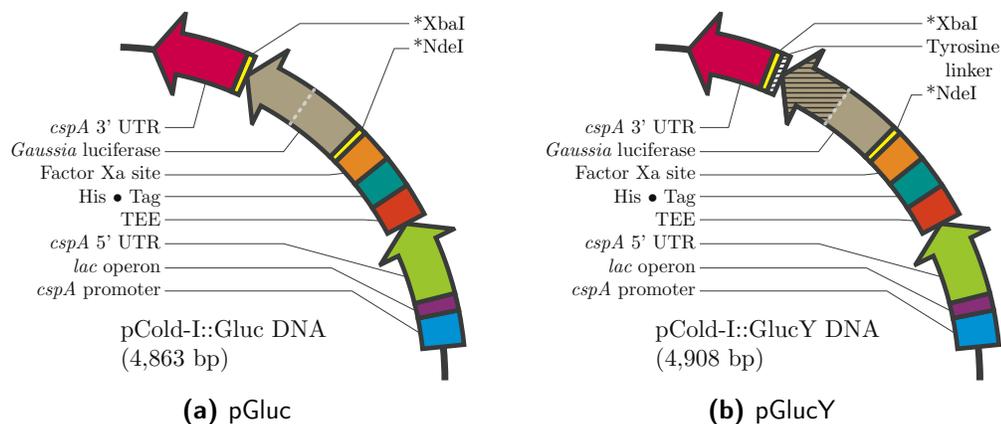
Figure 3.10. pCold-I cold-shock vector expression system plasmid map for Gluc with and without C-terminal tyrosine linker.

3.5.3 Expression and Purification from *Escherichia coli*

An expression study was performed to determine the optimal conditions for expressing Gluc from Origami^{TM} 2 *E. coli* in terms of total yield, purity with regard to auxiliary cellular proteins, and amount of degradation by proteases. As a preliminary step to culture expansion, small cultures were prepared for overnight growth at $37\,^{\circ}$C by inoculating 5 mL of Miller LB supplemented with 0.1 mg/mL ampicillin from glycerol stocks of Origami^{TM} 2 *E. coli* transformed with pGluc. Following overnight growth, the small cultures were centrifuged at $4,000 \times$g, resuspended in fresh media supplemented with 0.1 mg/mL ampicillin, and used to inoculate 50 mL cultures, which were each subjected to different sets of growth variables. In general, protein expression from the pCold-I plasmid proceeds as follows:

> › Cells are grown to the appropriate density.

> › Cells are cold-shocked in an ice bath for 1 hour to activate the *cspA* promoter.

> › Cells are induced with IPTG to activate the lac operon secondary checkpoint.

› Cells are grown under these expression-inducing conditions at 15 °C for the appropriate amount of time (usually on the order of 24 hours).

› Cells are collected by centrifugation, lysed according to user-preferred method, and protein is purified by IMAC utilizing the N-terminal 6× histidine tag.

In order to optimize the expression of Gluc, each step was approached as a separate variable. First, the final concentration of IPTG was varied from 0.1–2.0 mM, and the induction time carried out at these concentrations was varied from overnight (∼16 hours) to 60 hours. Next, the initial culture density was varied from an OD_{600nm} of 0.6–1.6, induced with the optimized amount of IPTG, and the induction time was varied over a narrower range showing optimal expression. The media used was also evaluated, using both LB and Terrific broth, as well as replenishing the media prior to induction with IPTG. (see Figure A1.2 on page 112)

In addition to optimizing the said parameters for expression, the lysis protocol was also optimized. Three different methods were evaluated: a softer chemical lysis using BugBuster® Protein Extraction Reagent (EMD Millipore, Billerica, Massachusetts), a French press (Emulsiflex C5, AVESTIN Inc., Ottawa, Ontario, Canada), and lysis by sonication (Model 500 Sonic Dismembrator, Fisher Scientific™, Pittsburgh, Pennsylvania). In terms of recovery and simplicity, sonication using a microtip probe for 5–10 minutes (depending on the viscosity of the resuspension of cells in lysis buffer) with a 0.5 second on/off pulse sequence—similar to the method described by Feliu et al. [163]—was chosen as the most optimal method of lysis.

The optimized expression protocol was determined to be growth to an OD_{600nm} of 1.6 in Terrific broth supplemented with 0.1 mg/mL ampicillin, replenishment of media following cold-shock on ice for 1 hour, induction with a final concentration of 0.1 mM IPTG, followed by growth at 15 °C overnight. The cells were collected by centrifugation at 4000 ×g and 4 °C for 20 minutes and resuspended in a lysis buffer—similar to that used by Rathnayaka et al. [151]—of 50 mM Tris · HCl pH 8.0, 150 mM sodium chloride, 10 mM imidazole, 1% (vol.) nonyl phenoxypolyethoxylethanol

(NP-40), 0.2% (vol.) polyoxyethylene (20) sorbitan monolaurate (Tween 20), and 10 mM 2-mercaptoethanol (β-ME). The cell suspension was supplemented with $1\times$ ProBlock$^{\text{TM}}$ Gold Bacterial Protease Inhibitor Cocktail (Gold Biotechnology Inc., St. Louis, Missouri) to prevent degradation following lysis and sonicated as described above. The insoluble material was removed by centrifugation at $10,000 \times$g and $4\,^{\circ}$C for 30 minutes and the supernatant was filtered by syringe through a 0.22 μm filter. The filtered crude protein was then incubated with Ni$-$NTA agarose (Qiagen, Valencia, California) at $4\,^{\circ}$C for 45 minutes, collected on a Pierce$^{\text{TM}}$ Centrifuge Column (Life Technologies, Grand Island, New York) by gravity flow, washed with 10 column volumes of lysis buffer followed by 20 column volumes of a wash buffer of 50 mM Tris $\cdot$ HCl pH 8.0, 150 mM sodium chloride, and 20 mM imidazole. The protein was then eluted with an elution buffer of 50 mM Tris $\cdot$ HCl pH 8.0, 150 mM sodium chloride, and 150 mM imidazole in 1 column volume increments.

The purified protein was then dialyzed into either a phosphate buffered saline (PBS) of 10 mM sodium phosphate pH 7.2 and 150 mM sodium chloride or a CD buffer of 5 mM potassium phosphate pH 7.8 and 25 mM ammonium sulfate, depending on the intended application following purification. Slide-A-Lyzer$^{\text{TM}}$ Dialysis Cassettes (Life Technologies, Grand Island, New York) with a molecular weight cutoff (MWCO) of 10 or 3.5 kDa were used for the full-size and truncated Gluc variants, respectively.

3.5.4 Characterization of Bioluminescence

Spectra Bioluminescence spectra were obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, California) equipped with a microplate reader accessory using the bio-/chemi-luminescence mode. All spectra measurements were carried out in 96-well black polystyrene non-binding microplates (Greiner Bio-One Inc., Monroe, North Carolina). Na-

tive coelenterazine was purchased from NanoLight$^{\text{TM}}$ Technology (Prolume Ltd., Pinetop, Arizona).

Briefly, 100 μL of 300 μM coelenterazine was injected into 200 μL of luciferase varying in concentration from 10 nM to 10 μM—depending on the decay kinetics of each bioluminescence reaction and PMT response—and the spectrum was recorded. Using the Cary Eclipse Scan software package, the spectrum was scanned from 400–650 nm with a 100-500 ms dwell time. Using the software CAT mode, 10 separate spectra were recorded, normalized, and averaged for each reaction.

Kinetic studies Michaelis–Menten and decay kinetics were obtained using a POLARstar$^{\circledR}$ Optima (BMG LABTECH GmbH, Ortenberg, Germany). The emission filter was set to "lens" (i.e. no filter) and total light was collected for all measurements. All kinetic measurements were carried out in 96-well black polystyrene non-binding microplates (Greiner Bio-One Inc., Monroe, North Carolina). Native coelenterazine was purchased from Nanolight$^{\text{TM}}$ Technology (Prolume Ltd., Pinetop, Arizona).

Michaelis–Menten data were obtained using a method modified from Tzertzinis et al. [164]. Briefly, 100 μL of native coelenterazine to the wells of 96-well microplates in concentrations varying from 200 nM to 200 μM. An equal volume of luciferase at a constant concentration was then injected into each well containing varying coelenterazine concentrations. The bioluminescent emission was quantified immediately following injection by integrating the signal for 10 seconds at maximum PMT gain. All measurements were done in triplicate.

To determine decay kinetics of the bioluminescence reaction for each Gluc variant, 100 μL of luciferase was added to a 96-well microplate at a concentration of 0.2–2 nM. The bioluminescent emission was quantified by integrating the signal in 2 second intervals for 3 minutes. At 10 seconds, 100 μL of 20 μM coelenterazine was injected to begin the reaction.

(a) substrate cooperativity **(b)** substrate inhibition at high concentration
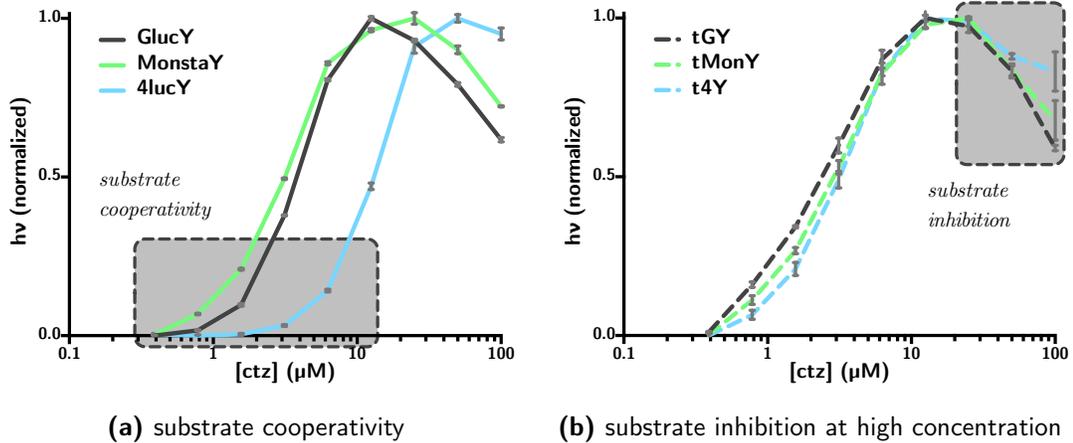
Figure 3.11. Characteristics of Michaelis–Menten plots for GlucY variants requiring a modified Michaelis–Menten model to compensate for substrate cooperativity and inhibition (see Equation (3.2)).

The kinetic data obtained for the Gluc variants exhibit two distinct features requiring a slightly modified Michaelis–Menten model in order to derive the typical kinetic parameters. The sigmoidal shape of the plots in Figure 3.11, apparent at lower concentrations and highlighted in Figure 3.11a is due to a level of positive substrate cooperativity [153, 165], which is observed in many of the marine luciferases [164]. As such, the traditional Michaelis–Menten model,

$$v = \frac{V_{\max}[S]}{K_m + [S]} \tag{3.1}$$

must be adapted to compensate for the positive substrate cooperativity (also known as the Hill equation [153]),

$$v = \frac{V_{\max}[S]^h}{K_{0.5}^h + [S]^h} \tag{3.2}$$

where h is the Hill coefficient, which quantifies the degree of cooperativity. $K_{0.5}^h$ is related to K_m, but also contains the Hill coefficient. It can effectively be computed as K_m^h [153, 166]. Additionally, the decrease in velocity at higher concentrations, highlighted in Figure 3.11b, is indicative of substrate inhibition through the formation of an inactive ternary complex [153, 165] and is observed by a sudden upward curvature near the y-intercept on a Lineweaver-Burk plot (data

not shown). The effect of this inhibition can be modeled by adding a dissociation constant for the inhibitory complex.

$$v = \frac{V_{\max}[S]}{K_m + [S](1 + \frac{[S]}{K_i})} = \frac{V_{\max}}{1 + \frac{K_m}{[S]} + \frac{[S]}{K_i}} \tag{3.3}$$

Using Excel 2011 (Microsoft Corporation, Redmond, Washington) and the scientific graphing and curve fitting software Prism 6 (GraphPad Software Inc., La Jolla, California), these equations were used to fit the kinetic data in Figure 3.11 on the previous page (see also Figure A1.3 on page 113 of *Appendix I*) and derive the parameters listed in Table 3.3 on page 59.

3.5.5   Circular Dichroism Spectroscopy

CD spectroscopy was performed on a Jasco J-815 Circular Dichroism Spectropolarimeter using a 0.1 cm path length quartz cuvette. Gluc variants were dialyzed into CD buffer and diluted to a concentration of 0.1 mg/mL. Spectra were acquired from 260–185 nm using the following set of instrument parameters:

Parameter	Value
Scan Mode	*Continuous*
Scan Speed	*50 nm/min*
Data Pitch	*0.5 nm*
Bandwidth	*2 nm*
D.I.T.	*2 s*
Accumulations	*5*

3.5.6   Computational Analysis of Structure

The locations of cystine disulfides in Gluc were predicted using DISULFIND[11] (Dipartimento di Ingegneria dell'Informazione, Università di Firenze, Firenze, Italy)

---

[11] http://disulfind.dsi.unifi.it/

[158] and DiANNA 1.1[12] (Clote Lab, Boston College, Chestnut Hill, Massachusetts) [159].

Hydrophobicity data for Gluc was obtained from the ProtScale tool[13] (ExPASy Swiss Institute of Bioinformatics (SIB) Resource Portal) [167, 168] and plotted according to the hydropathy scale developed by Kyte and Doolittle [169].

All analysis of CD spectra was performed using DichroWeb[14] (Department of Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, University of London, United Kingdom) [170]. The CDSSTR analysis program [171] was used to generate secondary structure assignments for Gluc variants in this study using reference sets 3, 4, 6, 7, SP175, and SMP180. The CDSSTR analysis program was chosen as it provided the best fit to experimental data, however, all analysis programs available through DichroWeb including SELCON3, CONTIN, VARSLC, and K2D were evaluated. Data input for analysis was limited to the wavelength range of 240–185 nm.

---

[12] `http://clavius.bc.edu/~clotelab/DiANNA/`
[13] `http://web.expasy.org/protscale/`
[14] `http://dichroweb.cryst.bbk.ac.uk/html/home.shtml`

Chapter 4

Expression of a Soluble Truncated *Vargula* Luciferase in *Escherichia coli*[15]

Overview

Due to the nature of many marine luciferases, they often present several issues when transitioned from their native eukaryotic expression environment into a more economical prokaryotic, i.e. bacterial, expression system. The lack of chaperone proteins to aid in the folding process as well as the reducing cytoplasmic environment in *E. coli* make correct folding of proteins with high cystine content difficult and can lead to the formation of incorrect tertiary structure and ultimately insoluble and potentially inactive protein. The marine luciferases commonly exhibit flash-type bioluminescence, which limits their utility in high-throughput applications where stability of emission is required for the duration of data collection. *Vargula* luciferase (Vluc) exhibits glow-type bioluminescence through a reaction dependent on its native substrate and molecular oxygen. However, expression and purification of this luciferase from bacterial systems has proven difficult. Here, the expression and purification of a tuncated form of Vluc from *E. coli* is demonstrated.

4.1 Perspective

Among the most common marine luciferases used as bioluminescent reporters is *Vargula* luciferase (Vluc), from the ostracod crustacean *Vargula hilgendorfii* also known as the "sea firefly." In 1989, Thompson et al. successfully cloned the cDNA

---

[15] *Supplementary Information* has been included as *Appendix II* on page 118.

for Vluc and expressed it in a mammalian cell system (monkey COS cells). The full sequence is 555 amino acids and could potentially be glycosylated at two different sites in its native organism [172, 173]. The native substrate for Vluc, *Vargula* luciferin (vargulin)—also called *Cypridina* luciferin as the same substrate is used by the *Cypridina noctiluca* luciferase (Cluc)—differs slightly from coelenterazine in the substituents around the imidazopyrazine skeleton, however, the bioluminescent reaction proceeds in a like manner through a dioxetanone intermediate (see Figure 4.1) emitting light around 462 nm (see Figure 4.2 on the next page).
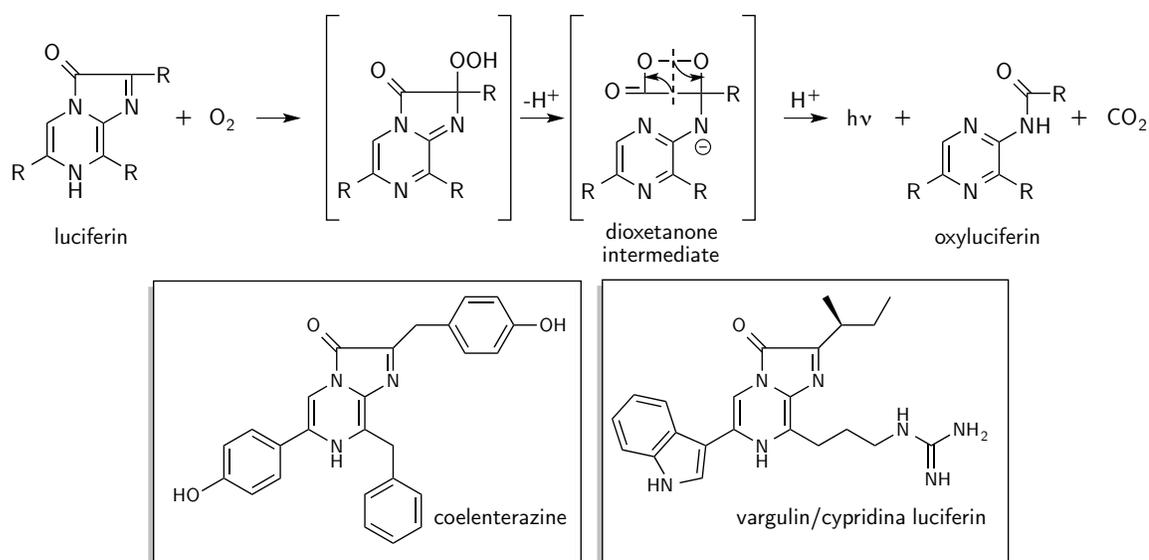
Figure 4.1. The light generating reaction for Vluc proceeds through a dioxetanone intermediate like other coelenterate photoproteins such as *Gaussia* and *Renilla* luciferase.

