DEVELOPMENT OF E-SRS (ENVIRONMENT-SENSING RESPONSE SYSTEM) AS A NOVEL METHOD TO DISTINGUISH GENETIC ENVIRONMENTS AND RESOLVE CLOSELY RELATED NUCLEIC ACID SEQUENCES

LEONG SHIANG RONG

(B.Sc.(Hons.), NUS)

A THESIS SUBMITTED FOR

THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF PHYSIOLOGY

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

This thesis and the work in it had benefited greatly from the kind support, advice and assistance of many individuals. I would first like to express my sincere gratitude and profound respect for my two Supervisors, Associate Professor Hooi Shing Chuan (main) and Associate Professor Soong Tuck Wah (co), without whose kind and generous support this bold and innovative (and therefore risky) project would never even have taken off.

Prof Hooi provided much strategic guidance in charting the progress of my project, as well as technical advice on the numerous such challenges that this project encountered. Every meeting with him enables me to reestablish a confidence that the research will eventually be workable. His willingness to look out for and help students despite his hectic schedule, has left in me a deep and lasting impression on what a good Supervisor should be like.

Prof Soong, or "Tuck" as he likes to be called, had been the kindest and most supportive Supervisor I had ever known of. Tuck supported me with generous laboratory space, equipments, supplies, contacts, and an income. His wealth of knowledge in molecular and cell biology helped me along on numerous occasions. He gave me a chance to write my first grant, which taught me what managing expenses and grant reports were like, something that most graduate students I know do not get to experience. His consultative leadership and empowering support gave me the courage to voice my opinions and take charge of my research, something for which I am deeply grateful.

I am grateful to many individuals who provided technical advice and assistance, particularly

Table of Contents

Summary

Existing limitations of conventional Nucleic Acid (NA) detection prompted us to conduct a Proof-of-Concept of a novel NA sensing platform called environment-Sensing Response System (e-SRS), which could deliver a physical response upon sensing its NA Target Sequence (NTS). e-SRS is a NA sensing and response system with two components: 1) An RNA based sensor that changes conformation and activates upon binding specific NTS; 2) A Response System that is triggered by the activated sensor to initiate some physical response, such as emitting a fluorescent signal to indicate presence of the NTS, or other biomolecular actions like induction of gene expression. Its modular nature, whereby the sensor is separate from the Response System, allows e-SRS flexibility in adapting to different formats and applications.

The ability to activate Response System after sensing enables the e-SRS sensor to serve as a signal transducer, which passes a signal of one form from the environment (e.g. presence of specific NA), to that of another form as produced by the Response System (e.g. activation of inducible expression system). A biomolecular signal transducer could function in more diverse ways than a biomolecular probe, and could be a powerful tool in research, diagnostics and therapy.

After initial tests, an early design known as e-SIGE was unworkable, likely because the sensor's RNA folding was designed without computational secondary structure prediction. The RNA folding likely did not occur as intended.

We redesigned the physical implementation to create the current e-SRS, adapting an existing allosteric ribozyme, the Maxizyme, as e-SRS sensor, employing computational secondary structure prediction. We were able to successfully test e-SRS in the test tube environment via Ribozyme Trans-cleavage Assays (RTA). Unsatisfied with RNA cleavage assays via Denaturing Polyacrylamide Gel Electrophoresis, we developed the Reporter Substrate (RS), which provided real time fluorescence reporting of e-SRS sensor activity, and served as a gene detection Response System. Our attempts to activate an inducible gene system as the Response System within cell lines were unsuccessful, likely due to interfering RNA secondary structure in the cellular environment.

e-SRS sensor with RS for fluorescence based real-time test tube detection and resolution of closely related RNA sequences was tested on 7 NTS from 2 categories: 1) 3 strains of Malaria parasites (Plasmodium falciparum), denoted as Mfs, Mfr1, and Mfr2; 2) 4 common serotypes of Dengue viruses, denoted as D1, D2, D3, and D4. We developed a computational algorithm in Perl that greatly automated the design and assessment of e-SRS Mz-based sensors.

Our seven e-SRS sensors were optimised to specifically detect their sNTS (19 to 24 nt synthesised RNA). Addition of a 24 nt "competitor nucleotide" (sNTS-Mfr2) allowed Mfr1 e-SRS sensor to distinguish a single nucleotide difference out of 24 nt between Mfs and Mfr1. For Malaria, we created long NTS (120 nt) from genomic sequences using NASBA (isothermal RNA amplification). Addition of antisense oligonucleotides allowed the detection of otherwise undetectable long NTS. Mfs and Mfr2 long NTS were specifically detected, while the same for Mfr1 required further work to establish.

List of Tables

List of Figures

List of Illustrations

List of Symbols

Term	Explanation
AS	Antisense oligonucleotides. Added to long NTS to improve access to the NTS by binding to sequences close to and flanking the NTS.
e-SRS	environment-Sensing Response System. A system that senses the environment (e.g. a specific mRNA or metabolite) and couples this sensing to the regulation of RS or PLG(s) expression(s).
iECFP	pcDNA4/TO/ECFP/myc-His A plasmid with the ECFP sequence cloned into the provided MCS. The starting "i" stands for "inducible".
iEGFP	pcDNA4/TO/EGFP/myc-His A plasmid with the EGFP sequence cloned into the provided MCS. The starting "i" stands for "inducible".
iEYFP	pcDNA4/TO/EYFP/myc-His A plasmid with the EYFP sequence cloned into the provided MCS. The starting "i" stands for "inducible".
NASBA	Nucleic Acid Sequence-Based Amplification. An isothermal amplification technique that amplifies DNA or RNA into RNA.
NNS	Nucleic acid Non-targeted Sequence. Typically an RNA sequence that is close to the NTS, and needs to be distinguished from the NTS by e-SRS.
NTS	Nucleic acid Target Sequence. An nucleic acid sequence that is to be targeted or identified by e-SRS.
NTS'	Reverse complement of NTS.
PCR	Polymerase Chain Reaction.
PLG	Payload gene. A "gene" (could be a protein coding gene, or a non-coding "gene", such as siRNA, antisense, Rz, or miRNA) whose expression that is regulated by e-SRS.
pTR	pcDNA6/TR plasmid that expresses the repressor protein, R1.
pPLG	pcDNA4/TO/myc-His plasmid with a PLG cloned into the provided MCS.
RCA	Ribozyme Cis-cleavage Assays.
RCAA	Ribozyme Cis-cleavage Allosteric Assay. RCA with an allosteric RNA trigger.
RS	Reporter Substrate.
RTA	Ribozyme Trans-cleavage Assays.
SB	Stability Buffer. Used in the stability analysis of RNA combinations of Mz+NTS+STS, Mz+NNS+STS and Mz+STS structures. This indicated how much "buffer" the most stable structure had for it to stay in its conformation, and thus how stable that particular structure was likely to be.
SB%	SB calculated as a percentage of the free energy value of the most stable structure.

$$SB\% = 100\% \times \frac{SB}{\text{Energy (most stable structure)}}$$

SAS	Sensor Arm Split. The MzL segment of the sensor arm is separated from the MzR segment at the SAS. The SAS is defined numerically as the number of sensor arm

Term	Explanation
	nucleotides that exist on the MzR.
SAS'	SAS' is the corresponding point of the SAS on the NTS that separates the segment that binds to MzL and the segment that binds to MzR.
sNTS	Short NTS.
STS	Substrate Target Sequence. An RNA sequence that is cleaved by the activated e-SRS sensor (typically existing on the RS, or the mRNA sequence of a repressor involved in repressing the PLG).
STS'	Reverse complement of STS.
tRES	tRNA expression system. This is an RNA expression plasmid where the RNA to be expressed is placed under a pol III promoter. The version used here is that of a modified human tRNA for Valine (tRNA^{Val}) with the last seven bases removed and replaced with a short linker to prevent 3' end processing.

CHAPTER 1   INTRODUCTION

1.1 Conventional methods of Nucleic Acid sequences detection

Conventional methods of detecting Nucleic Acid sequences (NA) are based on hybridisation of complementary fluorescent or radioactive oligonucleotides to target NA, such as fluorescence in situ hybridization (FISH). Various implementations of FISH includes probes consisting of DNA, RNA, 2'-O-methyl RNA, and protein nucleic acids [Dirks et al, 2003]. However, such methods utilising hybridisation alone have two intrinsic weaknesses: 1) They are often limited in their ability to detect small differences in sequence, especially single nucleotide differences (SND), and 2) They usually have poor Signal to Noise Ratio (SNR) as typically each hybridisation of probe and target results in only 1 signal generation event, meaning that each target transcript can at most give rise to a single weak signal. In addition, the more probes added, the more unhybridised probes exist, which contribute to a significant background.

The lack of signal amplification of conventional hybridisation probes often result in the need for a prior NA amplification step, such as Polymerase Chain Reaction (PCR), before they can be applied to gene detection. Many conventional methods of gene detection simply rely on amplification of targeted DNA using gene specific primers, followed by gel electrophoresis and staining with NA binding dyes such as ethidium bromide. Such methods are usually equipment/expertise demanding (e.g. thermal cyclers or melting curve analysis in quantitative

PCR) and are laborious and time consuming (e.g. gel running and staining). In addition, PCR detection of SND is also difficult to achieve.

There have been attempts to produce better fluorescent probes, such as the Pleiades probes, which contain a minor groove binder that improves specificity and reduces background fluorescence [Lukhtanov et al, 2007]; or the NABit probe that has green fluorescence only upon hybridisation [Kubota et al, 2007]. However, none of these have become mainstream methods. Moreover, they lack the power of amplification, hence limiting sensitivity.

The relatively recent introduction of Molecular Beacons (MB) has significantly improved specificity and SNR [Tyagi & Kramer, 1996]. A MB comprises of an oligonucleotide conjugated on opposite ends with a fluorescence reporter and a quencher that are constitutively close by, and which separate upon binding of target NA and thus generate significant fluorescent signal only when hybridised. However the same limitation of generating at most 1 fluorescent reporter per copy of target NA exist.

The environment-Sensing Response System (e-SRS) is a NA sensing and response system that could overcome the two intrinsic limitations of conventional hybridisation probes. When used as an in vitro gene detection tool with the Reporter Substrate (RS), each e-SRS sensor upon activation by the target NA can activate many copies of the RS, resulting in signal (e.g. fluorescence) amplification. While each e-SRS sensor activation is initiated by hybridisation to the target NA, precise conformational changes of the sensor (that are dependent on proper hybridisation) are required for eventual sensor activation, implying that requirement for correct hybridisation is higher than conventional probes, resulting in higher specificity. While

the typical SND detection precision of the MB can detect SND in 15 nt of probe region [Bonnet et al, 1999], we have shown that e-SRS is capable of detecting SND in up to 24 nt of probe region. As such, e-SRS with RS offers an alternative to conventional hybridisation probes for in vitro gene detection that has both signal amplification and better specificity.

1.2 Ribozymes

Nucleic acids do not merely passively "store" genetic information. Many of them, especially RNA, are possessed of enzymatic activities and thus are also "executers" of genetic instructions. An RNA that is able to catalyse the cleavage and ligation of the RNA backbone phosphodiester linkage is called a ribozyme (Rz) [Sun et al, 2000].

A Rz has a catalytic domain that usually requires Mg^{2+} or other divalent cations for its enzymatic interaction with the cleavage site on the substrate RNA. Specificity of the location of cleavage is usually determined by two factors. 1) A Rz usually has two binding arms that bind the sequences flanking the cleavage site on the substrate, via Watson-Crick base-pairing. 2) The cleavage site on the substrate usually has a consensus sequence requirement, which varies according to the type of Rz cleaving it.

Some common Rz types include the Hammerhead Rz, Hairpin Rz, and the Hepatitis delta virus Rz, whose cleavage sites and schematic diagrams are shown in **Table 1.2-1** and **Figure 1.2-1** respectively.

Table 1.2-1 Common ribozymes and their cleavage sites.

Ribozyme type	Cleavage site	Example of cleavage site
Hammerhead ribozyme	NUX^	GUC^ (the WT sequence)
Hairpin ribozyme	N^GNN	C^GUC

Ribozyme type	Cleavage site	Example of cleavage site
Hepatitis delta virus ribozyme	5' of G-U wobble pair on the P1 stem (G on substrate and U on Rz) located at 5' end of cleavage site.	^GUGGUUU

For cleavage sites, ^ represents the cleavage location; N: any nucleotide; X: A, U, or C.

| Hammerhead Rz | Hairpin Rz | Hepatitis delta virus Rz |
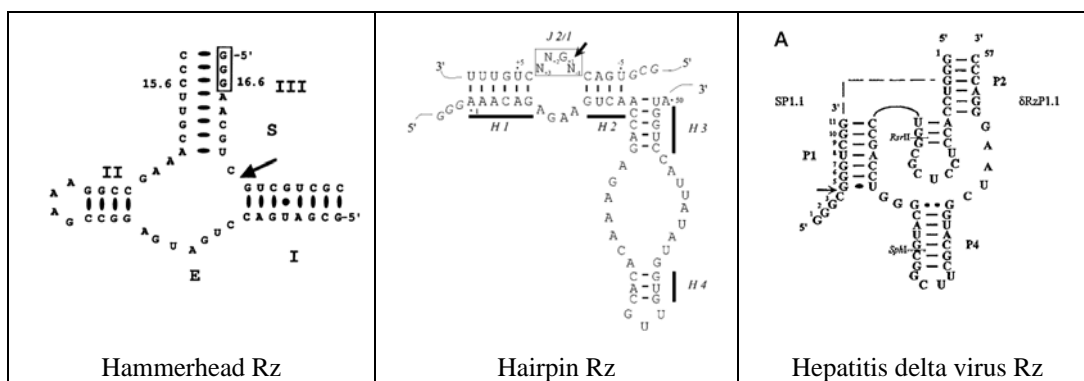
Figure 1.2-1 Adapted from [Sun et al, 2000]. Schematic diagrams of three common ribozymes.

Cleavage locations are indicated by arrow.

Aside from Watson-Crick base-pairing, nucleic acids are also able to interact with ligands via other molecular interactions along their three dimensional structures [Nimjee et al, 2005]. Aptamers are nucleic acids that bind strongly and specifically to their ligands. In fact, RNA aptamers binds their specific ligands with affinity matching and even exceeding (dissociation constants in low picomolar to nanomolar range) that of antibodies [Breaker, 2004].

The specific binding activity of aptamers can be selected in an in vitro evolution process known as SELEX (Systematic Evolution of Ligands by EXponential Enrichment) [Tuerk & Gold, 1990; Breaker, 2004]. The process starts with about 10^{14} to 10^{15} random sequences that are incubated with target molecules immobilised on a surface. Non-binding sequences are washed off, and binding sequences are amplified via RT-PCR and then in vitro transcribed to produce the candidates of the next selection round. After about 8-12 rounds of selection &

amplification, the final RNAs cloned & sequenced as highly selective RNA aptamers. The SELEX process is in vitro, fast, and can be used to generate aptamers for almost any target biomolecule.

Aptamers hold many advantages over protein antibodies [Nimjee et al, 2005], including: 1) In vitro selection is fast and applicable for any protein target. 2) A specific region of a ligand can be targeted. 3) Low or no evidence of immunogenicity. 4) Cross-species reactive aptamers can be generated using "toggle" strategy. 5) Uniform quality, regardless of synthesis batch. 6) Antidote (neutralising agent) to aptamer can be easily designed via Watson-Crick base pairing.

With both catalytic and specific binding capabilities, the fundamentals are present for nucleic acids to take on the role of biomolecular regulation. In fact, many examples already exist, both in nature and in the laboratory.

In nature, an RNA with both an aptamer domain and expression platform is called a riboswitch [Soukup & Soukup, 2004]. The aptamer domain (~70-170 nt) is usually located 5' of a nascent mRNA. The expression platform is usually 3' of the aptamer domain and can sometimes overlap with it. When the aptamer domain binds its specific ligand, it undergoes conformational changes that are transduced onto and activate the expression platform. The activated expression platform controls the expression of the gene downstream via allosteric modulation of the 5' UTR structure, such as by forming Rho-independent transcriptional terminators hairpins, by sequestering the ribosome binding site (RBS), or by ribozyme cleavage of the mRNA.

One example of a riboswitch in nature is the bacterial *glmS* mRNA that encodes the glutamine-fructose-6-phosphate amidotransferase enzyme, which uses fructose-6-phosphate and glutamine to generate Glucosamine-6-phosphate (GlcN6P), contains a riboswitch at its 5' UTR [Winkler et al, 2004]. This riboswitch regulates the *glmS* expression via negative feedback as the end product GlcN6P is the metabolite ligand for and activates the riboswitch, which then cleaves the *glmS* mRNA and reduces its translation. In fact, the ribozyme activity of the riboswitch is increased up to 1000 fold in the presence of its ligand. Such a ribozyme is known as an allosteric Rz as its activator (the ligand) is different from its substrate (*glmS* mRNA).

In the laboratory, an RNA aptamer had been designed against Human Coagulation Factor IXa (FIXa) that could induce dose-dependent reduction of coagulation to less than 1% of normal FIXa activity. This aptamer could thus serve as an anticoagulant. However, it is just as important to prevent excessive bleeding from an overly high dose of the aptamer, hence an antidote was created that consisted of a short RNA complementary to a large section of the aptamer. This antidote was able to neutralise more than 95% of the anticoagulation activity of a cholesterol modified aptamer within 10 minutes of its bolus injection [Rusconi et al, 2004].

In a potentially more far-reaching attempt, a metabolite (theophylline) specific aptamer had been designed to regulate the expression of another gene by sequestering its RBS in the presence of the metabolite [Bayer & Smolke, 2005]. These "programmable ligand-controlled riboregulators of eukaryotic gene expression" could distinguish between metabolites as similar as theophylline and caffeine, which differ by only a methyl group.

[Bergeron & Perreault, 2005] provided yet further demonstrations that Rz are amenable to Watson-Crick base-pairing based adjustment of their structures, and can be modularly adapted to suit the desired gene regulation. The wild type Hepatitis delta virus Rz cleaves any RNA that can bind correctly to its 7 nt P1 stem. [Bergeron & Perreault, 2005] changed its selectivity by adding a SOFA (specific on/off adapter) domain, that "locked" the natural P1 stem. SOFA only allowed P1 stem binding to a substrate RNA if the substrate had a 3' sequence that could bind to the biosensor (BS) region within SOFA (see Figure 1.2-2). This effectively improved the specificity of the Rz, and changed the specificity requirements from the short P1 stem, into a longer, specific sequence. This longer, specific sequence could be changed as desired by designing the BS region as the reverse complement of the desired target sequence (P1 stem complementarity is still required). The addition of SOFA therefore allows the Hepatitis delta virus Rz to become a much more specific RNA knock down tool.
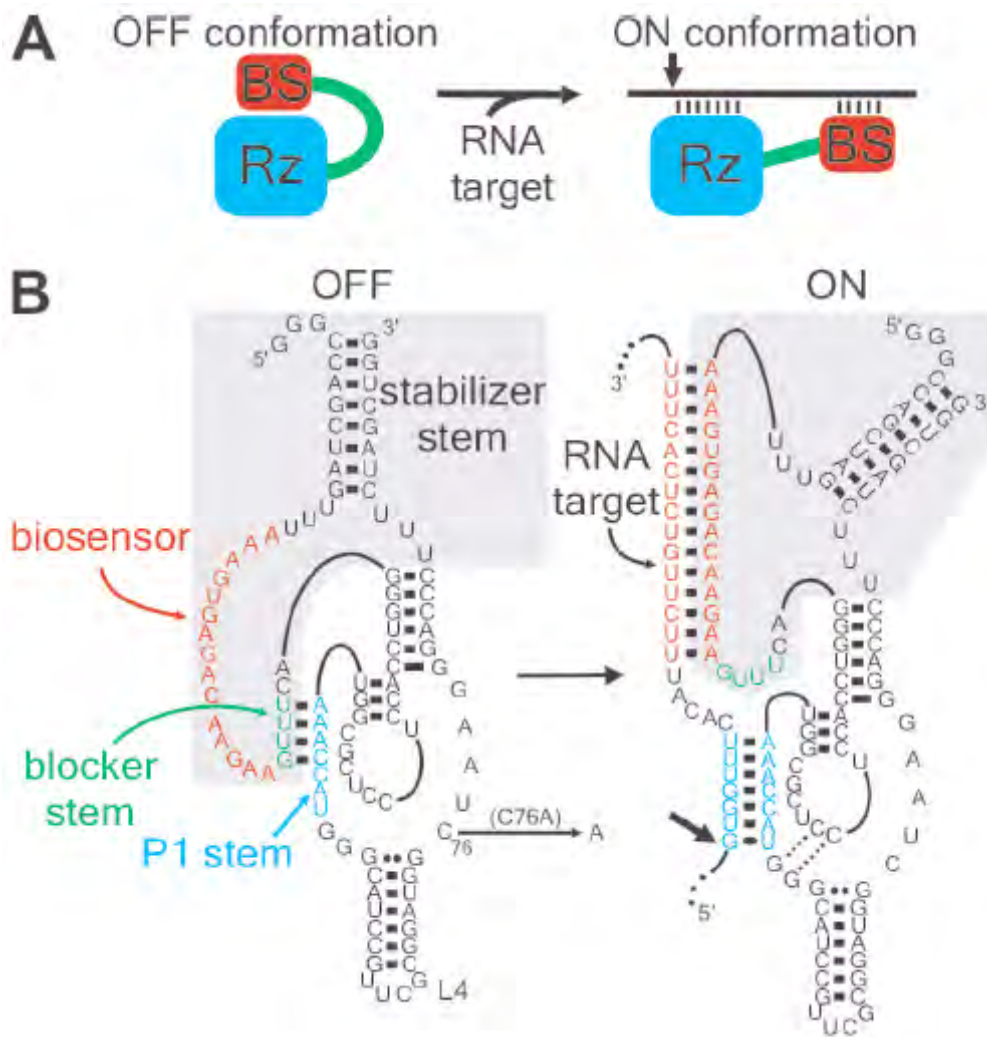
Figure 1.2-2 From [Bergeron & Perreault, 2005]. Schematic diagrams of the SOFA-ribozyme.
In A, the addition of the target RNA changes the Rz from an inactive "OFF" conformation into an active "ON" conformation. In B, the SOFA domain (highlighted in grey) was shown to lock the P1 stem, until the target RNA was added. The cleavage location on the substrate RNA was indicated by the bold arrow.

In summary, RNAs have characteristics that are amenable for development as biomolecular

sensors. Both in nature and in the laboratory, RNAs have been utilized, and even engineered

to detect biomolecules. However, most of these RNAs were used to sense metabolites or as

RNA knock down tools. To the best of our knowledge, there had been no significant work to

produce an RNA based sensor that could be designed to detect any desired RNA sequence and

to subsequently use that information to trigger a response system. This work therefore endeavoured to design such an RNA based sense and response sytem that we called the environment-Sensing Response System (e-SRS).

1.3 environment-Sensing Response System (e-SRS)

1.3.1 Motivation for e-SRS

The existing limitations of conventional Nucleic Acid (NA) detection, as well as a lack of a NA sensing platform that could actually deliver a physical action or response within the system being sensed, prompted us to search for a means to provide such a platform. The ability to react after sensing (as opposed to merely giving off a hybridisation signal like conventional hybridisation probes) enables the e-SRS sensor to serve as a signal transducer, which is able to convert a signal of one form from the environment (e.g. presence of a specific NA), to that of another form as produced by the Response System (e.g. activation of an inducible expression system). A biomolecular signal transducer could function in much more diverse ways than just a biomolecular probe, and could be a new and powerful tool in research, diagnostics and therapy.

Ribozymes (Rz), being able to both sense NA (via annealing), and respond (by catalytic cleavage of RNA substrates) stood out as prime candidates for the signal transducer in a sense and respond system. What was required was to design the Rz secondary structure such that target sensing could be used to induce a change in conformation that allows the Rz's catalytic abilities to be activated. Importantly, to be able to act as a signal transducer, we needed a

system whereby the NA sensed is not the NA cleaved (the typical case for natural Rz). In this way, the system would no longer be just another RNA knock-down tool.

The outcome of our innovation efforts was the environment-Sensing Response System (e-SRS). e-SRS is a nucleic acid sensing and response system with two components: 1) An RNA based sensor that changes conformation upon binding of specific Nucleic acid Target Sequence (NTS); 2) A Response System that is triggered by the activated sensor to initiate some physical response, such as emitting a signal (e.g. fluorescence) to indicate the presence of the NTS, or to activate some downstream action such as gene expression.

e-SRS aimed to provide a highly sequence specific (with resolution down to Single Nucleotide Differences, SND) molecular sensor (with signal amplification ability) to detect specific RNA or DNA sequences, which could be modularly coupled to various desired Response System. The modular use of different Response System, from unquenching of fluorescence RS to induction of gene expression, is novel and potentially allows a wide range of applications, ranging from gene detection (if the Response System is a RS) to killing of malignant cells (if the Response System is an inducible cytotoxic gene). In the existing final embodiment, we employed a fluorescence based RS as the Response System and applied e-SRS to the detection and resolution of closely related RNA sequences in test tubes.

1.3.2 Development of e-SRS

The initial concept was a system named Environment-Sensing Induced Gene Expression (e-SIGE). e-SIGE shared a similar principle as e-SRS, but was more complex in its physical

implementation. e-SIGE was jointly conceived by the author (Mr. LEONG Shiang Rong) and Mr. NG Kwang Loong Stanley on 14th March 2005 at Bioinformatics Institute (BII, A*STAR), independently of BII's core projects. The author and Mr. Ng KL communicated to Prof. Mishra SK (then Executive Director of BII) and to Exploit Technologies Pte Ltd (ETPL) for Invention Disclosure (ID) on 17th May 2005 (Invention Disclosure, ETPL, 2005). ETPL subsequently applied for a US provisional patent application on 08th June 2005.

The author also applied (as Collaborator/Inventor) jointly with A/Prof. Soong Tuck Wah (NUS, Physiology, as PI), A/Prof. Hooi Shing Chuan (NUS, Physiology, as Co-PI), Prof. Santosh K. Mishra (then BII-A*STAR, as Co-PI), and Mr. Ng Kwang Loong Stanley (then NGS-AGS, as Collaborator/Inventor), and was awarded the 1st SBIC Innovative Grant to develop e-SIGE as a bioimaging tool. The Project Title was "A Novel E-SIGE Technology to Bioimage Alternative Splicing Activity in Neuron and Muscle" (SBIC Innovative Grant, 2005). As such, the first two sections of this work endeavoured to develop an environment sensing response system based on the physical designs of e-SIGE.

After the initial tests of RNA test constructs derived from the e-SIGE design, it was realised that the e-SIGE design was not feasible for our implementation. As the e-SIGE RNA foldings were designed by eye without computational secondary structure prediction, the folding of RNA in ways other than as intended was a likely reason for the failure of the e-SIGE design. The subsequent use of RNA secondary structure prediction and adaptation of an existing allosteric ribozyme, the Maxizymes [Kuwabara et al, 1998], enabled us to successfully design and test the new and much simpler e-SRS in test tube environment. In addition, our

unsatisfactory experience with RNA cleavage assays via Denaturing Polyacrylamide Gel Electrophoresis (D-PAGE) prompted us to search for a new assay method. This was realised with our development of the RS, which provided real time fluorescence based reporting of the e-SRS sensor activity. Our eventual realisation of the sense and response system was thus a combination of the e-SRS sensor with the RS, for fluorescence based real-time detection of RNA in the test tube. A more detailed overview of e-SRS development is given in CHAPTER 2.

1.4 Dengue

Dengue fever (DF) is the most prevalent mosquito-borne viral illness in humans with an estimated yearly 100 million infections worldwide [Deen et al, 2006]. DF is characterized by high fever, chills, body aches and skin rash. Severity ranges from mild, flu-like symptoms to the more severe form, the dengue hemorrhagic fever (DHF). DHF affects around 250,000 to 500,000 individuals a year, with a mortality rate as high as 10% to 20%. However, much of these mortality is due to a lack of early diagnosis [Samuel & Tyagi, 2006]. With early diagnosis and appropriate management mortality can be kept below 0.5% [Oishi et al, 2007]. There is currently no vaccine for dengue [Simmons & Farrar, 2009]. Dengue is now endemic in more than 100 countries, with 2.5 billion people living in dengue endemic areas. With the dramatic expansion of dengue incidence in recent decades, the World Health Organization has classified dengue as a major international public health concern.

The causative agent is the mosquito-transmitted Dengue virus (DV). Four serologically

distinct serotypes DEN-1, DEN-2, DEN-3, and DEN-4 exist. Infection with one serotype confers lifelong immunity to the same serotype [Lanciotti et al, 1992] but increases the risk of more severe disease including DHF upon infection with a different serotype [Guzmán & Kourí, 2004]. DV has a positive strand RNA genome about 10,200 nt long that codes for three structural (capsid, membrane and envelope) and seven non-structural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5) [Guzmán & Kourí, 2004].

Diagnosis of DV infection plays important roles in prevention, treatment, initiation of control measures, and keeping of accurate epidemiological data. Traditional assays for detection of DV serotypes includes serum dilution plaque reduction neutralization, complement fixation, and hemagglutination inhibition. The serotype is then inferred by detection of at least fourfold increase or reduction in serotype specific antibodies. Unfortunately, specific diagnosis is often difficult due to widespread antibodies cross-reactivity to flaviviruses, especially between DV strains [Lanciotti et al, 1992; Guzmán & Kourí, 2004]. The need for paired serum samples also result in a delay in diagnosis and results are usually not straightforward.

For more definitive diagnosis, virus isolation using cell cultures or mosquitoes can be performed from patient serum collected during acute phase of disease or from mosquitoes. Sensitive methods include inoculation of adult A. aegypti or Toxorhynchites species mosquitoes and subsequently staining the mosquito brain tissues with fluorescent DV type-specific monoclonal antibodies [Samuel & Tyagi, 2006]. However, such methods take days to weeks and may not be successful due to insufficient viable virus in the inocula, virus-antibody complexes, and inappropriate handling of samples [Lanciotti et al, 1992].

There is therefore a strong motivation for the development of DV diagnostic assays that are rapid, specific and sensitive so as to improve prevention, treatment, initiation of control measures, and the keeping of accurate epidemiological data. Modern molecular biology based diagnostic methods have evolved in an attempt to meet this need. DV antigen detection has traditionally been of low sensitivity, but in recent years, ELISA based detection of DEN-3 antigen in serum has reached sensitivity of 90% [Guzmán & Kourí, 2004]. Of greater use for laboratory screening and molecular epidemiological studies are the RT-PCR based amplification of DV genome using either DV or serotype specific primers [Guzmán & Kourí, 2004; Lanciotti et al, 1992], followed by gel analysis or melting curve analysis of amplicons. Dengue RNA has been detected via PCR from serum, plasma, infected cells, mosquito larvae and mosquito pools. Today, a laboratory diagnosis of DV infection can be established by detection of genomic sequences via RT-PCR. However, as PCR requires complex equipment and sophisticated interpretation skills (for melting curve analysis), a simpler yet fast and specific detection format would be greatly useful in field detection or in less development areas. As such, e-SRS is tested here as a potentially rapid, simple, and specific way to distinguish between the 4 Dengue serotypes RNA sequences.

1.5 Malaria

Malaria is a global health concern for which the Who Health Organization estimates 300 to 500 million cases of infection with over one million deaths globally every year. The symptoms of malaria are caused by the asexual blood stages (rings, trophozoites, schizonts) of

the Malaria parasites, the Plasmodium species, which are thus the main target of chemotherapy.

There are 4 Plasmodium species that infect humans [Griffith et al, 2007]:

1) Plasmodium falciparum (Pf) predominates in sub-Saharan Africa, Hispaniola, Papua New Guinea (PNG), and includes both Chloroquine Resistant (CQR) and Sensitive (CQS) strains. CQR strains are in all endemic areas except Central America west of Panama Canal, Mexico, Hispaniola, & parts of China and Middle East. Pf accounts for slightly more than 50% of reported cases in US.

2) Plasmodium vivax (Pv), the second most common species, predominates in South Asia, Eastern Europe, Northern Asia, Central and most of South America, and are mostly CQS except in PNG and Indonesia. Pv accounts for about 25% of reported cases in US.

3) Plasmodium ovale occurs mostly in West Africa, occasionally in South East Asia and PNG., and are all CQS.

4) Plasmodium malariae occurs at low frequency in patchy distribution worldwide, and are CQS.

In Singapore, about 100 to 300 cases occur each year [Ministry of Health, 2004]. In 2006, the distribution of species causing infections are about 68.0 % for Plasmodium vivax, 26.0% for P. falciparum (26.0%), 2.2% for P. malariae, and 3.8% for mixed Plasmodium vivax and Plasmodium falciparum [Ministry of Health, 2006]. In 2009, there seemed to be a rise in Malaria infection, totalling 134 in the first 37 weeks, compared to only 119 for the same

period in 2008 [Ministry of Health, 2009].

In selecting the proper treatment for Malaria, Plasmodium species differentiation during diagnosis is essential [Mangold et al, 2005]. It is particularly important to differentiate P. falciparum infections from the rest, as such infections have potential to rapidly progress to severe illness or death, and are responsible for most of the deaths worldwide due to malaria. In the United States, P. falciparum accounts for about 95% of malaria deaths [Stoppacher & Adams, 2003].

The current gold standard for diagnosis is the microscopic examination of Giemsa-stained thick and thin blood films [Murray et al, 2008]. In Singapore, if diagnosis is considered likely and initial blood films are negative, they are to be repeated 12 hourly for 48 hours [Ministry of Health, 2004].

While diagnosis by microscopy is cheap [Jonkman et al, 1995] and can reach a sensitivity of about 5 to 10 parasites/µl [Murray et al, 2008], it is time-consuming and requires the availability of a well trained microscopist who can distinguish the various Plasmodium species. Such expertise is not always readily available, particularly in non-endemic regions. In addition, misdiagnosis in blood film examination occurs even with experienced microscopists, especially in cases of mixed infection or low parasitaemia [Kilian et al, 2000; Hänscheid, 2003].

There are three levels of diagnostic concerns in places where Malaria is not endemic, such as in Singapore:

1) Firstly, the physician needs to know as quickly as possible if the patient presenting

symptoms of Malaria is indeed infected with the plasmodium parasite. If this is the case, the patient needs to be quickly isolated from human-vector contact to limit the spread of the infecting plasmodium specie(s).

2) If malaria is confirmed, of particular urgency is the need to diagnose whether P. falciparum is present, as this is by far the most lethal of the 4 Plasmodium species that infect humans.

3) If P. falciparum is present, it would be useful to know if it is of the CQS or CQR strain as that would determine whether Chloroquine would be used.

Based on the urgency presented in the above three levels of diagnosis, it would be highly desirable to have a fast, kit based diagnostic method with simple reading of results that could be used in the general physician's office. This is particularly so for the containment of new infections in non-endemic regions.

Like in DV diagnosis, the need for more rapid, accurate, and simple assays have pushed for the development of molecular detection for Plasmodium diagnosis. One such development is that of Malaria Rapid Diagnostic Tests (MRDTs). MRDTs employ lateral-flow immunochromatographic methods, which are already commonly used in other diagnostic kits, such as pregnancy tests. The clinical sample liquid migrates via capillary action over the surface of a nitrocellulose membrane. If the appropriate antigen is present, it would be extracted and bound by a capture antibody immobilised on the nitrocellulose membrane. A detection antibody conjugated to an indicator (usually gold particles) in the mobile phase will then bind to the captured antigen and produce a visual signal on the membrane. Both

antibodies can be either monoclonal or polyclonal. The source of antigen used to induce these antibodies (i.e., native proteins, recombinant proteins, or peptides) can make significant differences in the performance of the MRDT [Murray et al, 2008].

Compared to the current gold standard of microscopy, MRDTs offer the much desired advantages of minimal operator training due to its ease of use, as well as rapid availability of results, usually in less than 1 hour [Murray et al, 2008]. However, MRDTs currently still suffer from limitations, such as reports of antigen persistence, particularly that of histidine-rich protein II (HRP-II), even after the clearance of malaria parasites [Mens et al, 2007], poor sensitivities at low but clinically relevant levels of parasitaemia, and the false negatives of certain strains that epitope diversity may lead to, especially in MRDTs that use monoclonal antibodies [Murray et al, 2008].

Another major molecular diagnostic approach is that of PCR based methods using either pan human infecting malaria or species specific primers, which have resulted in better species discrimination and sensitivity compared to both microscopic or immunochromatographic diagnosis. Of note, real-time PCR (as well as NASBA) assays have shown the potential to detect low levels of parasitaemia, with resolution up to about 0.02 parasites / µl blood [Schneider et al, 2005], identify mixed infections [Mangold et al, 2005], and allow for precise differentiation of species via melting curve analysis [Mangold et al, 2005]. However, such PCR based methods have their drawbacks, in particular the need for relatively expensive [Murray et al, 2008] and complex equipments (thermal cyclers) as well as specialist skills (melting curve analysis) when such requirements are often hard to meet in many of the

developing countries where Malaria is most devastating [Mens et al, 2007]. As such, PCR based methods are excluded from consideration as a field-ready rapid diagnostic kit for Malaria [McNamara et al, 2004].

In this work, we were thus motivated to show a Proof-of-Concept in the application of e-SRS as a potentially rapid and simple way to distinguish 3 strains of Plasmodium falciparum with closely related sequences (including single base differences for 2 of them): The Chloroquine sensitive strain, denoted here as Mfs (CVMNK haplotype for pfcrt gene), and two Chloroquine resistant strains denoted here as Mfr2, Mfr1 (CVIET and SVMNT haplotypes for the pfcrt gene). Of note, Mfr1 has only a single base difference with Mfs.

CHAPTER 2   OVERVIEW

2.1 Overview e-SRS development.

e-SRS development can be broadly viewed as 3 major phases, as shown in Figure 2.1-1 (including major problems and solutions/actions taken). The three phases were:

1) The initial design and testing of an intelligent gene expression platform known as e-SIGE (Environment-Sensing Induced Gene Expression) that would activate upon sensing specific Nucleic Acid sequences. As several possibilities exist for the actual molecular implementations, we selected a basic design that detects RNA, known as the IRC (Integrated RNA silencing Construct) and tested it in test tubes.

2) The design and testing of the present e-SRS (environment-Sensing Response System). As e-SIGE was complex and intial tests could not verify its functionality, we created

a much simpler design (e-SRS) using as sensor a modification of an existing allosteric ribozyme, the Maxizyme (Mz). A RS concept was conceived and served as a (gene reporting) response system that e-SRS could activate. Both in vitro and cell line assays were carried out.

3) Application of e-SRS in the detection of Nucleic Acids in test tubes using RSs. We selected 7 targets for detection: 3 strains of Malaria parasites (Plasmodium falciparum) and 4 common serotypes of Dengue viruses. A computational algorithm was created that greatly automated the design and assessment of these Mz-based sensors.

e-SIGE - Initial design for gene expression platform that activate upon sensing specific nucleic acid sequences.	Designed e-SIGE system [4.1] with sequences for a basic design, Integrated RNA silencing Construct (IRC). Aim 1a and 1b designed to test key elements of IRC mechanisms [4.2]. Synthesised Aim 1a and 1b contructs [4.2.2]. Problem: Assembly PCR yielded many sequence errors. Action: "Gene synthesis via oligonucleotide ligation". Tested IRC via Ribozyme Cis-cleavage Assays (RCA) [4.2.3]. RCA results analysed via Denaturing PAGE were mostly unexpected. Conclusions: 1) Initial e-SIGE sensor was unworkable. 2) New assay needed for analysing RCA that gave clear and immediate signals, and did not involve gel running. 3) Therefore, a new sensor and system had to be designed.

e-SRS with Maxizyme (Mz) based sensors to replace e-SIGE.	Designed e-SRS system [4.3] using modification of allosteric ribozyme, the Maxizyme (Mz). Created RS [4.3.2] to assay Ribozyme Trans-cleavage Assays (RTA).
	
	Successfully tested e-SRS RTA in vitro with RS [4.3.8].
	
	Tested e-SRS in cell lines to:
	    1. Activate inducible gene system [4.4.1].
	Problem: Mz-based sensors did not activate e-SRS.
	Action: Focused on using the simpler HH-2 ribozyme instead.
	Problem: Transfected HH-2 RNA did not activate e-SRS.
	Action: Expressed HH-2 with tRNA promoter, T7 RNA Pol promoter.
	    2. Activate RS [4.4.2].
	Problem: RS degraded when transfected.
	Action: Added nuclease protection to RS.
	Problem: RS degraded by DMEM despite nuclease protection.
	Action: Removed DMEM during transfection.
	
	Conclusions: 1) RS likely degraded in cells. 2) Therefore, it was best to develop e-SRS for in vitro applications.

e-SRS for in vitro nucleic acid detection using RS [4.5].	Designed Mz-based e-SRS sensors for 3 Malaria & 4 Dengue targets [4.5.2]. Problem: Too many Mz-based sensor designs to manually generate & assess. Action: Computational algorithm to greatly automate the design and assessment of Mz-based sensors. Tested e-SRS RTA [4.5.4, 4.5.5]. Problem: Mz-Mfr1 unable to distinguish 1 nt difference. Action: Added "competitor nucleotide" to improve specificity. Tested Malaria e-SRS via NASBA RTA [4.5.6]. Problem: NASBA products did not activate RTA. Action: Added antisense DNA to remove interfering secondary structures. Conclusions: 1) Malaria & Dengue RTA works. 2) Malaria NASBA RTA specific for 2 strains.

Figure 2.1-1 Overview of e-SRS development (relevant Sections in square brackets).

2.2 Initial designs of e-SRS.

Section 4.1 described a prior system known as e-SIGE (Environment-Sensing Induced Gene Expression), which eventually evolved and consolidated into the present e-SRS (Environment-Sensing Response System). e-SIGE shared similar principles as e-SRS, which was to couple detection of specific biomolecules to some physical response, such as emission of fluorescent signals or initiation of gene expression. However, e-SIGE was more complex in its physical implementation.

Section 4.1.1 explained the design of e-SIGE, including an overview of the activation mechanism of a basic design that detects RNA, known as the Integrated RNA silencing Construct (IRC), and how this basic design could conceptually be extended to detect multiple bio-molecules.

Section 4.1.2 described the RNA structure of IRC, including the various important segments and their sequence complementarities, which played putative roles in determining the folding of IRC.

Section 4.1.3 explained how the RNA segments described in the previous section could lead to the folding of IRC upon synthesis, and how the IRC would change in conformation upon activation by a nucleic acid it is supposed to recognise, the NTS.

Section 4.1.4 provided the sequence requirements for the various important segments of IRC, as well as actual RNA sequences for these segments.

2.3 *In vitro* testing of initial e-SRS constructs.

Section 4.2 described how we sought to test the activation of our e-SIGE IRC design, in the test tube environment, via RCA. As the IRC activation mechanism involved several RNA folding and ribozyme (Rz) cleavage processes, we had to first breakdown the whole process into smaller individual processes that could be more easily tested, via two aims (Aim 1a and Aim 1b).

Section 4.2.1 explained crucial elements of the folding process and how Aim 1a and Aim 1b were designed to fulfil the objectives of testing the activation of IRC. It detailed what the RNA test constructs and their respective controls should be, as well as the expected results of RCA.

Section 4.2.2 described how we attempted to synthesise the RNA constructs required for RCA in Aim 1a and Aim 1b. Our methodology for RCA was to first construct a DNA template inclusive of a T7 RNA polymerase promoter for each construct and then cloned this DNA template into plasmids. From these plasmids, PCR would produce many copies of in vitro transcription (IVT) templates. The IVT template would be gel purified and used in IVT to produce large amount of RNA constucts for RCA. Surprisingly, our attempts to construct the DNA templates via assembly PCR yielded many sequence errors, and we had to come up with a oligonucleotide ligation method that we called "Gene Synthesis via Oligonucleotide Ligation" to get our DNA templates constructed.

Section 4.2.3 described the results of RCA for Aim 1a and Aim 1b, which we analysed via Denaturing Polyacrylamide Gel Electrophoresis (D-PAGE). Much of these results were not

what we expected, and we concluded that the IRC design was unsuitable as the RNA-based sensor in e-SIGE. We also discovered a need for a new assay method (rather than D-PAGE) for analysing RCA as D-PAGE results were not very clear, possibly due to secondary structures in RNA that interfered with gel running.

2.4 Maxizymes based e-SRS.

Section 4.3 explained how we converted an existing allosteric ribozyme, the Maxizymes (Mz) into our new sensor for e-SRS (the new system that replaces e-SIGE). It elaborated on the two paradigms of e-SRS application, i.e., as a reporting system for detecting specific nucleic acids, and as a way to activate an inducible gene system upon detection of specific nucleic acids. It also described a RS system that could be used to better analyse (as compared to D-PAGE) RTA, and served as a reporting system in e-SRS application.

Section 4.3.1 explained how the Mz could be converted to an e-SRS sensor. The Mz is a heterodimer with one sensor arm, one catalytic arm, and is catalytic active only when both arms bind appropriately to their targeted specific RNA regions. The Mz was designed to recognise an RNA at one segment, and cleave the same RNA (i.e. cis-cleavage) at another point, for the specialised purpose of knocking down chimera genes. Our novel adaptation of the Mz separated the activating segment of the RNA from the cleaved segment, such that the activating RNA and cleaved RNA became separate RNA species (i.e., achieving trans-cleavage). This allowed us to turn the Mz from a knock down tool to a sensor for e-SRS. Section 4.3.2 introduced the RS system, which consisted of a short RNA that would give a

signal (e.g. fluorescence) when cleaved. The RS could be created by conjugating a fluorescent probe at one end, and a matching quencher at the other end.

Section 4.3.3 described the new e-SRS system with the Mz-based sensor.

Section 4.3.4 briefly described the commercial inducible gene expression system, T-REx (Invitrogen) that we used to implement the inducible gene system for e-SRS. T-REx provided a repressor protein, TetR that could suppress a co-transfected Payload Gene (PLG) until activation by the Mz based sensor (which would knock down TetR and thus relieve PLG repression).

Section 4.3.5 described the NTS selected to test the new Mz-based e-SRS and explained how 3 Substrate Target Sequence (STS) to be tested as RS were selected from the tetR mRNA sequence. As positive controls, a Hammerhead ribozyme (HH) that targeted and cleaved the each STS were also designed. The sequences of the three STS (RTS-1, RTS-2, and RTS-3) and the respective three HH (HH-1, HH-2, HH-3) were provided.

Section 4.3.6 explained how each Mz-based sensor was systematically designed and how the RNA secondary structures of each Mz-based sensor interacting with its NTS and STS were predicted (via software RNAstructure 4.5) and assessed to be specific for its NTS and STS. Strategies & guidelines for Mz design for achieving desired secondary structures were provided.

Section 4.3.7 provided the sequences of 3 Mz-based sensors (Mz-1, Mz-2, and Mz-3) that we designed.

Section 4.3.8 described the first RTA success obtained with the Mz-based sensor, Mz-2.

2.5 Cell line testing of Maxizyme based e-SRS.

Section 4.4 described our attempt to apply the Mz-based e-SRS in cell lines (HEK 293) for

two paradigms:

1) To activate an inducible gene expression system (provided by T-REx) with

   fluorescent reporter genes (ECFP and EGFP) as the PLG , if and only if NTS was

   present

2) To activate RS, if and only if NTS was present.

Section 4.4.1 described the test of paradigm 1 (e-SRS to activate inducible gene system). First,

we optimised the transfection ratio of Tet repressor R1 expressing pcDNA6/TR plasmid (pTR)

to PLG expressing plasmid (pPLG, which is a pcDNA4/TO/myc-His plasmid with a PLG

cloned into the provided MCS), using LacZ as the PLG. Using this ratio as a guide, we then

optimised the pTR to pPLG ratio using ECFP and EGFP as PLG, which turned out to be 8:1.

We thus established an e-SRS inducible gene system by transient transfection of 8:1 pTR and

pECFP or EGFP. One day after, Mz and NTS or HH were transfected. Fluorescence of the

PLGs were monitored for a few days. If e-SRS inducible gene system worked, the presence of

both Mz and NTS would activate the Mz to cleave and knock down the TetR mRNA, thus

relieving the repression of the PLG and allow fluorescence to be observed. No fluorescence

should be observed if NTS was not present. Despite using various concentration of RNA, we

were unable to observe activation of the PLGs.

As the Mz-2 design proved successful in test tubes, we focused on testing HH-2 (the wild type

Hammerhead ribozyme designed to cleavage the same STS as Mz-2), as it did not require an

NTS to activate, was simpler, and thus more likely to work. Since transfected HH-2 RNA did not work previously, we cloned HH-2 into a plasmid and transfected it to express HH-2 within cells. We employed a proven tRNA expression system (tRES), and later added a T7 RNA Pol III promoter to enhance the expression of HH-2 within cells. RT-PCR showed that HH-2 RNA was expressed within cells. However, we were still unable to activate PLGs.

Section 4.4.2 described the test of paradigm 2 (e-SRS to activate RS) in cells. The concept was similar to that of RTA in test tubes except that the RNA (including Mz, NTS, RS, and HH) were transfected into cells. Unlike the case of the inducible gene system (where e-SRS did not activate PLG under any conditions), RS was non-specifically activated under all conditions. This led us to suspect that RS was degraded in the process of transfection. We discovered that DMEM in transfection medium could degrade and thus activate RS. We tried adding to the RS nuclease resistant modifications, such as using phosphorothioate bonds and addition of 2'-O-methyl groups. We also tried removing DMEM from the transfection mix. However, these measures did not work and we concluded that the RS was also degraded after transfection, probably by nucleases present within the cells.

Based on the negative cell line results, and considering the time and resources available, we decided it was best to develop e-SRS for in vitro applications.

2.6 Applications in Nucleic Acid *in vitro* detection.

Section 4.5 described how we applied e-SRS in the detection of Nucleic Acids in test tubes via RTA using RS. We selected 7 new NTS for detection: The 1st category consisted of

Malaria parasites, which were 3 strains of Plasmodium falciparum, denoted here as Mfs, Mfr1, and Mfr2. The 2nd category consisted of Dengue viruses, which were 4 common serotypes denoted here as D1, D2, D3, D4. To design an Mz-based e-SRS sensor specific for each of these NTS, we created a computational algorithm that automated the design process. These e-SRS sensors were then optimised to specifically detect their NTS in vitro. This was successful for both Malaria and Dengue at the short NTS (sNTS) level. For Malaria, we were able to amplify using NASBA from genomic DNA and have specific detection for Mfs and Mfr2.

Section 4.5.1 listed the NTS sequence and strain information for both Dengue and Malaria. The NTS ranged from 19 to 24 nt.

Section 4.5.2 described in detail the automation process of Mz-based e-SRS sensor design. We developed in Perl 5.8.8 a computational algorithm that (using RNAstructure 4.5 for secondary structure prediction) greatly automated the 3 processes of: 1) designing sequences for specific e-SRS Mz-based sensor(s) for each NTS, 2) generating their resultant secondary structures, and 3) assessing these structures for appropriate specificity.

Section 4.5.3 listed the valid designs of Mz-based e-SRS sensors provided by our algorithm.

Section 4.5.4 described the results of Dengue sNTS detection.

Section 4.5.5 described the results of Malaria sNTS detection. Of note, the addition of a 24 nt "competitor nucleotide" (sNTS-Mfr2) allowed Mz-Mfr1 to distinguish a single base difference between Mfs and Mfr1.

Section 4.5.6 described how we used NASBA (Nucleic Acid Sequence-Based Amplification),

an isothermal amplification technique that amplifies DNA or RNA into RNA) to amplify a 120 nt segment from Malaria genomic DNA and detect them via RTA. We discovered that the use of antisense oligonucleotides (AS) allowed the otherwise undetectable NASBA products to be detected in RTA. Mfs and Mfr2 were specifically detected, while Mfr1 specific detection required further work to establish.

CHAPTER 3   MATERIALS AND METHODS

3.1 Cloning

3.1.1 PCR

3.1.1.1 Standard PCR

Standard PCR was used to amplify DNA targets (usually with known PCR conditions) for purposes such as cloning and IVT template generation. Reactions were scaled up or down as needed proportionately. Thermocycler used were the Biometra T3 Thermocycler and the Hybaid PCR Express.

3.1.1.1.1 Taq polymerase

PCR using Taq polymerase from Promega was normally employed for targets with no expected secondary structures. A 100 μl PCR reaction mix was set up as follows: Deionised H$_2$O to top up to 100 μl, 10 μl 10X PCR buffer, 4 μl dNTP mix (5 mM each), 0.2 μl Taq polymerase (5 U/μl, Promega), 8 μ.l Mg^{2+} (25 mM), 4 μl of each primers (5 μM), and ~10 to 60 ng of template DNA.

Thermocycling conditions were: Initial denaturation at 94 °C for 5 minutes, 30 cycles of amplification (denaturation at 94 °C for 30 seconds, annealing for 30 seconds, and extension at 72 °C with 1 minute for every 1000 bp of amplicon), and a final extension at 72 °C for 5 minutes.

3.1.1.1.2 Elongase Enzyme Mix

PCR using Elongase Enzyme Mix from Invitrogen was normally employed for targets with slightly greater requirement of sequence fidelity, such as those with predicted secondary structures. A 100 μl PCR reaction mix was set up as follows: Deionised H_2O to top up to 100 μl, 20 μl 5X PCR Buffer B (with 10 mM $MgSO_4$), 4 μl dNTP mix (5 mM each), 2 μl Elongase Enzyme Mix, , 4 μl of each primers (5 μM), and at least 0.05 ng of template DNA. Thermocycling conditions were: Initial denaturation at 94 °C for 5 minutes, 30 cycles of amplification (denaturation at 94 °C for 30 seconds, annealing for 30 seconds, and extension at 68 °C with 1 minute for every 1000 bp of amplicon), and a final extension at 68 °C for 5 minutes.

3.1.1.1.3 PfuTurbo DNA polymerase

PCR using PfuTurbo DNA polymerase from Stratagene was normally employed for targets with strict requirement of sequence fidelity, including those that turn out to be difficult to amplify accurately with Taq and Elongase enzymes, usually with predicted secondary structures. PCR for IVT template generation employed PfuTurbo.

A 100 μl PCR reaction mix was set up as follows: Deionised H_2O to top up to 100 μl, 10 μl 10X PfuTurbo reaction buffer (with 20 mM Mg^{2+}), 4 μl dNTP mix (5 mM each), 2 μl PfuTurbo DNA Polymerase, 4 μl of each primers (5 μM), and ~60 to 100 ng of template plasmid DNA. For the production of IVT template, 25 μl reaction volumes were used.

Thermocycling conditions were: Initial denaturation at 94 °C for 5 minutes, 35 cycles of amplification (denaturation at 94 °C for 30 seconds, annealing for 30 seconds, and extension

at 72 °C with 60 seconds for every 1000 bp of amplicon), and a final extension at 72 °C for 5 minutes. For the production of IVT template, annealing temperature of 55.2 °C was used.

3.1.1.1.4 **PfuUltra II Fusion HS DNA polymerase**

PCR using PfuUltra II Fusion HS DNA polymerase from Stratagene was normally employed for targets with greatest requirement of sequence fidelity, such as those that turn out to be difficult to amplify accurately with Taq and Elongase enzymes, usually with predicted secondary structures. PCR in the cloning of PLGs and HH-2-2_tRES employed PfuUltra II.

A 100 μl PCR reaction mix was set up as follows: Deionised H_2O to top up to 100 μl, 10 μl 10X PfuUltra II Reaction Buffer (with 20 mM Mg^{2+}), 5 μl dNTP mix (5 mM each), 2 μl PfuUltra II Fusion HS DNA Polymerase, 4 μl of each primers (5 μM), and ~40 ng of template DNA.

Thermocycling conditions were: Initial denaturation at 95 °C for 2 minutes, 30 cycles of amplification (denaturation at 95 °C for 20 seconds, annealing for 20 seconds, and extension at 72 °C with 15 seconds for every 1000 bp of amplicon, with minimum of 15 seconds), and a final extension at 72 °C for 3 minutes.

3.1.1.2 Gradient PCR

Gradient PCR was used for optimising annealing temperatures of new PCR targets. PCR reaction mixes were set up using one of the standard PCR procedures described above, typically either Taq polymerase (Promega) or Elongase Enzyme Mix (Invitrogen), depending on anticipated requirement of sequence fidelity. Thermocycler used was the Hybaid PCR

Express with gradient block.

Thermocycling conditions were similar to those of standard PCR for the specific enzyme system, except that 8 to 12 annealing temperature, spaced out 5 to 10 degrees around the predicted T_M of amplicon were used.

3.1.1.3 Colony screening via colony PCR

Colony screening of pGEM-T Easy vectors generally benefitted from blue/white selection, in which well-spaced white colonies were selected for colony PCR. Colony PCR was used to ensure DNA (cloned into plasmids) transformed into selected single bacteria colony were of the right PCR amplicon size before processing such colonies for DNA sequencing. Thermocyclers used were the Biometra T3 Thermocycler and the Hybaid PCR Express.

PCR reaction mix was set up using either the Taq polymerase (Promega) or Elongase Enzyme Mix (Invitrogen) protocols describe above with the following differences: 1) Reaction volume per tube was scaled to 12.5 μl, 13 μl, or 15 μl. 2) Template DNA were replaced with equal volumes of deionised H_2O. 3) For each reaction, DNA template was provided by dotting an autoclaved toothpick into the selected well-spaced colony and then rubbing the tip of the toothpick in the PCR reaction mix for the colony. The same toothpick was then dabbed onto a defined grid on an Luria-Bertani (LB) agarose with ampicillin (100 μg/ml) master plate which would be incubated at 37 °C overnight. After incubation, master plate would be sealed with parafilm and stored at 4 °C.

Thermocycling conditions were similar to those of standard PCR for the specific enzyme

system, except that 35 cycles of amplification were used.

Results of PCR were analysed via agarose gel electrophoresis and clones with the specific colony PCR amplicon size would be selected. Selected clones would be retrieved from the master plate (after it had grown) with an autoclaved toothpick, inoculated for plasmid miniprep and then sent for DNA sequencing. Where desired, a glycerol bacteria stock would be created for clones with the right DNA sequence for future use.

3.1.2 Agarose gel electrophoresis

DNA were mixed with about 20% volume of 6X DNA loading buffer for loading into wells. Electrophoresis buffer used was TAE (Tris-acetate-EDTA, with 40 mM Tris acetate, 1 mM EDTA.). Agarose gels were prepared by dissolving agarose in TAE buffer, with Ethidium Bromide added at 100 μg / 100 ml gel volume. After gel running, DNA band(s) were visualised under UV illumination.

3.1.3 Gel purification of DNA

After Agarose gel electrophoresis, DNA bands of the desired size as visualised under UV illumination were excised using a clean razor blade and purified using the QIAquick Gel Extraction Kit from Qiagen, essentially following the manufacturer's microcentrifuge protocol. After a gel was dissolved, one gel volume of Isopropanol was added if the DNA was shorter than 500 bp. Before washing with Buffer PE, 0.5 ml of Buffer QG was added to each QIAquick column and centrifuged for 1 minute to remove all traces of agarose. For the elution step, after adding water and before centrifugation, QIAquick spin columns were allowed to

stand for 1 minute to improve yield DNA concentration. DNA for RNA related use (e.g. as in-vitro transcription template) were eluted with 30 µl of DEPC treated water per column, otherwise 30 µl of deionised water was used. Purified DNA samples were stored at -20 °C.