Through structural alignment, it was apparent upon cloning the cDNA for Vluc that, like many of the marine luciferases including *Gaussia*, *Metridia*, and *Cypridina*, the protein contained two homologous domains with significant homology to the photoprotein aequorin from the jellyfish *Aequorea victoria* [172]. In 1996, Maeda et al. showed that a fusion of protein A with the N-terminal homologous domain of Vluc (P28-C312) could be expressed in a mammalian system and retain ∼40% of bioluminescent activity as compared to the full-size form [175].
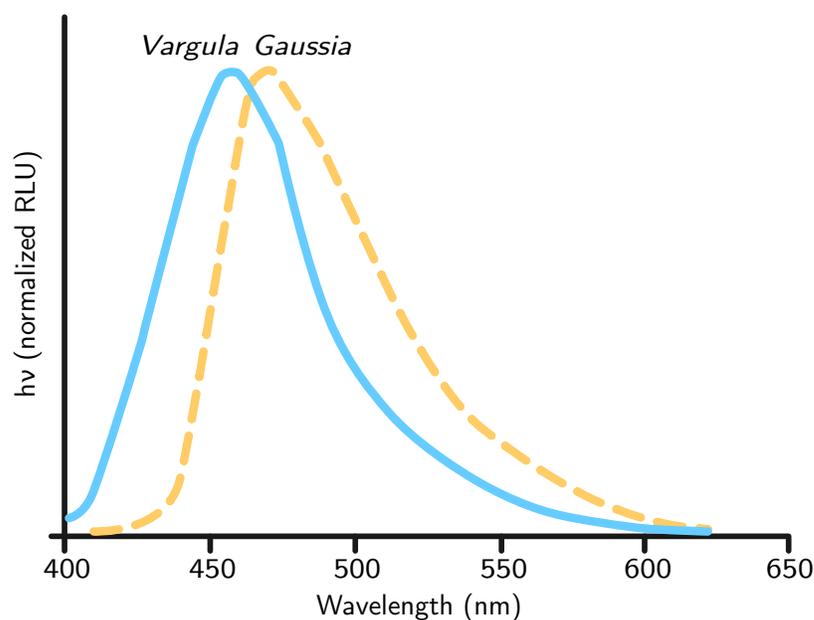
Figure 4.2. The bioluminescence spectrum of *Vargula* luciferase as compared to *Gaussia* luciferase. Reproduced from Dobbs et al. [174].

To date, the production of Vluc in bacterial systems has remained elusive. While soluble production of Vluc in *E. coli* was achieved by Inouye and Sahara through cold shock expression of a fusion with a solubilizing partner of a synthetic IgG-binding domain, the fusion showed no bioluminescence, indicating that proper folding and correct cystine formation necessary for an active structure did not occur [149]. Moreover, the addition of a solubility partner through genetic fusion adds extra bulk to the protein size, which may have undesirable effects in downstream applications, such as binding interference or reduction of reporter activity, potentially requiring another processing step to remove the partner are re-purify.

Currently, there are three main bioluminescent substrates readily available from commercial sources. They are D-luciferin (also called firefly luciferin or beetle luciferin), coelenterazine, and *Cypridina* luciferin (also called *Vargula* luciferin or vargulin). The beetle luciferases require ATP in their bioluminescent reaction. However, the coelenterate luciferase and Vluc/Cluc reactions only require the substrate and molecular oxygen, greatly simplifying the reaction setup/requirements. It has been shown that there is no cross-reactivity observed between these three luciferin

substrates and their corresponding proteins [176, 177]. Therefore, the bacterial production of a Vluc variant would be very useful in developing multiplex analytical systems. Additionally the glow type bioluminescence of Vluc could be used to develop multiplexed systems based on time resolution of the signals.

The truncated form of Vluc (tVluc), being the N-terminal homologous domain, contains 16 cysteine residues as opposed to 34 in the full-size form. This could make production in a bacterial system more feasible while simultaneously making the luciferase easier to work with in downstream applications requiring bioconjugation or cellular delivery. In this study, the bacterial expression of tVluc in a soluble and active form from *E. coli* is demonstrated, and the luciferase is generally characterized according to its bioluminescent properties. It is believed that this work will help guide the further development of Vluc and Cluc luciferase variants that can be expressed and purified easily from a bacterial system.

4.2   Results

4.2.1   Molecular Cloning of Truncated *Vargula* Luciferase

The pColdI::tVluc vector (ptVluc, see Figure 4.3b on the following page) was created by inserting a codon optimized gene for tVluc into the pCold-I cold shock expression vector (see Figure 4.3a on the next page).

(a) pCold-I   **(b)** ptVluc
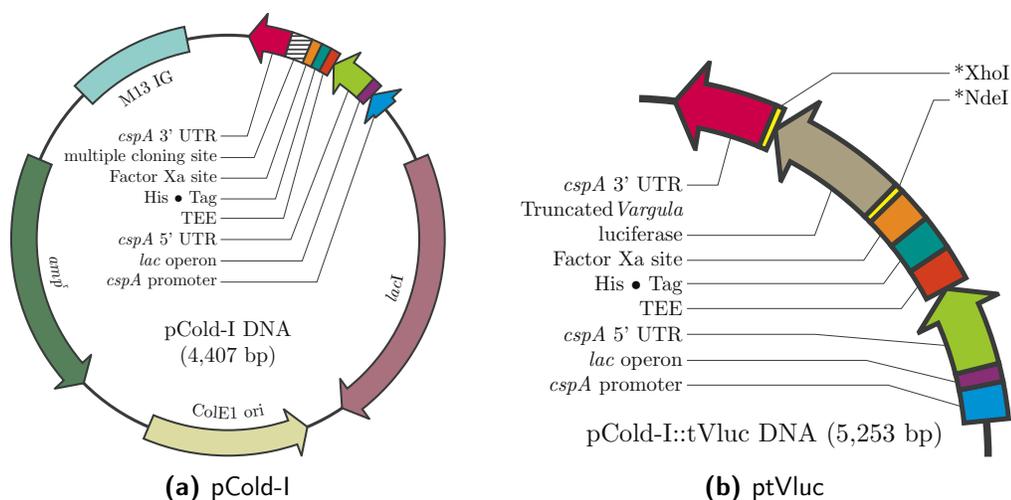
Figure 4.3. The pCold-I Cold Shock Expression System introduces an N-terminal $6\times$ histidine tag followed by a factor Xa cleavage site for tag removal following purification. The gene of interest inserted into the multiple cloning site is under the control of the *cspA* promoter and *lac* operon for expression control. The vector imparts ampicillin resistance through the amp^r gene encoding β-lactamase.

4.2.2   Expression and Purification of Truncated *Vargula* Luciferase from *Escherichia coli*

tVluc was expressed and purified from the soluble fraction of *E. coli* strains Origami™ 2 (see Figure 4.4a on the following page) and SHuffle® Express (see Figure 4.5 on the next page). Purification from SHuffle® Express produced a soluble tVluc product that was far purer than that obtained with Origami™ 2.

SDS-PAGE analysis on a 4-20% gradient gel of the purification of tVluc from Origami™ 2 was used to calculate the observed molecular weight of the over-expressed bands. Four prominent bands were chosen from the elution fractions of the tVluc purification. The migration of each band in the Precision Plus Protein™ Dual Color Standard (Bio-Rad Laboratories Inc., Hercules, California) was measured using Photoshop CS6 and plotted with automated curve fitting in Prism 6 to extrapolate a precise molecular weight value for the over-expressed bands (see Figure 4.4b on the following page). The most prominent over-expressed band was calculated to be 34 kDa, approximately the theoretical molecular weight of tVluc. Additionally,

(a) 4-20% gradient SDS-PAGE

(b) analysis of protein migration
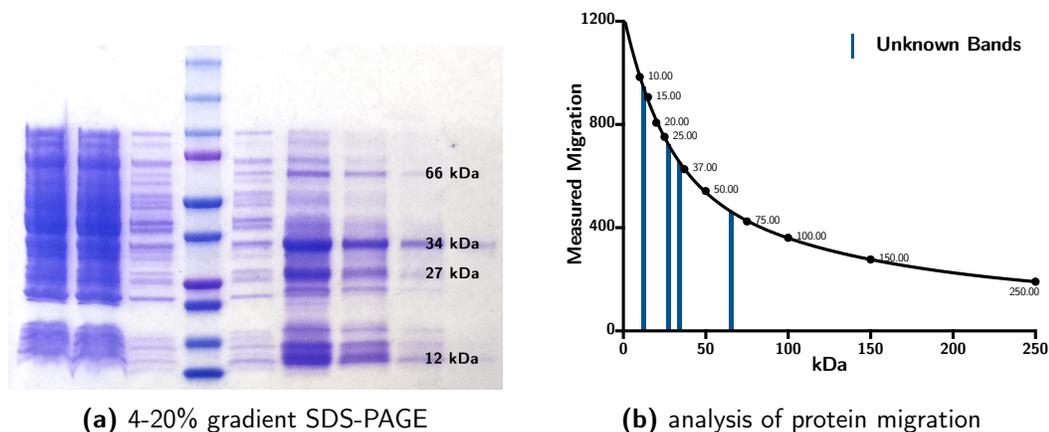
Figure 4.4. Molecular weight analysis of tVluc expressed in Origami$^{\text{TM}}$ 2 *E. coli*. Analysis of band migration was done using Adobe Photoshop CS6 and GraphPad Prism 6. The main over-expressed band was calculated to be 34 kDa. The theoretical molecular weight of tVluc is 33.3 kDa. The 66 kDa band may be evidence of dimerization. The Precision Plus Protein$^{\text{TM}}$ Dual Color Standard used to create generate the model for migration is located in lane 4 of the gel in part (a) of the figure.

a 66 kDa band was observed, which could be the result of dimerization between individual tVluc monomers (see Figure 4.4a).

As further confirmation for the successful purification of tVluc from *E. coli*, a Western blots were performed using an anti-6× histidine epitope tag antibody (Life Technologies, Grand Island, New York) and an anti-*Cypridina* luciferase epitope antibody (antibodies-online Inc., Atlanta, Georgia). The anti-6× histidine epitope
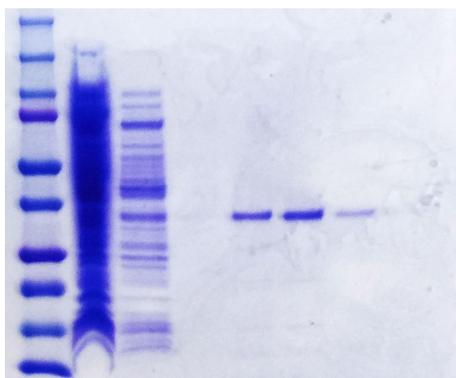
Figure 4.5. Purification of tVluc from SHuffle$^{\circledR}$ Express *E. coli*. Lanes from left: 1) Precision Plus Protein$^{\text{TM}}$ Dual Color Standard; 2) flowthrough; 3) lysis buffer wash; 4) wash buffer wash; 5–8) elution fractions.

tag antibody blot confirmed that the most over-expressed band observed at ∼34 kDa in Figure 4.4a on the preceding page was tVluc produced from the pCold-I cold shock expression system (see Figure 4.6). The anti-*Cypridina* blot did to give any readable signal. It is likely that the epitope targeted by the antibody was removed with truncation.
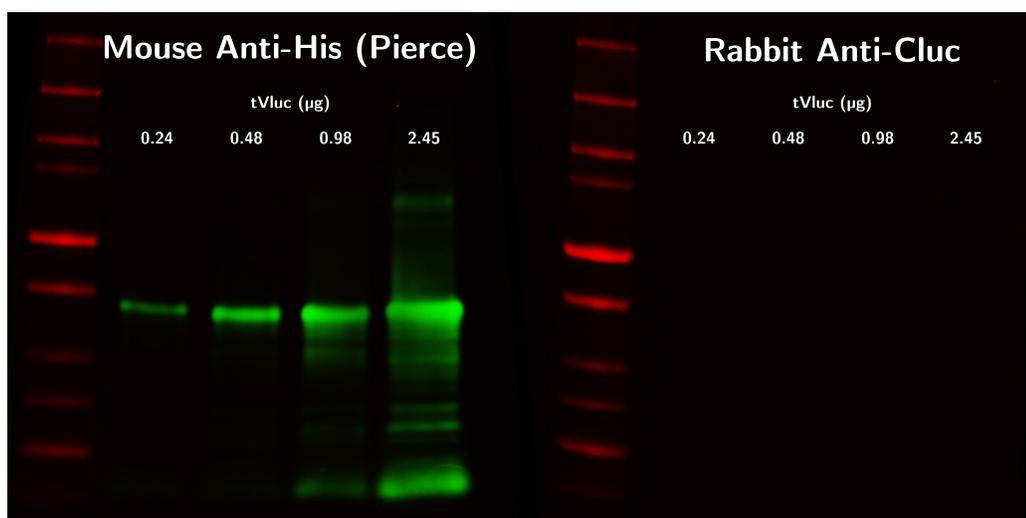
Figure 4.6. Western blots of tVluc expressed in Origami$^{\text{TM}}$ 2 *E. coli* using 6× histidine tag (left) and Cluc (right) primary antibodies. No binding was observed for the Cluc antibody, potentially due to epitope removal with truncation.

4.2.3 Bioluminescent Characteristics of Truncated *Vargula* Luciferase

The purified tVluc exhibited a glow-type bioluminescent emission with a half-life of over two hours (see Figure 4.7 on the following page). The reaction is quick to initiate upon injection of the vargulin substrate and well suited for high-throughput analytical applications where signal stability over time is necessary to compensate for substrate injection into a large number of samples simultaneously to reduce assay time.
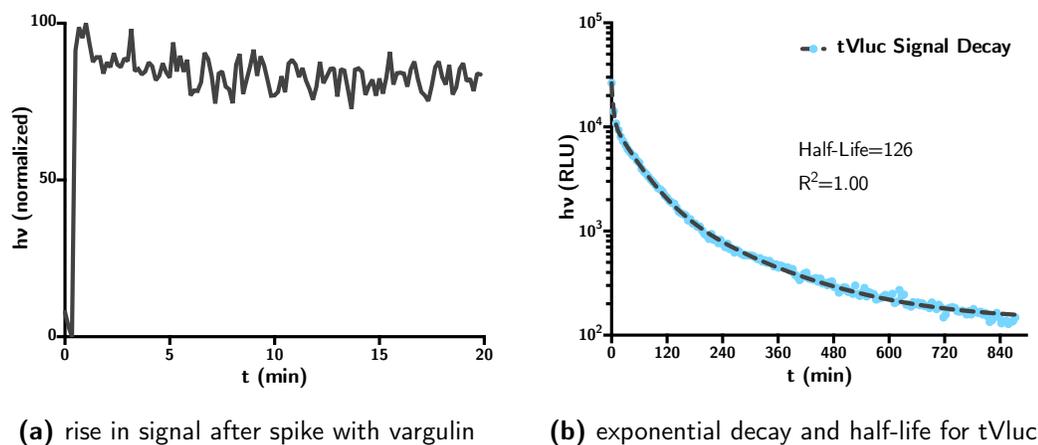
(a) rise in signal after spike with vargulin

(b) exponential decay and half-life for tVluc

Figure 4.7. The extended emission kinetics of tVluc with vargulin substrate. The half-life is over two hours.

4.2.4 Secondary Structure Analysis of Truncated *Vargula* Luciferase by Circular Dichroism Spectroscopy

The circular dichroism (CD) spectrum for tVluc was obtained from 260–180 nm (see Figure 4.8a) and analyzed by the CDSSTR analysis package to generate secondary structure assignments for the luciferase (see Figure 4.8b).
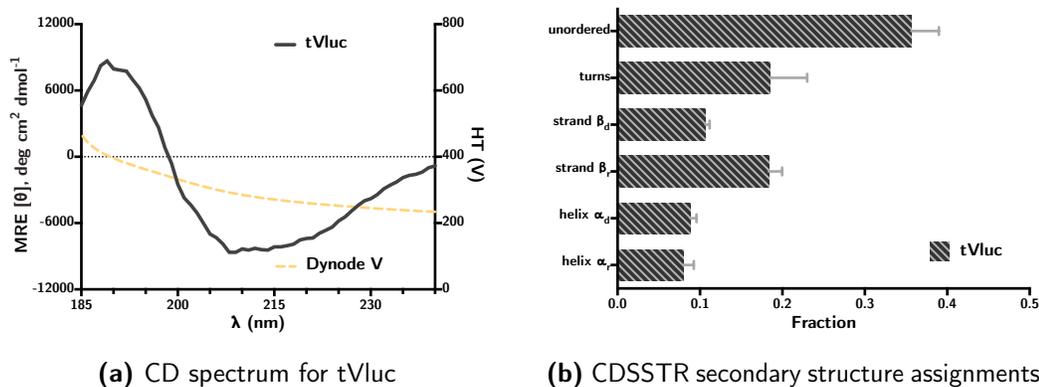
(a) CD spectrum for tVluc

(b) CDSSTR secondary structure assignments

Figure 4.8. CD spectrum for tVluc and secondary structure assignments made using the CDSSTR analysis program from DichroWeb.