3.1.4 Ligation for TA cloning

Gel purified PCR amplified DNA were usually ligated into pGEM-T Easy vectors (Promega) via TA ligation. Each 5 µl ligation mix contained: 2.5 µl pGEM-T Easy 2X Rapid Ligation Buffer, 1.75 µl gel purified DNA, 0.25 µl pGEM-T Easy vectors (12.5 ng), and 0.5 µl T4 DNA Ligase (1.5 Weiss units). Ligation mix was incubated at either room temperature for 2 hours, or 16 °C overnight.

3.1.5 Bacteria Transformation

Competent cells DH10B, a substrain of E. coli that is RecA and RecB negative, was used to replicate plasmid DNA.

3.1.5.1 Via Heat Shock

For each transformation, 1 tube of 125 to 200 µl E. coli was first thawed on ice and kept at 4 °C. 5 µl of pGEM-T Easy ligation mix or plasmid DNA was mixed gently into thawed bacteria in autoclaved 1.5 ml microcentrifuge tube and incubated on ice for 30 minutes. Mixture was then incubated at 42 °C in a water–bath for 45 seconds, and quickly returned to ice for 2 minutes. 900 µl of LB media was added, and the mixture was incubated at 37°C with shaking at 220 rpm for about 1.5 hour. E. coli was then pelleted by centrifugation on a desktop

microcentrifuge at 8000 rpm for 2 minutes. Supernatant was removed until about 100 – 200 μl remained. Bacteria were resuspended by flicking and then plated onto an LB agarose plate with appropriate antibiotics (e.g. 100 μg/ml ampicilin for pGEM-T Easy vectors) and IPTG and X-Gal added for blue/white screening (e.g. for pGEM-T Easy vectors). Plate was then incubated at 37 °C overnight.

3.1.5.2 Via Electroporation

For each transformation, one 0.5 ml microcentrifuge tube of 20 μl electroporation competent E. coli was first thawed on ice and kept at 4 °C. 0.3 μl out of a 5 μl pGEM-T Easy ligation mix or plasmid DNA was mixed gently into thawed bacteria and incubated on ice for 1 minutes, and then transferred into a sterile disposable electroporation cuvette which had been pre-chilled on ice. The mixture was then quickly electroporated. The parameters for electroporation were Resistance: 200 Ohms, Capacitance: 25 μF, Voltage: 2.5 kV. 500 μl of LB media was quickly added, and the mixture was incubated at 37°C with shaking at 220 rpm for about 1 hour. Various volumes (usually 100 μl) of the mixture were plated onto an LB agarose plate with appropriate antibiotics (e.g. 100 μg/ml ampicilin for pGEM-T Easy vectors) and IPTG and X-Gal added for blue/white screening (e.g. for pGEM-T Easy vectors). Plate was then incubated at 37 °C overnight.

3.1.6 Miniprep Plasmid Purification

Minipreped DNA were usually used for DNA sequencing of selected bacteria colonies that passed colony screening.

Selected colonies (usually from the master plate) were inoculated for overnight culture in 2 to 5 ml LB media supplemented with appropriate antibiotics (e.g. 100 μg/ml ampicillin for pGEM-T Easy vectors). Tubes were incubated at 37 °C with shaking at 220 rpm for 9-16 hours. Plasmid purification was performed using the Wizard Plus SV Minipreps DNA Purification System (Promega), essentially following the manufacturer's centrifugation protocol. DNA was eluted with 50 μl of deionised water per column. DNA concentrations were measured by the NanoDrop ND-1000 full-spectrum (220-750nm) spectrophotometer, using 1 μl of DNA sample. Purified DNA samples were stored at -20 °C.

3.1.7 Midiprep Plasmid Purification

Midiprep DNA were usually used for transfection. Each selected colony (usually from the glycerol bacteria stock or master plate) was inoculated for starter culture in 2 to 5 ml LB media with appropriate antibiotics (e.g. 100 μg/ml ampicillin for pGEM-T Easy vectors). Starter cultures were incubated at 37°C with shaking at about 300 rpm for 8 hours. Overnight cultures were prepared by mixing 50 to 100 μl of the starter cultures with 50 ml LB media with appropriate antibiotics in 250 ml conical flasks, and incubated at 37 °C with shaking at 300 rpm for 12 to 16 hours. Plasmid purification was performed using the Qiagen Plasmid Midi Kit (QIAGEN), essentially following the manufacturer's centrifugation protocol. Unless otherwise stated, all centrifugation was done using the BECKMAN COULTER Avanti JA-25 Centrifuge with the JA-14 rotor for 6000 g and JA-20 rotor for 20,000 g, with non-polycarbonate Oakridge centrifuge tubes at 4 °C. Initial bacteria cultures were split into

25 ml fractions for pelleting After DNA precipitation by isopropanol, DNA pellets were washed and transferred to 2 ml microcentrifuge tubes by two volumes of 0.75 ml of 70% ethanol, and then centrifuged at 15,000 g for 10 minutes. Supernanants were quickly discarded to minimise DNA losses. DNA pellets were air-dried by standing at room temperature for about 10-20 minutes. 50 μl of deionised water per 25 ml of bacteria culture processed was added to each tube to redissolve DNA, at either room temperature for 30 minutes or 4 °C overnight. DNA concentrations were measured by the NanoDrop ND-1000 full-spectrum (220-750nm) spectrophotometer, using 1 μl of DNA sample. Purified DNA samples were stored at -20 °C.

3.1.8 Glycerol Bacteria Stock

To create a glycerol bacteria stock, an overnight LB media bacteria culture with appropriate antibiotics (e.g. 100 μg/ml ampicillin for pGEM-T Easy vectors), such as that used in miniprep of the desired clone would be thoroughly mixed with sterile 87% glycerol to achieve a final concentration of glycerol of at least 20%. Glycerol bacteria stocks were stored at -20 °C.

3.1.9 Gene Synthesis via Oligonucleotide Ligation.

Gene Synthesis via Oligonucleotide Ligation was a method that we employed to create double stranded DNA of a desired (usually synthetic) DNA sequence ("gene"). We divided our desired DNA sequence into overlapping fragments (overlaps of around 15 nt), with the 3' fragment having an extra 3' A, and ordered each fragment as unpurified synthesized single

stranded oligonucleotides called "Ligation Oligonucleotides". Ligation Oligonucleotides are therefore the starting material for Gene Synthesis via Oligonucleotide Ligation, which on proper overlapping with each other, form the double stranded sequence of the desire gene.

For each construct, the Ligation Oligonucleotides (each 10 μM) were mixed in an annealing buffer (Annealing Buffer, pH 7.89: 10mM Tris, pH 7.5 - 8.0, 50mM NaCl, 1mM EDTA) at 95 °C for 2 mins, then cooled to 25 °C at 0.02 °C/s using a thermocycler. This created double stranded DNAs with the sequence of the desired gene, although still with breaks in the sugar phosphate backbones, called "Annealed Oligonucleotides".

Approximately 1 μg of the annealed DNA was ligated overnight at 16 °C with 400U of T4 DNA Ligase (NEB) in a total volume of 10 μl. The Annealed Oligonucleotides that were fully ligated in all their sugar phosphate backbone breaks became the double stranded DNA version of the desired gene, and were also known as "Ligated Annealed Oligonucleotides". The Ligated Annealed Oligonucleotides were subsequently extracted from the ligation mix via agarose gel electrophoresis, excision of the desired sized band, and then gel purification. The gel purified Ligated Annealed Oligonucleotides was then cloned via TA cloning into pGEM-T Easy vector, minipreped with DH10B, and sequenced for accuracy.

3.1.10 DNA Sequencing

Plasmid DNA and gel purified DNA samples were sent for DNA sequencing at National Neuroscience Institute (NNI) Core Analytical Laboratory Services. Sequencing was done using the ABI Prism 3100 Genetic Analyzer with the BigDye Terminator v3.1 Cycle

Sequencing Kit (Applied Biosystems). A 20 μl sequencing reaction mix had the following components: 8 μl Terminator Ready Reaction Mix, 3.2 pmol primer, template DNA, and deionised H_2O to top up to 20 μl. The amount of DNA template used followed the guidelines as shown in **Table 3.1-1**.

For a gel purified PCR DNA sample, a 6 μl sequencing sample was sent to NNI as follows: 1 μl of sequencing primer (5 μM), 0.5 μl of DMSO, template DNA, and deionised H_2O to top up to 6 μl.

For a plasmid DNA sample, a 12 μl sequencing sample was sent to NNI as follows: 1 μl of sequencing primer (5 μM), 0.5 μl of DMSO, template DNA, and deionised H_2O to top up to 12 μl.

For sequencing of constructs cloned into pGEM-T Easy vectors, either the M13 (24nt) Fwd or M13 (22nt) Rvs primers as shown in **Table 3.1-2** would be used, sometimes both if the construct was long or if the presence of secondary structures was expected.

Table 3.1-1 Amount of DNA template used in sequencing reactions.

Sample Type	Template length	Quantity (ng)
PCR product	100 - 200 bp	1 – 3
PCR product	200 - 500 bp	3 - 10
PCR product	500 - 1000 bp	5 - 20
PCR product	1000 - 2000 bp	10 - 40
PCR product	> 2000 bp	40 - 100
Plasmid	-	200 - 500

Table 3.1-2 Sequencing primers used.

Primer	Sequence (5' to 3')	Length (nt)
M13 (24nt) Fwd	CGCCAGGGTTTTCCCAGTCACGAC	24
M13 (22nt) Rvs	TCACACAGGAAACAGCTATGAC	22

3.1.11 Nucleic acid concentration determination

Measurement of DNA or RNA concentration was done mainly by the NanoDrop ND-1000 full-spectrum (220-750nm) spectrophotometer, usually using 1 μl of DNA or RNA sample, always ensuring that the liquid sample column was formed. Measuring samples were discarded and not reused. Blanking samples used were the solutions that the DNA or RNA samples were diluted in (deionised water or DEPC treated water). The detection limits were from 2 ng/μl to 3700 ng/ul (dsDNA), 3000 ng/ul (RNA), 2400 ng/ul (ssDNA). If measured concentrations were above detection limits, nucleic acid solutions were diluted and measured until a steady concentration was obtained well within detection limits. via UV spectrophotometry. Absorbance values at 260 nm (OD_{260}), 280 nm (OD_{280}), and 230 nm (OD_{230}) were taken. Ratio of OD_{260} to OD_{280} were checked to be around or above 1.8 for DNA and 2.0 for RNA for a sample to be used subsequently. Nucleic acid concentrations in ng/μl were calculated as OD_{260} x dilution factor x [50 (dsDNA) or 40 (RNA) or 33 (ssDNA)].

3.2 Cell Culture

3.2.1 Cell Culture Materials

3.2.1.1 Cell lines

Two types of cell lines were used in this study. Initially, PC12 (a pheochromocytoma of the rat adrenal medulla) with high endogenous expression of Cav2.1 (37b) was used. Subsequently, HEK 293 (a transformed cell line from cultured Human Embryonic Kidney cells) with no

endogenous expression of Cav2.1 was used.

3.2.1.2 Cell line maintenance

PC12 cells were cultured with RPMI 1640 medium supplemented with 5% fetal bovine serum (FBS), 10% fetal bovine serum (HS), 1% Sodium pyruvate, and 1% Penicillin / Streptomycin (P/S), adjusted to pH 7.1 (with Concentrated HCl and/or 10 N NaOH). HEK293 cells were cultured with Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin / Streptomycin (P/S).

All cells were in T25 or T75 culture flasks with 3.5 or 12 ml of media, kept in incubation at 37°C, with 5% CO_2, and passaged when confluency reached about 80% or more (around 2 to 4 days) by trypsinisation (Trypsin/EDTA composition: 0.25% Trypsin, 0.02% EDTA in PBS) and replateing at approximately 10% confluency.

3.2.2 Cell counting

Cell counting was done by making appropriate dilutions (usually 10 to 20 times) in PBS of trypsinised and resuspended cells, and counting under light microscope using hemocytometer. Cells in 2 opposite large squares were counted such that at least 80 cells were counted for each sample. Cell concentration was given by the average number of cells counted in each large square multiplied by 10^4 and the dilution factor.

3.2.3 Transfection

Cells were harvested for transfection seeding about 2 days after passage, with a maximum

confluency of 75% in order to ensure cells were in the log phase of growth for best

transfection efficiency. Cells were seeded in culture vessels (6, 12, 24, and 96 well plates) in

normal culture medium less antibiotics. Seeding density of cells were between 1 to 2 $\times 10^5$

cells per well for a 24 well plate, and adjusted proportionately according to well surface areas

for other plates (as shown in **Table 3.2-1**). Seeded cells were allowed to grow for 1 to 2 days

until they were about 75% to 95% confluent before transfection.

Table 3.2-1 Approximate surface areas of culture vessels.

No. of wells in plate	Approximate surface area per well (cm^2)
96	0.3
24	2.0
12	4.0
6	10.0

Plasmid DNA and RNA were transfected into cells using Lipofectamine 2000 from Invitrogen,

essentially following the manufacturer's protocol for plasmid DNA and siRNA transfection

respectively. The following reagent volumes were used for transfection of a single well for a

24 well plate, and were adjusted proportionately according to well surface areas for other

plates (as shown in Table 3.2-1).

For plasmid DNA transfection, DNA was diluted in 50 μl of Opti-MEM (from Invitrogen)

with no serum added. 2.5 μl of Lipofectamine 2000 per 1 μg of DNA transfected was diluted

in 50 μl of Opti-MEM and incubated for 5 to 25 minutes at room temperature. The diluted

DNA and diluted Lipofectamine 2000 were then mixed and incubated for about 30 minutes

(between 20 minutes and 6 hours) to create the transfection complex.

For RNA transfection, RNA was diluted in 12.5 μl of Opti-MEM with no serum added. 2.0 μl

of Lipofectamine 2000 per 40 pmol of RNA transfected was diluted in 12.5 μl of Opti-MEM and incubated for 5 to 25 minutes at room temperature. The diluted RNA and diluted Lipofectamine 2000 were then mixed and incubated for about 30 minutes (between 20 minutes and 6 hours) to create the transfection complex.

Cells were transfected by gently pipetting the transfection complex onto the edge of well, followed by several gentle pipetting up and down of existing medium onto the edge of well to ensure proper mixing of the entire transfection complex into the well while minimising disturbance to cells. Medium were usually replaced with normal culture medium with antibiotics 6 to 24 hours after transfection.

For harvesting of cells, medium was removed, cells were washed in PBS, trypsinised, pelleted, and washed again with PBS. Cell pellets were frozen at –80°C if not immediately used for the next step.

3.2.4 Fixation of cells

Cells were either fixed within their culture vessels or for some samples grown on coverslips, were removed to an empty well of similar culture vessels. One volume (Vol) was defined as about 0.6 times that of the normal volume of culture medium used for such culture vessels, and was able to just cover the cells in a thin layer. One Vol for a well in a 24 well plate would be about 0.3 ml. To reduce the loss of cells, all introduction and removal of media in wells were done by gentle pipetting at walls of wells, minimising contact with cells to be fixed. All introduction of media and aspiration of formaldehyde were performed with a Gilson Pipetman

P200.

For fixation, culture media (if present) were aspirated and cells were gently washed with 1 Vol 1X PBS. Cells were then fixed by covering them with 1 Vol of 3% formaldehyde for about 10 minutes. Formaldehyde was removed and cells were washed twice with 1 Vol of 1X PBS.

If cells were to be observed in wells, they were overlaid with about 1 Vol of 1X PBS or deionised water. For storage, such plates would be sealed with parafilm and kept at 4 °C. To view cells grown on coverslips on microscope slides, the coverslips were mounted on slides in 2.5 μl of FluorSave Reagent (from Calbiochem), dried at room temperature, and stored at 4 °C.

3.3 Inducible gene system, T-REx

For the e-SRS inducible gene system, we used an inducible gene expression system, T-REx from Invitrogen. The T-REx system comprised of 2 plasmids and an extra control vector:

1) pcDNA6/TR was a 6662 bp high copy number plasmid that over-expressed the Tet repressor (tetR), a repressor protein via a human cytomegalovirus immediate-early (CMV) promoter.

2) pcDNA4/TO/myc-His was an approximately 5.1kb high copy number plasmid that expressed a PLG via a CMV promoter (with 2 tetracycline operator 2 sites embedded), which could be constitutively repressed by an appropriate ratio of cotransfected tetR. tetR repression could be relieved by application of tet of about 1 μg/ml in culture medium.

3) pcDNA4/TO/myc-His/lacZ was a 8198 bp high copy number plasmid containing the

   LacZ gene that expressed β-galactosidase as a PLG within a pcDNA4/TO/myc-His,

   and was provided as a control vector.

In the e-SRS inducible gene system, the repressor R1 was provided by the T-REx tetR

reporessor protein. R1 would suppress a PLG cloned into pcDNA4/TO/myc-His. Activation of

the Mz based e-SRS sensor would knock down R1 (relieving PLG repression), and thus

induce PLG expression.

3.3.1 Cloning of PLGs

We cloned three PLGs, EGFP (enhanced green fluorescent protein), ECFP (enhanced cyan

fluorescent protein), and EYFP (enhanced yellow fluorescent protein) into the 5151 bp

version A of pcDNA4/TO/myc-His.

pcDNA4/TO/myc-His A was cut twice by restriction enzyme digest at two unique sites within

the provided Multiple Cloning Site (MCS), Xho I and Xba I. The restriction sites for Xho I

and Xba I were C|TCGAG and T|CTAGA respectively, with "|" indicating the cut position. In

the choice of restriction sites, care was taken to ensure they were not present in the intended

PLGs.

Since the PLG sequences cloned into pcDNA4/TO/myc-His A did not contain Xho I and Xba

I, we used PCR to add short tags to both ends of all PLG sequences that contained the

restriction sites for Xho I (5' end) and Xba I (3' end). The plasmid templates for EGFP, ECFP,

and EYFP were respectively pEGFP-Cav2.1EFa (a plasmid that expressed the Cav2.1EFa or

NTS-37a with an EGFP tag), pECFP-C1, and pEYFP-C1 (both from CLONTECH Laboratories, Inc.). For both ECFP and EYFP, gradient PCR was performed using PfuUltra II Fusion HS DNA Polymerase from Stratagene, with 12 annealing temperatures spread between 49.1 °C and 64.2 °C. Each 13 μl reaction volumes contained about 5.4 ng of template DNA and primers were as shown in **Table A 1**.

The entire range of annealing temperature used was successful and a similar PCR was done for EGFP using 26 μl reaction volume with 5 ng of template DNA.

PCR products of the right sizes were gel purified, and were subjected to double restriction digest using Xho I and Xba I, similarly to pcDNA4/TO/myc-His A. A restriction digest mix was set up as follows: Deionised water to top up to final volume, 0.1 volume of 10X NEBuffer 2 (NEB), 0.01 volume of 100X BSA (NEB), DNA (about 1 to 2 μg of gel purified PCR DNA or 4 μg of pcDNA4/TO/myc-His A), 20 U of Xho I restriction enzyme (NEB), and 20 U of Xba I restriction enzyme (NEB). The restriction digest volumes were 40 μl for EGFP, 50 μl for ECFP and EYFP, and 10 μl for pcDNA4/TO/myc-His A. All restriction digest mixes were incubated at 37 °C for 1 hour, and then 65 °C for about 20 to 25 minutes to heat inactive the restriction enzymes. All restriction digest products were gel purified for products of the right sizes.

Gel purified digested inserts (EGFP, ECFP, EYFP) were ligated into gel purified digested

pcDNA4/TO/myc-His A using T4 DNA Ligase. Each 20 μl ligation mix contained: Deionised

water to top up to final volume, 2 μl 10X T4 DNA Ligase Buffer (NEB), gel purified cut

pcDNA4/TO/myc-His A vectors (~0.6 to 2 nM), gel purified cut inserts (~1.8 to 6 nM), and 1

μl T4 DNA Ligase (400 U, NEB). Molar ratio of insert to vector was designed to be about 3:1.

Ligation mix was incubated at 16 °C overnight (~17 to 21 hours).

Ligation mix were transformed into E. coli via electroporation (ECFP and EYFP) using about

0.55 to 0.75 ng DNA in 20 μl competent cells or heat shock (EGFP) using about 54 ng DNA

in 200 μl competent cells. Colony screening was performed using forward primers CMV

Fwd_1 (CGCAAATGGGCGGTAGGCGTG) and reverse primer BGH Rvs_1

(TAGAAGGCACAGTCGAGG) which primes respectively at the CMV Forward priming site

and BGH Reverse priming site to produce a 1131 bp amplicon. Promising clones were

sequenced in both directions using CMV Fwd_1 and BGH Rvs_1, and clones of the right

sequence and orientation were used for subsequent experiments.

The inducible EGFP, ECFP, and EYFP plasmids thus created were named

pcDNA4/TO/EGFP/myc-His A, pcDNA4/TO/ECFP/myc-His A and

pcDNA4/TO/EYFP/myc-His A, abbreviated respectively as iEGFP, iECFP, and iEYFP.

3.3.2 PLG induction via addition of Tetracycline

TetR represses the transcription of PLG cloned into pcDNA4/TO/myc-His by forming a

homodimer that binds with extremely high affinity to each tetracycline operator (TetO$_2$)

sequence (two of which were embedded in the promoter of the PLG in pcDNA4/TO/myc-His). In the presence of tetracycline (Tet) each TetR binds one molecule of Tet, which induces a change in conformation of the TetR homodimer that causes it to lose the ability to bind TetO$_2$, dissociate from TetO$_2$ and thus allows normal transcription of the PLG. Addition of Tet thus induces PLG expression and could be provided by addition to cell culture medium at the final medium concentration of 1 μg/ml.

100 mg of Tet was thoroughly dissolved in 100 ml of sterile water and stored at -20 °C, with aluminium foil wrapping to protect from light. To prevent degradation via multiple freeze thaw cycles, 1 ml aliquots were made and used for each experiment. For induction, old medium was removed and 1 μl of Tet (1 mg/μl) was added to each ml of fresh medium used. Tet inductions were performed at least one day after transfection of inducible gene system, and the effects were assayed at least one day after induction.

3.4 RNA methods

3.4.1 Total RNA Isolation from HEK 293 cells

HEK 293 cells in 6 well plates were harvested for RNA isolation by trypsinisation. Briefly, media was aspirated and cells were washed with 1X PBS. Cells were incubated in 0.5 ml T/E at 37 °C for 1 min, and then neutralised with 1.5 ml of culture medium. Cells were pipetted up and down 10 times with 1 ml Gilson pipette to facilitate dissociation. Cells were transferred to 2 ml microcentrifuge tubes and pelleted for 3 minutes at 300 g. Cells were resuspended in 1X PBS and repelleted for 3 minutes at 300 g. Supernatants were removed and cells were stored

at -80 °C.

On the day of RNA isolation, pelleted cells were thawed and loosened by gentle flicking. RNA was isolated, using RNeasy Mini Kit (Qiagen), essentially following the manufacturer's microcentrifuge protocol for "Purification of Total RNA from Animal Cells Using Spin Technology". 10 μl of β-Mercaptoethanol was added to every ml of Buffer RLT prepared for use. Cells in each tube were homogenised by adding 600 μl of Buffer RLT and passing through a 22½ gauge needle at least 10 times, until homogenous lysate was obtained. After the washing steps with Buffer RPE, each RNeasy spin column was placed in new 2 ml collection tube and centrifuged at full speed for 1 minute to prevent carry over of Buffer RPE. RNA was eluted with 50 μl RNase free water per column. RNA concentrations were measured by the NanoDrop ND-1000 full-spectrum (220-750nm) spectrophotometer, using 1 μl of RNA sample. RNA samples were stored at –80 °C.

3.4.2 RT-PCR

Two steps conventional RT-PCR was performed to confirm the expression of HH-2-1_tRES in transfected HEK 293 cells, using β-Actin (βA) as a positive control. Reverse transcription was performed using Superscript II Reverse Transcriptase (RT) from Invitrogen.

3.4.2.1 1st strand cDNA synthesis

A 10 μl 1st strand cDNA synthesis reaction mix was set up as follows. First, a 6 μl reaction mix was set up as follows: DEPC treated water to top up to 6 μl, 1 μl each of 5 μM gene specific primers HH-2-1 tRES Rvs L1 (for HH-2-1 tRES) and Actin_B Rvs (for βA), 0.5 μl

dNTPs (10 mM), and 2.5 μl total RNA (300 to 1725 ng) or DEPC treated water for RT-PCR

samples and water control respectively. Primer information was as shown in **Table 3.4-1**. This

reaction mix was denatured at 65 °C for 5 minutes, followed by immediate incubation on ice.

To this mix was added: 2 μl 5X First-Strand Buffer, 1 μl 0.1 M DTT, and 0.5 μl DEPC treated

water. This reaction mix was incubated at 42 °C for 2 minutes. Immediately after, 0.5 μl of

SuperScript II RT (Invitrogen) or DEPC treated water was added for RT-PCR samples and

"minus RT" (-RT) control respectively. This final 10 μl reaction mix was incubated at 42 °C

for 50 minutes for 1st strand cDNA synthesis and heat inactivated at 70 °C for 15 minutes. 1st

strand cDNAs were stored at –20 °C.

3.4.2.2 PCR

PCR of HH-2-1_tRES and βA was performed with each 12.5 μl PCR reaction mix set up as

follows: Deionised H$_2$O to top up to 12.5 μl, 2.6 μl 5X PCR Buffer B (with 10 mM MgSO$_4$),

0.52 μl dNTP mix (5 mM each), 0.26 μl Elongase Enzyme Mix, 0.52 μl of each primers (5

μM), and 0.5 μl cDNA (as obtained above). Primer information was as shown in **Table 3.4-1**.

Thermocycling conditions were: Initial denaturation at 94 °C for 30 seconds, 30 cycles of

amplification (denaturation at 94 °C for 30 seconds, annealing at 55.5 °C for 30 seconds, and

extension at 68 °C for 39 seconds), and a final extension at 68 °C for 5 minutes.

The expected product sizes were 134 bp for HH-2-1_tRES and 643 bp for βA. RT-PCR

products were analysed by agarose gel electrophoresis in 1.5% gel at 120 V for about 20

minutes.

Table 3.4-1 Primers used in RT-PCR.

Primers	Role	Sequence (5' to 3')	Length (nt)
HH-2-1_tRES Fwd L1	HH-2-1_tRES forward primer.	ACCGTTGGTTTCCGTAGTGTAGTGGTTAT CACGTTCGCCTAACACGCGA	49
HH-2-1_tRES Rvs L1	HH-2-1_tRES reverse primer.	AAAATATGAAACTTTCGGCCTTTCGGCCT CATCA	34
Actin_B Fwd	βA forward primer.	GACCCAGATCATGTTTGAG	19
Actin_B Rvs	βA reverse primer.	AGGAGGAGCAATGATCTTG	19

3.4.3 Ribozyme Cis-cleavage Assays

For the initial designs of e-SRS sensors, RNA sensor constructs were designed to self-cleave, i.e. cis-cleave. This process can occur either 1) without trigger, known as RCA, as in the cases of T1a, HP-LR-CS, and DT-LR-CS; or 2) with an allosteric RNA trigger, known as Ribozyme Cis-cleavage Allosteric Assay (RCAA)as in the case of T1b.

3.4.3.1 In Vitro Transcription with RCA or RCAA

In Vitro Transcription (IVT) and RCA or RCAA processes occurred concurrently with the species incubated at 37°C for various durations. All RCA test constructs underwent IVT and RCA with no other RNA species in the same environment. T1b and T1b(-) underwent IVT and RCAA together with sNTS-37a. IVT DNA templates were gel purified PCR products except for sNTS-37a and sNTS-37b, which were Annealed Oligonucleotides, obtained as described in Gene Synthesis via Oligonucleotide Ligation (Section 3.1.9).

IVT was performed using the MAXIscript Kit from Ambion, essentially following the manufacturer's protocol. A 6 μl IVT reaction mix was set up as follows: DEPC treated or

Nuclease free water to top up to 6 μl, 0.6 μl 10X Transcription Buffer (9 U), 1.2 μl NTP mix (10 mM each), 0.6 μl T7 Enzyme Mix, and 1 to 2 μl IVT DNA template (to final concentrations of ~5 nM for T1a, C1a, C1a(-), T1b and T1b(-), ~16 nM for sNTS-37a and sNTS-37b, ~3 nM for HP-LR-CS, ~9 nM for DT-LR-CS).

IVT mixes were incubated for about 3 hours at 37 °C. Reactions were terminated by addition of an equal volume of 2X Gel Loading Buffer II provided in the MAXIscript Kit (95% Formamide, 0.025% xylene cyanol, 0.025% bromophenol blue, 18 mM EDTA, 0.025% SDS) and stored at -20 °C.

3.4.3.2 Denaturing Polyacrylamide Gel Electrophoresis (D-PAGE)

The end products of RCA and RCAA were analysed on Denaturing Polyacrylamide Gel Electrophoresis (D-PAGE). Before loading into gel, the products (with Gel Loading Buffer II added) were denatured by incubation at 95 °C for 3 mins, and then quickly cooled on ice (4 °C).

D-PAGE was conducted at in 1X TBE (Tris, Boric Acid, EDTA) gel running buffer, at 100 V for about 100 minutes, using 1.5, 0.75, and 0.5 mm 15% polyacrylamide TBE Urea (7 M) denaturing gels. A stock D-PAGE gel casting solution of 7 M Urea, 15% polyacrylamide was prepared by dissolving urea (42 g per 100 ml) in deionised water at 60 °C with magnetic stirring for about 50 minutes, followed by addition of 0.1 volume of 10X TBE, 0.375 volume of 40% polyacrylamide solution, and deionised water to top up to final volume. For casting of 1.5 mm gels, 10 ml gel casting solution was used, while 0.75 and 0.5 mm gels used 5 ml gel

casting solutions. Gels were solidified by addition of 4 μl TEMED and 0.01 volume 10% APS (100 mg ammonium persulfate in 1 ml deionised water).

To improve gel solidification, gels were allowed to set for about 45 minutes at room temperature, and then left over night at 4 °C, covered with lens paper soaked in 1X TBE and wrapped in Saran wrap.

After D-PAGE, gels were stained in SYBR Gold for about 45 to 55 minutes with gentle rocking and visualised under UV. SYBR Gold staining solutions were prepared by diluting the stock 10,000X SYBR Gold stain from Invitrogen into 1X concentration in 1X TBE, within a plastic container protected form light. Staining solutions were kept at 4 °C and reused several times as long as staining results were satisfactory.

3.4.4 RTA for Mz-based e-SRS

For the Mz based e-SRS sensors, RNA sensor constructs were designed to cleave a different RNA strand, i.e. trans-cleave, upon activation. This process of trans-cleavage can occur either with or without triggers (e.g. the control HH ribozymes), and was tested in RTA.

RTA protocols were modified from Mz [Tanabe et al, 2000]. Basically, Rz (either Mz or HH) were incubated with the substrate (STS), and activator, at 37 °C in RTA buffer, and results were monitored over time. The activator could be NTS, Nucleic acid Non-targeted Sequence (NNS, typically an RNA sequence that is close to the NTS, and needs to be distinguished from the NTS by e-SRS), or nothing (i.e. water).

3.4.4.1 RS RTA

All RS used for fluorometer measurement were labelled at 5' with 6-FAM, and at 3' with BHQ1.

3.4.4.1.1 RTA for Mz-1, Mz-2, Mz-3, and Mz-4

For the cases of Mz-1, Mz-2, Mz-3, and Mz-4, RTA buffer was 50 mM Tris (pH 8.0), 10 mM MgCl$_2$. Concentration of reactants were 0.2 μM Mz or HH, 2 μM NTS/NNS, 1 μM STS (RTS-1, RTS-2, RTS-3, RTS-4 respectively). 5 μl aliquots would be manually removed at intended time points. For each aliquot, RTA was terminated quickly by addition of 50 μl of 50 mM EDTA and incubation on ice. Aliquots were placed in 96 well plates and RTA results were assessed by fluorometer reading via the Tecan GENios Plus. Excitation and emission filter wavelengths were set at 485 nm and 520 nm respectively.

3.4.4.1.2 RTA for Mz-based e-SRS sensors for Dengue and Malaria

For the remaining Mz-based e-SRS sensors, RTA were assessed by the fluorometer Tecan Infinite M200. 50 μl RTA reaction mixes were prepared in 96 well plates. 25 μl per reaction master mix(es) of STS and RTA buffer was prepared on ice and protected from light, and quickly dispensed into the remaining RTA components (25 μl) already prepared in 96 well plates. Upon constitution of RTA mixes, the 96 well plate was protected from light and quickly placed within the Tecan Infinite M200, which had been prepared and warmed to 37 °C, so that measurements could start quickly. Measurement intervals were about 20 to 30 seconds and the RTA would typically span about 1.5 to 3 hours, until the fluorescent readings

had peaked. Other parameters used were as shown in **Table 3.4-2**.

Table 3.4-2 RTA parameters for Tecan GENios Plus.

Parameter	Value
Mode	Fluorescence Top Reading
Excitation Wavelength	490
Emission Wavelength	525
Excitation Bandwidth	9
Emission Bandwidth	20
Gain	100
Number of Flashes	3
Integration Time	20

RTA buffer contained 50 mM Tris (pH 8.0), and concentrations of reactants included 0.2 μM

Mz or HH, and 1 μM STS (RTS-2). The remaining conditions were as shown in Table 3.4-3.

Antisense oligonucleotides (AS) were used for NASBA RTA (Section 3.4.4.2).

Table 3.4-3 RTA conditions for Mz-based e-SRS sensors for Dengue and Malaria.

Mz	NTS type	MgCl$_2$ concentration	NTS/NNS concentration	Other components
D1, D2, D3, D4.	sNTS	10 mM	0.2 μM	Nil.
Mfs	sNTS	2 mM	0.2 μM	Nil.
Mfr1	sNTS	3 mM	0.2 μM	2 μM sNTS-Mfr2.
Mfr2	sNTS	3 mM	0.2 μM	Nil.
Mfs	NASBA NTS	2 mM	3 μl NASBA NTS	AS.
Mfr1	NASBA NTS	3 mM	3 μl NASBA NTS	AS, 2 μM sNTS-Mfr2.
Mfr2	NASBA NTS	3 mM	3 μl NASBA NTS	AS.

3 μl of NASBA NTS with 15 μM of each AS (Mal Rvs 194-215 for Mfr2 / Mal Rvs 206-214 for Mfs and Mfr1, and Mal Rvs 242-271 for all) in a 5 μl volume was denatured at 70 °C for 2 mins followed by quick chill on ice. This 5 μl mix was then used for NASBA RTA.

3.4.4.2 NASBA RTA

3.4.4.2.1 **Genomic DNA**

Genomic DNA of 3 Plasmodium falciparum strains as shown in **Table 3.4-4** were kindly

provided by Dr Cheng Qin, Laboratory Head of Molecular studies, Australian Army Malaria

Institute, Brisbane.

Table 3.4-4 Plasmodium falciparum strains from which genomic DNA were obtained.

pfcrt haplotypes	Pf Strain	Chloroquine	Denoted here as
SVMNT	S99	Resistant	Mfr1
CVIET	Dd2	Resistant	Mfr2
CVMNK	3D7	Sensitive	Mfs

3.4.4.2.2 NASBA reaction

NASBA reactions were performed using reagents from the NucliSENS EasyQ Basic Kit v2 from bioMérieux. Reactions were scaled up or down as needed proportionately.

A NASBA Enzyme Solution was prepared as follows: A red Enzyme Sphere was mixed with 45 μl of Enzyme Diluent and incubated at room temperature for 5 minutes. The solution was then mixed by tapping and spun down. A Reagent Premix was prepared as follows: A blue Reagent Sphere was mixed with 64 μl of Reagent Sphere Diluent and vortexed until solution cleared. If solution remained cloudy, it was incubated at 65 °C for 5 minutes and vortexed again. Solutions with precipitates were not used. Both the Reagent Premix and Enzyme Solution were sufficient for 16 NASBA reactions with 10 μl reaction volumes. They were aliquoted as required for 1 freeze-thaw cycle each within 2 weeks and stored at -80 °C quickly after preparation. NASBA KCl Water Mix (80 mM KCl) was prepared as follows: 16 ul KCl Stock Solution was mixed with 14 μl NASBA Water (nuclease free) and stored at -20 °C for multiple uses.

A 10 μl NASBA reaction mix was set up as follows. First, a 7 μl NASBA Primer-Reagent mix was set up as follows: NASBA Water to top up to 7 μl, 1.25 μl NASBA KCl Water Mix, 3.33 μl Reagent Premix, and 0.2 μl each of 10 μM NASBA primers Mal Fwd 03 and Mal Rvs 01.

Primer information was as shown in Table 3.4-5. The NASBA Primer-Reagent mix was treated like the Reagent mix until it was clear of precipitate (otherwise it was not used).

To NASBA Primer-Reagent mix (within 30 minutes of its preparation) was added: 0.5 µl NASBA template (about 6 x 10^9 copies). This reaction mix was incubated at 65 °C for 2 minutes, and 41 °C for 2 minutes. Immediately after, 2.5 µl Enzyme Solution was added to complete the 10 µl NASBA reaction mix. This was given a quick spin, mixed by tapping, given a quick spin again, and incubated at 41 °C for 90 minutes. NASBA products were stored at –80 °C.

Table 3.4-5 Oligonucleotides used in NASBA.

Oligonucleotides	Role	Sequence (5' to 3')	Length (nt)
Mal Fwd 03	NASBA forward primer.	AATTCTAATACGACTCACTATAGGGAGAA GGTATTTTAAGTATTATTTATTTAAGTGT A	59
Mal Rvs 01	NASBA reverse primer.	AAGTTGTGAGTTTCGGATGTTAC	23
Mal Rvs 194-215	5' AS for Mfr2.	CATACACTTAAATAAATAATAC	22
Mal Rvs 206-214	5' AS for Mfs, Mfr1.	ATACACTTA	09
Mal Rvs 242-271	3' AS for Mfs, Mfr1, Mfr2.	AACTATAGTTACCAATTTTGTTTAAA GTTC	30

3.4.4.2.3 NASBA product denaturation via antisense oligonucleotides (AS) action

NASBA products were used as NTS or NNS (known as nNTS) in NASBA RTA. For a 50 µl NASBA RTA, a 5 ul nNTS denatured with antisense oligonucleotides (AS) was used. Reactions were scaled up or down as needed proportionately.

A 5 µl nNTS was set up as follows: 3 µl NASBA product, 0.75 µl each of 100 µM 5' and 3'

AS (final concentration of 15 μM), and 0.5 μl of DEPC treated water. AS information was as shown in Table 3.4-5. This nNTS mix was denatured at 70 °C for 2 minutes, followed by immediate incubation on ice before use.

3.5 Bioimaging

Fluorescence microscopy was carried out with parameters as shown in **Table 3.5-1**. Live cells were viewed in their culture vessels using the Olympus FV500 or Olympus IX81. Cells grown on coverslips and mounted on microscope slides in FluorSave Reagent (from Calbiochem) were viewed using the Zeiss LSM.

Table 3.5-1 Excitation and emission parameters for fluorophores.

Fluorophores	Microscope	Excitation (nm)		Emission (nm)	
		Laser	Fluorophores	Filter	Fluorophores
ECFP	Olympus FV500	405	433, 453	ECFP	475, 501
	Zeiss LSM	405		LP 475	
EGFP	Olympus FV500	488	488	EGFP	507
	Olympus IX81	450 - 480		>515 (FITC)	
6-FAM	Olympus IX81	450 - 480	495	>515 (FITC)	520
EYFP	Olympus FV500	515	513	EYFP	527
Cy5	Zeiss LSM	633	650	LP 650	670
	Olympus IX81	545 - 580		>610 (WIY)	

3.6 Software used

RNAstructure 4.5 was used to predict secondary structure of RNA contructs. Perl 5.8.8 was used to implement the automation of the design of e-SRS Mz-based sensors.

CHAPTER 4   RESULTS

4.1 Initial designs of e-SRS.

Section 4.1 elaborates on the original designs and motivation of e-SRS.

4.1.1 Design Overview of Environment-Sensing Induced Gene Expression (e-SIGE)

e-SIGE aimed to achieve similar objectives (as an environment sensing response system) as

e-SRS, but was more complex in its physical implementation. e-SIGE was a generic platform

that could be further distinguished into such forms as r-SIGE, p-SIGE, m-SIGE, l-SIGE,

s-SIGE, and d-SIGE that sense RNA (and thus potentially DNA) sequence, protein,

metabolite, lipid, sugar, and drug, respectively. In this work, e-SIGE is taken to refer

essentially to r-SIGE. As detailed in the 2005 Invention Disclosure to ETPL (Invention

Disclosure, ETPL, 2005), e-SIGE is an environment sensing gene expression platform that

was designed to achieve the following objectives:

1) e-SIGE plasmid can be transfected into all cells of any tissues without problem of

    side effects in non-targeted cells. No specific delivery to any cell was needed because

    e-SIGE construct discriminate cells with unique target RNA sequence(s).

2) If e-SIGE sensor identifies the specified unique target RNA sequence(s), it can

    initiate expression of engineered genetic system.

3) If e-SIGE sensor does not identify any specified unique target RNA sequence(s), it

can initiate auto-degradation of the e-SIGE plasmid.

4) e-SIGE plasmid can also be degraded as desired by administering drug or agents

Within the e-SIGE sensor is one or more RNA domain(s) known individually as a Environment-sensing induced Conformation changing domain (EC), which couples the sensing of specified component(s) (essentially RNA in this work) of the intracellular environment with the initiation of e-SIGE sensor conformational change. Due to the ability of RNA to be potentially designed to bind specifically almost any biomolecular compound (Hermann, T. and Patel, D.J., 2000), ECs can potentially be created to sense for virtually any kind of biomolecule, including (but not limiting to) nucleic acids (RNA or DNA) sequences, proteins, peptides, metabolites, lipids, sugars, synthetic small molecular compounds, such as drugs, and dyes, etc.

Of the five embodiments of e-SIGE described in the Invention Disclosure, this work focused on the first embodiment, where the sequence to be detected, the NTS activates RNA silencing via the e-SIGE sensor known as the Integrated RNA silencing Construct (IRC).

To summarise, e-SIGE (like e-SRS) has two main components: 1) An RNA based sensor (the IRC in this case) that changes conformation upon binding of specific NTS; 2) A Response System that is triggered by the activated sensor to initiate some physical response, such as driving the expression of an exogenous gene known as the PLG or emitting a signal (e.g. fluorescence) to indicate the presence of the NTS. Both components can be encoded in plasmids that are transfected into cells. For simplicity, both components are encoded in a plasmid called the e-SIGE plasmid.

4.1.1.1 Mechanism of e-SIGE activation

The mechanism of e-SIGE activation is summarized in **Figure 4.1-1** and can be described as follows: (1) Upon transfection of the e-SIGE plasmid into a cell, IRC is constitutively transcribed and is present in the cell in an inactive conformation. IRC consists of various domains, including Ribozyme Catalytic (RC) domains and Cleavage Site (CS). Each RC is a Ribozyme Catalytic domain that is able to recognize and cleave at its specific sequence, i.e., its own CS. Each Cleavage Site harbours a particular sequence that its RC can recognize and cleave at only when it is single stranded (i.e. when IRC is in active conformation). (2) A repressor (R1) is also constitutively transcribed and inhibits the transcription of the PLG(s). (3) Upon recognition of NTS based on (but not limiting to) Watson-Crick complementary base pairing principle, IRC is activated via conformational changes, which involves two sequential RNA cleavage events. (4) RC_2 first cleaves at CS_2, releasing the <u>A</u>ntisense and <u>R</u>ibozyme <u>R</u>NA (ARR). ARR mediates silencing of R1 via antisense and ribozyme cleavage. (5) RC_1 next cleaves at CS_1, releasing the <u>s</u>hort <u>h</u>airpin si<u>R</u>NA (SHR). This short hairpin siRNA mediates silencing of R1 via RNAi pathway. (7) Upon the knockdown of R1, PLG is expressed.
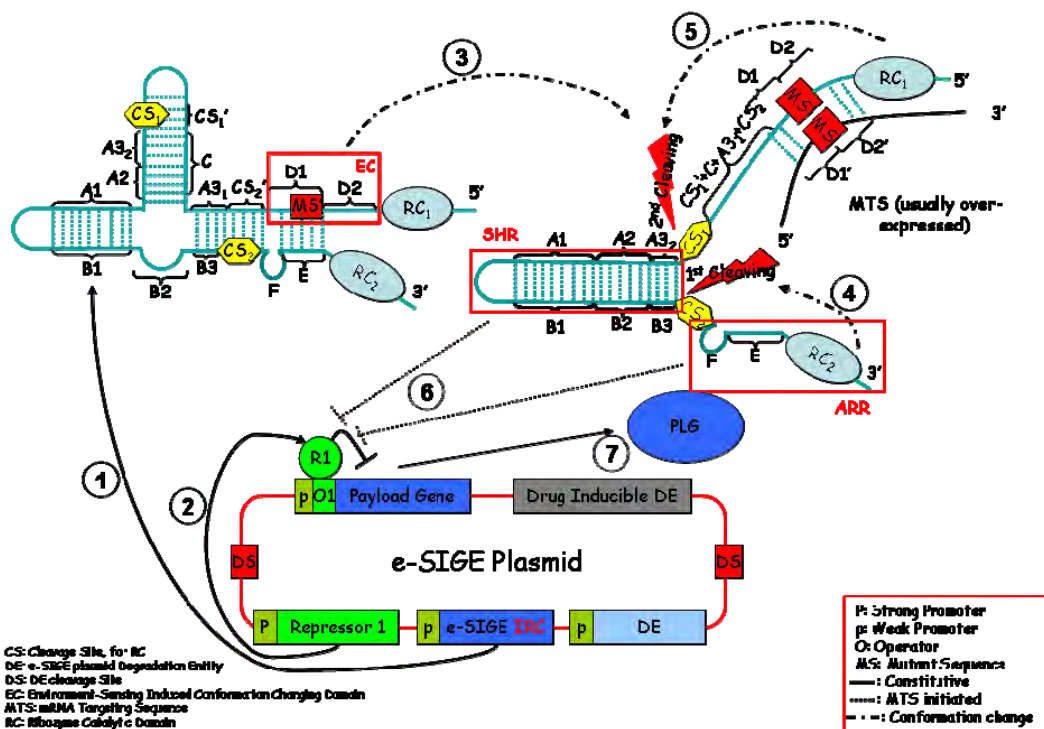
Figure 4.1-1 Mechanism of e-SIGE (taken from Innovative Grant application [SBIC Innovative Grant, 2005].

NTS activates RNA-based sensor to initiate expression of PLG in 7 steps. (1) Sensor is constitutively transcribed and exists in an inactive conformation. (2) R1 is constitutively transcribed and inhibits the transcription of the PLG. (3) Recognition of NTS results in an activation of sensor via conformational changes, which triggers two sequential cleavage events. (4) RC_2 first cleaves at CS_2, releasing the ARR. (5) RC_1 next cleaves at CS_1, SHR. (6) SHR and ARR mediate silencing of R1. (7) In the absence of R1, PLG is expressed.

4.1.1.2 Extensibility of e-SIGE

Due to the modular nature of e-SIGE activation, the e-SIGE sensor and thus the entire platform can be extended in a modular nature as illustrated in Figure 4.1-2. Each EC is a specific sequence of RNA that can sense a specific bio-molecule, e.g. EC1 can sense mRNA, EC2 can sense Protein. Such combinatorial e-SIGE requires all of EC1, EC2, and EC3 to sense their targets (independently) to activate the silencing domains comprising of ARR1, ARR2, ARR3.
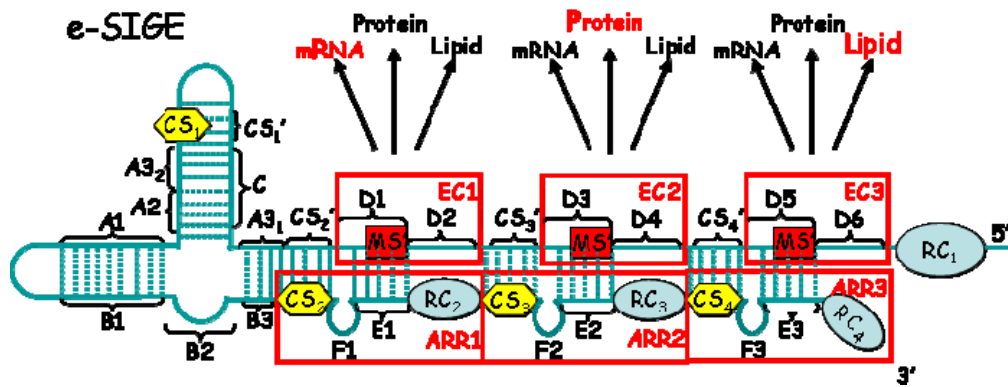
Figure 4.1-2 Extensibility of e-SIGE.

4.1.2 Structure of e-SIGE components

4.1.2.1 RNA Segments of e-SIGE components

Referring to Figure 4.1-1, the structure of e-SIGE components can be broken down into modular segments (and sub-segments) of RNA as follows:

Segment A – Sense sequence of the RTS_I segment of the RNA silencing Target Sequence (RTS) (Found in Inactivating Repressor (R1) and e-SIGE plasmid Degradation Entity (DE)) (no homolog in Homo sapiens transcriptome).

1) Segment A1 – 3' part of A, ~ 10 bp

2) Segment A2 – Middle part of A, ~ 5 bp

3) Segment $A3_1$ (=$A3_2$) – 5' of C, ~ 4 bp

4) Segment $A3_2$ (=$A3_1$) – 5' part of A2, ~ 4 bp

Segment B – Antisense sequence of STS (no homolog in Homo sapiens transcriptome).

1) Segment B1 – 5' part of B, ~ 10 bp

2) Segment B2 – Middle part of B, ~ 5 bp

3) Segment B3 – 3' part of B, ~ 4 bp

4) Segment B2 + B3 = C

Segment C – Reverse complement of $A3_2$ + A2.

Segment D (NTS') – Reverse complement sequence of NTS.

1) Segment D1– 3' part of D, ~ 10 bp (which can have flexible length)

2) Segment D2 – 5' part of D, ~ 10 bp (which can have flexible length)

Segment E – Reverse complement of D1, found on the 3' part of NTS.

Segment F – Reverse complement of RTS_A (F'), which is found on the RTS (which can have flexible length). This binding of segment F to RTS_A can mediate antisense silencing of the RTS containing gene.

NTS - Nucleic Acid Target Sequence.

Segment D1'– 5' part of NTS, ~ 10 bp (which can have flexible length)

Segment D2' – 5' part of NTS, ~ 10 bp (which can have flexible length)

RTS - RNA silencing Target Sequence.

1) RTS_I (= A) – 3' of RTS, ~ 19 bp

2) RTS_A (= Reverse complement of F) – Middle part of RTS, may be part of the RTS_I (flexible length, no homolog in Homo sapiens transcriptome).

3) RTS_C – 5' of RTS, serves as recognition and cleavage sequence of RC_2.

4) RTS_{I-A} – Linker sequence between the RTS_I and RTS_A (which can have flexible length).

5) RTS_{A-C} – Linker sequence between RTS_A and RTS_C (this length may affect RC_2 catalysis efficiency).

Segment MS – Mutant Sequence within NTS that is specific to the mutant copy and not found in the wild type copy of the targeted mRNA. It could be more than one nucleotide in length, but is usually one nucleotide.

Segment MS' - Reverse complement of MS, could exist at the middle or 3' end of D2, or 5' end of D1.

Segment RC - Ribozyme Catalytic (RC) domain that is able to recognize and cleave at a specific sequence. There are two RC within IRC: RC_1 and RC_2, which may or may not bear the same sequences. RC_1 exists at the 5' end of the construct, while RC_2 exists at the 3' end of the construct.

Segment CS – Cleavage Site (CS) that harbours a particular sequence that the RC can recognize and cleave at only when it is single stranded (i.e. when IRC is in active conformation). There are two CS within IRC: CS_1 and CS_2, which may or may not bear the same sequences, and are recognized by RC_1 and RC_2, respectively. CS_1 is located at the 5' of A3 (while remaining on the same strand of the hairpin stem), paired with the complementary CS_1' on the opposite stem. CS_2 is located at the 3' of B3, 5' of F, paired with the complementary CS_2' on the opposite stem.

4.1.2.2 Complementarities of Segments

The following segments exhibit Watson-Crick complementary base pairing:

1) $A \leftrightarrow B$

2) $A1 \leftrightarrow B1$

3) $A2 \leftrightarrow B2$

4) $A3 \leftrightarrow B3$

5) $(A3 + A2) \leftrightarrow C$

6) $D1 \leftrightarrow E$

7) $CS1 \leftrightarrow CS1'$

8) $CS2 \leftrightarrow CS2'$

4.1.3 Design of RNA folding mechanism in e-SIGE

4.1.3.1 e-SIGE Sensor (IRC) Conformation upon Synthesis

As mRNA synthesis occurs from 5' to 3', the segments will be synthesized in the following order: RC_1, D2, D1, CS_2', $A3_1$, C, CS_1', CS_1, $A3_2$, A2, A1, B1, B2, B3, CS_2, F, E, and RC_2. C will be synthesized before B2 and B3, while A2 and $A3_2$ will be synthesized in between. Although the sequences of C and B2 + B3 are identical, $A3_2$ + A2 will base-pair with C first, leaving B2 unpaired, and B3 pair to $A3_1$ immediately after synthesis. Since D1 is also paired with E, CS_2 is not available to the RCs as well. Hence, r-SIGE IRC is inactive upon synthesis.

4.1.3.2 Proposed RNA folding mechanism for NTS activation of the e-SIGE IRC into functional siRNA

The activation process of IRC can be structured into the following RNA folding steps.

1) Activation of EC.

   Presence of NTS provided specifically by the targeted (possibly mutant or specific splice variant) mRNA will allow first D2 and then D1 (after successful competition

with E) to base-pair with NTS, forming a transient [IRC – NTS] hetero-duplex complex T1. The formation of T1 releases F and E into an unbound conformation that hinges at CS_2.

2) Cleavage of CS_2 by RC_2.

Conformation of T1 exposes CS_2 as the 3' end of a double stranded region, which thus increases significantly the probability that CS_2 will spontaneously unpair from CS_2'. This can be facilitated by introduction of mismatches between CS_2 and CS_2'. Once this occurs, there is a chance that RC_2 will recognize and cleave the single stranded CS_2, forming the transient IRC complex T2. The formation of T2 results in the release of Segments F, E, and RC_2, as the Antisense and Ribozyme RNA (ARR).

ARR binds to the RTS_A segment of RTS via Segment F. This binding alone can mediate antisense silencing of the RTS containing gene. In addition, as the RC_2 recognition and cleavage site, RTS_C has been engineered into RTS, the RC_2 domain of ARR can also mediate the silencing of the RTS containing gene by cleaving at RTS_C.

3) Formation of A-B hairpin stem (consisting of A3, A2, A1, and B1, B2, B3 on each strand).

The formation of T2 exposes B3 as the 3' end of a double stranded region (as well as of the present construct), and thus increases significantly the probability that B3 will spontaneously unpair from $A3_1$. This can be facilitated by introduction of mismatches between B3 and $A3_1$. Once this occurs, there is a chance that the free B2 + B3 will

compete successfully with C for pairing with A3 + A2, forming the transient IRC

complex T3, which contains the A-B hairpin stem. This can be facilitated by

introduction of mismatches between A3 + A2 and C.

4) Cleavage of CS_1 by RC_1.

The formation of T3 forms the A-B hairpin stem that hinges at CS_1. This exposes

CS_1' as the 5' end of a double stranded region, which thus increases significantly the

probability that CS_1 will spontaneously unpair from CS_1'. This can be facilitated by

introduction of mismatches between CS_1 and CS_1'. Once this occurs, there is a

chance that RC_1 will recognize and cleave the single stranded CS_1, releasing the A-B

hairpin stem, which is a short hairpin siRNA that will mediate the RNAi knockdown

of the RTS containing gene.

4.1.4 IRC e-SIGE segment sequences.

In order to satisfy RNA folding requirements of IRC, the conditions as shown in Table 4.1-1

were imposed for generating sequences for the various RNA segments of IRC e-SIGE listed in

Section 4.1.2.1. Several variants were planned for most segments in order to allow for RNA

folding optimisation.

Table 4.1-1 Conditions for IRC e-SIGE segments.

Segments	Conditions
A1, A2, A3 (as 1 segment)	Random sequence (fixed 60% GC content)
	3 variants.
	20 nt in length.
	Blasted against Human, Mouse, and Rat Refseq to ensure no matches.

Segments	Conditions
B3	2-5nt in length.
	4 variants.
C	0,1,2 opening mismatches
	3 variants.
E, D2	Sequence determined by NTS.
	2 variants.
	E: 5, 8nt in length.
	D2: 15, 12nt in length.
F	Random sequence (fixed 60% GC content)
	3 variants.
	12, 15, 17nt in length.
	Blasted against Human, Mouse, and Rat Refseq to ensure no matches.
NTS	Cav2.1 alternatively spliced sequences (Exon 37a or 37b) from Human, mouse and rat.
CS (1 and 2)	Ribozyme cleavage signal flanked by specific sequences.
	3 variants per ribozyme.
	5-7nt in length

Suitable segment sequences (with several variants in many cases) are shown in **Table A 2**.

Core ribozyme sequences and their consensus cleavage site sequences were assembled by consulting various references [Ananvoranich and Perreault, 1998; Bergeron et al, 2005; Berzal-Herranz et al, 1992; Chowrira et al, 1991; Chowrira et al, 1994; Fujitani et al, 1993; Hertel et al, 1996; Pérez-Ruiz et al, 1999; Quan, 1999; Wrzesinski, 2001].

4.2 *In vitro* testing of initial e-SRS constructs.

The first phase of development was to test the ability of the IRC to fold and activate as we designed, in the simple test tube environment. This section details the design and synthesis of e-SIGE IRC test constructs for in vitro tests.

4.2.1 e-SIGE test constructs required to show appropriate RNA folding in NTS activation of IRC.

4.2.1.1 Key elements in RNA folding steps in NTS activation of IRC.

As the IRC activation by NTS is a complex process, it was desirable to breakdown the activation process into individual RNA folding steps (as detailed in section 4.1.3.2) and identity crucial elements and design simpler test constructs for groups of these elements. By analysing the steps in section 4.1.3.2, two key elements for the IRC activation process was identified:

1) Cleavage of CS_1 and CS_2 had to occur readily if and only if the CS were single stranded (SS), and not if they were double stranded (DS).

   IRC activation steps have two irreversible check points given by the cleavage of CS_1 and CS_2. The paradigm would fail if these CS were to be readily cleaved even when they were DS (i.e., IRC not activated). Likewise, the CS had to be readily cleaved when they were SS (i.e., IRC activated).

2) NTS (but not NNS) must be able to activate the EC by displacing segment E, which must be sufficient to result in cleavage of CS_2, in order to start the whole process of activation.

4.2.1.2 e-SIGE IRC activation test constructs.

In order to verify the two key elements in IRC activation, two groups of test constructs were designed as follows. The design of test constructs created new segments based on the original

IRC segments. In order to keep a systematic list of segment names, the nomenclature in Table

4.2-1 was used.

Table 4.2-1 Nomenclature for segments.