4.3 Discussion

The cold shock expression system chosen for production of tVluc in *E. coli* slows down the rate of protein production, thereby encouraging proper folding and reducing aggregation [157]. This expression vector was used with three different strains of *E. coli* for production of tVluc.

When analyzing the expression of tVluc from NEB Express by SDS-PAGE, the bands were extremely faint, began eluting from the column at 40 mM imidazole, and were completely removed before finishing the 60 mM imidazole elution (data not shown). This was taken as an indication that the N-terminal $6\times$ histidine tag may have been folding in a way that sequestered its binding to the Ni−NTA agarose and potentially disrupted the active structure. There was no bioluminescent activity observed in the presence of vargulin.

Production of tVluc from Origami$^{\text{TM}}$ 2 gave a much more prominent over-expressed band in the soluble fraction, which tolerated immobilized-metal affinity chromatography (IMAC) well and stayed bound to the Ni−NTA agarose until eluted with a high concentration of imidazole (see Figure 4.4a on page 86). This improvement was attributed to the thioredoxin reductase ($trxB^-$) and glutathione oxidoreductase (gor^-) mutations present in Origami$^{\text{TM}}$ 2, which may have been the factor in proper disulfide formation and folding of active tVluc.

While tVluc was obtained from Origami$^{\text{TM}}$ 2 in an active form from the soluble fraction, its purity was not acceptable. This strain is derived from K12 *E. coli*, and it is possible that it exhibits higher protease activity that other strains. As such, as third strain of *E. coli* was also used for expression of tVluc. SHuffle$^{®}$ Express is an *E. coli* B strain, which in addition to being a $trxB^-/gor^-$ mutant, also constitutively expresses a chromosomal copy of a disulfide bond isomerase DsbC, which is usually found in the periplasm of Gram-negative bacteria to aid in the rearrangement of incorrectly formed disulfides [178]. Expression from this strain produced active tVluc in an exceptionally pure from from the soluble fraction (see Figure 4.5 on page 86).

The online secondary structure prediction server Phyre^2 was used to model the structure of tVluc [179]. The server was able to model ~65% of residues with a >90% confidence level (data not shown). While these sorts of predictions should be taken lightly, it is an interesting correlation that the predicted model exhibited ~8% α_r character and ~20% β_r character. Given that the CDSSTR secondary structure assignments from the CD data obtained match this prediction closely (see Figure 4.8b on page 88), it could indicated that tVluc is at least folding in a consistent and predictable manner within the bacterial cytoplasmic environment. However, a large portion of the secondary structure was still classified as disordered. As this can contribute to potential aggregation and precipitation of protien, further mutation and truncation may be necessary to remove this disorder and create a more stable variant of tVluc.

4.4 Conclusion

In conclusion, it has been demonstrated that a truncated form of *Vargula* luciferase (tVluc) can be successfully expressed in an *E. coli* bacterial system given that appropriate care is taken to ensure the cytoplasmic environment and expression conditions are amenable to disulfide formation for proper folding. The tVluc was obtained from the soluble fraction in pure form and exhibited a glow-type bioluminescent emission with a half-life of over two hours. The truncated form of the luciferase is slightly smaller than *Renilla* luciferase (39.5 kDa), thus tVluc could be used in biosensor-based applications where size is a critical factor [113]. The expression of tVluc in *E. coli* will be a guide for further production of *Vargula* and *Cypridina* luciferase variants which will have unprecedented value in the design and implementation of multiplexed reporter systems.

4.5 Methods

4.5.1 Molecular Cloning

The 555 amino acid sequence for Vluc from the NCBI GenBank (accession number `AAA30332`, "`luciferase [Vargula hilgendorfii]`") was used to create the tVluc sequence for expression in *E. coli*. The signal peptide, which was identified using the online prediction tool provided by the Center for Biological Sequencing Analysis at the Technical University of Denmark[16] [162], was predicted to be the first 15 amino acids of the native sequence and is not present in the truncated sequence.

The tVluc sequence used in this study is 302 amino acids long with a theoretical molecular weight of 33.3 kDa, and consists of the Pro28–Cys312 segment of the native sequence with an N-terminal 6× histidine tag and factor Xa cleavage site. The sequence synthesized commercially (GeneCopoeia Inc., Rockville, Maryland) codon optimized for expression in an *E. coli* host and inserted into the pCold-I Cold Shock Expression System vector (Takara Bio. Inc., Japan) using the NdeI (`CA*TATG`) and XhoI (`C*TCGAG`) restriction sites of the multiple cloning site (see Figure 4.3 on page 85). The pCold-I::tVluc (ptVluc) plasmid was transformed into the cloning strain NEB5-α (New England Biolabs, Ipswich, Massachusetts) for propagation and storage.

4.5.2 Expression and Purification from *Escherichia coli*

After propagating the ptVluc by growing at 37 °C, the plasmid was purified using a QIAprep Spin Miniprep Kit (Qiagen, Valencia, California) and transformed into three different expression strains, NEB Express (New England Biolabs, Ipswich, Massachusetts), Origami™ 2 (Novagen - EMD Millipore, Billerica, Massachusetts) and SHuffle® Express (New England Biolabs, Ipswich, Massachusetts). As a preliminary step to culture expansion, small cultures were prepared for overnight growth

---

[16] `http://www.cbs.dtu.dk/`

at $37\,^\circ$C by inoculating 5 mL of Miller LB supplemented with 0.1 mg/mL ampicillin from glycerol stocks of tVluc for each of the three expression strains. Following overnight growth, the small cultures were centrifuged at $4,000\times$g, resuspended in fresh media supplemented with 0.1 mg/mL ampicillin, and used to inoculate large 250 mL cultures.

Protocol for NEB Express Large 250 mL cultures were grown to an OD_{600nm} of 0.6 in Miller LB supplemented with 0.1 mg/mL ampicillin before being cold-shocked on ice for 1 hour. Cultures were then induced with a final concentration of 1 mM IPTG, followed by growth at $15\,^\circ$C for 48 hours. The cells were collected by centrifugation at $4000\times$g and $4\,^\circ$C for 20 minutes and resuspended in a lysis buffer of 50 mM sodium phosphate pH 8.0, 300 mM sodium chloride, 10 mM imidazole, and 0.05% (vol.) polyoxyethylene (20) sorbitan monolaurate (Tween 20). The insoluble material was removed by centrifugation at $17,000\times$g and $4\,^\circ$C for 20 minutes and the supernatant was filtered by syringe through a 0.22 μm filter. The filtered crude protein was then incubated with 500 μL of Ni$-$NTA agarose (Qiagen, Valencia, California) per culture at $4\,^\circ$C for 2 hours, collected on a Pierce^{TM} Centrifuge Column (Life Technologies, Grand Island, New York) by gravity flow, and washed with 2 column volumes of a wash buffer of 50 mM sodium phosphate pH 8.0, 300 mM sodium chloride, and 20 mM imidazole. The protein was then eluted in 500 μL increments with a stepwise gradient of imidazole in 50 mM sodium phosphate pH 8.0, 300 mM sodium chloride (four column volumes were collected at each stepped concentration).

Protocol for Origami^{TM} 2 Large 250 mL cultures were grown to an OD_{600nm} of 1.6 in Terrific broth supplemented with 0.1 mg/mL ampicillin, replenishment of media following cold-shock on ice for 1 hour, induction with a final concentration of 0.1 mM IPTG, followed by growth at $15\,^\circ$C overnight. The cells were collected by centrifugation at $4000\times$g and $4\,^\circ$C for 20 minutes and resuspended in a lysis

buffer of 50 mM Tris · HCl pH 8.0, 150 mM sodium chloride, 10 mM imidazole, 1% (vol.) nonyl phenoxypolyethoxylethanol (NP-40), 0.2% (vol.) Tween 20, and 10 mM 2-mercaptoethanol (β-ME). The cell suspension was supplemented with 1× ProBlock^{TM} Gold Bacterial Protease Inhibitor Cocktail (Gold Biotechnology Inc., St. Louis, Missouri) to prevent degradation following lysis and sonicated as described above. The insoluble material was removed by centrifugation at $10,000 \times$g and $4\,^\circ$C for 30 minutes and the supernatant was filtered by syringe through a 0.22 μm filter. The filtered crude protein was then incubated with 250 μL of Ni−NTA agarose per culture at $4\,^\circ$C for 1–2 hours, collected on a centrifuge column by gravity flow, washed with 10 column volumes of lysis buffer followed by 20 column volumes of a wash buffer of 50 mM Tris · HCl pH 8.0, 150 mM sodium chloride, and 20 mM imidazole. The protein was then eluted with an elution buffer of 50 mM Tris·HCl pH 8.0, 150 mM sodium chloride, and 150 mM imidazole in 1 column volume increments.

Protocol for SHuffle$^{®}$ Express Expression SHuffle$^{®}$ Express proceeds similarly to the protocol for Origami^{TM} 2, except the large cultures are grown to an OD$_{600nm}$ of 1.2 and induced with a final concentration of 0.5 mM IPTG. The protein was only bound to the Ni−NTA agarose for 30–45 minutes and subsequently purified in the same manner.

All purifications were analyzed by SDS-PAGE using 4–20% gradient Mini-PROTEAN$^{®}$ TGX^{TM} Precast Gels ((Bio-Rad Laboratories Inc., Hercules, California) under denaturing conditions with running buffers containing sodium dodecyl sulfate (SDS).

The purified tVluc was then dialyzed into either a phosphate buffered saline (PBS) of 10 mM sodium phosphate pH 7.2 and 150 mM sodium chloride or a CD buffer of 5 mM potassium phosphate pH 7.8 and 25 mM ammonium sulfate, depending on the intended application following purification. Slide-A-Lyzer^{TM} Dialysis Cassettes (Life Technologies, Grand Island, New York) with a molecular weight cutoff (MWCO) of 10 kDa were used.

4.5.3   Characterization of Bioluminescence

Spectra   The bioluminescence spectrum for tVluc was obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies, Santa Clara, California) equipped with a microplate reader accessory using the bio-/chemi-luminescence mode. Measurements were carried out in a 96-well black polystyrene non-binding microplate (Greiner Bio-One Inc., Monroe, North Carolina). *Cypridina* luciferin was purchased from NanoLight[TM] Technology (Prolume Ltd., Pinetop, Arizona).

Briefly, 100 μL of 300 μM vargulin was injected into 200 μL of luciferase at a concentration of $\sim$1.5 μM and the spectrum was recorded. Using the Cary Eclipse Scan software package, the spectrum was scanned from 400–650 nm with a 100-500 ms dwell time. Using the software CAT mode, 10 separate spectra were recorded, normalized, and averaged to produce the final spectrum.

Kinetic studies   Decay kinetics were obtained using a POLARstar[®] Optima (BMG LABTECH GmbH, Ortenberg, Germany). The emission filter was set to "lens" (i.e. no filter) and total light was collected for all measurements. All kinetic measurements were carried out in 96-well black polystyrene non-binding microplates (Greiner Bio-One Inc., Monroe, North Carolina). *Cypridina* luciferin was purchased from Nanolight[TM] Technology (Prolume Ltd., Pinetop, Arizona). Briefly, 100 μL of luciferase was added to a 96-well microplate at a concentration of 1.4 μM. The bioluminescent emission was quantified by integrating the signal in 10 second intervals for 20 minutes. At 30 seconds, 100 μL of 100 μM vargulin was injected to begin the reaction.

4.5.4   Circular Dichroism Spectroscopy

CD spectroscopy was performed on a Jasco J-815 Circular Dichroism Spectropolarimeter using a 0.1 cm path length quartz cuvette. tVluc was dialyzed into CD buffer and diluted to a concentration of 0.05 mg/mL. The spectrum was acquired

from 260–185 nm using the following parameter set:

Parameter	Value
Scan Mode	*Continuous*
Scan Speed	*50 nm/min*
Data Pitch	*0.5 nm*
Bandwidth	*2 nm*
D.I.T.	*2 s*
Accumulations	*5*

4.5.5   Computational Analysis of Structure

All analysis of the CD spectrum was performed using DichroWeb[17] (Department of Crystallography, Institute of Structural and Molecular Biology, Birkbeck College, University of London, United Kingdom) [170]. The CDSSTR analysis program [171] was used to generate secondary structure assignments for tVluc using reference sets 3, 4, 6, 7, SP175, and SMP180. The CDSSTR analysis program was chosen as it provided the best fit to experimental data, however, all analysis programs available through DichroWeb including SELCON3, CONTIN, VARSLC, and K2D were evaluated. Data input for analysis was limited to the wavelength range of 240–185 nm.

---

[17] `http://dichroweb.cryst.bbk.ac.uk/html/home.shtml`

Chapter 5

Prospective Research

5.1 Perspective

As discussed in Chapter 1, *Introduction / MicroRNA Detection*, microRNA and more generally nucleic acids have great potential to serve as novel biomarkers for several diseases ranging from cancer, to metabolic disorders, to viral infections. The purpose of the work discussed in this dissertation is to create novel reporter elements and biosensors that can be applied to these emerging clinical applications of nucleic acid biomarkers, and used to develop new diagnostic techniques that work in tandem with the extremely powerful bioinformatic approaches to personalized medicine that are emerging today.

This task comes with many inherent challenges. One of the major difficulties in biosensor design arising from the necessity for reliable a reproducible bioconjugation techniques that utilize robust chemistries that can be applicable to a variety of design strategies. Luckily, the field of bioconjugate techniques is rapidly growing thanks to many collaborative and ingenious advances made by bioanalytical research groups all across the world [161].

New variants of luciferases with novel bioluminescent properties are needed to develop the next generation of biosensors that will help answer some of the questions posed by these new diagnostic needs. And in conjunction with clever new chemistries, more robust biosensors can be created.

This chapter is titled *Prospective Research* as much headway has already been made on many of the projects and it is not difficult to see the future that lay ahead of each or the impact that this variety of work can have on bioanalytical research. The following sections are brief roadmap discussions of individual projects which have stemmed from the work already presented in this dissertation.

5.2 Protein Modification through Tyrosine Residues Using Diazonium Salts

One specific hurtle we have faced in our biosensor designs is the desire to incorporate highly-active luciferases into current sensor designs, specifically, as it pertains to the work discussed in this dissertation, into BSLPs [113]. The method used for chemical modification and onjugation of luciferase to the SLP oligonucleotide (see Chapter 2 Section 2.4.2 on page 44) worked well for the attachment of Rluc. However, when applied to Gluc, the chemical modification with SANH (now called by its commercial name S-HyNic) to introduce aldehyde-reactive hydrazine functionality rendered the luciferase inactive.

The lack of structural data available for Gluc greatly increases the level of difficulty involved in using it as a reporter in bioconjugation methods. The addition of a C-terminal hinge with reactive cystine residue by Inouye et al. [152] was a great stride forward in this regard, however thiol chemistry can become problematic in proteins already having high cysteine content and requiring precise disulfide bridge formation for functionality. While Inouye et al. addressed this somewhat in the hinge design, it becomes an unpredictable factor when used in conjunction with truncated mutants like those discussed in Chapter 3 on page 54.

To address this unpredictable variable of cysteine modification in Gluc variants, we have explored new chemistries for the modification of luciferases that are more

predictable in their result, more reliable, and are performed easily with a broader range of available functionality. As previously mentioned, lysine residues are an obvious choice for bioconjugation as the primary amine makes an excellent nucleophilic handle for reaction with widely available reagents containing succinimidyl ester moieties. Lysine residues are, however, commonly abundant in protein sequences. This can lead to over-modification of the protein making a variety of difficult to characterize multimers that can potentially crash out of solution or lose their binding or reporting functionality. Tyrosine residues, on the other hand, are much less common than lysine residues and do not have the same level of structural importance as cysteine residues; thus the phenolic moiety could provide a very specific handle for bioconjugation.

Gavrilyuk et al. described the synthesis and utility of a bench-stable diazonium salt 4-Formylbenzene diazonium hexafluorophos- phate (FBDP), which can be used to introduce a bioorthogonal aldehyde ortho to the phenol group on a tyrosine residue [160]. This bioorthogonal aldehyde can then be used in a secondary reaction like that described using S-HyNic in Chapter 2 Section 2.4.2 on page 44 (see Figure 5.1).
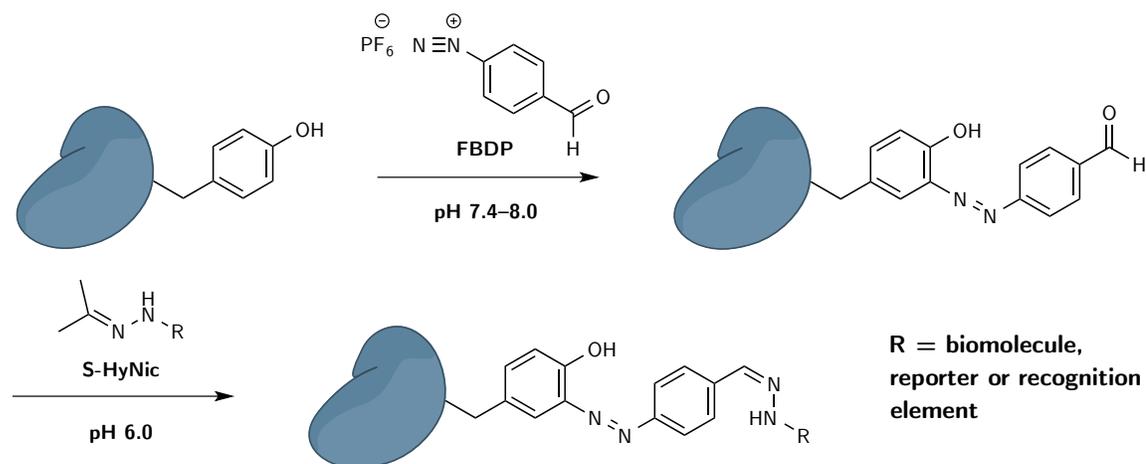
Figure 5.1. Tyrosine residues may be modified with FBDP to introduce a bioorthogonal aldehyde for subsequent modification with S-HyNic.