Symbol	Meaning	Types	Examples
Underscore "_", as used in X_M.	Denotes a modification M of segment X.	X_R: Randomised sequence of X.	C_R means a segment with the randomised sequence of C.
		X_D: "Dead" CS sequence. CS with sequence modified such that it is not recognised by its Rz	CS1_D means CS1 sequence altered such that it is not cleaved by its RC1.
Dash "-", as in X-V.	Denotes a variant V of X.	X-1: Variant number 1 of X.	HP-1 means variant 1 of HP (there were 3 variants of HP, differing slightly in binding arm sequence.
		NTS-A: NTS sequence of A.	NTS-D1 means NTS sequence of D1.
		X-Rn/X-Mm/X-Hs: Segment X for species Rat, Mouse, Human respectively.	sNTS-37a1-Rn/Mm means sNTS-37a of variant 1a, representing both Rat and Mouse variants.
Prime """, as in X'.	Denotes reverse complemented sequence of X. Thus, X and X' can have perfect base-pairing.	X': Reverse complement of X.	CS2-1' means the reverse complement of variant 1 of CS2.
Vertical bar "|", as in X|Y|Z.	Denotes a field separator, to delineate consecutive segments, especially when its use could improve clarity.	X|Y|Z: Segments X, Y, Z joined consecutively in that order.	`CS|Linker|RC` means the segments `CS`, `Linker`, and `RC` joined consecutively in that order.

4.2.1.2.1 Aim 1a constructs to show that RC1 can cis-cleave CS1 if and only if CS1 was single stranded.

Aim 1a served to verify the first key element that CS must be readily cleaved if and only if it

was SS. Specifically, the constructs (as illustrated in Figure 4.2-1) were designed to show that

RC1 can cis-cleave CS1 if and only if CS1 was SS.
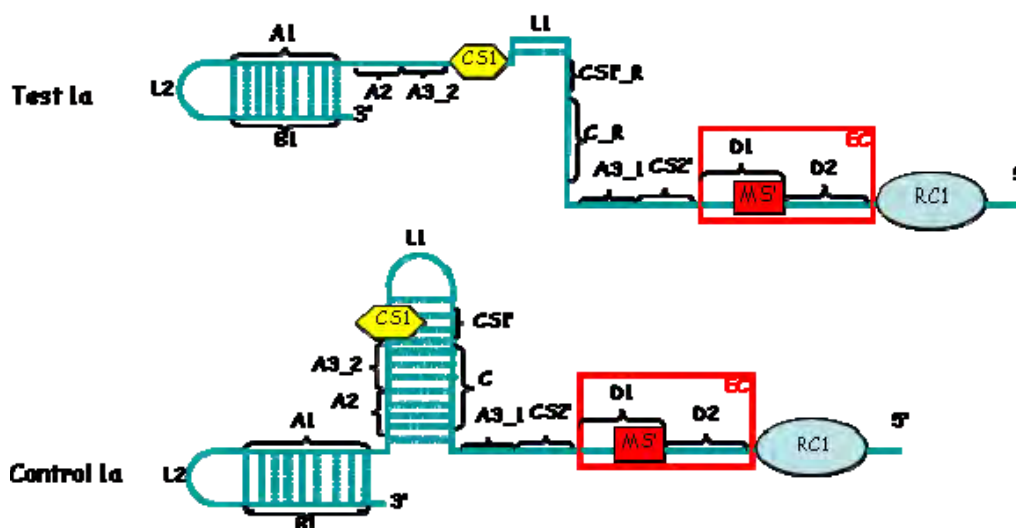
Figure 4.2-1: Illustration of Aim 1a constructs.
Aim 1a was to show that RC1 could cis-cleave CS1 if and only if CS1 was SS.

The list of Aim 1a constructs are as follows:

1) Test 1a (T1a)

This construct aimed to emulate the crucial parts of complex T3, where A3 + A2 has

been unpaired from C, in order to show that RC1 cleaves CS1 readily IF CS1 is SS.

In T3 CS1' was exposed as the 5' end of a double stranded region. The IRC activation

model assumed critically that such conformation of CS1 increased significantly the

probability that CS1 would spontaneously unpair from CS1', allowing RC1 to cleave

CS1 readily. T1a allowed a further simplification of the T3 model by allowing CS1 to

be SS by default. The expected result would be that T1a would be readily cleaved.

Note that for simplicity, segments B2 and B3 was not included.

T1a was able to emulate unpaired A3 + A2 as the pairing segment C in IRC had been

replaced with segment C_R, which being a randomised sequence of C would not base pair with A3 + A2. Likewise, CS1 would be SS by default as CS1' had been replaced by CS1'_R.

2) Control 1a (C1a)

This construct aimed to show that when CS1 is DS, it cannot be cleaved readily by RC1. It therefore would provide a basal level of cleavage in Test 1a and Control 1a due to self unwinding of the C-A3_2A2 stem, or unwinding starting at the middle of the loop.

This is possible as this construct is Test 1a but with A3+A2 paired and CS1 DS by default, as in the case of the unactivated IRC.

3) Control 1a(-) (C1a(-))

This construct aimed to show the basal level of "cleaved band" intensity due to systematic error, e.g. background intensity due to fragmented nucleotides.

This is possible as this construct is Control 1a with CS_D instead of CS, and would not be cleaved by RC1.

4) HP-LR-CS

This construct aimed to provide a +ve control to ensure reaction conditions are suitable for Hairpin (HP) ribozyme (Rz) cleavage.

This is possible as this construct contained only HP (hairpin Rz), LR (Linker: Rz), and CS (for HP, same as those used in T1a, C1a, C1a(-)). If reaction conditions were suitable, HP should cis-cleave readily at the nearby CS, producing two fragments

smaller than the original HP-LR-CS. This assay would also show how much of the

CS would be cleaved by a wild type HP in a given time, and allow us to judge the

cleavage efficiency of our test constructs.

Based on the experimental requirements outlined above, segments for Aim 1a constructs

were designed as given in **Table A 3**. Based on the Aim 1a constructs given in **Table A 3**, the

experiments and predicted results are outlined in Table 4.2-2.

Table 4.2-2 Experiments for Aim 1a constructs.

Lane No.	Constructs	Experiment	Expected Fragments	Fragment Size	Comments
1	Test 1a	IVT + RCA	Uncleaved	167	High % cleavage.
			5' Fragment	119	
			3' Fragment	48	
2	Control 1a	IVT + RCA	As Lane 1.		Low % cleavage.
3	Control 1a(-)	IVT + RCA	Uncleaved	167	Lowest % cleavage
					Serves as negative control.
4	HP-LR-CS	IVT + RCA	Uncleaved	72	Highest % cleavage.
			5' Fragment	63	
			3' Fragment	9	Serves as positive control.

IVT, In-vitro Transcription; RCA, Ribozyme Cis-Cleavage Assay.

4.2.1.2.2 **Aim 1b constructs to show that sNTS-37a but not sNTS-37b can activate EC and lead to cleavage of CS2.**

Aim 1b served to verify the second key element that NTS (but not NNS) must be able to

activate the EC by displacing segment E, which must be sufficient to result in cleavage of CS2,

in order to start the whole process of activation.

Specifically, the constructs (as illustrated in Figure 4.2-2) were designed to show that

sNTS-37a but not sNTS-37b (the NNS in this case) can trigger activation of EC by displacing

segment E and lead to cleavage of CS2.
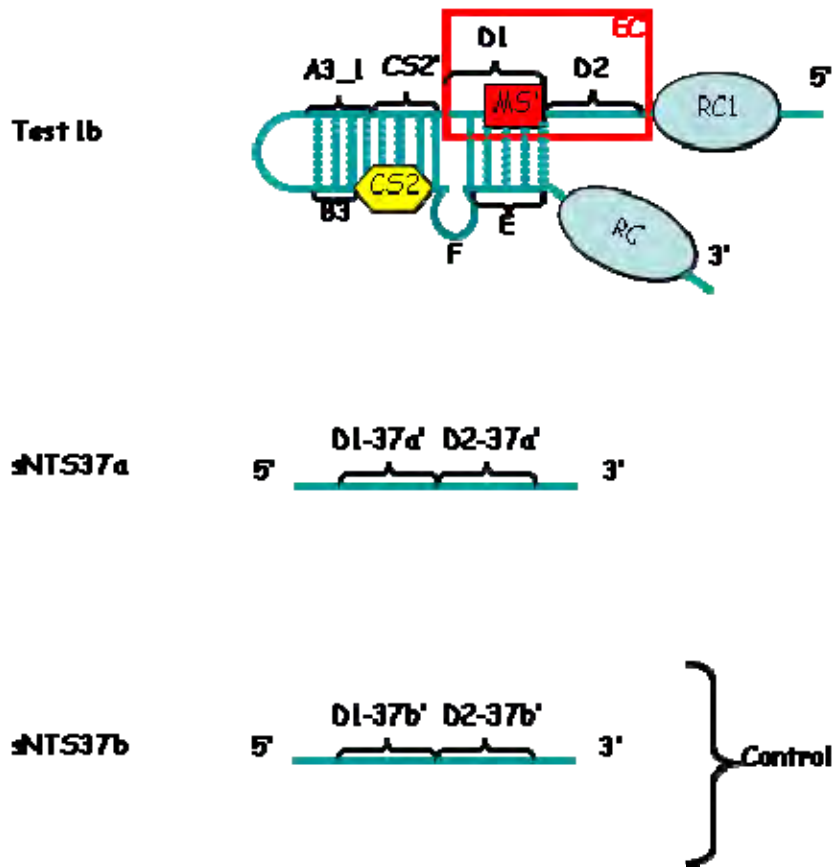
Figure 4.2-2: Illustration of Aim 1b constructs.

Aim 1b was to show that a short hairpin RNA (with RC1 to A3_1 on 5', and B3 to RC2 on 3')
would be cleaved if and only if sNTS was present.

The list of Aim 1b constructs are as follows:

1) Test 1b (T1b)

This construct aimed to emulate the crucial parts of IRC that were responsible for the

formation of complex T1, where 1) D2 and then D1 (after successful competition

with E) base-pair with NTS, 2) releasing F and E into an unbound conformation that

hinged at CS2 leading to cleavage of CS2 by RC2, in order to show that sNTS-37a

but not sNTS-37b can activate EC and lead to cleavage of CS2.

In T1, CS2 was exposed as the 3' end of a double stranded region. The IRC EC activation model assumed critically that such conformation of CS2 increased significantly the probability that CS2 would spontaneously unpair from CS2', allowing RC2 to cleave CS2 readily. If this assumption was true, the expected result would be that T1b would be readily cleaved if and only if sNTS-37a is present, and the same cannot be achieved by substitution of sNTS-37b. Note that for simplicity, segments between A3_1 and B3 (which were not responsible for the activation of EC) were not included.

This construct thus provides a simple way to test the formation of complex T2 and ARR due to EC activation by sNTS.

2) sNTS-37a

This construct aimed to emulate the NTS37a containing mRNA by representing a crucial 22 nt segment of it that differed slightly from NTS37b. It should be able to trigger unwinding of D1-E stem in Test 1b.

3) sNTS-37b

This construct aimed to emulate the NTS37b containing mRNA by representing a crucial 22 nt segment of it that differed slightly from NTS37a. It should not be able to trigger unwinding of D1-E stem in Test 1b.

4) Test 1b(-) (T1b(-))

This construct (used together with sNTS-37a) aimed to show the basal level of "cleaved band" intensity due to systematic error, e.g. background intensity due to

fragmented nucleotides.

This is possible as this construct is Test 1b with CS_D instead of CS, and would not be cleaved by RC2.

5) DT-LR-CS

This construct aimed to provide a +ve control to ensure reaction conditions are suitable for Delta (DT) ribozyme (Rz) cleavage.

This is possible as this construct contained only DT (<u>D</u>el<u>t</u>a Rz), LR (<u>L</u>inker: <u>R</u>z), and CS (for DT, same as those used in T1b, T1b(-)). If reaction conditions were suitable, DT should cis-cleave readily at the nearby CS, producing two fragments smaller than the original DT-LR-CS. This assay would also show how much of the CS would be cleaved by a wild type DT in a given time, and allow us to judge the cleavage efficiency of our test constructs.

Based on the experimental requirements outlined above, segments for Aim 1b constructs were designed as given in Table 4.2-3.

Table 4.2-3 Aim 1b construct segments.

Construct	Segment	Variants	Sequence
Test 1b (T1b)	RC1	HP-1	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
	D2D1	1a	TACATGTCCTTATAGTGAAT
	CS2'	DT-1	aaaccac
	A3_1	1	AA
	L3	1	TTCAAGAGA
	B3	1	TT
	CS2	DT-1	gtggttt
	F	1	CCAATACTTACGCGTCG
	E	1	ATTCACTATA
	RC2	DT-1	GGGTCCACCTCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTAAGGGACCC
Test 1b (-) (T1b(-))	RC1	HP-1	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
	D2D1	1a	TACATGTCCTTATAGTGAAT
	CS2_D'	DT-1	aaaggaa
	A3_1	1	AA
	L3	1	TTCAAGAGA
	B3	1	TT
	CS2_D	DT-1	ttccttt
	F	1	CCAATACTTACGCGTCG
	E	1	ATTCACTATA
	RC2	DT-1	GGGTCCACCTCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTAAGGGACCC
sNTS-37a1-Rn/Mm (sNTS-37a)	D1'	1a	ATTCACTATA
	D2'	1a	AGGACATGTA
sNTS-37b1-Rn/Mm (sNTS-37b)	D1'	1b	ATGCCTTACC
	D2'	1b	CGGACATGTA
DT-LR-CS	L3	1	TTCAAGAGA
	B3	1	TT
	CS2	DT-1	gtggttt
	LR	DT-1	AT
	RC2	DT-1	GGGTCCACCTCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTAAGGGACCC

Based on the Aim 1b constructs given in Table 4.2-3, the experiments and predicted results are

outlined in Table 4.2-4.

Table 4.2-4 Experiments for Aim 1b constructs.

Lane No.	Constructs	Experiment	Expected Fragments	Fragment Size	Comments
1	Test 1b sNTS-37a	IVT + RCAA	Uncleaved	183	High % cleavage.
			5' Fragment	92	
			3' Fragment	91	
			sNTS	22	
2	Test 1b sNTS-37b	IVT + RCAA	As Lane 1.		Low % cleavage.
3	Test 1b(-) sNTS-37a	IVT + RCAA	Uncleaved	183	Lowest % cleavage.
			sNTS	22	Serves as negative control.
4	DT-LR-CS	IVT + RCA	Uncleaved	79	Highest % cleavage.
			5' Fragment	13	Serves as positive control.
			3' Fragment	66	

IVT, In-vitro Transcription.

4.2.2 Synthesis of e-SIGE test constructs.

4.2.2.1 e-SIGE test constructs sequences.

The required Aim 1a and 1b constructs (Test 1a, Control 1a, Control 1a(-), HP-LR-CS, Test 1b, Test 1b(-), sNTS37a, sNTS37b, DT-LR-CS) were first constructed de novo (using gene synthesis via oligonucleotide ligation) as double stranded DNA templates with protruding 3' A and then inserted via TA-cloning into pGEM-T Easy vectors. As large quantity of RNA test constructs (longer than commercially available for synthesis) were required for RCA, each of the construct had an IVT promoter appended at 5' end so that their RNA transcripts could be produced in bulk via IVT. The 5' appended T7 RNA Pol III promoter minimum sequence (-17 to +2, -17 TAATACGACTCACTATA -1 +1 GG +2) was the minimum sequence required for efficient transcription. Transcribed RNA starts at +1, hence adding "GG" to 5' end of construct sequences. Sequences of the test constructs with IVT promoter added are given in Table 4.2-5. Construct sequences cloned into pGEM-T Easy vectors were verified by DNA sequencing using vector primers.

Table 4.2-5 Sequences of Aim 1a and 1b test constructs with 5' IVT promoter.

Constructs	Sequence (5' to 3')	Length (nt)
T1a	<u>TAATACGACTCACTATAGG</u>AAACAGagaaGTCAACCAGAGAAACACAC GTTGTGGTATATTACCTGGTAtacatgtccttatagtgaatAAACCAC AACTTGTACGGAGTCACCAAATAGGATTCAAGAGATGACagtcCTGTT TAATGGATCCCGTTCGGCAGCTTCAAGAGAGCTGCCGAAC	184
C1a	<u>TAATACGACTCACTATAGG</u>AAACAGagaaGTCAACCAGAGAAACACAC GTTGTGGTATATTACCTGGTAtacatgtccttatagtgaatAAACCAC AAGGGATCCATTAAACAGgactGTCATTCAAGAGATGACagtcCTGTT TAATGGATCCCGTTCGGCAGCTTCAAGAGAGCTGCCGAAC	184
C1a(-)	<u>TAATACGACTCACTATAGG</u>AAACAGagaaGTCAACCAGAGAAACACAC GTTGTGGTATATTACCTGGTAtacatgtccttatagtgaatAAACCAC AAGGGATCCATTAAACAGcgaaGTCATTCAAGAGATGACttcgCTGTT TAATGGATCCCGTTCGGCAGCTTCAAGAGAGCTGCCGAAC	184
HP-LR-CS	<u>TAATACGACTCACTATAGG</u>AAACAGagaaGTCAACCAGAGAAACACAC GTTGTGGTATATTACCTGGTAAGATCTTGACagtcCTGTTT	89
T1b	<u>TAATACGACTCACTATAGG</u>AAACAGagaaGTCAACCAGAGAAACACAC GTTGTGGTATATTACCTGGTAtacatgtccttatagtgaatAAACCAC AATTCAAGAGATTGTGGTTTccaatacttacgcgtcgATTCACTATAG GGTCCACCTCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTA AGGGACCC	200
T1b(-)	<u>TAATACGACTCACTATAGG</u>AAACAGagaaGTCAACCAGAGAAACACAC GTTGTGGTATATTACCTGGTAtacatgtccttatagtgaatAAAGGAA AATTCAAGAGATTTTCCTTTccaatacttacgcgtcgATTCACTATAG GGTCCACCTCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTA AGGGACCC	200
sNTS-37a	<u>TAATACGACTCACTATAGG</u>ATTCACTATAAGGACATGTA	39
sNTS-37b	<u>TAATACGACTCACTATAGG</u>ATGCCTTACCCGGACATGTA	39
DT-LR-CS	<u>TAATACGACTCACTATAGG</u>TTCAAGAGATTGTGGTTTATGGGTCCACC TCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTAAGGGACCC	96

<u>Underlined</u> sequence, Appended T7 promoter minimum sequence (-17 to +2) required for efficient transcription. Transcribed RNA starts at +1, hence adding "GG" to 5' end of construct sequences.

4.2.2.2 De novo synthesis of constructs DNA template via PCR.

We initially attempted to synthesise our required constructs via assembly PCR. Instructions

for assembly PCR were as given in [Rydzanicz et al, 2005]. A web-based tool

(http://publish.yorku.ca/~pjohnson/AssemblyPCRoligomaker.html) was used to design assembly PCR oligonucleotides. However, it turned out that DNA templates synthesized by this technique often contained missing regions. After many optimization attempts, including optimization of annealing temperatures, primer concentrations, assembly oligonucleotide sequences, addition of DMSO, and switch from the use of Taq polymerase (Promega) to use of high fidelity DNA polymerases (Elongase Enzyme Mix from Invitrogen and PfuTurbo DNA polymerase from Stratagene), we were still unable to synthesise proper products, and had to try a different approach.

The only success in DNA template synthesis via PCR was for HP-LR-CS. HP-LR-CS was purchased as a single stranded synthesized oligonucleotide without special purification. A forward primer (containing T7 promoter sequence) and reverse primer was used to amplify and construct the full length double stranded DNA template in two rounds of PCR using Taq polymerase (Promega). The list of oligonucleotides used in synthesis of HP-LR-CS is given in Table 4.2-6. The PCR product was gel purified and cloned into pGEM-T Easy vector, miniprep with DH10B, and sequenced for accuracy. However, attempts to use this method for T1a (even with PAGE purification of synthesized oligonucleotide) produced similar problems as in assembly PCR.

Table 4.2-6 Oligonucleotides used in synthesis of HP-LR-CS.

Oligonucleotide	Length (nt)	Sequence (5' to 3')
HP-LR-CS (SN)	70	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTG GTAAGATCTTGACagtcCTGTTT
HP-LR-CS Fwd	30	TAATACGACTCACTATAGGAAACAGagaaG
HP-LR-CS Rvs	25	AAACAGgactGTCAAGATCTTACCA

4.2.2.3 De novo synthesis of DNA templates of constructs using gene synthesis via oligonucleotide ligation.

Upon exhaustion of all available avenues to synthesise the DNA template via PCR based methods, we tried a new technique we termed "gene synthesis via oligonucleotide ligation". The designed constructs were broken into overlapping fragments (known as "ligation oligos" with overlaps of around 15 nt), with the 3' fragments having an extra 3' A, as shown in Figure 4.2-3 and Figure 4.2-4.

We were able to obtain clones with correct sequence for every construct. In addition, this technique yielded a large percentage of plasmids with correct inserts. As "gene synthesis via oligonucleotide ligation" totally avoided the process of PCR from incomplete templates (i.e. semi-double stranded DNA templates that have single stranded regions), as used in assembly PCR, we suspect that highly structure constructs like ours might be inappropriate for PCR from incomplete templates.

Ligation oligos and their sequences for the respective test constructs (with IVT promoter added) are given in Table 4.2-7 and **Table A 4**.
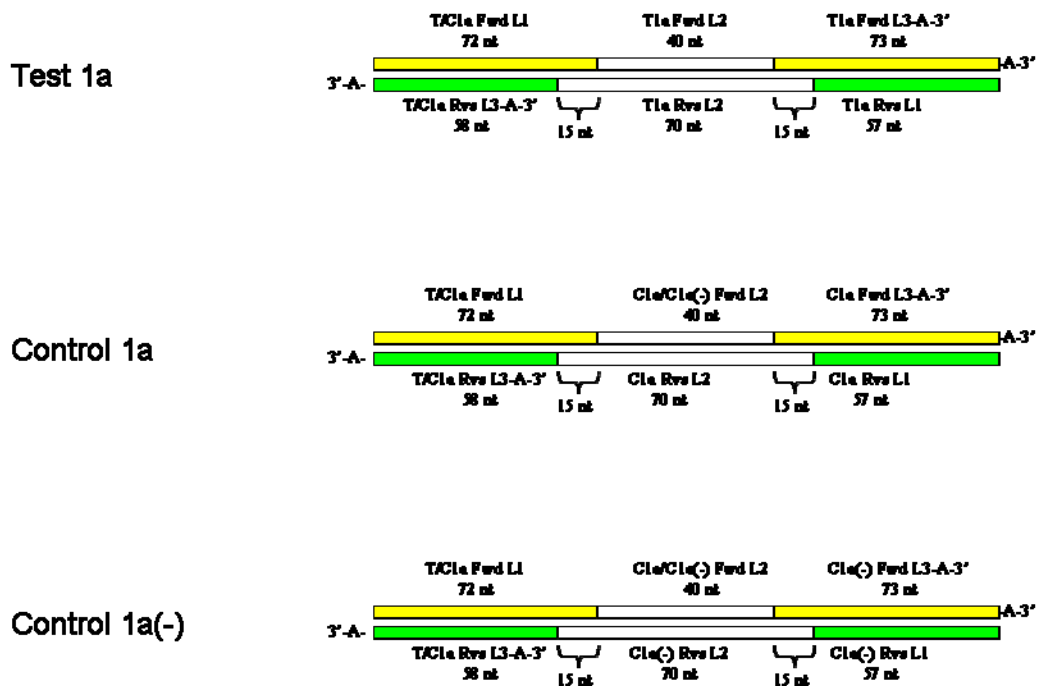
Figure 4.2-3 Ligation oligos configurations for Aim 1a constructs.
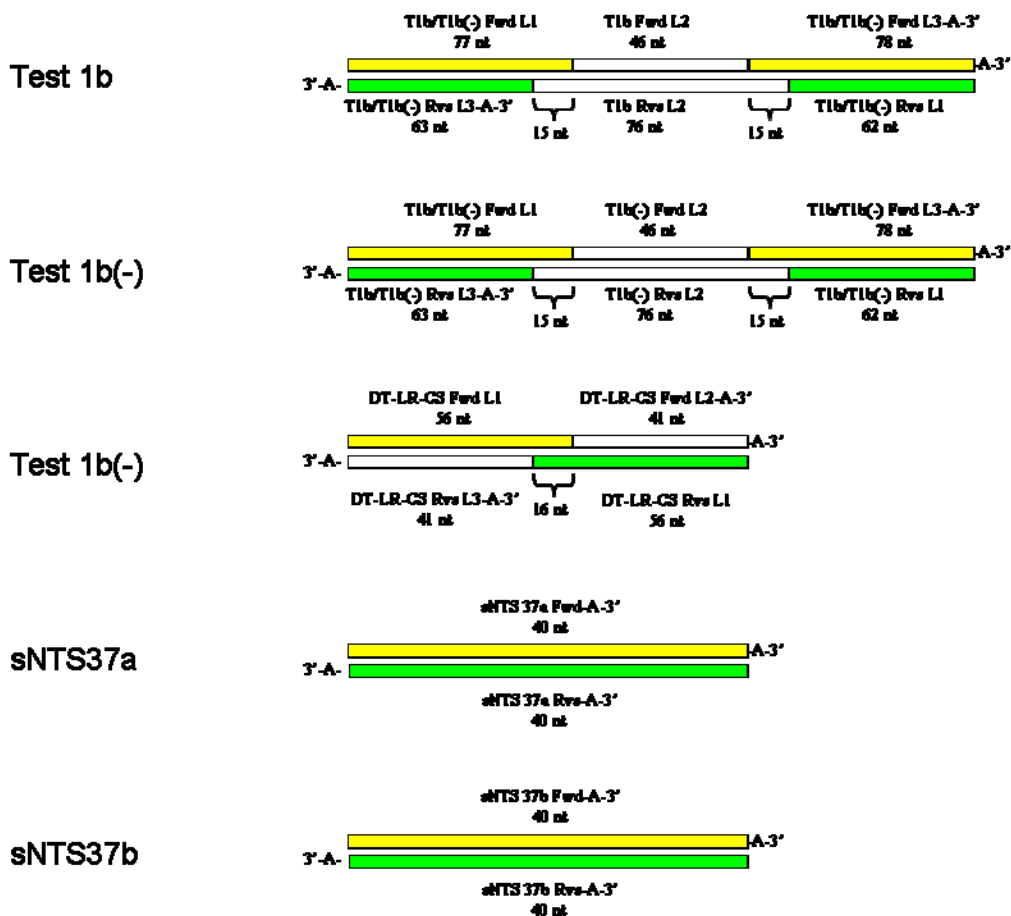
Figure 4.2-4 Ligation oligos configurations for Aim 1b constructs.

Table 4.2-7 Ligation oligos for Aim 1a and 1b constructs.

Construct	Ligation Oligo (Sense)	Length (nt)	Ligation Oligo (Antisense)	Length (nt)
T1a	T/C1a Fwd L1	72	T1a Rvs L1	57
	T1a Fwd L2	40	T1a Rvs L2	70
	T1a Fwd L3-A-3'	73	T/C1a Rvs L3-A-3'	58
C1a	T/C1a Fwd L1	72	C1a Rvs L1	57
	C1a/C1a(-) Fwd L2	40	C1a Rvs L2	70
	C1a Fwd L3-A-3'	73	T/C1a Rvs L3-A-3'	58
C1a(-)	T/C1a Fwd L1	72	C1a(-) Rvs L1	57
	C1a/C1a(-) Fwd L2	40	C1a(-) Rvs L2	70
	C1a(-) Fwd L3-A-3'	73	T/C1a Rvs L3-A-3'	58
T1b	T1b/T1b(-) Fwd L1	77	T1b/T1b(-) Rvs L1	62
	T1b Fwd L2	46	T1b Rvs L2	76
	T1b/T1b(-) Fwd L3-A-3'	78	T1b/T1b(-) Rvs L3-A-3'	63
T1b(-)	T1b/T1b(-) Fwd L1	77	T1b/T1b(-) Rvs L1	62
	T1b(-) Fwd L2	46	T1b(-) Rvs L2	76
	T1b/T1b(-) Fwd L3-A-3'	78	T1b/T1b(-) Rvs L3-A-3'	63
DT-LR-CS	DT-LR-CS Fwd L1	56	DT-LR-CS Rvs L1	56
	DT-LR-CS Fwd L2-A-3'	41	DT-LR-CS Rvs L2-A-3'	41
sNTS-37a	sNTS 37a Fwd-A-3'	40	sNTS 37a Rvs-A-3'	40
sNTS-37b	sNTS 37b Fwd-A-3'	40	sNTS 37b Rvs-A-3'	40

4.2.2.4 IVT template synthesis.

After obtaining the test constructs in the form of inserts within plasmids, we needed to generate large quantities of in vitro transcription (IVT) templates for the production of RNA test constructs. Our first attempt was to amplify the constructs via PCR using insert primers. We were unsuccessful in sequencing these PCR products to verify their sequence fidelity as most products were quite short (~ 200 nt or less).

As we could not be sure that IVT templates obtained via PCR from plasmid inserts were free

of PCR errors, we tried to obtain IVT templates by restriction digesting the inserts out of the plasmids (the circular plasmids could not be used as templates as run-off transcriptions were to be employed). However, after considerable optimization, using as much as 20 U of EcoR V on 3 μg of plasmid, we were unable to purify enough released template for IVT.

We returned to using PCR using the Elongase enzyme to generate IVT template, and this time sequenced the PCR products by cloning them back into pGEM-T Easy vectors. DNA sequencing of these clones could be performed, but most of the IVT templates showed missing regions similar to those in assembly PCR DNA template generation, except for the two positive controls, HP-LR-CS and DT-LR-CS, which could be successfully amplified without error via PCR with Elongase. Some of these errors might have been due to recombination within DH10B, however, there were many cases of single nucleotide mutations, which were likely due to the PCR process.

To rule out the possibility of recombination induced errors, we used M13 vector primers to yield products large enough to sequence directly after PCR, with PfuTurbo enzyme (to reduce mutations) to amplify the remaining IVT templates (aside from sNTS-37a and sNTS-37b) from plasmids via PCR. DNA sequencing of the gel purified PCR products showed no errors in sequence.

As sNTS-37a and sNTS-37b IVT templates were very short (~40 bp), we decided to use their synthesized oligonucleotides that have been annealed as IVT templates.

Since DNA sequencing showed that PCR using PfuTurbo was able to produce reliable IVT template, we finalised on using PfuTurbo PCR to generate large quantity of IVT template.

After PCR, products were separated on agarose gel, and the correct sized bands were excised and gel purified for use as IVT template. Primers (listed in Table 4.2-8 and Table 4.2-9) used for PfuTurbo PCR consisted of one vector based primer (depending on orientation of cloned insert), and a reverse primer located at the 3' end of each construct.

Table 4.2-8 Primers used to synthesise IVT template for Aim 1a and 1b constructs.

Constructs	Forward Primer	Reverse Primer
T1a	pGEMTE Rvs_1	Test/Control 1a Rvs_2
C1a	pGEMTE Fwd_1	Test/Control 1a Rvs_2
C1a(-)	pGEMTE Fwd_1	Test/Control 1a Rvs_2
HP-LR-CS	pGEMTE Rvs_1	Test/Control 1a Rvs_2
T1b	pGEMTE Fwd_1	T1b/DT Rvs_1
T1b(-)	pGEMTE Fwd_1	T1b/DT Rvs_1
DT-LR-CS	pGEMTE Fwd_1	T1b/DT Rvs_1

Table 4.2-9 Sequences and characteristics of primers used in IVT template synthesis for Aim 1a and 1b constructs.

Primer	Length (nt)	Tm	Sequence (5' to 3')
pGEMTE Fwd_1	15	46.6 °C	CGCGGGAATTCGATT
pGEMTE Rvs_1	15	41.7 °C	GCCGCGAATTCACTA
Test/Control 1a Rvs_2	22	56.8 °C	GTTCGGCAGCTCTCTTGAAGCT
T1b/DT Rvs_1	18	54.1 °C	GGGTCCCTTAGCCATCCG

4.2.3 RNA cleavage assays of e-SIGE test constructs.

4.2.3.1 Optimisation of Denaturing Polyacrylamide Gel Electrophoresis (D-PAGE).

Using the gel purified IVT templates synthesised (as described above), RNA versions of Aim 1a and 1b constructs were transcribed by In Vitro Transcription (IVT) using the Ambion MAXIscript Kit, following the prescribed protocol. The IVT and RCA processes occurred

concurrently with the species incubated at 37°C for various durations. Except where stated, each test construct was allowed to undergo IVT and RCA with no other RNA species in the same environment. Subsequently, the RCA products were subjected to denaturing polyacrylamide gel electrophoresis (D-PAGE) and the gel was stained with SYBR Gold and visualised under UV.

We were unable to obtain RNA ladders for sizes below 100 nt, and had to estimate our gel bands sizes from single stranded DNA (ssDNA) ladders. This left some uncertainty in our gel interpretation as RNA migrates slower than DNA.

We had problems getting good resolution of RNA bands on our D-PAGE, and partially improved resolution by optimising gel thickness (1.5 mm, 0.75 mm, and 0.5 mm), reducing voltage and increasing gel running time, and allowing gels to solidify at 4 °C overnight before use.

4.2.3.2 Aim 1a constructs to show that RC1 can cis-cleave CS1 if and only if CS1 was single stranded.

After optimisation of RCA protocols and D-PAGE running conditions, our designs of Test 1a and HP-LR-CS seemed to yield the Rz cleavage intended. The optimised RCA results of Test 1a and HP-LR-CS are shown in Figure 4.2-5, and the observations given in Table 4.2-10.
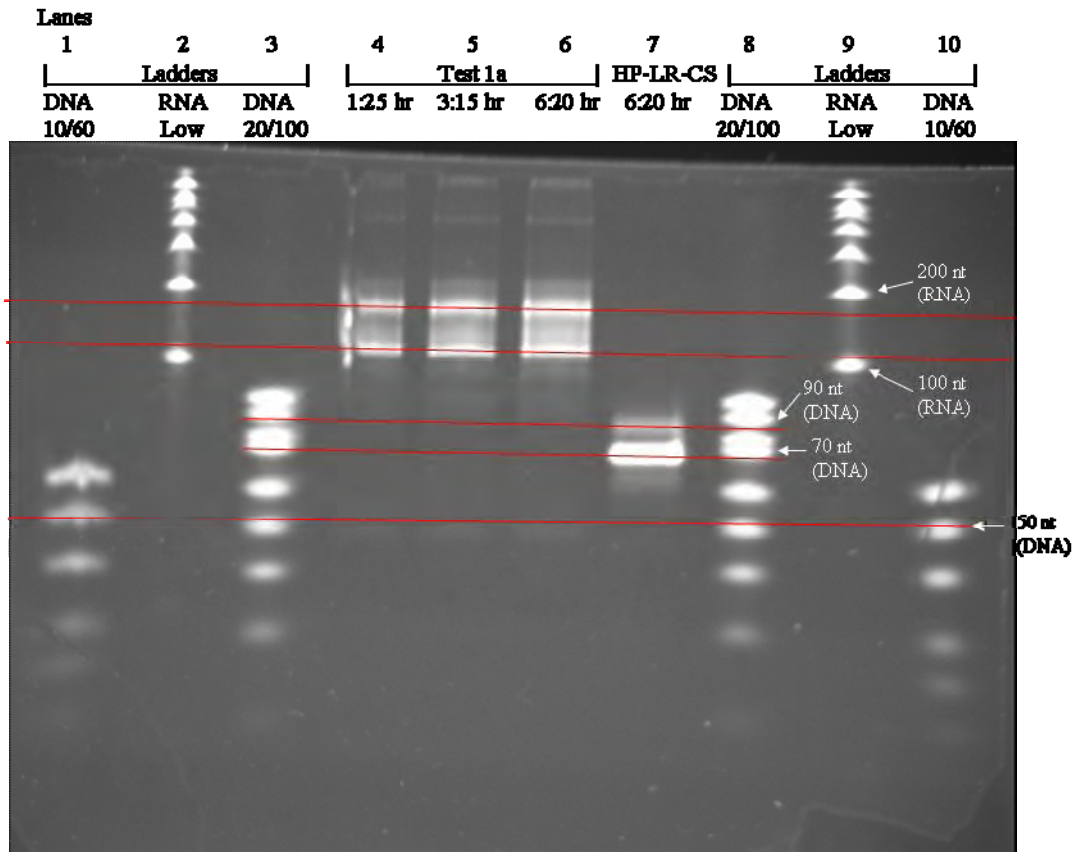
Figure 4.2-5 RCA results for Test 1a and HP-LR-CS.
Gel was stained with SYBR Gold for 45 min. Time stated refers to IVT incubation duration.

Table 4.2-10 Observations for Aim 1a Constructs.

Lanes	Samples	Observations, Comments.
1,10	DNA 10/60	Single stranded DNA ladder with 8 bands: 60, 50, 40, 30, 25, 20, 15, 10 nt.
2,9	DNA 20/100	Single stranded DNA ladder with 9 bands: 100, 90, 80, 70, 60, 50, 40, 30, 20 nt.
3,8	RNA Low	Single stranded RNA ladder with 7 bands: 1000, 800, 600, 400, 300, 200, 100 nt.
4-6	Test 1a 1:25, 3:15, 6:20 hr	
	Possible bands:	Major band(s) observed: 3
	Uncleaved 167 nt	<200 nt – Could be uncleaved 167 nt.
	5' Fragment 119 nt	>100 nt – Could be cleaved 5' 119 nt fragment.
	3' Fragment 48 nt	<50 nt – Could be cleaved 3' 48 nt fragment.
7	HP-LR-CS 6:20 hr	
	Possible bands:	Major band(s) observed: 2
	Uncleaved 72 nt	<90 nt – Could be uncleaved 72 nt.
	5' Fragment 63 nt	<70 nt – Could be cleaved 5' 63 nt fragment.
	3' Fragment 9 nt	
		Note that ladders used for sizing in this case were DNA ladders, which migrate faster than RNA. As such, actual sizes of product bands should be smaller than the ladders here indicate (i.e. 80-90 band could actually be 72 nt). 3' fragment of 9 nt was not apparent here, possibly due to small fragment size.

As shown in Figure 4.2-5 and Table 4.2-10., after optimisation, RCA of Test 1a and HP-LR-CS seemed to yield the intended Rz cleavage. However, there was significant smearing in Test 1a.

4.2.3.3 Aim 1b constructs to show that sNTS-37a but not sNTS-37b can activate EC and lead to cleavage of CS2.

Upon the successful RCA testing of T1a and HP-LR-CS, we proceeded to optimise RCA for

the remaining Aim 1a and Aim 1b constructs. The optimised RCA results were shown in

Figure 4.2-5, and the observations given in **Table 4.2-10**. Both T1b and T1b(-) constructs

underwent IVT and RCAA together with sNTS-37a.
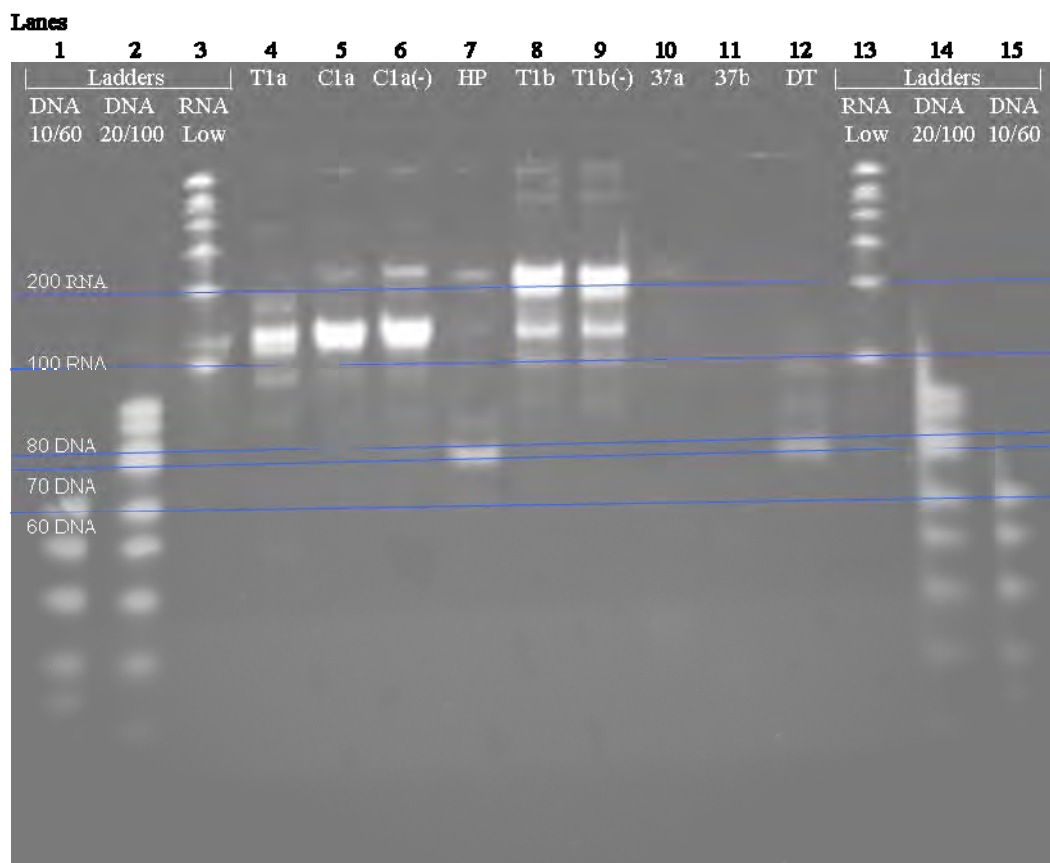
Figure 4.2-6 RCA results for remaining A1a and Aim 1b Constructs.
IVT incubation duration was 40 min.

Table 4.2-11 Observations for remaining Aim 1a and Aim 1b Constructs.
Time stated refers to IVT incubation duration.

Lanes	Samples	Observations, Comments.
1,15	DNA 10/60	Single stranded DNA ladder with 8 bands: 60, 50, 40, 30, 25, 20, 15,10 nt.
2,14	DNA 20/100	Single stranded DNA ladder with 9 bands: 100, 90, 80, 70, 60, 50, 40, 30, 20 nt.
3, 13	RNA Low	Single stranded RNA ladder with 7 bands: 1000, 800, 600, 400, 300, 200, 100 nt.

Lanes	Samples		Observations, Comments.
4	T1a		
	0:40 hr		
	Possible bands:		Major band(s) observed: 4
	Uncleaved	167 nt	<200 nt – Could be uncleaved 167 nt.
	5' Fragment	119 nt	>100 nt – Could be cleaved 5' 119 nt fragment. There appeared to be 2 closely-spaced bands, which might correspond to variants in secondary structures.
	3' Fragment	48 nt	
			~ 90 nt – Could be dimer (96 nt) of 3' 48 nt fragment.
			<50 nt – Could be cleaved 3' 48 nt fragment. Band was very faint, but was visible under long exposure.
5	C1a		
	0:40 hr		
	Possible bands:		Major band(s) observed: 3
	Uncleaved	167 nt	>200 nt – Could be a dimer (238 nt) of 5' 119 nt fragment. This seemed supported by the absence of this band for T1a, for which the 5' 119 nt had less potential pairing region and so less likely to dimerise.
	5' Fragment	119 nt	
	3' Fragment	48 nt	>100 nt – Could be cleaved 5' 119 nt fragment.
			~ 90 nt – Could be dimer (96 nt) of 3' 48 nt fragment.
			The expected products sizes were as for T1a (Lane 4), but with much lower intensity, due to the proposed ability of DS CS1 to prevent cleavage. However, observed products seemed no less intense, meaning DS region meant to protect the CS1 might not be long enough.
6	C1a(-)		
	0:40 hr		
	Possible bands:		Major band(s) observed: 3
	Uncleaved	167 nt	>200 nt – Could be a dimer (238 nt) of 5' 119 nt fragment.
			>100 nt – Could be cleaved 5' 119 nt fragment.
			~ 90 nt – Could be dimer (96 nt) of 3' 48 nt fragment.
			The bands observed were essentially similar to that of C1a, which should not have been possible since CS1 here has been mutated to a non-functional sequence ("dead").

Lanes	Samples		Observations, Comments.
7	HP-LR-CS 0:40 hr		
	Possible bands:		Major band(s) observed: 2
	Uncleaved	72 nt	<90 nt – Could be uncleaved 72 nt. Band was very faint, but was visible under long exposure.
	5' Fragment	63 nt	
	3' Fragment	9 nt	<70 nt – Could be cleaved 5' 63 nt fragment.
			Note that ladders used for sizing in this case were DNA ladders, which migrate faster than RNA. As such, actual sizes of product bands should be smaller than the ladders here indicate (i.e. 80-90 band could actually be 72 nt). 3' fragment of 9 nt was not apparent here, possibly due to small fragment size.
8	T1b 0:40 hr		
	Possible bands:		Major band(s) observed: 3
	Uncleaved	183 nt	~200 nt – Could be uncleaved 183 nt. This was a thick band that appeared to be 2 closely-spaced bands, the higher band could correspond to dimers of the uncleaved with sNTS-37a (222 nt).
	5' Fragment	92 nt	
	3' Fragment	91 nt	
	sNTS-37a	22 nt	~119 nt – Could be dimers of cleaved 5' fragment with sNTS-37a (114 nt). The size was similar to T1a cleaved 5' 119 nt fragment. ~90 nt – Could be cleaved 5' & 3' fragments (92, 91 nt). The low intensity of this band need not imply low cleavage if the majority of 5' fragments dimerised with sNTS-37a to form ~119 nt band.
			sNTS-37a of 22 nt was not apparent here, possibly due to small fragment size. In addition, some of these could have dimerised with the uncleaved as well as cleaved 5' fragment.

Lanes	Samples		Observations, Comments.
9	T1b(-)		
	0:40 hr		
	Possible bands:		Major band(s) observed: 3
	Uncleaved	183 nt	~200 nt – Could be uncleaved 183 nt and dimers of the uncleaved with
	sNTS-37a	22 nt	sNTS-37a (222 nt).
			~119 nt – Could be dimers of cleaved 5' fragment with sNTS-37a (114 nt).
			~90 nt – Could be cleaved 5' & 3' fragments (92, 91 nt).
			The bands observed were essentially identical to those of T1b, which should not have been possible since CS here has been mutated to non functional sequence ("dead").
10	sNTS 37a		
	0:40 hr		
	Possible bands:		Major band(s) observed: Nil
	sNTS-37a	22 nt	sNTS-37a of 22 nt was not apparent here, possibly due to small fragment size.
11	sNTS 37b		
	0:40 hr		
	Possible bands:		Major band(s) observed: Nil
	sNTS-37b	22 nt	sNTS-37a of 22 nt was not apparent here, possibly due to small fragment size.

Lanes	Samples	Observations, Comments.
12	DT-LR-CS 0:40 hr	
	Possible bands:	Major band(s) observed: 3
	Uncleaved 79 nt	<100 nt – Could be dimers of uncleaved and cleaved 3' fragment (92 nt).
	5' Fragment 13 nt	
	3' Fragment 66 nt	~90 nt – Could be uncleaved 79 nt.
		~75 nt – Could be 3' fragment 66 nt. The relative higher intensity of this band compared to the other two implied high cleavage.
		5' fragment of 13 nt was not apparent here, possibly due to small fragment size.
		Note that ladders used for sizing for the cases of ~90 nt and ~75 nt bands were DNA ladders, which migrate faster than RNA. As such, actual sizes of product bands should be smaller than the ladders here indicate (i.e. ~90 nt band could actually be 79 nt, and ~75 nt band could be 66 nt).

Cleavages for T1a, C1a, HP-LR-CS (i.e. all constructs with HP Rz cleavage) seemed complete within the 0:40 hr incubation. Cleavage for T1b seemed significant within the 0:40 hr incubation although it less than that for DT-LR-CS (which uses the same DT Rz and CS). In addition, previous time course optimisation for T1b seemed to show similar level of cleavage from 20 min to 6 hr. This indicated that the secondary structure of T1b could have protected the CS2 to some degree.

These results thus showed that T1a and T1b were cleavable as expected, and which were corroborated by the correct cleaving of the two positive controls, HP-LR-CS and DT-LR-CS. However, the constructs that were not to be cleaved (at least not to high levels), namely C1a, C1a(-), T1b, and T1b(-) were also cleaved. For C1a, cleavage of comparable levels to T1a indicated that the secondary structures meant to protect CS1 were insufficient, perhaps due to the DS region meant to protect the CS1 not being long enough. What was more disturbing

were the cleavage activities for C1a(-) and T1b(-), both of which were designed with dead CS that should not have been cleavable at all.

The above unexpected RCA results were likely indications that that RNA foldings did not occur as intended, which was likely as the foldings were designed by eye without computational design. As such, these initial designs were unsuitable for use as the RNA-based sensor in e-SIGE.

In addition, secondary structures in ssRNA may interfere with gel running, leading to inaccurate separation of otherwise correct sized fragments. Despite spending significant amount of time and resources optimising D-PAGE protocol, it was often difficult to accurately infer the sizes or even number of species of RCA products, which was an important requirement if we were to accurately pin point cleavage errors in order to improve construct designs.

In view of the above difficulties, and our limited time and resource, we decided to make two targeted changes:

1) We needed to design a new sensor. Since de novo design of properly activating RNA structure was difficult, there should be a higher chance of success if we were to adapt from the structures of existing allosteric ribozymes in the literature for the design of our e-SRS sensor.

2) We needed to adopt or create a new methodology of RCA reporting that did not involve gel running and that gave unambiguous signals of cleavage occurrence.

4.3 Maxizymes based e-SRS.

4.3.1 New e-SRS sensor based on the Maxizyme (Mz)

Our three guiding principles for obtaining a new design for the e-SRS sensor were to 1) select from existing allosteric Rz (those that recognise at one site and cleavage at another) that 2) had been shown to work both in vitro and in cell lines, and 3) which would require only minor modifications to suit our sensor purposes, for ease of implementation.

Eventually, we decided on the Maxizyme (Mz), which were a class of allosteric synthetic Rz shown to work both in vitro and in cell lines [Kuwabara et al, 1998; Taira et al, 2004]. These two facts fulfilled our first two guiding principles.

Each Mz is a heterodimer of 2 modified Hammerhead Rz (HH) such that each of the two HH have no catalytic activity individually, and can only be catalytically active when paired as a heterodimer, with the appropriate (i.e. "active") 3D conformation. As shown in Figure 4.3-1, the Mz has one sensor arm, and a catalytic arm. To attain the active 3D conformation, the sensor arm has to bind to a target RNA (of defined sequence), and the catalytic arm has to bind to a substrate RNA. Appropriate binding of the sensor and catalytic arms allows the catalytic core (at the centre of the Mz) to capture a Mg^{2+} that activates the catalytic capability of the Mz, at which point, the substrate RNA bound to the catalytic arm will be cleaved.

Traditional Use of Maxizyme
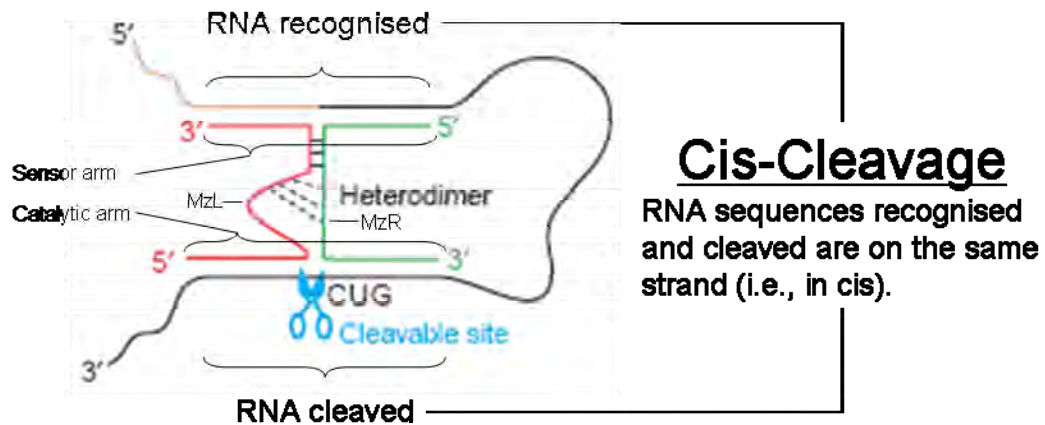
Figure 4.3-1 Traditional use of Mz as an allosteric knock down tool.

The Mz was an ingenious tool designed to recognise an RNA at one point, and cleave the same RNA (i.e. cis-cleavage) at another point (often far away), for the specialised purpose to knock down chimera genes which do not have a Rz recognition sequence near the junction of the chimera gene. The Rz recognition sequence of Mz is that of HH, which is the triplet NUX, where N stands for A, G, C, or U, and X stands for A, U, or C (i.e. any base but G). One common Rz recognition sequence for HH is the triplet GUC.

The rationale for developing such a mechanism is to knock down chimera genes is that for a chimera gene, the only point of specific recognition is around the chimeric junction, since sequences on either side (alone) of the junction also matches the respective parent genes that created the chimera gene. If a Rz is targeted at the non-chimera specific region of the chimera RNA, the normal mRNA that comprises that part of the chimera gene would also be cleaved, resulting in damage to cells.

The advantage that the Mz thus brought to bear was to enable recognition of a chimera gene specifically at the chimeric junction, while allowing cleavage of the RNA to occur elsewhere on the RNA where a Rz recognition sequence could be found. In this case, the normal RNA will not be cleaved as it does not have the chimera specific junction sequence, which the Mz must bind in order to become active. An example is the BCR-ABL fusion mRNA associated with the Philadelphia chromosome that causes chronic myelogenous leukemia (CML). For the b2a2 chimera gene (BCR exon 2 fused with ABL exon 2 via reciprocal chromosomal translocations), there were no recognition sequences for HH within 2-3 nucleotides of the chimeric junction. One GUC triplet was located 45 nucleotides from the junction, which could then be specifically targeted by a Mz. If a normal HH was used to cleave this GUC, normal ABL RNA would also be cleaved as it shared this part of the sequence of with the b2a2 mRNA, with resultant damage to cells.

Our novel adaptation of the Mz required only a minor physical alteration, which was to separate the activating segment of the substrate RNA from the cleaved segment, such that the activating RNA and cleaved RNA were now two separate RNA species. Since the main alteration to the existing Mz mechanism involved only the interaction with substrates (from being cis-cleaved to trans-cleaved), rather than the general structure of the Mz itself, we were quite confident that the Mz foldings should not be affected. By following closely to original designs (for both Mz and most required features of substrates), we felt that it should not take long to tell if the major paradigm shift of substrate cis-cleavage to trans-cleavage was feasible. This fulfilled our third guiding principle.

While the physical alteration was minor, the paradigm shift in changing the cis-cleavage mechanism to a trans-cleavage mechanism was significant, as instead of being a knock down tool, the Mz could now be used as a signal transducer, i.e. a sensor that upon detection of a specific RNA, is able to initiate certain actions (e.g. activation of a gene expression system by knocking down a repressor as previously conceptualised in e-SIGE, or cleavage of a RS (see next section) to release fluorescence for bioimaging). To our knowledge, such a modification of Mz utility has never been reported, and that Mz has always been designed and used as a RNA knock down tool.

4.3.2 RS to provide new methodology of RCA reporting

In order to have a new system of reporting successful substrate cleavage in RCA that 1) did not involve gel running and 2) that gave unambiguous signals of cleavage occurrence, we decided to create a substrate would release fluorescence immediately upon cleavage. Since such a substrate could report on the activity of its ribozyme, we termed it a "Reporter Substrate".

All RNA substrates to be cleaved would be designed with a fluorescent probe (e.g., 6-FAM) conjugated at one end, where the fluorescence signal would be constitutively quenched by a quencher (e.g., BHQ-1) conjugated at the other end. Upon activation of Rz, and thus cleavage of RS, the probe would be separated from its quencher and thus release intense fluorescence that could be assessed quantitatively by a fluorescence plate reader. Each activated Mz-based sensor can cleave many RS as long as the target nucleic acid is present, resulting in

amplification of signal (i.e. multiple fluorescent reporter per copy of target nucleic acid detected).

The e-SRS RS gene detection is thus suitable for all forms of existing gene detection applications, such as detection of pathogenic genes, mRNA, small RNA expression, etc. This RS system had the additional advantages of enabling both real time and quantitative measurements of cleavage events.

4.3.3 New design of e-SRS based on Maxizymes

With the creation of a new sensor and RCA reporting methodology, we essentially created a new system of e-SRS (environment-Sensing Response System). The new system had a much simpler and tested sensor (thus more robust), and was designed to be easily applied in a modular way for bioimaging.

The new e-SRS (as shown in the schematic diagram in **Figure 4.3-2**) is a nucleic acid sensing and response system with two components: 1) A Maxizyme based sensor that changes conformation and becomes active upon binding of specific NTS; 2) A Response System that is triggered by the activated sensor to initiate some physical response, such as emitting a signal (e.g. fluorescence) to indicate the presence of the NTS, or to activate some downstream action such as gene expression.

The new RNA substrate to be cleaved could either be a RS or other nucleic acids, such as a repressor mRNA, and is known as STS, to distinguish it from that of the old (e-SIGE) version.

Conceptually, the new e-SRS could have two major applications:

1) RS based bioimaging, which was the imaging of targeted nucleic acids via the cleaving of RS.

2) Induction of PLG , which was the induction of a PLG in cells containing a target nucleic acid.
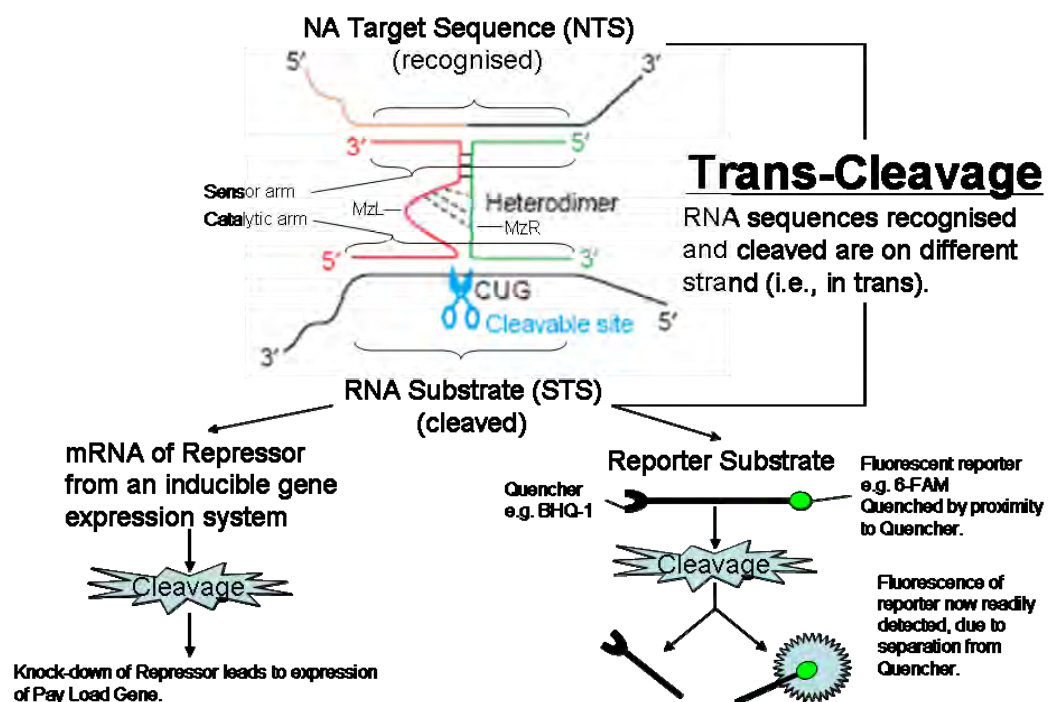
Figure 4.3-2 The new e-SRS uses a Mz based sensor and is easily applied in bioimaging. Note that the Reporter Substrate when uncleaved is most likely not straight but exist in various dynamically changing shapes that put the quencher in close proximity with the FAM.