We have already synthesized FBDP following the procedure published by Gavrilyuk et al. [160] (see also Figure 5.2). Briefly, the reaction proceeds as follows:

1. A 0.7 M aqueous solution of sodium nitrate (1.2 eq) is added to a 0.5 M suspension of 4-aminobenzaldehyde polymer (1.0 eq) in 12 N HCl at $-15\,°C$.

   The 4-aminobenzaldehyde polymer suspension is chilled in an ice/acetone bath. The aqueous sodium nitrate solution is chilled in an ice bath and added drop-wise to the polymer suspension to obtain better temperature control.

2. The solution is stirred for 1.5 hours at $-15\,°C$.

3. A 60% aqueous solution of hexafluorophosphoric acid HPF_6 (1.7 eq) is added drop-wise, keeping the temperature at $-15\,°C$.

   Before the addition of HPF_6, the solutio should be orange-brown. Filter out any brown solid if it exists before proceeding.

4. The solution is stirred for 1 hour at $-15\,°C$ followed by 30 minutes at room temperature.

5. Filter to collect the white-yellow solid and wash with 10 mL of cold water and 10 mL of cold ethyl acetate.

6. Store the solid product at $-20\,°C$ under argon.
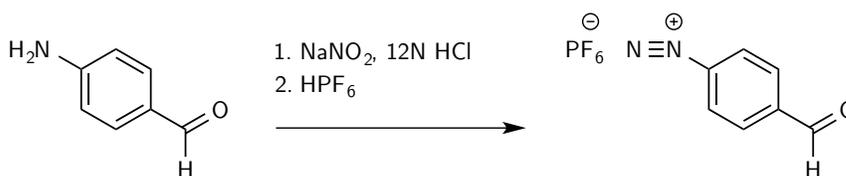
Figure 5.2. The facile synthesis of FBDP.

To date, we have successfully used FBDP synthesized according the the described protocol to modify the tyrosine residue incorporated into the GlucY variants described in Chapter 3 on page 54. The FBDP modification can be monitored by UV-Vis spectroscopy (see Figure 5.3) and has also been verified using a colorimetric Aldehyde Quantification Assay Kit (Abcam, Cambridge, Massachusetts).
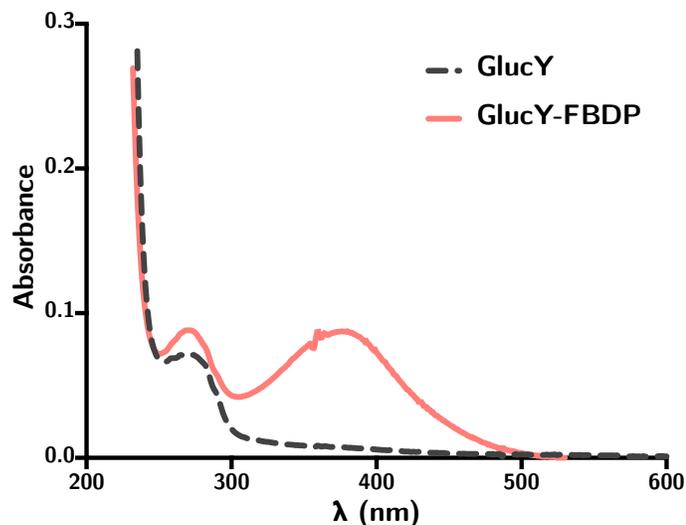
Figure 5.3. Modification of GlucY with FBDP yields a functionalized product that absorbs strongly around 380 nm. This allows for a convenient method for monitoring the reaction progress.

5.3 Modification of the Bioluminescent Stem-Loop Probe Design

While the BSLP probe described in Chapter 2 on page 36 is a novel concept, the quenching of the bioluminescent signal could be handled in a more elegant and/or complete manner. The possibility of using multiple quencher assemblies was discussed as a means to lower background emission and improve probe sensitivity. Additionally, the use of different quenchers that could potentially have better spectral or electronic overlap with coelenterazine emission are being evaluated; BHQ-0 (Biosearch Technologies Inc., Petaluma, California) is currently being evaluated in the BLSP design as a promising alternative to DABCYL.

Multiple-quencher assemblies and alternative energy transfer-based quenchers may reduce the background to some degree, however, a completely alternative approach to energy transfer-based quenching is possible through exploiting the inherent enzyme/substrate behavior of the luciferase, i.e. the use of a competitive chemical inhibitor of the substrate binding pocket of the luciferase could drastically lower
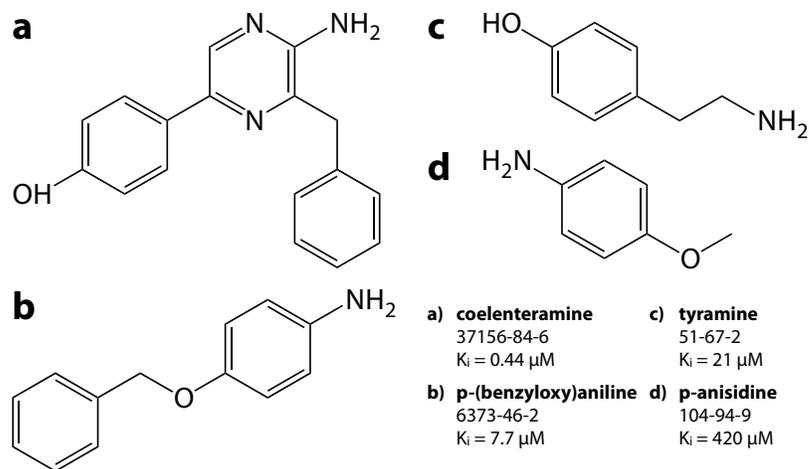
Figure 5.4. Known molecules to employ as active-site inhibitor quenchers. CAS and Kd listed for each. Many more additional compounds exist in the literature.

the background and improve biosensor sensitivity by eliminating the bioluminescent reaction altogether when the BSLP is in the closed state.

Several studies have identified mechanistic inhibitors that can act on the luciferase active site with varying degrees of affinity (see Figure 5.4 [180–186]. These inhibitors bear structural similarity to the native substrate and thus quench bioluminescence via a competitive process which disrupts enzymatic catalysis. In a preliminary study, we have demonstrated this effect using 4-(benzoxy)aniline as a competitive inhibitor in solution (see Figure 5.5a on the following page). Tethering an inhibitor to the SLP using a flexible linker will bring it closer to the active site in the closed conformation (see Figure 5.5b on the next page). Tethering of active site inhibitors has been reported in the literature for gated control of channel proteins [187].

Inclusion of a sufficiently soluble linker between the SLP and the inhibitor will provide added flexibility to ensure proper orientation and binding in the active site. We plan to attach the linker to the inhibitor compounds through aliphatic and aromatic primary amines present in their structures (see Figure 5.4 on the preceding page). A polyethylene glycol (PEG) linker containing an fluorenylmethyloxycarbonyl (FMOC) protected amine and a carboxylic acid (COOH) will be used. This linker is available in
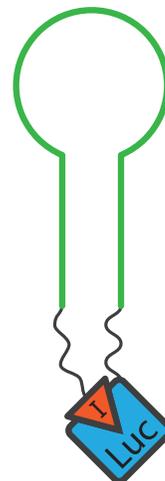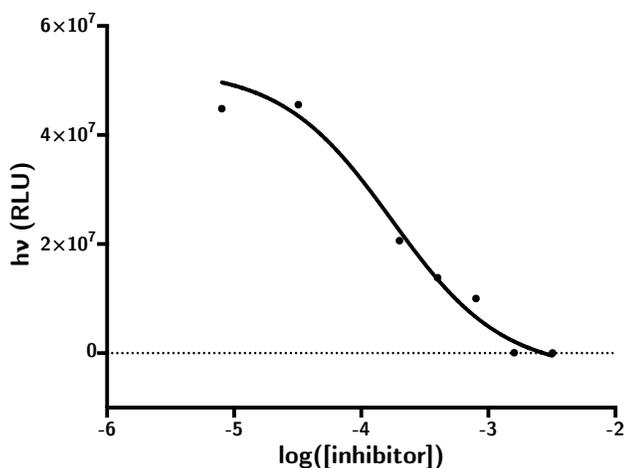
(a) Competitive inhibition of Rluc by 4-(benzoxy)aniline (compound *b* in Figure 5.4 on the previous page).

(b) BSLP constructed with an active site inhibitor.

Figure 5.5. Competitive chemical inhibition of luciferase bioluminescence.

varying sizes from several companies (Quanta BioDesign, Ltd., Powell, Ohio; BioBlocks, Inc., San Diego, California; and Sigma-Aldrich Co., St. Louis, Mossouri) allowing for the optimal length linker to be selected based on experimental results.

There are many well established chemistries that may be used to attach the inhibitor to the linker:

1. N,N'-disuccinimidyl carbonate (DSC) chemistry [188–190]

2. n-propanephosphonic acid anhydride (T3P) [191–196]

3. ethyl 2-cyano-2-(hydroxyimino)acetate (Oxyma)-based uronium coupling agent (COMU) [197–200]

4. 2,4,6-triacyloxy-1,3,5-triazine. [201–203]

The inhibitor-linker products will be characterized by ^{1}H-NMR as well as HPLC to verify their purity. The inhibitor-linker molecule will be conjugated to the SLP via well-established EDC chemistry. The conjugation of luciferase to the SLP can proceed as previously mentioned (Figure 13).

To date, we have successfully attached the inhibitor coelenteramine (compound *a* in Figure 5.4 on page 101) to the PEG linker {2-[2-(Fmoc-amino)ethoxy]ethoxy} acetic acid (`CAS Number 166108-71-0`) using COMU, and removed the Fmoc group by de-protection using piperidine (see Figure 5.6). The product was confirmed by ^{1}H-NMR.
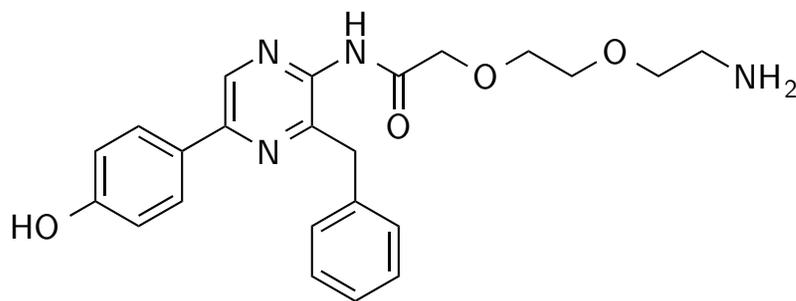
Figure 5.6. PEG linker attached to coelenteramine (compound *a* in Figure 5.4 on page 101) with amine de-protected and ready for subsequent attachment to carboxy-functionalized SLP.

We are currently using this newly synthesized inhibitor-linker molecule to construct an inhibitor-based BSLP. Using an amine/carboxy-modified oligonucleotide as the scaffold, the amine terminus will be modified with S-HyNic to present a hydrazine functionality, followed by attachment of the inhibitor-linker to the carboxy terminus using EDC/NHS. Finally, the FBDP-modified GlucY variants discussed in Chapter 3 will be attached through the hydrazine functionality. We believe this will be the next generation of BSLP and offer superior sensitivity for difficult to detect targets such as latent HIV RNAs.
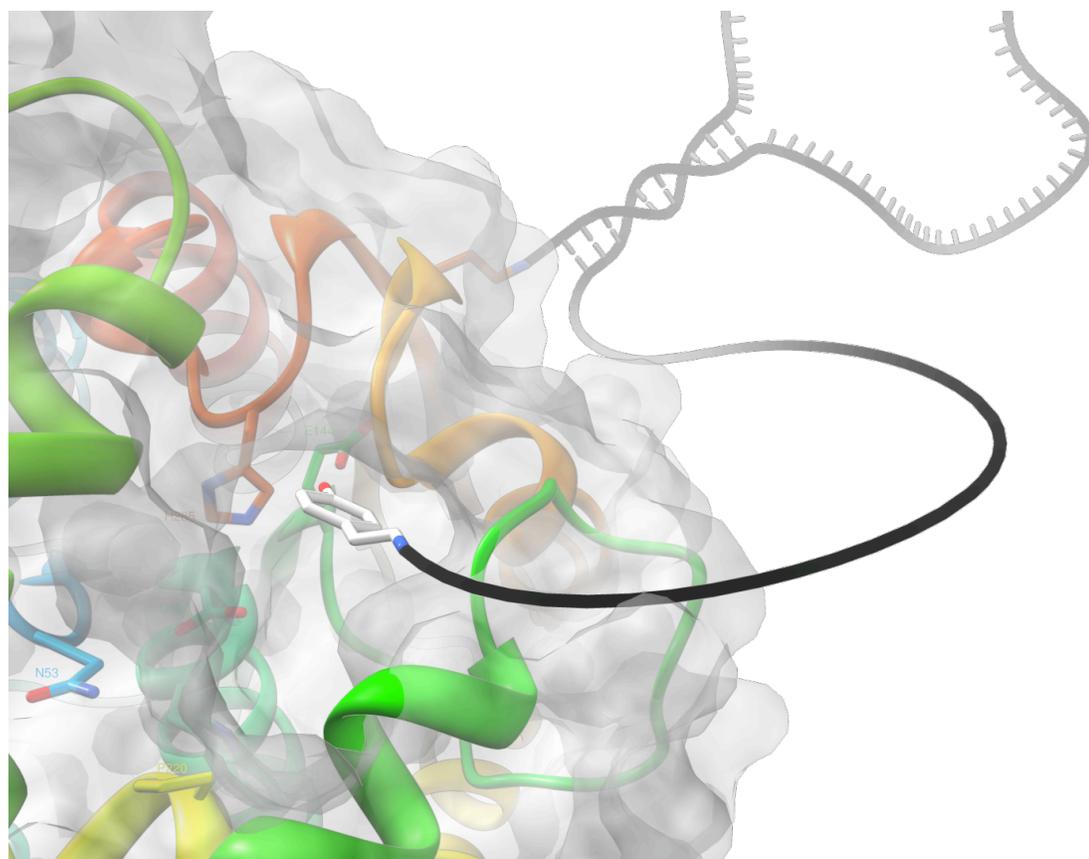
Figure 5.7. An artistic representation of a BSLP based on a tethered mechanistic inhibitor of the enzymatic active site.

5.4 *Gaussia* Luciferase-Modified Antibodies for Detection of the Latent Reservoir in HIV-1 Infection

Antiretroviral therapy (ART) has greatly reduced the mortality associated with HIV-1 infection, however, the infection is not cured by this treatment. Persistency of the infection is perpetuated by a latent reservoir of resting CD4$^+$ T cells, which retain the latent provirus of HIV-1 as an integrated component of the host genome. The decay rate half-life of the latent reservoir of HIV$^+$ patients undergoing ART has been shown to be nearly four years, and as such, it has been estimated that if the latent reservoir consists of just 1×10^5 cells, total eradication could take as long as 60 years [204], thus effectively requiring patients to endure lifelong ART therapy. This aspect of HIV-1 infection constitutes the primary clinical challenge in the effort to develop a treatment which results in eradication of the latent reservoir (i.e. a cure for the infection) [204, 205].

The method known as the quantitative viral outgrowth assay (QVOA), is currently the most reproducible and reliable method available for quantitatively measuring the latent reservoir (see Figure 5.8 on the next page) [206]. The method is quite involved, however, and requires ~2 weeks of culture time during which CD4$^+$ T cells derived from patient peripheral blood mononuclear cells (PBMC) are serially diluted and activated with the phytohaemagglutinin (PHA) to reverse latency and restart the production of infectious HIV-1, which is then further propagated in lymphoblasts from healthy donors and detected by enzyme-linked immunosorbent assay (ELISA). The process is considered expensive, laborious, and lengthy in the time required to obtain a final result [205].

Reactivation of latent HIV-1 provirus is being explored as a potential avenue—in conjunction with other therapies such as T cell vaccination—for reservoir eradication [205]. As such, methods are required to rapidly quantitate the latent reservoir in an

economical fashion scalable to the demands of large clinical studies. It has been shown that upon reinitiating the active HIV-1 infection, the levels of endogenous interferon gamma (IFNγ) are increased. We have proposed a method for assessing the size of the latent reservoir using a modified QVOA in which reactivated cells are directly lysed allowing for the released IFNγ to be captured and detected on a solid phase platform in a sandwich-style ELISA method utilizing a GlucY-modified IFNγ antibody (see Figure 5.9 on the following page).