4.3.4 Inducible Gene Expression System

For ease of implementation, we decided to use a commercial inducible gene expression system, T-REx (Invitrogen). It provided the repressor R1 (via the TetR repressor protein) that could suppress a co-transfected PLG until activation by the Mz based sensor (which would

knock down R1 and thus relieve PLG repression).

The T-REx system comprises of 2 plasmids: 1) pcDNA6/TR, which was a pCMV driven TetR (as R1) over-expression plasmid; 2) pcDNA4/TO/myc-His, which was a pCMV (with 2 tetracycline operator 2 sites embedded) driven expression plasmid that expressed a PLG that can be constitutively repressed by an appropriate ratio of cotransfected TetR. TetR repression can be relieved by application of Tet of about 1 μg/ml in culture medium.

However, instead of using tetracycline to relieve repression of PLG, an activated Mz-based sensor designed to cleave the tetR mRNA can knock down tetR expression and hence induce PLG expression if and only if the desired target nucleic acid was detected. Hence e-SRS can induce the expression of a desired exogenous gene when a target gene is detected. In such an embodiment, e-SRS can be used, amongst many other possibilities, to selectively kill malignant cells by targeting an mRNA specific to the malignant cell (e.g. a gene with a mutation), with the PLG being a cytotoxic gene.

4.3.5 NTS/NNS and selection of STS, STS

In this new e-SRS, we decided to keep the NTS and corresponding NNS of before, i.e. the NTS was sNTS-37a and its corresponding NNS was sNTS-37b. We also included a b1a2 model 2 Mz [Tanabe et al, 2000] from the Mz literature as a positive control. Its NTS and NNS were sNTS-b1a2_NTS and sNTS-b1a2_NNS respectively. NTS/NNS sequences were as shown in Table 4.3-1.

Table 4.3-1 Sequences of NTS and NNS.

Constructs	Sequence (5' to 3')	Length (nt)

Constructs	Sequence (5' to 3')	Length (nt)
sNTS-37a	AUUCACUAUAAGGACAUGUA	20
sNTS-37b	AUGCCUUACCCGGACAUGUA	20
sNTS-b1a2_NTS	UCCAUGGAGACGCAGAAGCCCUUC	24
sNTS-b1a2_NNS	AUCUGGAAGAAGCCCUUC	18

For both applications (i.e., RS based bioimaging and induction of PLG) of e-SRS, a STS was required. In order to have a STS that could satisfy both applications, we decided to select the STS from our R1 mRNA, i.e., the tetR mRNA.

3 STS, RTS-1, RTS-2, RTS-3 (RTS meaning Repressor Target Sequence) were chosen from the tetR mRNA sequence. RTS-b1a2 for the b1a2 model 2 Mz was taken from the Mz literature [Tanabe et al, 2000]. The selection criteria for our STS were as follows:

1) A region of tetR mRNA containing HH recognition sequence (GUC, CUC, AUC).

2) Predicted to be accessible (not hidden in secondary structure) using RNA folding software, RNAstructure 4.5

3) Had no Blast hit against Human, Mouse, and Rat Refseq genes (excluding predicted/hypothetical genes).

In order to test the effectiveness of each STS, we had had to design one Mz based sensor for each STS (Mz-1, Mz-2, Mz-3). In addition, for each STS, we also designed a corresponding wild type Hammerhead Ribozyme (HH-1, HH-2, HH-3, HH-b1a2) that was supposed to cleave it with high efficiency to compare the cleavage efficiency of our Mz based sensors.

Sequences of the STS and HH designed are shown in Table 4.3-2.

Table 4.3-2 Sequences of STS and HH.

Constructs	Sequence (5' to 3')	Length (nt)
RTS-1	GCAUCAAGUCGCUAAAG	17
RTS-2	UAUGAAACUCUCGAAAA	17

Constructs	Sequence (5' to 3')	Length (nt)
RTS-3	CUCGAAA<u>AUC</u>AAUUAGC	17
RTS-b1a2	AAAACCUU<u>CUC</u>GCUGGACCCA	21
HH-1	CUUUAGCCUGAUGAGGCCGAAAGGCCGAAACUUGAUGC	38
HH-2	UUUUCGACUGAUGAGGCCGAAAGGCCGAAAGUUUCAUA	38
HH-3	GCUAAUUCUGAUGAGGCCGAAAGGCCGAAAUUUUCGAG	38
HH-b1a2	UGGGUCCAGCCUGAUGAGGCCGAAAGGCCGAAAGAAGGUUUU	42

<u>Underlined</u> sequence, HH recognition sequence. Cleavage occurs immediately 3' of recognition sequence.

4.3.6 Design of Mz based sensors

4.3.6.1 Conditions for a specific Mz based sensor

We designed a specific Maxizyme for each of the above STS sequences (RTS-1, RTS-2, RTS-3). Each Mz is comprised of a Mz Left (MzL) and a Mz Right (MzR) strand, defined according to the convention shown in the configuration of the b1a2 model 2 Mz [Tanabe et al, 2000], as shown in Figure 4.3-3. Mz were designed to have secondary structures that were active if and only if the specific NTS and STS were both present. This meant that if the NTS was absent, with or without the NNS, the secondary structure of the Mz would be in an inactive form.
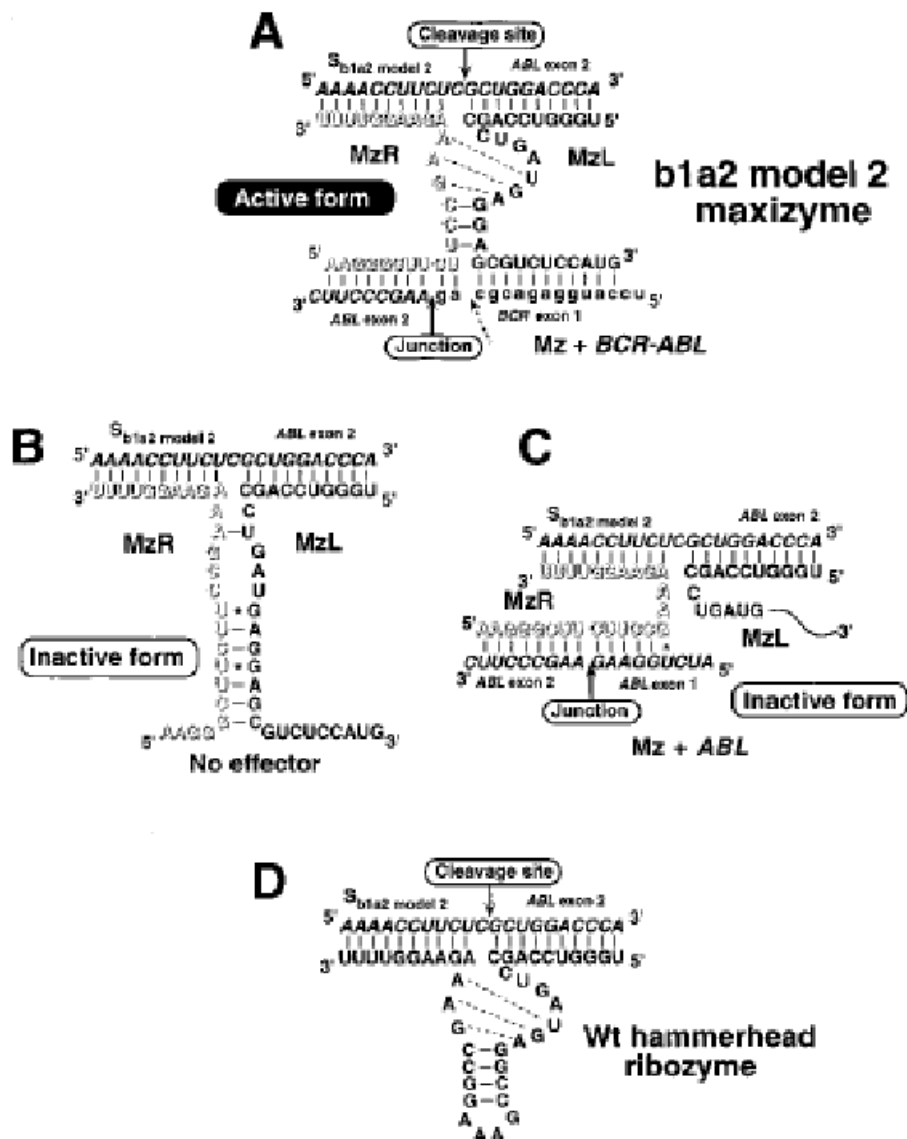
Figure 4.3-3 Illustration of Mz activation, obtained from [Tanabe et al, 2000], Figure 6.
(A) b1a2 model 2 Mz is active in the presence of BCR-ABL fusion (b1a2) mRNA (i.e., the NTS) at the bottom and the substrate RNA (i.e., the STS) at the top. (B) Mz is inactive in the absence of the NTS, even though the STS is present (top). (C) Mz is also inactive in the presence of the NNS (the ABL mRNA in this case, at the bottom). (D) Active wild-type HH with STS (top).

With RNA sequences of Mz, NTS/NNS, and STS defined, secondary structures were predicted using the software, RNAstructure 4.5 for three combinations of RNA species:

1) Mz with NTS and STS (Mz+NTS+STS).

2) Mz with NNS and STS (Mz+NNS+STS).

3) Mz and STS with nothing else (Mz+STS).

Each combination of RNA species would sometimes have more than one predicted secondary structure (some may be active and some maybe inactive). Each predicted structure would have a corresponding free energy, which indicated its stability – the lower (more negative) the free energy, the more stable the structure. In general, the more stable a secondary structure, the greater the proportion of such structures amongst a pool of RNA structures that may dynamically fluctuate between various secondary structures. We considered a particular Mz design as valid (specific for its NTS) when the conditions in Table 4.3-3 were achieved.

Table 4.3-3 Conditions for a Mz to be considered specific for its NTS.

Mz and STS in the presence of	Predicted free energy for	
	Inactive form	Active form
Nothing else	Lowest	High or not predicted
NNS	Lowest	High or not predicted
NTS	High or not predicted	Lowest

Values (e.g. "Lowest" or "High", etc) were defined relative to values in the same row, i.e. for the same combination of RNA species.

4.3.6.2 RNA regions that make up a Mz based sensor

Structurally, each Mz is comprised of two RNA strands – a Mz Left (MzL) and a Mz Right (MzR) strand. However, the objective of Mz design is to create a Mz that will fold into appropriate secondary structures as described in Table 4.3-3. Hence, in Mz design, the Mz is instead conceptually divided into regions (contributed by both MzL and MzR) that are relevant in the determination of secondary structures.

As described in section 4.3.1, the key for a Mz to attain an active 3D conformation is for the catalytic core to capture a Mg^{2+} that then activates the catalytic capability of the Mz. This process requires the catalytic core to first be available in a conducive conformation, which is

facilitated when the sensor arm binds to an RNA with the intended NTS. The binding of the

catalytic arm to an RNA with the STS may also facilitate the adoption of the catalytic core,

but this alone (without the binding of NTS by the sensor arm) is ensured (during the Mz

design process) not to cause the activation of catalytic core. The catalytic core is a conserved

RNA loop region enclosed by the Stem II (a mid section RNA stem), and the junction of the

catalytic arm-STS duplex, both of which are double stranded in an active Mz.

Therefore, as shown in Figure 4.3-4, tor the purpose of design, the Mz can be conceptually

divided into 4 regions:

1) Sensor arm

2) Stem II

3) Catalytic core

4) Catalytic arm
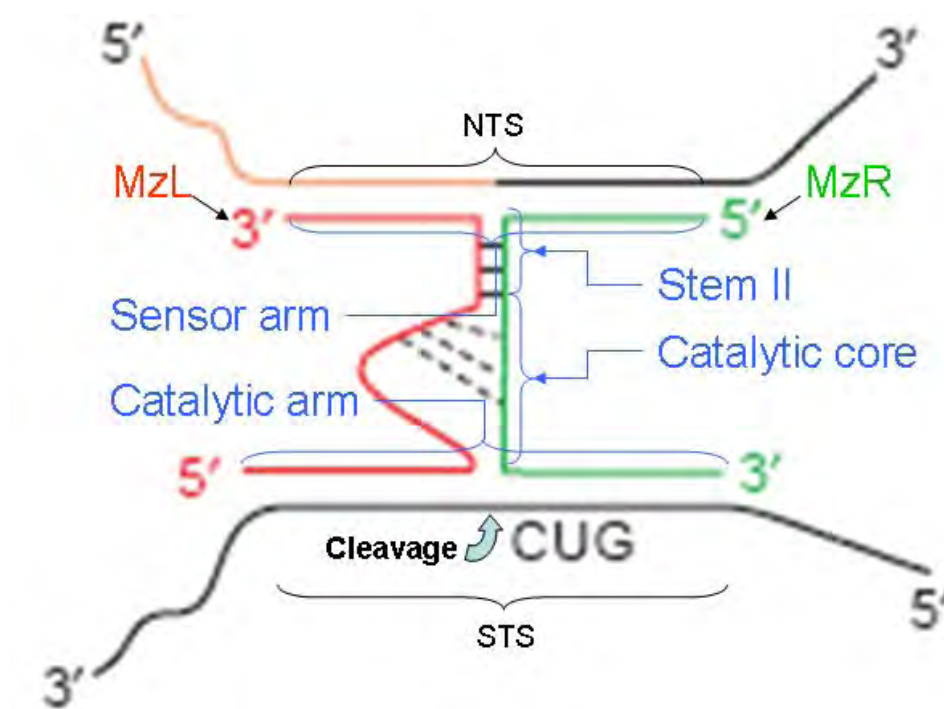
Figure 4.3-4 Regions in e-SRS design of Mz.

4.3.6.2.1 **Sensor arm**

The sensor arm sequence is determined by the NTS, and is the reverse complement of the NTS (i.e. NTS'). The MzL segment of the sensor arm is separated from the MzR segment at the Sensor Arm Split (SAS). The SAS is defined numerically as the number of sensor arm nucleotides that exist on the MzR. This would also define the corresponding number for MzL as the total length of the sensor arm is predefined by the length of NTS. As such, the sensor arm sequence can be divided into a 5' segment (MzR) and 3' segment (MzL) by the SAS. The corresponding point of the SAS on the NTS that separates the segment that binds to MzL and the segment that binds to MzR is known as the SAS'.

By varying the SAS, the resulting Mz secondary structure can be altered, giving different specificities for the given NTS (and STS).

4.3.6.2.2 **Stem II and catalytic core**

For sequences of Stem II and catalytic core, we compared 6 functional Mz from the literature (Table 4.3-4) to identify a consensus for these two segments. Based on these observations, we derived the catalytic core consensus sequence and some potential Stem II sequences (which had no consensus sequence) for our Mz as shown in Table 4.3-5. As Stem II sequences were highly variable (in the literature examples), and were often modified in our Mz design in order to attain the right secondary structures, only the pair used eventually for our Mz for in vitro detection was shown.

Table 4.3-4 Mz selected from the literature.

Mz Target	Strand	Sequence (5' to 3')
PML-RARα	MzL	GGGCAGGAACUGAUGAGGUUCCCCGGCGCCAC
(Tanabe et al, 2000)	MzR	UCUCAAUGGCUGCAUCGAAACCUCACUU
	STS	AAGUGAGGUC^UUCCUGCCC
	NTS	**GUGGCGCCGGGGAGGCAG**ccauugaga
sDLST	MzL	GGCACUGUCUCUGAUGAGAUGCUGAAAAC
(Tanabe et al, 2000)	MzR	UAGCAGGAGCGUCGAAAAAUGAGAAG
	STS	CUUCUCAUUUC^AGACAGUGCC
	NTS	guuuuucag**CUGCUCCUGCUA**
b1a2 model 1	MzL	GUUAUGCUUACUGAUGAGGUUCUUCAGGGG
(Tanabe et al, 2000)	MzR	GAAGGGCUGCUGCAUCGAAAGUGUUAUCU
	STS	AGAUAACACUC^UAAGCAUAAC
	NTS	uccauggagacgcagAAGCCCUUC
b1a2 model 2	MzL	UGGGUCCAGCCUGAUGAGGAGCGUCUCCAUG
(Tanabe et al, 2000. Soda	MzR	GAAGGGCUUCUUCCGAAAGAAGGUUUU
et al, 2004)	STS	AAAACCUUCUC^GCUGGACCCA
	NTS	uccauggagacgcagAAGCCCUUC
Bradeion β	MzL	ACUUCACUGAUGAGAGCCUCAGAGGAAUC
(Tanabe et al, 2000)	MzR	AUACUCCUUGUCAUAUCGAAACUCCCU
	STS	AGGGAGUC^UGAAGU
	NTS	uaugaggaacaguaGGAGUCUCCUUAG
P53 mutant	MzL	GAGGAUGGCUGAUGAGCGAAAGGUCUG
(Kong et al, 2005)	MzR	AGUUUCCACUGAUGAGCGAAACUCCGG
	STS	CAGACCUA^UGGAAACU
	NTS	ccggaguCCCAUCCUC

Highlighted sequence, Catalytic core and Stem II sequence; <u>Underlined</u> sequence, Catalytic core sequence; ^, Cleavage position on the STS; Red font, Rz recognition sequence.

Table 4.3-5 Catalytic core and Stem II sequences.

Region	Strand	Sequence (5' to 3')
Catalytic core	MzL	CUGAUGAG
	MzR	CGAA
Stem II	MzL	UGA
	MzR	AGCA

4.3.6.2.3 **Catalytic arm**

The catalytic arm sequence is determined by the STS, and is the reverse complement of the STS (i.e. STS'). The MzL segment of the catalytic arm is separated from the MzR segment at the Catalytic Arm Split. The Catalytic Arm Split is defined numerically as the number of catalytic arm nucleotides that exist on the MzR. This would also define the corresponding number for MzL as the total length of the catalytic arm is predefined by the length of STS. As such, the catalytic arm sequence can be divided into a 5' segment (MzL) and 3' segment (MzR) by the Catalytic Arm Split. The corresponding point of the Catalytic Arm Split on the STS that separates the segment that binds to MzL and the segment that binds to MzR is known as the Catalytic Arm Split'.

By varying the Catalytic Arm Split, the resulting Mz secondary structure can be altered, giving different specificity for the given NTS (and STS).

4.3.6.3 Joining RNA regions to create a Mz based sensor and predicting secondary structures of RNA combinations.

When the sequences for the 4 regions (sensor arm, Stem II, catalytic core, and catalytic arm) have been obtained, a Mz based sensor can be obtained by creating the MzL and MzR

sequences.

4.3.6.3.1 **Creating a MzL sequence**

To create a MzL sequence, join from 5' to 3', the following regions:

1) 5' segment of the catalytic arm.

2) Catalytic core sequence for MzL (CUGAUGAG).

3) Stem II sequence for MzL (e.g. UGA)

4) 3' segment of sensor arm.

4.3.6.3.2 **Creating a MzR sequence**

To create a MzR sequence, join from 5' to 3', the following regions:

1) 5' segment of sensor arm.

2) Stem II sequence for MzR (e.g. AGCA).

3) Catalytic core sequence for MzR (CGAA).

4) 3' segment of the catalytic arm.

4.3.6.3.3 **Folding combinations of RNA strands in RNAstructure 4.5**

As explained in section 4.3.6.1, each potential Mz has to be assessed for specificity via secondary structure prediction using RNAstructure 4.5, before it would be sent for RNA synthesis and subsequent experimental validation. Three combinations of RNA species (Mz with NTS and STS; Mz with NNS and STS; Mz and STS with nothing else) had to be tested to fulfil the conditions in **Table 4.3-3**.

However, RNAstructure v4.5 can only fold a single strand of RNA at a time. In order to

predict the secondary structures of multiple RNA strands at the same time, we had to join

separate strands into one long strand using the unfolded linker sequence "gaaa". "gaaa" was a

commonly used linker sequence that Taira et al used to connect their Mz and other RNA for

computational foldings. By using lower cases ("gaaa" instead of "GAAA"), these linkers were

prevented from participating in base pairing in RNAstructure 4.5.

It was initially thought that more repeats of "gaaa" would be better, as that might produce

more obvious boundaries of each strands as well as confer more flexibility to the positioning

of different strands. However, after trying up to 10 repeats of "gaaa", it was discovered that

the more repeats introduced, the less stable (higher free energy value predicted) the folded

structures, indicating that such additions might create more spurious effects on the RNA

folding, and thus best to be kept to a minimum. Also the secondary structure output was

actually easier to visualise with 1 repeat, than with 10 repeats of gaaa. Hence, we eventually

kept to using just one repeat of "gaaa" between adjacent RNA strands.

4.3.6.3.4 Order of joining RNA strands

RNAstructure 4.5 does not predict pseudoknot secondary structures. A pseudoknot (**Figure**

4.3-5) is an RNA secondary structure that has at least two stem-loop structures in which each

stem has one half embedded in the enclosed loop of the other stem.
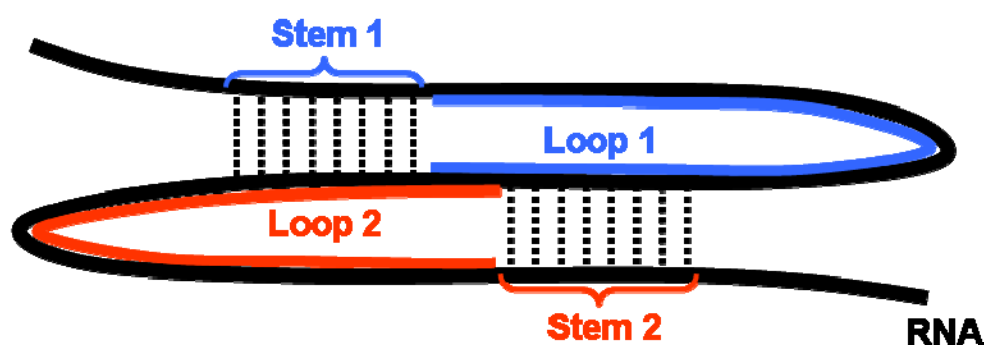
Figure 4.3-5 A pseudoknot RNA secondary structure.

As a result, the order in which the strands were joined in the three combinations must be properly sequenced in order for it to be possible to predict the active structure without requiring the formation of pseudoknots. The strand sequences used for our Mz design were as shown in Table 4.3-6.

Table 4.3-6 Strand sequences used in Mz based sensor design.

RNA combinations	Strand sequences (5' to 3')
Mz with NTS and STS.	STS, MzL, NTS, MzR.
Mz with NNS and STS.	STS, MzL, NNS, MzR.
Mz and STS with nothing else.	STS, MzL, MzR.

Note that for strand sequences, each comma indicated the addition of a linker "gaaa".

4.3.6.4 Assessment and modification of Mz based sensor secondary structures using RNAstructure 4.5

4.3.6.4.1 Assessment of active and inactive structures

Upon obtaining the secondary structures of the three RNA combinations for each potential Mz design, we needed to assess whether such potential designs met the conditions for a valid Mz design, as given in Table 4.3-3. To be able to perform such assessments, we needed to define what constituted active and inactive secondary structures.

We placed priority on specificity over sensitivity as the strength of e-SRS lies in its ability to

distinguish between closely related nucleic acid sequences. Sensitivity, on the other hand, could be improved by using better detection or amplification systems. As such, there was a need to avoid accepting false positives of specific Mz designs, i.e. where RNA combinations without NTS (intended and predicted to be inactive) were actually active. Hence, it was decided to establish the bare minimum criteria needed for defining a structure to be active – as long as the Mz can cleave STS. For the purpose of our work, we defined the sufficient conditions for an active secondary structure as follows:

1) The NUX triplet must pair correctly. This means that the first and second bases of NUX must pair respectively with the MzR bases two and one position 3' of the Catalytic Arm Split. For the 17 nt STS and 16 nt catalytic arm used here, these positions would be Length minus 7 and Length minus 8 on the combined RNA (where Length refers to total length of combined RNA, including linkers).

   The third base of NUX must be unpaired.

2) 1st base 3' of the NUX on the STS (whose 5' bond is cleaved) must be paired to MzL at a position of the catalytic arm one base 3' of the Catalytic Arm Split. For the 17 nt STS and 16 nt catalytic arm used here, with a "gaaa" linker between STS and MzL, this position would be 28 on the combined RNA.

3) Loop structure of the catalytic core (for binding Mg^{2+}) must be correct. This means that

   a. The first 7 of the 8 MzL catalytic core bases must be unpaired.

   b. The 8th base must pair with the first base of MzR catalytic core.

c. The last 3 of the 4 bases of the MzR catalytic core must be unpaired.

In summary, in e-SRS Mz based sensor design for the three proposed STS (RTS-1, RTS-2, RTS-3), the sufficient conditions for an active secondary structure required the pairing conditions of only 15 bases (counting only one base of a matched pair) to be defined as shown in Table 4.3-7. For each secondary structure predicted, the conditions in Table 4.3-7 were checked by visually examining the secondary structures figures created by RNAstructure 4.5.

Table 4.3-7 Sufficient conditions for an active secondary structure.

S/no.	Base	Base position	Base position of partner
1	1st base of NUX.	008	L-7
			MzR catalytic arm.
2	2nd base of NUX.	009	L-8
			MzR catalytic arm.
3	3rd base of NUX.	010	0
4	Base 3' of NUX.	011	028
			MzL catalytic arm.
5	1st of MzL catalytic core.	029	0
6	2nd of MzL catalytic core.	030	0
7	3rd of MzL catalytic core.	031	0
8	4th of MzL catalytic core.	032	0
9	5th of MzL catalytic core.	033	0
10	6th of MzL catalytic core.	034	0
11	7th of MzL catalytic core.	035	0
12	8th of MzL catalytic core.	036	L-12
			(1st base of MzR catalytic core.)
13	2nd of MzR catalytic core.	L-11	0
14	3rd of MzR catalytic core.	L-10	0
15	4th of MzR catalytic core.	L-9	0

L, Length of combined RNA (including linkers); 0, No pairing partner, i.e., base is unpaired.; L-n means "Length minus n", for e.g. L-8 means "Length minus 8". Note that "Base position" here referred to base position on the combined RNA, including linkers.

4.3.6.4.2 Strategies & guidelines for achieving desired structures.

Deciding on the appropriate sequences of MzL and MzR to use given the required NTS and STS sequences was a rather time consuming "trial and error" based task. In general, we would

start with Catalytic Arm Split and SAS at around the middle of the catalytic or sensor arms, and perfectly matched three base pairs long Stem II. Based on the secondary structures generated, we would then try to guess the best adjustments to make in the Catalytic Arm Split, SAS, or Stem II sequence, and then fold the new structures. The process would repeat until we obtain valid Mz designs.

Through this process, we obtained a few strategies or pointers that appeared useful, as follows:

1) The first focus should be on creating a Stem II that is unstable in absence of NTS. This was because the main challenge was usually getting Mz+STS to fold with no Stem II while Mz+STS+NTS does. Mz+STS+NNS usually folded in an inactive manner.

2) To create unstable Stem II in absence of STS:

   a. Design one or both half of Stem II to base pair with either part of the catalytic core or sensor arms, by designing some complementary sequences in the respective areas.

   b. Create mismatches in Stem II.

   c. Start with 3 base pair Stem II, then increase (not necessarily evenly on both strand) if needed.

3) Sometimes, sequences designed to form a stem simply refused to do so. This could be due to the loop not being favourable. A loop looking like GAAA is probably more favourable. Hence, inserting an unpaired G at the area of the intended loop could

precipitate the formation of the desired stem-loop structure.

4) Should all else fail, change the SAS and start again with above strategies. The
    Catalytic Arm Split was usually not changed as there is usually only one cleavage site
    on STS, where changing the STS would then require changing the STS sequence.

4.3.7 Sequences of Mz based sensors

For each STS, a specific Mz based sensor was designed according to the methodology above.
The MzL and MzR sequences and their respective STS are as shown in Table 4.3-8. All RNA
oligonucleotides (including STS, MzL, MzR, NTS, NNS) were synthesised commercially
with HPLC purification.

Table 4.3-8 MzL and MzR sequences.

STS	Constructs	Sequence (5' to 3')	Length (nt)
RTS-1	MzL-1	CUUUAGCCUGAUGAGUCAuauagugaau	28
RTS-2	MzL-2	UUUUCGACUGAUGAGUGAuauagugaau	28
RTS-3	MzL-3	GCUAAUUCUGAUGAGUCAuauagugaau	28
RTS-b1a2	MzL-b1a2	UGGGUCCAGCCUGAUGAGGAGCGUCUCCAUG	31
RTS-1	MzR-1	UACAUGUCCuAGGACGAAACUUGAUGC	27
RTS-2	MzR-2	UACAUGUCCuAGCACGAAAGUUUCAUA	27
RTS-3	MzR-3	UACAUGUCCuAGGACGAAAUUUUCGAG	27
RTS-b1a2	MzR-b1a2	GAAGGGCUUCUUCCGAAAGAAGGUUUU	27

Highlighted sequence, Catalytic core and Stem II sequence; Underlined sequence, Catalytic
core sequence.

4.3.8 In vitro test of Mz based sensors

We tested the new Mz based RNA sensors in test tubes via RTA, using sNTS or NNS (sNTS
or sNNS) that were short ~20 nt RNA that represented the region on NTS or NNS to be
recognised by Mz. RTA protocols were modified from Mz [Tanabe et al, 2000]. Basically, Rz

(either Mz or HH) were incubated with the substrate (STS), and activator (NTS, NNS, or nothing), at 37 °C in RTA buffer (50 mM Tris (pH 8.0), 10 mM $MgCl_2$) over a duration, during which aliquots would be removed at various time points and RTA terminated by addition of EDTA.