The bioconjugation of antibodies with other reporter biomolecules can be a difficult task, and is often fraught with heterogeneous and incomplete labeling. We aim to improve the reproducibility and reliability of antibodies labeled with bioluminescent reporters, which should also improve upon the sensitivity and lower limit of detection of assays utilizing them by creating an IFNγ antibody labeled with GlucY through the reactive tyrosine linker. The single attachment point should minimize adverse effects of bioconjugation on the bioluminescent activity and make it easier to control reaction stoichiometry with regards to labeling of the antibody. Currently, we have modified IFNγ antibody with GlucY and are characterizing it in the early proof-of-concept stages of the assay described in Figure 5.9 on the next page.
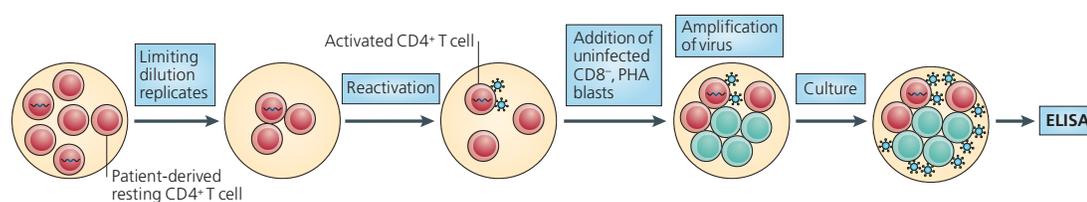
Figure 5.8. The current quantitative viral outgrowth assay (QVOA) method. Figure reprinted from Archin et al. [206].
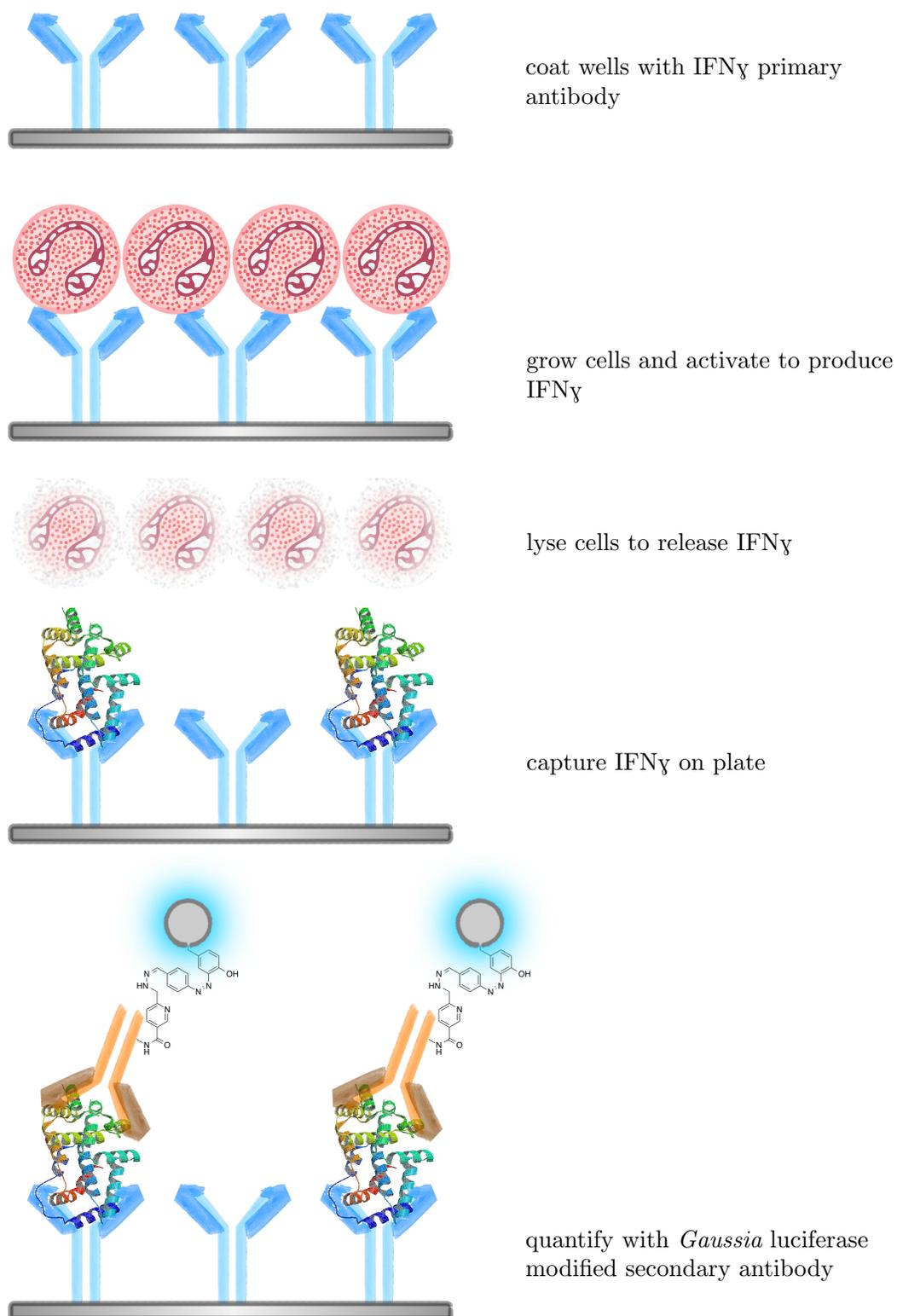
Figure 5.9. GlucY-modified IFNγ antibody for detection of latent HIV infection following cellular activation.

5.5 Self-Amplifying RNAs as an Alternative Reporter System

We believe that bioluminescent reporters have already been demonstrated to be very powerful analytical tools for a variety of applications, and that there is great promise for their future use. However, we feel it is also prudent to also consider other, less obvious ways in which bioluminescent reporters may be utilized in bioanalytical applications. In a relatively new project, we are investigating the use of self-amplifying RNAs containing luciferase genes as deliverable reporters for targeted cellular imaging applications.

The use of mRNA as a therapeutic molecule has only recently grown in consideration. The relative instability of mRNA in comparison to DNA has made plasmid DNA (pDNA) a more attractive option. However, recent advances in viral and, more importantly, non-viral delivery of RNA to cells has opened the doors to a new field of research (see Figure 5.10) [207]. In many ways, the use of mRNA as a therapeutic molecule surpasses that of plasmid DNA. Perhaps the greatest advantage is that mRNA does not need to cross the nuclear barrier once delivered to the cytoplasm. Additionally, mRNA presents fewer immunogenic complexities due to the lack of un-methylated cytosine-phosphate-guanine (CpG) motifs present in bacterial and some viral DNA [207].
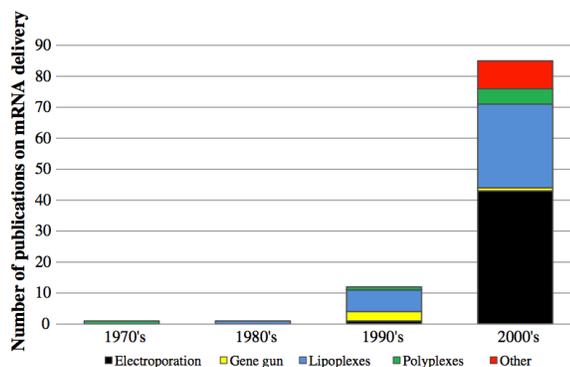
Figure 5.10. Number of publications on mRNA delivery in cells by non-viral means. Figure reprinted from Tavernier et al. [207].

Self-amplifying RNAs for luciferase reporters will be made by creating a chimeric RNA sequence containing both a luciferase cDNA sequence and the sequence for the alphavirus polyprotein responsible for RNA replication. Upon delivery to the cytoplasm of a cell, the viral machinery takes over, begins replicating itself, and subsequently producing mRNA for the luciferase reporter protein. The cell then translates this mRNA and produces active reporter within the cytoplasm, which can be imaged using a variety of techniques.

The alphavirus genome is a single-stranded polycistronic RNA sequence consisting of two open reading frames (see Figure 5.11). The first reading frame codes for four non-structural proteins (NSP1-4) responsible for RNA replication—these proteins are expressed as a single polyprotein and later cleaved by viral (cysteine-targeting) and host proteases. The second reading frame codes for five structural proteins (also expressed as a polyprotein) responsible for viral encapsidation and budding, and is translated from a subgenomic 26S RNA promoter. To create a self-replicating RNA, the structural proteins encoded by the second reading frame are removed and replaced by the luciferase cDNA sequence.

Figure 5.11. Single-stranded polycistronic RNA sequence of the alpha virus genome. The structural proteins responsible for virulence have been removed and replaced with a gene of interest, in this case a luciferase reporter. *This figure was reproduced from* [208].

The self-amplifying sequence was produced as a DNA plasmid using standard molecular biology techniques. The cassette was created by cloning the Gluc gene from plasmid constructs already established in our laboratory and appending into the alphavirus NSP sequence using overlap PCR. The plasmids created were amplified in *E. coli* and purified using a QIAprep Spin Miniprep Kit (Qiagen, Valencia, California). The DNA sequence was then linearized by restriction digest with XhoI (`GC*GGCCGC`)

and transcribed into RNA using the mMESSAGE mMACHINE® SP6 Transcription Kit (Life Technologies, Grand Island, New York).

This finished self-amplifying RNA product will be incorporated into cellular delivery systems and subsequently used for targeted in vivo bioluminescence imaging (IVIS).

5.6   Final Remarks

The research presented in this dissertation has highlighted the importance of nucleic acids as biomarkers, especially in the modern clinical research environment, which is utilizing bioinformatics and high-throughput technologies to discover new correlations between human disease, our genetics, and the genetic basis of infectious and biological afflictions. As new biomarkers are discovered are constantly being discovered the need for clinical diagnostics regarding these markers is essential. The utility of bioluminescent proteins in creating such diagnostic systems has been demonstrated with regard to a microRNA target present in human serum and associated with breast cancer. As the discovery of new biomarkers will drive the need for more robust sensors, the design and production of new *Gaussia* luciferase variants with a variety of kinetic and spectral characteristics was discussed, ultimately creating smaller versions of the luciferase that is easier to implement in bioconjugation strategies and is more amenable to cellular delivery applications. Additionally, a truncated form of *Vargula* luciferase intended for high-throughput applications was discussed, which, through carful consideration of molecular conditions, was expressed and purified from the soluble fraction of a more economical bacterial host. The consideration of these factors has led to the exploration of more robust bioconjugation techniques and reliable chemistries for the reproducible production of high-quality biosensors.

Appendix I[18]

Table A1.1. Gluc RADAR analysis

No. of Repeats	Total Score	Length	Diagonal	BW-From	BW-To	Level
2	198.08	54	68	43	97	1
43-97	(98.96/50.11)		DRGKLPGKKlPLEVLKEMEANAR.KAGCTRGCLICLSHIKCTPKMKKFIPGRCHTY			
114-168	(99.12/47.08)		DIPEIPGFK.DLEPMEQFIAQVDlCVDCTTGCLKGLANVQCSDLLKKWLPQRCATF			
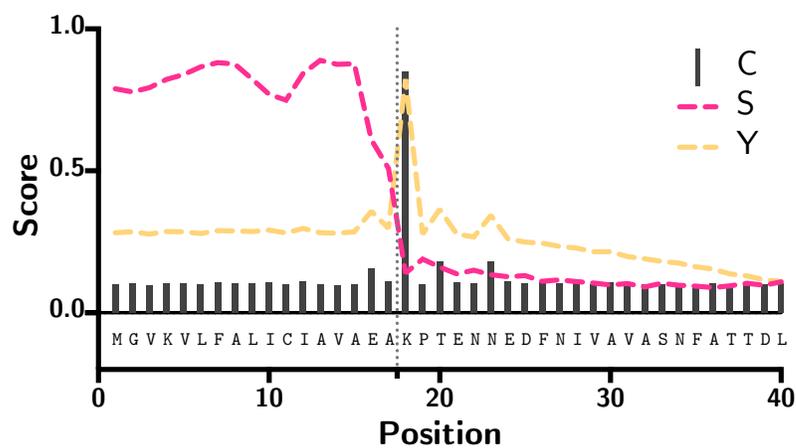
Figure A1.1. SignalP 4.1 Prediction for Gluc.

(a) elutions
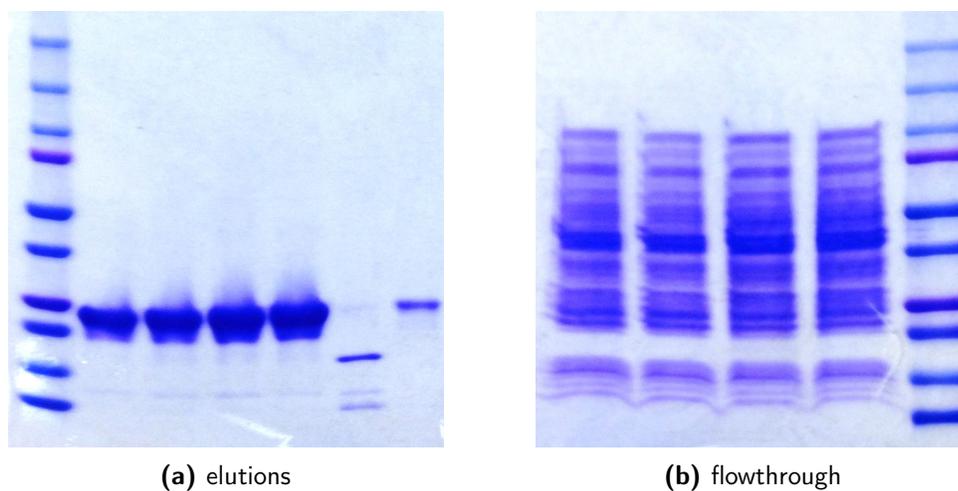
(b) flowthrough

Figure A1.2. Elutions (left) and flowthrough (right) from the optimized expression protocol for Gluc.
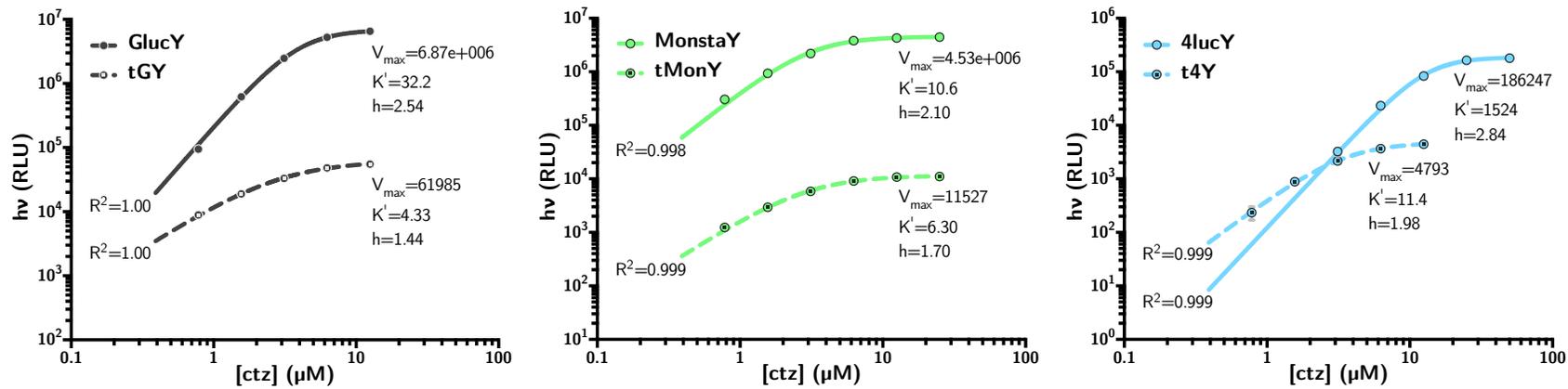
Figure A1.3. Michaelis–Menten plots for GlucY variants

GlucY · ● **tGY** ○

$V_{max}=6.87e{+}006$
$K'=32.2$
$h=2.54$

$R^2=1.00$

$R^2=1.00$

$V_{max}=61985$
$K'=4.33$
$h=1.44$

MonstaY ● **tMonY** ⊙

$V_{max}=4.53e{+}006$
$K'=10.6$
$h=2.10$

$R^2=0.998$

$R^2=0.999$

$V_{max}=11527$
$K'=6.30$
$h=1.70$

4lucY ● **t4Y** ⊙

$V_{max}=186247$
$K'=1524$
$h=2.84$

$R^2=0.999$

$R^2=0.999$

$V_{max}=4793$
$K'=11.4$
$h=1.98$

hν (RLU)

[ctz] (µM)

(a) CD spectra for full-size native variants

(b) CD spectra for Gluc and GlucY

(c) CD spectra for full-size and truncated WT Gluc

(d) CD spectra for full-size and truncated Monsta

(e) CD spectra for full-size and truncated 4luc

(f) averaged CD spectra for full-size and truncated Gluc variants plotted with range for each dataset

Figure A1.4. CD comparison spectra overlays

Table A1.2. Primer sequences: the M43I mutation was carried out using the QuikChange kit; the Monsta and 4luc mutations were carried out using the Q5® kit. The codons changed for each mutation are underlined. (Note that the M43I mutation is included in the L40P_f primer and that this mutation differs from the M43V mutation found in Gluc4 produced by Degeling et al. [148]). The tyrosine hinge was introduced using the Q5® kit. The forward primer (luc_f) is the same for both the full-size and truncated reactions. The lowercase portion contains the C-terminal tyrosine insert and is not complementary to pGluc.
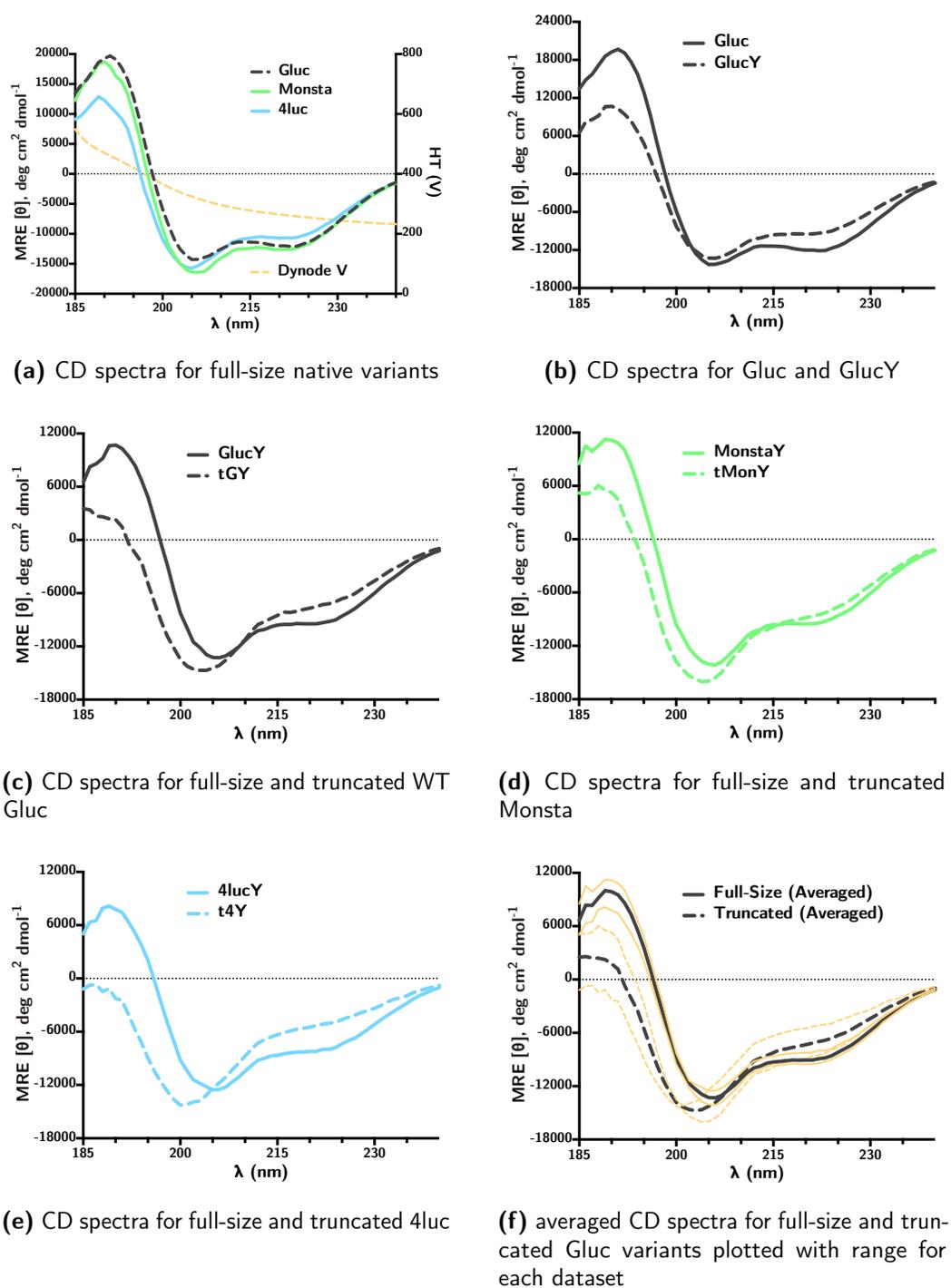
Name	(bp)	T_m (°C)	Sequence (printed 5'->3')
for Sanger sequencing			
pCold_f	19	69.1	ACGCCATATCGCCGAAAGG
pCold_r	20	31.7	GGCAGGGATCTTAGATTCTG
for Monsta mutations [147] - F89W/I90L/H95E/Y97W			
F89W/I90L_f	35	85.5	GCCGAAAATGAAAAAATGGCTCCCGGGCCGTTGTC
F89W/I90L_r	40	85.1	GTGCATTTGATATGGCTCAGACAGATCAGGCAACCACGGG
H95E/Y97W_f	41	84.4	GGGCCGTTGTGAGACCTGGGAAGGTGATAAAGAATCTGCAC
H95E/Y97W_r	53	84.3	GGGATAAATTTTTTCATTTTCGGCGTGCATTTGATATGGCTCAGACAGATCAG
for M43I mutation - M43I			
M43I_f	41	79.6	CTGGAAGTTCTGAAAGAAATTGAAGCAAACGCACGTAAAGC
M43I_r	41	79.6	GCTTTACGTGCGTTTGCTTCAATTTCTTTCAGAACTTCCAG
for 4luc mutations [148] - L30S/L40P/(*M43I*)			
L30S_f	34	86.6	CTGGACGCTGACCGTGGTAAATCGCCGGGCAAAA
L30S_r	43	83.2	GTCGGTCGTAGCAAAGTTACTCGCCACTGCCACGATGTTAAAG
L40P_f	32	77.8	CTGCCGCTGGAAGTTCCGAAAGAAATTGAAG
L40P_r	27	77.3	TTTTTTGCCCGGCAGTTTACCACGGTC
for tyrosine hinge mutations			
lucY_f	44	56	ctccgtctaccccgccgtactaatTAATCTCTGCTTAAAAGCAC
GlucY_r	43	59	acggggtcggcggggtagacagagaTCTAGAATCACCACCTGC
tGlucY_r	43	58	acggggtcggcggggtagacagagaCGGGATATCGACAATAGC

```
Gluc    |AAG54095.1  1   .MGVKVLFALICIAVAEAKP.TENNEDFNIVAVASNFATTDL........
Mluc164 |AAR17541.1  1   .MDIKVVFTLVFSALVQAKS.TEFDPNIDIVGLEGKFGITNLETDLFTIW
Mluc39  |ABW06650.1  1   .MDIKVLFALICIALVQANP.TENNDHINIVGIEGKFGITDLETDLFTIW
Mluc7   |AJC98141.1  1   .MDIKFIFALVCIALVQANP.TVNN.........................
Mpluc1  |BAG48249.1  1   MMEIKVLFALICFALVQANP.TENKDDIDIVGVEGKFGTTDLETDLFTIV
Mpluc2  |BAG48250.1  1   .MGVKLIFAVLCVAAAQAATINENFEDIDVVAIGGSFALD..........


Gluc    |AAG54095.1  41  ..........................DADRGK...LPGKKLPLEVLKEMEA
Mluc164 |AAR17541.1  49  ETMEVMIKADIADTDRASNFVATETDANRGK...MPGKKLPLAVIMEMEA
Mluc39  |ABW06650.1  49  ETNRMISTDN..........EQANTDSNRGK...MPGKKLPLAVLIEMEA
Mluc7   |AJC98141.1  24  ........................DVNRGK...MPGKKLPLEVLIEMEA
Mpluc1  |BAG48249.1  50  EDMNVISRDTN..........LVNSDADRGK...MPGKKLPLEVLIEMEA
Mpluc2  |BAG48250.1  40  .......................VDANRGGHGGHPGKKMPKEVLVEMEA


Gluc    |AAG54095.1  63  NARKAGCTRGCLICLSHIKCTPKMKKFIPGRCHTYEGDKESAQGGIGEAI
Mluc164 |AAR17541.1  96  NAFKAGCTRGCLICLSKIKCTAKMKVYIPGRCHDYGGDKKTGQAGIVGAI
Mluc39  |ABW06650.1  86  NAFKAGCTRGCLICLSKIKCTAKMKKYIPGRCHDYGGDKKTGQAGIVGAI
Mluc7   |AJC98141.1  46  NAFKAGCTRGCLICLSKIKCTAKMKQYIPGRCHDYGGDKKTGQAGIVGAI
Mpluc1  |BAG48249.1  87  NARKAGCTRGCLICLSKIKCTAKMKVYIPGRCHDYGGDKKTGQAGIVGAI
Mpluc2  |BAG48250.1  66  NAKRAGCHRGCLICLSHIKCTKKMKKFIPGRCHSYEGDKDSAQGGIGEEI


Gluc    |AAG54095.1  113 VDIPEIPGFKDLEPMEQFIAQVDLCVDCTTGCLKGLANVQCSDLLKKWLP
Mluc164 |AAR17541.1  146 VDIPEISGFKEMAPMEQFIAQVDRCASCTTGCLKGLANVKCSELLKKWLP
Mluc39  |ABW06650.1  136 VDIPDISGFKEMGPMEQFIAQVDRCTDCTTGCLKGLANVKCSELLKKWLP
Mluc7   |AJC98141.1  96  VDIPEISGFKEMEPMEQFIAQVDLCADCTTGCLKGLANVKCSELLKKWLP
Mpluc1  |BAG48249.1  137 VDIPEISGFKELGPMEQFIAQVDLCADCTTGCLKGLANVKCSALLKKWLP
Mpluc2  |BAG48250.1  116 VDMPEIPGFKDKEPMDQFIAQVDLCVDCTTGCLKGLANVHCSALLKKWLP


Gluc    |AAG54095.1  163 QRCATFASKIQGQVDKIKGAGGD.
Mluc164 |AAR17541.1  196 DRCASFADKIQKEVHNIKGMAGDR
Mluc39  |ABW06650.1  186 DRCASFADKIQSEVHNIKGLAGDR
Mluc7   |AJC98141.1  146 DRCASFADKIQKEAHNIKGLAGDR
Mpluc1  |BAG48249.1  187 DRCASFADKIQSEVDNIKGLAGDR
Mpluc2  |BAG48250.1  166 SRCKTFASKIQSQVDTIKGLAGDR
```

Figure A1.5. Alignment of Gluc with *Metridia* luciferase isoforms. Cysteine residues which are conserved between the internal homologous structural domains are colored red, while those not conserved are colored blue.

Table A1.3. Primary sequences of Gluc variants with tyrosine linker.

GlucY primary sequence (∧ marks point of truncation; linker is in italics):

```
        10         20         30         40         50         60
MNHKVHHHHH HIEGRHMKPT ENNEDFNIVA VASNFATTDL DADRGKLPGK KLPLEVLKEM

        70         80         90        100        110        120
EANARKAGCT RGCLICLSHI KCTPKMKKFI PGRCHTYEGD KESAQGGIGE AIVDIPEIPG
                                                                ∧
       130        140        150        160        170        180
FKDLEPMEQF IAQVDLCVDC TTGCLKGLAN VQCSDLLKKW LPQRCATFAS KIQGQVDKIK

       190        200
GAGGDSR*SLS* *TPPTPSPSTP* *PY*
```

MonstaY primary sequence (∧ marks point of truncation; linker is in italics):

```
        10         20         30         40         50         60
MNHKVHHHHH HIEGRHMKPT ENNEDFNIVA VASNFATTDL DADRGKLPGK KLPLEVLKEM

        70         80         90        100        110        120
EANARKAGCT RGCLICLSHI KCTPKMKKWL PGRCETWEGD KESAQGGIGE AIVDIPEIPG
                                                                ∧
       130        140        150        160        170        180
FKDLEPMEQF IAQVDLCVDC TTGCLKGLAN VQCSDLLKKW LPQRCATFAS KIQGQVDKIK

       190        200
GAGGDSR*SLS* *TPPTPSPSTP* *PY*
```

4lucY primary sequence (∧ marks point of truncation; linker is in italics):

```
        10         20         30         40         50         60
MNHKVHHHHH HIEGRHMKPT ENNEDFNIVA VASNFATTDL DADRGKSPGK KLPLEVPKEI

        70         80         90        100        110        120
EANARKAGCT RGCLICLSHI KCTPKMKKFI PGRCHTYEGD KESAQGGIGE AIVDIPEIPG
                                                                ∧
       130        140        150        160        170        180
FKDLEPMEQF IAQVDLCVDC TTGCLKGLAN VQCSDLLKKW LPQRCATFAS KIQGQVDKIK

       190        200
GAGGDSR*SLS* *TPPTPSPSTP* *PY*
```

Appendix II[19]

---

[19] *Supplementary Information* to Chapter 4 on page 81, "Expression of a soluble truncated *Vargula* luciferase in *Escherichia coli*"

Table A2.1. Primary sequence of tVluc.

```
        10         20         30         40         50         60
MNHKVHHHHH HIEGRHMPPS STPTVPTSCE AKEGECIDTR CATCKRDILS DGLCENKPGK

        70         80         90        100        110        120
TCCRMCQYVI ECRVEAAGYF RTFYGKRFNF QEPGKYVLAR GTKGGDWSVT LTMENLDGQK

       130        140        150        160        170        180
GAVLTKTTLE VAGDVIDITQ ATADPITVNG GADPVIANPF TIGEVTIAVV EIPGFNITVI

       190        200        210        220        230        240
EFFKLIVIDI LGGRSVRIAP DTANKGLISG ICGNLEMNDA DDFTTDADQL AIQPNINKEF

       250        260        270        280        290        300
DGCPFYGNPS DIEYCKGLME PYRAVCRNNI NFYYYTLSCA FAYCMGGEER AKHVLFDYVE

TC
```

References

[1] Eric A Hunt et al. "MicroRNA Detection: Current Technology and Research Strategies". In: *Annual Review of Analytical Chemistry* 8.3 (2015), pp. 3.1–3.21.

[2] R. C. Lee, R. L. Feinbaum, and V. Ambros. "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14". In: *Cell* 75.5 (1993), pp. 843–54.

[3] Victor Ambros. "microRNAs: tiny regulators with great potential." In: *Cell* 107 (2001), pp. 823–826.

[4] Minju Ha and V Narry Kim. "Regulation of microRNA biogenesis." In: *Nature reviews. Molecular cell biology* 15 (2014), pp. 509–524.

[5] Emily F Finnegan and Amy E Pasquinelli. "MicroRNA biogenesis: regulating the regulators." In: *Critical reviews in biochemistry and molecular biology* 48 (2013), pp. 51–68.

[6] Julia Winter et al. "Many roads to maturity: microRNA biogenesis pathways and their regulation." In: *Nature cell biology* 11 (2009), pp. 228–234.

[7] Richard W Carthew and Erik J Sontheimer. "Origins and Mechanisms of miRNAs and siRNAs." In: *Cell* 136 (2009), pp. 642–655.

[8] V Narry Kim, Jinju Han, and Mikiko C Siomi. "Biogenesis of small RNAs in animals." In: *Nature reviews. Molecular cell biology* 10 (2009), pp. 126–139.

[9] Martin Jinek and Jennifer A Doudna. "A three-dimensional view of the molecular machinery of RNA interference". In: *Nature* 457 (2009), pp. 405–412.

[10] Haifeng Dong et al. "MicroRNA: function, detection, and bioanalysis." In: *Chemical reviews* 113 (2013), pp. 6207–6233.

[11] Colin C Pritchard, Heather H Cheng, and Muneesh Tewari. "MicroRNA profiling: approaches and considerations." In: *Nature Reviews Genetics* 13 (2012), pp. 358–369.

[12] Mariàngels de Planell-Saguer and María Celina Rodicio. "Detection methods for microRNAs in clinic practice." In: *Clinical biochemistry* 46 (2013), pp. 869–878.

[13] Bianca C Bernardo et al. "A microRNA guide for clinicians and basic scientists: background and experimental techniques." In: *Heart, lung & circulation* 21 (2012), pp. 131–142.

[14]   Mitsuo Kato, Nancy E Castro, and Rama Natarajan. "MicroRNAs: potential mediators and biomarkers of diabetic complications." In: *Free radical biology & medicine* 64 (2013), pp. 85–94.

[15]   Ewelina Nalejska, Ewa Mączyńska, and Marzena Anna Lewandowska. "Prognostic and predictive biomarkers: tools in personalized oncology." In: *Molecular diagnosis & therapy* 18 (2014), pp. 273–284.

[16]   Shlomit Gilad et al. "Serum microRNAs are promising novel biomarkers." In: *PLoS ONE* 3 (2008), e3148.

[17]   Jin Wang et al. "Tumor-associated circulating microRNAs as biomarkers of cancer." In: *Molecules (Basel, Switzerland)* 19 (2014), pp. 1912–1938.

[18]   Mehmet Seven et al. "The role of miRNAs in cancer: from pathogenesis to therapeutic implications." In: *Future oncology (London, England)* 10 (2014), pp. 1027–1048.

[19]   C. L. Bartels and G. J. Tsongalis. "MicroRNAs: novel biomarkers for human cancer". In: *Clin Chem* 55.4 (2009), pp. 623–31.

[20]   Colles Price and Jianjun Chen. "MicroRNAs in cancer biology and therapy: Current status and perspectives". In: *Genes & Diseases* (2014), pp. 1–11.

[21]   Daniel J Hogan et al. "Anti-miRs competitively inhibit microRNAs in Argonaute complexes." In: *PLoS ONE* 9 (2014), e100951.

[22]   S. Rossi and G. A. Calin. "Bioinformatics, non-coding RNAs and its possible application in personalized medicine". In: *Adv Exp Med Biol* 774 (2013), pp. 21–37.

[23]   H. A. Ebhardt, A. Fedynak, and R. P. Fahlman. "Naturally occurring variations in sequence length creates microRNA isoforms that differ in argonaute effector complex specificity". In: *Silence* 1.1 (2010), p. 12.

[24]   Julia Starega-Roslan et al. "Structural basis of microRNA length variety." In: *Nucleic acids research* 39 (2011), pp. 257–268.

[25]   Edyta Koscianska et al. "Northern blotting analysis of microRNAs, their precursors and RNA interference triggers." In: *BMC molecular biology* 12 (2011), p. 14.

[26]   Blake N Johnson and Raj Mutharasan. "Biosensor-based microRNA detection: techniques, design, performance, and challenges." In: *The Analyst* 139 (2014), pp. 1576–1588.

[27]   Sang Woo Kim et al. "A sensitive non-radioactive northern blot method to detect small RNAs". In: *Nucleic acids research* 38 (2010), e98.

[28]   Gurman Singh Pall et al. "Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot." In: *Nucleic acids research* 35 (2007), e60–e60.

[29] Shakti H Ramkissoon et al. "Nonisotopic detection of microRNA using digoxigenin labeled RNA probes." In: *Molecular and Cellular Probes* 20 (2006), pp. 1–4.

[30] Anna Válóczi et al. "Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes." In: *Nucleic acids research* 32 (2004), e175.

[31] Dr Zhiguo Wang and Dr Baofeng Yang. *Northern Blotting and Its Variants for Detecting Expression and Analyzing Tissue Distribution of miRNAs BT - MicroRNA Expression Detection Methods.* Case. 2010.

[32] Wei Wu et al. "Simple and nonradioactive detection of microRNAs using digoxigenin (DIG)-labeled probes with high sensitivity." In: *RNA (New York, N.Y.)* 20 (2014), pp. 580–584.

[33] T. W. Nilsen. "Splinted ligation method to detect small RNAs". In: *Cold Spring Harb Protoc* 1 (2013).

[34] Patricia A Maroney et al. "A rapid, quantitative assay for direct detection of microRNAs and other small RNAs using splinted ligation." In: *RNA (New York, N.Y.)* 13 (2007), pp. 930–6.

[35] Shirley Tam et al. "Robust global microRNA expression profiling using next-generation sequencing technologies." In: *Laboratory investigation; a journal of technical methods and pathology* 94 (2014), pp. 350–358.

[36] James Q Yin, Robert C Zhao, and Kevin V Morris. "Profiling microRNA expression with microarrays." In: *Trends in biotechnology* 26 (2008), pp. 70–76.

[37] M. Castoldi et al. "A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA)". In: *RNA* 12.5 (2006), pp. 913–20.

[38] I. Beuvink et al. "A novel microarray approach reveals new tissue-specific signatures of known and predicted mammalian microRNAs". In: *Nucleic Acids Res* 35.7 (2007), e52.

[39] Wei Li and Kangcheng Ruan. "MicroRNA detection by microarray." In: *Analytical and bioanalytical chemistry* 394 (2009), pp. 1117–1124.

[40] Hui Wang, Robert A Ach, and Bo Curry. "Direct and sensitive miRNA profiling from low-input total RNA." In: *RNA (New York, N.Y.)* 13 (2007), pp. 151–159.

[41] R. A. Ach, H. Wang, and B. Curry. "Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods". In: *BMC Biotechnol* 8 (2008), p. 69.

[42] A. Git et al. "Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression". In: *RNA* 16.5 (2010), pp. 991–1006.

[43] S. Pradervand et al. "Concordance among digital gene expression, microarrays, and qPCR when measuring differential expression of microRNAs". In: *Biotechniques* 48.3 (2010), pp. 219–22.

[44] Fumiaki Sato et al. "Intra-platform repeatability and inter-platform comparability of microRNA microarray technology." In: *PLoS ONE* 4 (2009), e5540.

[45] Monya Baker. "MicroRNA profiling: separating signal from noise." In: *Nature methods* 7 (2010), pp. 687–692.

[46] J. M. Lee, H. Cho, and Y. Jung. "Fabrication of a structure-specific RNA binder for array detection of label-free microRNA". In: *Angew Chem Int Ed Engl* 49.46 (2010), pp. 8662–5.

[47] D. Duan et al. "Label-free high-throughput microRNA expression profiling from total RNA". In: *Nucleic Acids Res* 39.22 (2011), e154.

[48] Y. Shen et al. "Label-free microRNA profiling not biased by 3' end 2'-O-methylation". In: *Anal Chem* 84.15 (2012), pp. 6361–5.

[49] Peter T Nelson et al. "Microarray-based, high-throughput gene expression profiling of microRNAs." In: *Nature methods* 1 (2004), pp. 155–161.

[50] T. Ueno and T. Funatsu. "Label-free quantification of microRNAs using ligase-assisted sandwich hybridization on a DNA microarray". In: *PLoS ONE* 9.3 (2014), e90920.

[51] Mirco Castoldi et al. "Expression Profiling of MicroRNAs by Quantitative Real-Time PCR: The Good, the Bad, and the Ugly". In: *PCR Technology*. CRC Press, 2013, pp. 307–322.

[52] Rashmi S Goswami et al. "Optimization and analysis of a quantitative real-time PCR-based technique to determine microRNA expression in formalin-fixed paraffin-embedded samples." In: *BMC biotechnology* 10 (2010), p. 47.

[53] Wei Meng et al. "Comparison of microRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues." In: *PLoS ONE* 8 (2013), e64393.

[54] C Becker et al. "mRNA and microRNA quality control for RT-qPCR analysis." In: *Methods (San Diego, Calif.)* 50 (2010), pp. 237–243.

[55] C. Chen et al. "Real-time quantification of microRNAs by stem-loop RT-PCR". In: *Nucleic Acids Res* 33.20 (2005), e179.

[56] Pavan Kumar, Brian H Johnston, and Sergei A Kazakov. "miR-ID: a novel, circularization-based platform for detection of microRNAs." In: *RNA (New York, N.Y.)* 17 (2011), pp. 365–380.

[57] Pauline Chugh and Dirk P Dittmer. "Potential pitfalls in microRNA profiling." In: *Wiley interdisciplinary reviews. RNA* 3 (2012), pp. 601–616.

[58] Vladimir Benes and Mirco Castoldi. "Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available." In: *Methods (San Diego, Calif.)* 50 (2010), pp. 244–249.

[59] Swanhild U Meyer et al. "Profound effect of profiling platform and normalization strategy on detection of differentially expressed microRNAs–a comparative study." In: *PLoS ONE* 7 (2012), e38946.

[60] T. Kang et al. "Ultra-Specific Zeptomole MicroRNA Detection by Plasmonic Nanowire Interstice Sensor with Bi-Temperature Hybridization". In: *Small* (2014).

[61] J. M. Lee and Y. Jung. "Two-temperature hybridization for microarray detection of label-free microRNAs with attomole detection and superior specificity". In: *Angew Chem Int Ed Engl* 50.52 (2011), pp. 12487–90.

[62] Federica Degliangeli, Pier Paolo Pompa, and Roberto Fiammengo. "Nanotechnology-Based Strategies for the Detection and Quantification of MicroRNA." In: *Chemistry (Weinheim an der Bergstrasse, Germany)* 20.31 (2014), pp. 9476–9492.

[63] Ezat Hamidi-Asl et al. "A review on the electrochemical biosensors for determination of microRNAs." In: *Talanta* 115 (2013), pp. 74–83.

[64] Ali Akbar Jamali et al. "Nanomaterials on the road to microRNA detection with optical and electrochemical nanobiosensors". In: *TrAC-Trends in Analytical Chemistry* 55 (2014), pp. 24–42.

[65] L Zhang et al. "Detection of cancer biomarkers with nanotechnology". In: *American Journal of Biochemistry and Biotechnology* 9.1 (2013), pp. 71–89.

[66] Gergely Lautner and Róbert E Gyurcsányi. "Electrochemical Detection of miRNAs". In: *Electroanalysis* 26.6 (2014), pp. 1224–1235.

[67] J. Y. Park and S. M. Park. "DNA hybridization sensors based on electrochemical impedance spectroscopy as a detection tool". In: *Sensors (Basel)* 9.12 (2009), pp. 9513–32.

[68] Huimin Deng et al. "A highly sensitive microRNA biosensor based on hybridized microRNA-guided deposition of polyaniline". In: *Biosensors and Bioelectronic* 60.c (2014), pp. 195–200.

[69] Z. Gao et al. "A label-free biosensor for electrochemical detection of femtomolar microRNAs". In: *Anal Chem* 85.3 (2013), pp. 1624–30.

[70] S. Karkare and D. Bhatnagar. "Promising nucleic acid analogs and mimics: characteristic features and applications of PNA, LNA, and morpholino". In: *Appl Microbiol Biotechnol* 71.5 (2006). Karkare, Shantanu Bhatnagar, Deepak eng Review Germany 2006/05/10 09:00 Appl Microbiol Biotechnol. 2006 Aug;71(5):575-86. Epub 2006 May 9., pp. 575–86.

[71] Y. Ren et al. "A highly sensitive and selective electrochemical biosensor for direct detection of microRNAs in serum". In: *Anal Chem* 85.9 (2013), pp. 4784–9.

[72] Z. Gao and Y. Peng. "A highly sensitive and specific biosensor for ligation- and PCR-free detection of MicroRNAs". In: *Biosens Bioelectron* 26.9 (2011), pp. 3768–3773.

[73] Z. Ge et al. "Hybridization chain reaction amplification of microRNA detection with a tetrahedral DNA nanostructure-based electrochemical biosensor". In: *Anal Chem* 86.4 (2014), pp. 2124–30.

[74] M. Lin et al. "Target-responsive, DNA nanostructure-based E-DNA sensor for microRNA analysis". In: *Anal Chem* 86.5 (2014), pp. 2285–8.

[75] Y. Wen et al. "DNA Nanostructure-based Interfacial engineering for PCR-free ultrasensitive electrochemical analysis of microRNA". In: *Sci Rep* 2 (2012), p. 867.

[76] E. J. Chapman et al. "Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step". In: *Genes Dev* 18.10 (2004), pp. 1179–86.

[77] P. Ramnani et al. "Electronic detection of microRNA at attomolar level with high specificity". In: *Anal Chem* 85.17 (2013), pp. 8061–4.

[78] T. Kilic, S. Nur Topkaya, and M. Ozsoz. "A new insight into electrochemical microRNA detection: a molecular caliper, p19 protein". In: *Biosens Bioelectron* 48 (2013), pp. 165–71.

[79] Mahmoud Labib et al. "Three-mode electrochemical sensing of ultralow microRNA levels." In: *Journal of the American Chemical Society* 135.8 (2013), pp. 3027–3038.

[80] M. Wanunu et al. "Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors". In: *Nat Nanotechnol* 5.11 (2010), pp. 807–14.

[81] N. Khan et al. "Quantitative analysis of microRNA in blood serum with protein-facilitated affinity capillary electrophoresis". In: *Anal Chem* 83.16 (2011), pp. 6196–201.

[82] Yurong Yan et al. "A Simple and Highly Sensitive Electrochemical Biosensor for microRNA Detection Using Target-Assisted Isothermal Exponential Amplification Reaction". In: *Electroanalysis* 25 (2013), pp. 2354–2359.

[83] Y. Yu et al. "Ultrasensitive electrochemical detection of microRNA based on an arched probe mediated isothermal exponential amplification". In: *Anal Chem* 86.16 (2014), pp. 8200–5.

[84] X Wu et al. "Dual signal amplification strategy for enzyme-free electrochemical detection of microRNAs". In: *Sensors and Actuators B* 203 (2014), pp. 296–302.

[85] H. Yin et al. "An electrochemical signal 'off-on' sensing platform for microRNA detection". In: *Analyst* 137.6 (2012), pp. 1389–95.

[86] G. J. Zhang et al. "Label-free direct detection of MiRNAs with silicon nanowire biosensors". In: *Biosens Bioelectron* 24.8 (2009), pp. 2504–8.

[87] X. Zhang et al. "Programming nanopore ion flow for encoded multiplex microRNA detection". In: *ACS Nano* 8.4 (2014), pp. 3444–50.

[88] J Zhang and D Cui. "Nanoparticle-based Optical Detection of MicroRNA". In: *Nano Biomedicine and Engineering* 5.1 (2013), pp. 1–10.

[89] L. Jiang et al. "Rapid and multiplex microRNA detection on graphically encoded silica suspension array". In: *Biosens Bioelectron* 61 (2014), pp. 222–6.

[90] W. Liu, X. Zhou, and D. Xing. "Rapid and reliable microRNA detection by stacking hybridization on electrochemiluminescent chip system". In: *Biosens Bioelectron* 58 (2014), pp. 388–94.

[91] T. Liu et al. "Label-free and ultrasensitive electrochemiluminescence detection of microRNA based on long-range self-assembled DNA nanostructures". In: *Microchimica Acta* 181.7-8 (2014), pp. 731–736.

[92] Pu Zhang et al. "An electrochemiluminescent microRNA biosensor based on hybridization chain reaction coupled with hemin as the signal enhancer." In: *The Analyst* 139.11 (2014), pp. 2748–2753.

[93] H. Dong et al. "Label-free and ultrasensitive microRNA detection based on novel molecular beacon binding readout and target recycling amplification". In: *Biosens Bioelectron* 53 (2014), pp. 377–83.

[94] Federica Degliangeli et al. "Absolute and direct microRNA quantification using DNA-gold nanoparticle probes." In: *Journal of the American Chemical Society* 136.6 (2014), pp. 2264–2267.

[95] A. J. Qavi et al. "Anti-DNA:RNA antibodies and silicon photonic microring resonators: increased sensitivity for multiplexed microRNA detection". In: *Anal Chem* 83.15 (2011), pp. 5949–56.

[96] D Zhang et al. "Streptavidin-enhanced surface plasmon resonance biosensor for highly sensitive and specific detection of microRNA". In: *Microchimica Acta* (2013).

[97] Y. C. He et al. "The rapid detection of microRNA based on p19-enhanced fluorescence polarization". In: *Chem Commun (Camb)* 50.47 (2014), pp. 6236–9.

[98] H. Arata et al. "Rapid and sensitive microRNA detection with laminar flow-assisted dendritic amplification on power-free microfluidic chip". In: *PLoS ONE* 7.11 (2012), e48329.

[99] X. Gao et al. "Visual detection of microRNA with lateral flow nucleic acid biosensor". In: *Biosens Bioelectron* 54 (2014), pp. 578–84.

[100] H. V. Tran et al. "An electrochemical ELISA-like immunosensor for miRNAs detection based on screen-printed gold electrodes modified with reduced graphene oxide and carbon nanotubes". In: *Biosens Bioelectron* 62 (2014), pp. 25–30.

[101] H. Cao et al. "A carbon nanotube/quantum dot based photoelectrochemical biosensing platform for the direct detection of microRNAs". In: *Chem Commun (Camb)* (2014).

[102] F. Haque et al. "Solid-State and Biological Nanopore for Real-Time Sensing of Single Chemical and Sequencing of DNA". In: *Nano Today* 8.1 (2013), pp. 56–74.

[103] E. Arefian et al. "Analysis of microRNA signatures using size-coded ligation-mediated PCR". In: *Nucleic Acids Res* 39.12 (2011), e80.

[104] G. K. Geiss et al. "Direct multiplexed measurement of gene expression with color-coded probe pairs". In: *Nat Biotechnol* 26.3 (2008), pp. 317–25.

[105] Eva van Rooij. "The art of microRNA research." In: *Circulation Research* 108 (2011), pp. 219–234.

[106] E. L. van Dijk et al. "Ten years of next-generation sequencing technology". In: *Trends Genet* 30.9 (2014), pp. 418–426.

[107] M. L. Metzker. "Sequencing technologies - the next generation". In: *Nat Rev Genet* 11.1 (2010), pp. 31–46.

[108] R. M. Marin, M. Sulc, and J. Vanicek. "Searching the coding region for microRNA targets". In: *RNA* 19.4 (2013), pp. 467–74.

[109] M. Reczko et al. "Functional microRNA targets in protein coding sequences". In: *Bioinformatics* 28.6 (2012), pp. 771–6.

[110] S. M. Peterson et al. "Common features of microRNA target prediction tools". In: *Front Genet* 5 (2014), p. 23.

[111] W. Ritchie, J. E. Rasko, and S. Flamant. "MicroRNA target prediction and validation". In: *Adv Exp Med Biol* 774 (2013), pp. 39–53.

[112] S. Tarang and M. D. Weston. "Macros in microRNA target identification: a comparative analysis of in silico, in vitro, and in vivo approaches to microRNA target identification". In: *RNA Biol* 11.4 (2014), pp. 324–33.

[113] E. A. Hunt and S. K. Deo. "Bioluminescent stem-loop probes for highly sensitive nucleic acid detection". In: *Chemical communications* 47.33 (2011), pp. 9393–5.

[114] S. Tyagi and F. R. Kramer. "Molecular beacons: Probes that fluoresce upon hybridization". In: *Nature biotechnology* 14.3 (1996), pp. 303–308.

[115] S. A. Marras, F. R. Kramer, and S. Tyagi. "Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes". In: *Nucleic acids research* 30.21 (2002), e122.

[116] K. Wang et al. "Molecular engineering of DNA: molecular beacons". In: *Angewandte Chemie* 48.5 (2009), pp. 856–70.

[117] L. Tan et al. "Molecular beacons for bioanalytical applications". In: *The Analyst* 130.7 (2005), pp. 1002–5.

[118] G. Goel et al. "Molecular beacon: a multitask probe". In: *J Appl Microbiol* 99.3 (2005), pp. 435–42.

[119] X. Liu and W. Tan. "A fiber-optic evanescent wave DNA biosensor based on novel molecular beacons". In: *Anal Chem* 71.22 (1999), pp. 5054–9.

[120] L. Wang et al. "Locked nucleic acid molecular beacons". In: *J Am Chem Soc* 127.45 (2005), pp. 15664–5.

[121] K. Martinez et al. "Locked nucleic acid based beacons for surface interaction studies and biosensor development". In: *Anal Chem* 81.9 (2009), pp. 3448–54.

[122] X. Li, E. Morgenroth, and L. Raskin. "Quantitative rRNA-targeted solution-based hybridization assay using peptide nucleic acid molecular beacons". In: *Appl Environ Microbiol* 74.23 (2008), pp. 7297–305.

[123] Y. W. Lin et al. "Fluorescence detection of single nucleotide polymorphisms using a universal molecular beacon". In: *Nucleic acids research* 36.19 (2008), e123.

[124] T. N. Grossmann, L. Roglin, and O. Seitz. "Triplex molecular beacons as modular probes for DNA detection". In: *Angewandte Chemie* 46.27 (2007), pp. 5223–5.

[125] E. Socher et al. "Low-noise stemless PNA beacons for sensitive DNA and RNA detection". In: *Angewandte Chemie* 47.49 (2008), pp. 9555–9.

[126] C. H. Lu et al. "Increasing the sensitivity and single-base mismatch selectivity of the molecular beacon using graphene oxide as the "nanoquencher"". In: *Chemistry* 16.16 (2010), pp. 4889–94.

[127] B. Dubertret, M. Calame, and A. J. Libchaber. "Single-mismatch detection using gold-quenched fluorescent oligonucleotides". In: *Nature biotechnology* 19.4 (2001), pp. 365–70.

[128] C. J. Yang, H. Lin, and W. Tan. "Molecular assembly of superquenchers in signaling molecular interactions". In: *J Am Chem Soc* 127.37 (2005), pp. 12772–3.

[129] Joong Hyun Kim, Dimitrios Morikis, and Mihrimah Ozkan. "Adaptation of inorganic quantum dots for stable molecular beacons". In: *Sensors and Actuators B: Chemical* 102.2 (2004), pp. 315–319.

[130] J. Li et al. "Design of a room-temperature phosphorescence-based molecular beacon for highly sensitive detection of nucleic acids in biological fluids". In: *Anal Chem* 83.4 (2011), pp. 1356–62.

[131] A. M. Loening et al. "Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output". In: *Protein Eng Des Sel* 19.9 (2006), pp. 391–400.

[132] A. Witkowski, S. Ramanathan, and S. Daunert. "Bioluminescence binding assay for biotin with attomole detection based on recombinant aequorin". In: *Analytical chemistry* 66.11 (1994), pp. 1837–1840.

[133] K. A. Cissell et al. "Bioluminescence-based detection of microRNA, miR21 in breast cancer cells". In: *Anal Chem* 80.7 (2008), pp. 2319–25.

[134] V. N. Kim and J. W. Nam. "Genomics of microRNA". In: *Trends in genetics : TIG* 22.3 (2006), pp. 165–73.

[135] C. E. Stahlhut Espinosa and F. J. Slack. "The role of microRNAs in cancer". In: *Yale J Biol Med* 79.3-4 (2006), pp. 131–40.

[136] H. Grosshans and W. Filipowicz. "Molecular biology: the expanding world of small RNAs". In: *Nature* 451.7177 (2008), pp. 414–6.

[137] S. M. Johnson et al. "RAS is regulated by the let-7 microRNA family". In: *Cell* 120.5 (2005), pp. 635–47.

[138] K. A. O'Donnell et al. "c-Myc-regulated microRNAs modulate E2F1 expression". In: *Nature* 435.7043 (2005), pp. 839–43.

[139] M. V. Iorio et al. "MicroRNA gene expression deregulation in human breast cancer". In: *Cancer Res* 65.16 (2005), pp. 7065–70.

[140] M. S. Kumar et al. "Impaired microRNA processing enhances cellular transformation and tumorigenesis". In: *Nat Genet* 39.5 (2007), pp. 673–7.

[141] P. P. Medina, M. Nolde, and F. J. Slack. "OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma". In: *Nature* 467.7311 (2010), pp. 86–90.

[142] A. Silahtaroglu and J. Stenvang. "MicroRNAs, epigenetics and disease". In: *Essays Biochem* 48.1 (2010), pp. 165–85.

[143] B. A. Tannous et al. "Codon-optimized Gaussia luciferase cDNA for mammalian gene expression in culture and in vivo". In: *Mol Ther* 11.3 (2005), pp. 435–43.

[144] T. Suzuki et al. "Real-time bioluminescence imaging of a protein secretory pathway in living mammalian cells using Gaussia luciferase". In: *FEBS Lett* 581.24 (2007), pp. 4551–6.

[145] S. Inouye and Y. Sahara. "Identification of two catalytic domains in a luciferase secreted by the copepod Gaussia princeps". In: *Biochem Biophys Res Commun* 365.1 (2008), pp. 96–101.

[146] C. A. Maguire et al. "Gaussia luciferase variant for high-throughput functional screening applications". In: *Anal Chem* 81.16 (2009), pp. 7102–6.

[147] S. B. Kim et al. "Superluminescent variants of marine luciferases for bioassays". In: *Anal Chem* 83.22 (2011), pp. 8732–40.

[148] M. H. Degeling et al. "Directed molecular evolution reveals Gaussia luciferase variants with enhanced light output stability". In: *Anal Chem* 85.5 (2013), pp. 3006–12.

[149] S. Inouye and Y. Sahara. "Soluble protein expression in E. coli cells using IgG-binding domain of protein A as a solubilizing partner in the cold induced system". In: *Biochem Biophys Res Commun* 376.3 (2008), pp. 448–53.

[150] A. R. Goerke et al. "Cell-free metabolic engineering promotes high-level production of bioactive Gaussia princeps luciferase". In: *Metab Eng* 10.3-4 (2008), pp. 187–200.

[151] T. Rathnayaka et al. "Biophysical characterization of highly active recombinant Gaussia luciferase expressed in Escherichia coli". In: *Biochim Biophys Acta* 1804.9 (2010), pp. 1902–7.

[152] S. Inouye, J. Sato, and Y. Sahara-Miura. "Recombinant Gaussia luciferase with a reactive cysteine residue for chemical conjugation: expression, purification and its application for bioluminescent immunoassays". In: *Biochem Biophys Res Commun* 410.4 (2011), pp. 792–7.

[153] Athel Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. 4th Edition. John Wiley & Sons, 2012.

[154] S. V. Markova et al. "The smallest natural high-active luciferase: cloning and characterization of novel 16.5-kDa luciferase from copepod Metridia longa". In: *Biochem Biophys Res Commun* 457.1 (2015), pp. 77–82.

[155] A. Heger and L. Holm. "Rapid automatic detection and alignment of repeats in protein sequences". In: *Proteins* 41.2 (2000), pp. 224–37.

[156] M. A. Andrade, C. Perez-Iratxeta, and C. P. Ponting. "Protein repeats: structures, functions, and evolution". In: *J Struct Biol* 134.2-3 (2001), pp. 117–31.

[157] G. L. Rosano and E. A. Ceccarelli. "Recombinant protein expression in Escherichia coli: advances and challenges". In: *Front Microbiol* 5 (2014), p. 172.

[158] A. Ceroni et al. "DISULFIND: a disulfide bonding state and cysteine connectivity prediction server". In: *Nucleic acids research* 34.Web Server issue (2006), W177–81.

[159] F. Ferre and P. Clote. "DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification". In: *Nucleic acids research* 34.Web Server issue (2006), W182–5.

[160] J. Gavrilyuk et al. "Formylbenzene diazonium hexafluorophosphate reagent for tyrosine-selective modification of proteins and the introduction of a bioorthogonal aldehyde". In: *Bioconjug Chem* 23.12 (2012), pp. 2321–8.

[161] O. Boutureira and G. J. Bernardes. "Advances in chemical protein modification". In: *Chem Rev* 115.5 (2015), pp. 2174–95.

[162] O. Emanuelsson et al. "Locating proteins in the cell using TargetP, SignalP and related tools". In: *Nat Protoc* 2.4 (2007), pp. 953–71.

[163] J. X. Feliu, R. Cubarsi, and A. Villaverde. "Optimized release of recombinant proteins by ultrasonication of E. coli cells". In: *Biotechnol Bioeng* 58.5 (1998), pp. 536–40.

[164] G. Tzertzinis, E. Schildkraut, and I. Schildkraut. "Substrate cooperativity in marine luciferases". In: *PloS one* 7.6 (2012), e40099.

[165] Robert A. Copeland. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*. United States: Wiley, John & Sons, 2000.

[166] P. Wang et al. "Structural insight into mechanisms for dynamic regulation of PKM2". In: *Protein Cell* 6.4 (2015), pp. 275–87.

[167] P. Artimo et al. "ExPASy: SIB bioinformatics resource portal". In: *Nucleic acids research* 40.Web Server issue (2012), W597–603.

[168]   Elisabeth Gasteiger et al. "Protein identification and analysis tools in the ExPASy server". In: *The Proteomics Protocols Handbook*. Ed. by John M. Walker. Springer, 2005, pp. 571–607.

[169]   J. Kyte and R. F. Doolittle. "A simple method for displaying the hydropathic character of a protein". In: *J Mol Biol* 157.1 (1982), pp. 105–32.

[170]   L. Whitmore and B. A. Wallace. "Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases". In: *Biopolymers* 89.5 (2008), pp. 392–400.

[171]   N. Sreerama and R. W. Woody. "Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set". In: *Anal Biochem* 287.2 (2000), pp. 252–60.

[172]   E. M. Thompson, S. Nagata, and F. I. Tsuji. "Cloning and expression of cDNA for the luciferase from the marine ostracod Vargula hilgendorfii". In: *Proc Natl Acad Sci U S A* 86.17 (1989), pp. 6567–71.

[173]   Y. Maeda et al. "Expression of a bifunctional chimeric protein A-Vargula hilgendorfii luciferase in mammalian cells". In: *Biotechniques* 20.1 (1996), pp. 116–21.

[174]   Megan Dobbs et al. *Simultaneous dual-emission detection of luciferase reporter assays*. Application Note 233. `http://www.bmglabtech.com/en/applications/`: BMG LABTECH, Mar. 2013.

[175]   Y. Maeda et al. "Truncation of Vargula luciferase still results in retention of luminescence". In: *J Biochem* 119.4 (1996), pp. 601–3.

[176]   C. Wu, C. Suzuki-Ogoh, and Y. Ohmiya. "Dual-reporter assay using two secreted luciferase genes". In: *Biotechniques* 42.3 (2007), pp. 290, 292.

[177]   C. A. Maguire et al. "Triple bioluminescence imaging for in vivo monitoring of cellular processes". In: *Mol Ther Nucleic Acids* 2 (2013), e99.

[178]   P. W. Haebel et al. "The disulfide bond isomerase DsbC is activated by an immunoglobulin-fold thiol oxidoreductase: crystal structure of the DsbC-DsbDalpha complex". In: *EMBO J* 21.18 (2002), pp. 4774–84.

[179]   L. A. Kelley et al. "The Phyre2 web portal for protein modeling, prediction and analysis". In: *Nat Protoc* 10.6 (2015), pp. 845–58.

[180]   J. C. Matthews, K. Hori, and M. J. Cormier. "Substrate and substrate analogue binding properties of Renilla luciferase". In: *Biochemistry* 16.24 (1977), pp. 5217–20.

[181]   H. Nakamura et al. "Design, synthesis, and evaluation of the transition- state inhibitors of coelenterazine bioluminescence: probing the chiral environment of active site". In: *J Am Chem Soc* 123.7 (2001), pp. 1523–4.

[182] H. Nakamura et al. "Syntheses of the mechanism-based inhibitors of coelenterazine bioluminescence". In: *Bioluminescence and Chemiluminescence*. Ed. by JF Case et al. World Scientific Publishing Company Incorporated, 2001, pp. 127–130.

[183] D. S. Auld et al. "Characterization of chemical libraries for luciferase inhibitory activity". In: *J Med Chem* 51.8 (2008), pp. 2372–86.

[184] K. J. Herbst, M. D. Allen, and J. Zhang. "The cAMP-dependent protein kinase inhibitor H-89 attenuates the bioluminescence signal produced by Renilla Luciferase". In: *PloS one* 4.5 (2009), e5642.

[185] L. P. da Silva and J. C. da Silva. "Kinetics of inhibition of firefly luciferase by dehydroluciferyl-coenzyme A, dehydroluciferin and L-luciferin". In: *Photochem Photobiol Sci* 10.6 (2011), pp. 1039–45.

[186] P. I. Ho et al. "Reporter enzyme inhibitor study to aid assembly of orthogonal reporter gene assays". In: *ACS Chem Biol* 8.5 (2013), pp. 1009–17.

[187] T. J. Morin and W. R. Kobertz. "Tethering chemistry and K+ channels". In: *J Biol Chem* 283.37 (2008), pp. 25105–9.

[188] T. Miron and M. Wilchek. "A simplified method for the preparation of succinimidyl carbonate polyethylene glycol for coupling to proteins". In: *Bioconjug Chem* 4.6 (1993), pp. 568–9.

[189] Yunxia Hu et al. "Ferritin–polymer conjugates: grafting chemistry and integration into nanoscale assemblies". In: *Advanced Functional Materials* 20.20 (2010), pp. 3603–3612.

[190] M. Pignatto, N. Realdon, and M. Morpurgo. "Optimized avidin nucleic acid nanoassemblies by a tailored PEGylation strategy and their application as molecular amplifiers in detection". In: *Bioconjug Chem* 21.7 (2010), pp. 1254–63.

[191] Constantin I Chiriac, Fulga Tanasa, and Marioara Onciu. "A new direct synthesis of cinnamic acids from aromatic aldehydes and aliphatic carboxylic acids in the presence of sodium borohydride". In: *Tetrahedron letters* 44.17 (2003), pp. 3579–3580.

[192] Michael E Kopach et al. "Practical synthesis of chiral 2-morpholine:(4-benzyl-morpholin-2-(S)-yl)-(tetrahydropyran-4-yl) methanone mesylate, a useful pharmaceutical intermediate". In: *Organic Process Research & Development* 13.2 (2009), pp. 209–224.

[193] Federico Scaravelli et al. "Efficient method to prepare diethylphosphonacetamides". In: *Tetrahedron Letters* 51.39 (2010), pp. 5154–5156.

[194] J. R. Dunetz et al. "General and scalable amide bond formation with epimerization-prone substrates using T3P and pyridine". In: *Org Lett* 13.19 (2011), pp. 5048–51.

[195] Joshua R Dunetz et al. "Multikilogram synthesis of a hepatoselective glucokinase activator". In: *Organic Process Research & Development* 16.10 (2012), pp. 1635–1645.

[196] Donna S MacMillan et al. "Evaluation of alternative solvents in common amide coupling reactions: replacement of dichloromethane and N, N-dimethylformamide". In: *Green Chemistry* 15.3 (2013), pp. 596–600.

[197] A. El-Faham et al. "COMU: a safer and more effective replacement for benzotriazole-based uronium coupling reagents". In: *Chemistry* 15.37 (2009), pp. 9404–16.

[198] A. El-Faham and F. Albericio. "COMU: a third generation of uronium-type coupling reagents". In: *J Pept Sci* 16.1 (2010), pp. 6–9.

[199] C. A. Chantell, M. A. Onaiyekan, and M. Menakuru. "Fast conventional Fmoc solid-phase peptide synthesis: a comparative study of different activators". In: *J Pept Sci* 18.2 (2012), pp. 88–91.

[200] R. Subiros-Funosas et al. "COMU: scope and limitations of the latest innovation in peptide acyl transfer reagents". In: *J Pept Sci* 19.7 (2013), pp. 408–14.

[201] Giampaolo Giacomelli, Andrea Porcheddu, and Lidia D Luca. "[1, 3, 5]-Triazine: a versatile heterocycle in current applications of organic chemistry". In: *Current Organic Chemistry* 8.15 (2004), pp. 1497–1519.

[202] BP Bandgar and SS Sawant. "Novel and Gram-Scale Green Synthesis of Flutamide". In: *Synthetic communications* 36.7 (2006), pp. 859–864.

[203] Grzegorz Blotny. "Recent applications of 2, 4, 6-trichloro-1, 3, 5-triazine and its derivatives in organic synthesis". In: *Tetrahedron* 62.41 (2006), pp. 9507–9522.

[204] D. Finzi et al. "Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy". In: *Nat Med* 5.5 (1999), pp. 512–7.

[205] G. M. Laird et al. "Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay". In: *PLoS Pathog* 9.5 (2013), e1003398.

[206] N. M. Archin et al. "Eradicating HIV-1 infection: seeking to clear a persistent pathogen". In: *Nat Rev Microbiol* 12.11 (2014), pp. 750–64.

[207] Geertrui Tavernier et al. "mRNA as gene therapeutic: how to control protein expression". In: *Journal of Controlled Release* 150.3 (2011), pp. 238–247.

[208] A. J. Geall et al. "Nonviral delivery of self-amplifying RNA vaccines". In: *Proc Natl Acad Sci U S A* 109.36 (2012), pp. 14604–9.

VITA

Eric was born in 1986 to his parents Bryan and Susan Hunt, and raised in rural central Indiana with his younger sister, Lori. In classic "Hoosier" fashion, Eric spent his summers largely outdoors, playing soccer, going camping with his family, working as a farmhand stocking barns, and participating in the elusive detassling of corn. Growing up, Eric was a curious kid – fascinated by the smattering of electronic components commonplace to his father's career as an electrical engineer – and early on developed an interest for the underlying mechanisms of how things in life worked, which often got him into trouble with his parents when it involved disassembling something expensive or mixing potentially dangerous chemicals.

In 2004, he graduated from Westfield High School before attending Indiana University Purdue University Indianapolis on a full academic scholarship from the Bepko Scholars & Fellows Program to study Chemistry. It was here, in his freshman year, he met his future wife Angela. Throughout his undergraduate career Eric was involved in biochemical research and spent the summer of 2007 at the University of Kentucky in the labs of Dr. Sylvia Daunert as part of an NSF REU program. 2008 was a big year, Eric earned his B.S. in chemistry, decided to start graduate school to obtain a Ph.D. in biochemistry with Dr. Sapna Deo, and finally got up the courage to ask Angela to marry him. In 2009, Eric received the NSF Graduate Research Fellowship which supported his predoctoral studies for three years. In 2010, he and Angela were wed and shortly thereafter moved across the country to Miami, where Dr. Deo had taken a new position in the Department of Biochemistry & Molecular Biology. In Miami, Eric received an NSF G–K12 fellowship as part of the University of Miami Science Made Sensible Program, and taught middle school science for one year in Brownsville. Eric has mentored several undergraduate and high school students throughout his graduate career and is passionate about education. In his free time, Eric likes to read, travel, cook and sing with his wife (sometimes at the same time), and play guitar in a shoegaze fashion involving far too many pedals.