At the initial stages of RTA testing (before the advent of the RS system) RTA products were analysed via D-PAGE with pre-cast TBE-Urea gels from Biorad. However, repeated runs of RTA did not show the expected cleavage of RTS-1 and that of RTS-b1a2 (taken from the literature to serve as positive control). A major reason suspected was that the cleaved STS was in too small a quantity to be detected on D-PAGE stained with SYBR Gold.

To resolve this issue, we came up with the RS system and synthesised the 4 STS with a fluorescent probe (e.g., 6-FAM) conjugated at one end, where the fluorescence signal would be constitutively quenched by a quencher (e.g., BHQ-1) conjugated at the other end. Upon activation of Mz, and thus cleavage of STS, the probe would be separated from its quencher and thus release intense fluorescence that could be assessed quantitatively by a fluorescence plate reader.

Of the 3 STS selected, RTS-2 performed well (was able to detect the 20 nt NTS specifically against NNS, a similar sequence) and eventually, cell line tests focused on its corresponding Mz-2 and HH-2.

Figure 4.3-6 shows combined results of Mz-2 from two RTA experiments. Fluorescence values for each RTA time point were normalised with respect to values at 0 min, and expressed as a percentage of HH-2 fluorescence signal at 60 min. Both experiments showed

similar desired trends, which was that HH-2 (Yellow) gave the maximum fluorescence (i.e.,

substrate cleavage), followed closely by Mz with sNTS (Blue), while Mz with sNNS (Red) or

nothing (Black) gave little or no substrate cleavage. Essentially, Mz-2 was activated if and

only if sNTS was present.
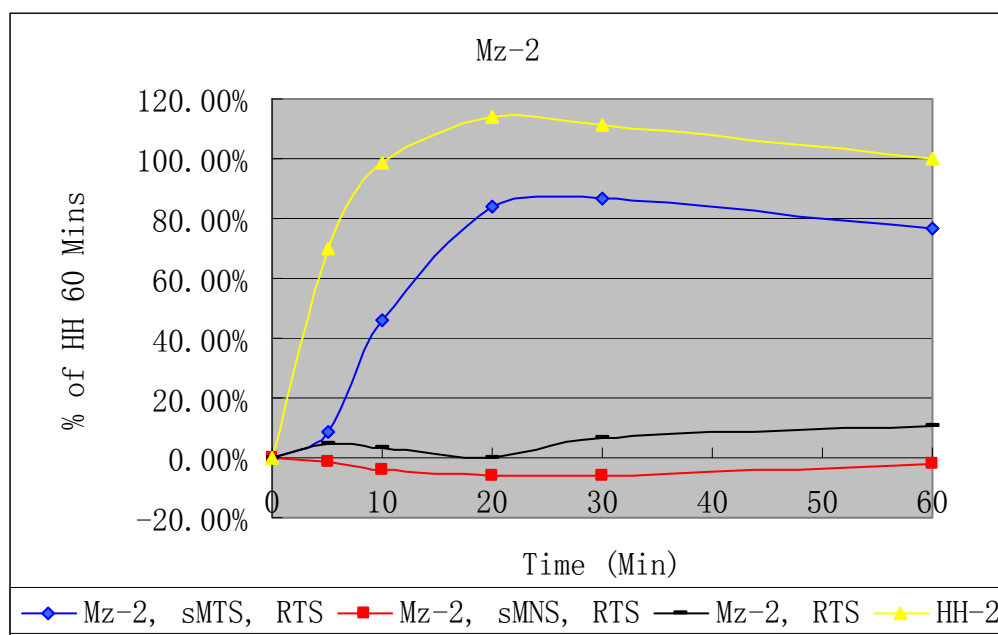
Figure 4.3-6 RTA results for Mz-2 from two experiments.

4.4 **Cell line testing of Maxizyme based e-SRS.**

In this Section, we tested the functional Mz based e-SRS sensors in cell line (HEK293) for

two paradigms:

1) To activate inducible gene expression system if and only if NTS is present, using the

   T-REx system with reporter genes (ECFP or EGFP) as PLG.

2) To activate RS if and only if NTS is present.

For paradigm 1, the cell line tests were unsuccessful, despite optimisations of 1) inducible

gene system; 2) Expression of RNA (including using a tRES, and addition of a T7 RNA Pol

III promoter).

For paradigm 2, we attempted to transfect both e-SRS sensors and RS into cells to allow detection of specific RNA within cells. However, the RS degraded when transfected into culture vessels with cells grown and were thus not usable. The problem was not resolved by the use of nuclease resistant modifications, such as phosphorothioate bonds and addition of 2'-O- methyl groups. These issues lead us to redirect e-SRS for in vitro development.

4.4.1 Inducible Gene System.

4.4.1.1 Optimisation of the Inducible Gene System for the ratio of pTR to inducible PLG.

As explained in Section 4.3.4, for ease of implementation, we used a commercial inducible gene expression system, T-REx (Invitrogen). The repressor, R1 was provided by the pTR, and the PLG (activated by the Mz based sensor knock down of R1) would be cloned into the provided the pPLG.

As each pPLG has two operator sites for R1 to bind in order to suppress the cloned PLG, greater ratio of pTR to pPLG transfected implied stronger PLG suppression and smaller ratio of pTR to pPLG trasnfected implied weaker PLG suppression. There is thus a need to establish an optimum ratio of pTR to pPLG in transfecting the inducible gene system, for which there are two main considerations:

1) When the Mz-based sensors are inactive, R1 must be expressed at high enough level to ensure PLG is suppressed.

2)   When Mz-based sensors are active, the level of R1 must be low enough for it to be

knocked down sufficiently to allow detectable expression of PLG.

4.4.1.2 pcDNA4/TO/myc-His/lacZ as PLG in PC-12.

As a quick optimisation of ratio of pTR to pPLG, we used the provided

pcDNA4/TO/myc-His/lacZ as the pPLG with the PLG being LacZ. We tested a range of ratio

of pTR to pPLG from 1:1 to 9:1, with the ratio for pTR incrementing by 1 per step (thus

testing nine different ratios). Total mass of transfected DNA was kept at 800 ng for all wells.

On Day -1, PC-12 cells were seeded at 3×10^6 cells / well in 24 well plate and grown with

RPMI+++.

The next day, Day 0, the inducible gene system (i.e. pTR and pPLG) were transfected at

various ratio of pTR to pPLG. The next day, Day 1, cells from each well were split into two

wells and allowed to grow for another day. On Day 2 media were replaced with fresh

RPMI+++ and cells for one well of each ratio were induced by an addition of tet (1μg/ml) in

the media.

On Day 3, formaldehyde fixation was carried out for all wells, followed by B-Galactosidase

Staining (using the protocol from Invitrogen B-Galactosidase Staining Kit in general). Wells

were covered and left overnight in 37 °C incubator for further staining.

On Day 4, after staining had been carried out at 37 °C for ~ 22 hr, the intensity of blueness of

cells, which represented the degree of LacZ activity in those cells, were assessed visually via a

light microscope, and assigned a numerical score. The score for intensity of blueness range

form 0 to 10, where 0 represented minimal intensity of blueness (that which could be seen in cells transfected with only EGFP, which was a very slight blueness in the background but not in the cells), and 10 represented the intensity of blueness in wells transfected with only pPLG (with no pTR, and hence freely expressing LacZ without repression). The data were as shown in **Table 4.4-1**.

The conclusion derived from this optimisation was that a ratio of 8:1 pTR to pPLG seemed like the lowest ratio for stable repression with tet (1μg/ml) inducibility of pPLG in PC-12 cells.

Table 4.4-1 Optimisation of pTR to pPLG (pcDNA4/TO/myc-His/lacZ) for PC-12.

Ratio of pTR to pPLG.		No induction.		Induction with tet (1μg/ml).		Comments
pTR	pPLG	Well	Intensity of blueness.	Well	Intensity of blueness.	
1	1	1	8	13	6	
2	1	2	8	14	6	
3	1	3	8	15	6	
4	1	4	8	16	4	
5	1	5	8	17	4	
6	1	6	8	18	3	
7	1	7	8	19	3	
8	1	8	8	20	2	Seemed to be lowest ratio for stable repression with tet (1μg/ml) inducibility of pPLG.
9	1	9	5	21	1	
1	0	10	0	22	0	Only 800 ng of pTR was transfected.
0	1	11	10	23	10	Only 800 ng of pPLG was transfected.
-	-	12	0	24	0	Only 800 ng of pEGFP was transfected. All cells exhibited green fluorescence (GF), although ~30% showed particularly bright GF.

4.4.1.3 ECFP and EYFP as PLG in HEK293.

With a guide on the optimum ratio of pTR to pPLG, we proceeded to test fluorescent proteins (FP) as PLG, which was our objective for a Proof-of-Concept for the inducible gene expression system paradigm of e-SRS. For FP, we tested both ECFP (enhanced cyan fluorescent protein) and EYFP (enhanced yellow fluorescent protein). We iECFP and iEYFP by restriction digest of the respective FP sequences from pECFP and pEYFP and cloning them within the MCS of pcDNA4/TO/myc-His A.

Using a similar transfection schedule as the pcDNA4/TO/myc-His/lacZ above, we optimised for ratio of pTR to pPLG for 6:1, 7:1, 8:1, 9:1.

It was discovered that yellow fluorescence was hard to detect, and that 8:1 ratio of pTR to pPLG gave us the best result for iECFP, where the amount of cyan fluorescence (CF, detected in the ECFP Channel) observed was readily distinguished on Day 3 for wells with tet (1μg/ml) induction (on Day 1), compared to those without tet induction (Figure 4.4-1). It was thus decided to use 8:1 pTR to iECFP for subsequent experiments.

Tet (1μg/ml).	
Channels: ECFP.	Channels: ECFP, DIC.

Tet not added.	
Channels: ECFP.	Channels: ECFP, DIC.
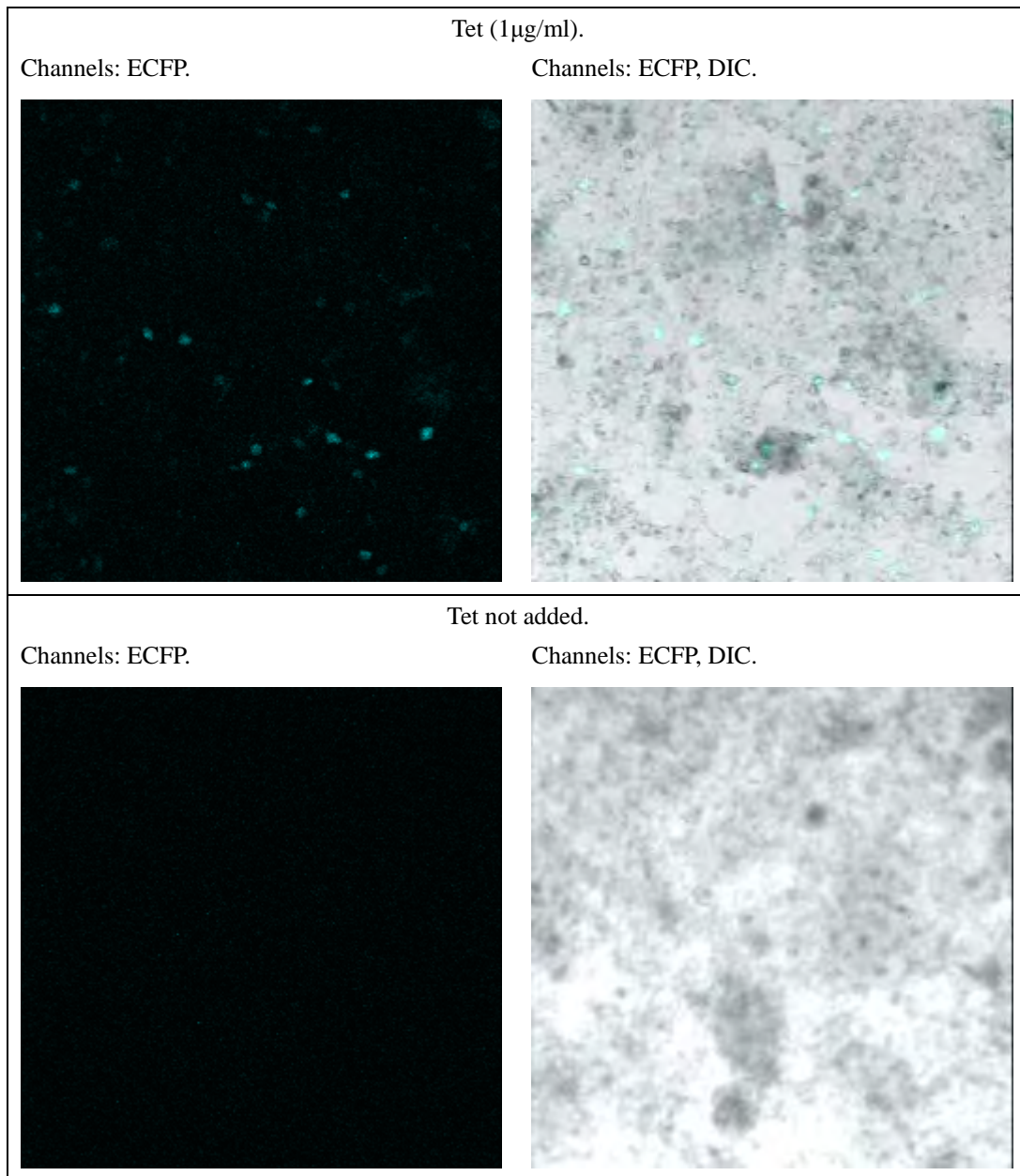
Figure 4.4-1 8:1 ratio of pTR to iECFP with and without Tet (1μg/ml) induction.
DIC, Differential Interference Contrast.

4.4.1.4 Test of Mz-1,2,3 to activate inducible system.

After determining a functional ratio of pTR to pPLG for iECFP, we proceeded to test the

ability of Mz-1, Mz-2, Mz-3 to activate iECFP. At this stage, before the use of RSs, we had

yet to determine which of these Mz based sensor had the best ability to specifically activate in the presence of NTS-37a. As such, we started with testing Mz-1.

4.4.1.4.1 Test of Mz-1 (with pEGFP-Cav2.1EFa / sNTS-37a, or pmCherry-Cav2.1EFb / sNTS-37b) and HH-1 to activate inducible system (8:1 pTR to pcDNA4/TO/ECFP/myc-His A) in HEK293.

We conducted the first cell line test of e-SRS using HEK293 with iECFP as the PLG, Mz-1 as the Mz-based e-SRS sensor, and pEGFP-Cav2.1EFa (which expresses the full length NTS-37a with an EGFP tag) or sNTS-37a as NTS, and pmCherry-Cav2.1EFb (which expresses the full length NTS-37b with an mCherry tag) or sNTS-37b as NNS. In addition, we used HH-1, the wild-type Hammerhead Rz which recognised the same target as the activated Mz-1. For more details on the relationships between the various Mz based sensors, targets, and HH, please see Sections 4.3.5 and 4.3.7.

Parameters tested were as follows:

Test ability to activate 500 ng of iECFP (8:1 pTR to pcDNA4/TO/ECFP/myc-His A) using

1) Mz-1 (20 pmol) with

   a. pEGFP-Cav2.1EFa (500 ng)

   b. pEGFP-Cav2.1EFa (500 ng) and sNTS 37a (20 pmol, 100 pmol)

2) Mz-1 (20 pmol) with

   a. pmCherry-Cav2.1EFb (500 ng)

   b. pmCherry-Cav2.1EFb (500 ng) and sNTS 37b (20 pmol)

3) HH-1 at

a. 20 pmol

b. 100 pmol

We used a similar transfection schedule as the optimisation of iECFP above, we tested the ability of Mz-1 to activate a 8:1 ratio of pTR to pPLG.

On Day -1, HEK293 cells were seeded at 2 x 10^5 cells / well in 24 well plate and grown with DMEM (10 % FBS, w/o P/S). The next day, Day 0, the inducible gene system (i.e. pTR and pPLG), and pEGFP-Cav2.1EFa or pmCherry-Cav2.1EFb were transfected.

On Day 1, wells designated as positive control by tet were induced by an addition of tet (1µg/ml) in the media. Wells designated for RNA transfection (i.e. sNTS/sNNS or Mz-1/HH-1) were also transfected. On Day 2 media were replaced with fresh DMEM (10 % FBS, 1% P/S), with tet (1µg/ml) where required.

On Day 3, cells were viewed via confocal microscopy for Cyan and Green fluorescence (CF and GF, as detected in the ECFP and EGFP Channels respectively), and pictures were taken (some representative pictures were as shown in **Figure 4.4-2**).

The data showed that presence of NTS, sNTS, and HH-1 (both at 20 pmol and 100 pmol) did not seem to produce a visually significant difference for ECFP, compared to Mz-1 alone or Mz-b1a2 alone (Mz-b1a2 does not recognise any NTS and also does not target the R1 mRNA, thus acting as a negative control).

The conclusion derived was that the failure of NTS, sNTS, and HH-1 to increase ECFP indicated failure of Mz-1 to specifically activate the inducible system in HEK293.

Mz-1 (20 pmol), NTS-37a (500 ng).
Channels: ECFP, DIC.   Channels: EGFP, DIC.

Mz-1 (20 pmol), sNTS-37a (100 pmol).
Channels: ECFP, DIC.   Channels: EGFP, DIC.

Mz-1 (20 pmol).
Channels: ECFP, DIC.   Channels: EGFP, DIC.
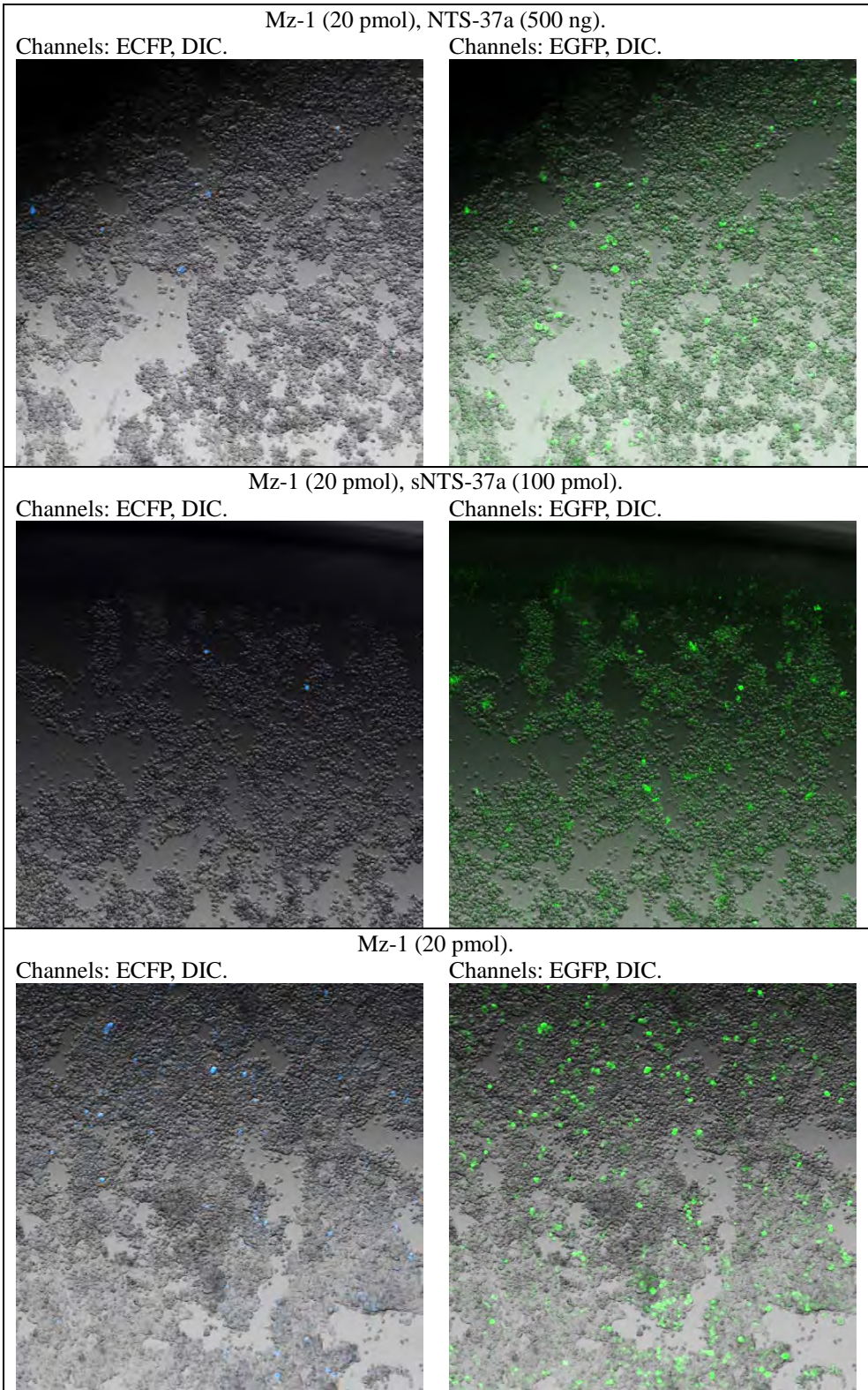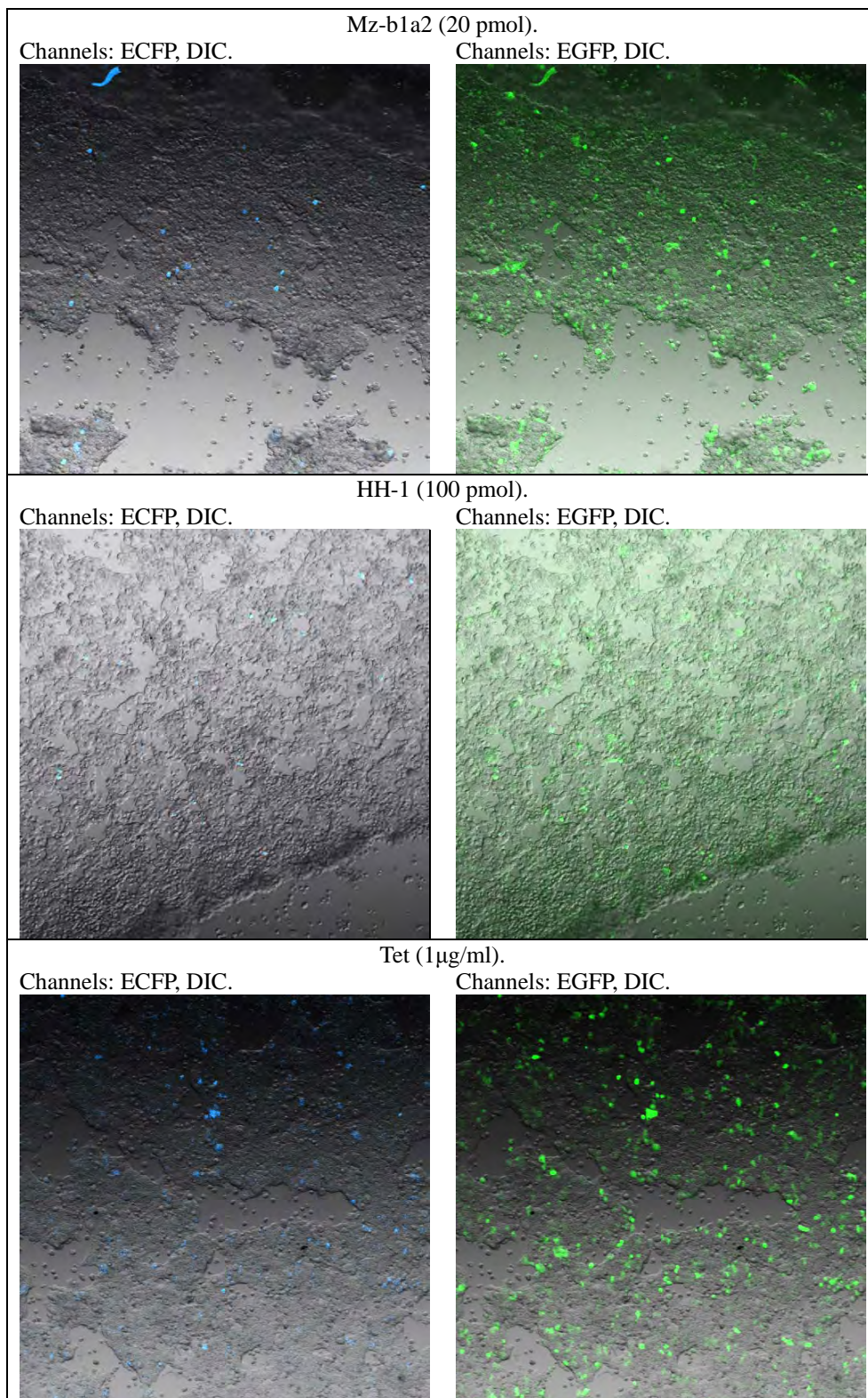
Figure 4.4-2 Mz-1 was unable to specifically activate the inducible system in HEK293.

4.4.1.4.2 **Test of Mz-1,2,3 (with sNTS 37a) and HH-1,2,3 to activate inducible system (iECFP) in HEK293.**
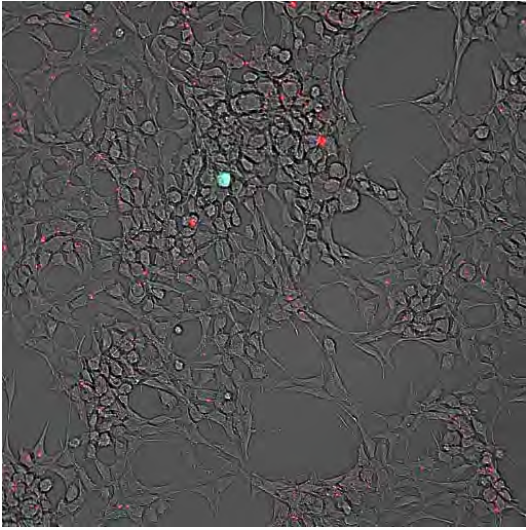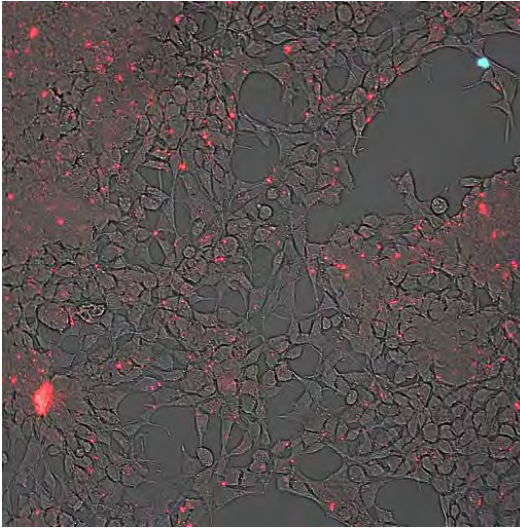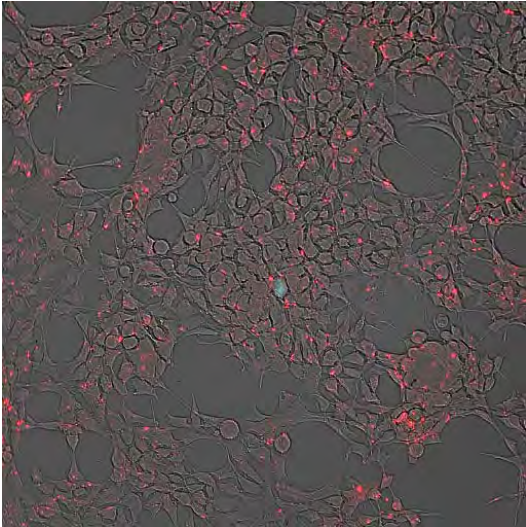
As the full length NTS used in the previous experiment had to be properly expressed after trasnfection and could have secondary structures that complicated the process of detection by the Mz based e-SRS sensors, we decided to first reduce the complexity involved and optimise other aspects of the detection process before reverting to full length NTS. As such we focused on using transfected sNTS (sNTS-37a and sNTS-37b).

Using a similar experimental protocol as the previous experiment, we tested different parameters as shown in Table 4.4-2. Briefly, on Day -1, HEK293 cells were seeded at a density equivalent to about 1×10^5 cells / well in 24 well plate. On DAY 0, the inducible gene system were transfected. On Day 1, RNA transfection as well as tet (1μg/ml) induction were carried out. On Day 2 old media were replaced with fresh media. Confocal microscopy were used to evaluate iECFP activation starting from Day 2, and pictures were taken on at least one day.

Despite the testing of many conditions, we were unable to obtain significant activation of iECFP by either the Mz based sensors or HH used. Only once on Day 4 for one of the experiments did the overall ECFP activation for Mz-3 and HH-3 (all RNA at 13.3 pmol) seem to fit desired patterns. Unfortunately, these results were not repeatable on subsequent testing of a greater range of concentration for HH-3 (some representative pictures were as shown in Figure 4.4-3). iECFP was inducible for cells induced with Tet (1 μg/ml).

Table 4.4-2 Paremeters and conditions tested for iECFP.

Parameters	Conditions tested
Mz-1.	6.7, 8, 13.3, 20.0, 26.7, 33.3 pmol.
Mz-2, Mz-3.	8, 13.3, 26.7 pmol.
HH-1, HH-2.	8, 13.3, 26.7, 40 pmol.
HH-3.	8, 13.3, 26.7, 53.3, 106.7 pmol.
sNTS-37a / sNTS-37b.	6.7, 8, 13.3, 20.0, 26.7, 33.3 pmol.
Ratio of pTR:pPLG.	8:1, 7:1, 3:1, 1:1.
Mass of inducible system transfected for equivalent of one well in 24 well plate.	1, 2 μg.

HH-3 (26.7 pmol).
Channels: ECFP, Cy5, DIC.

HH-3 (53.3 pmol).
Channels: ECFP, Cy5, DIC.

HH-3 (106.7 pmol).
Channels: ECFP, Cy5, DIC.

| Tet (1μg/ml) | Tet not added. |
| Channels: ECFP, DIC. | Channels: ECFP, DIC. |
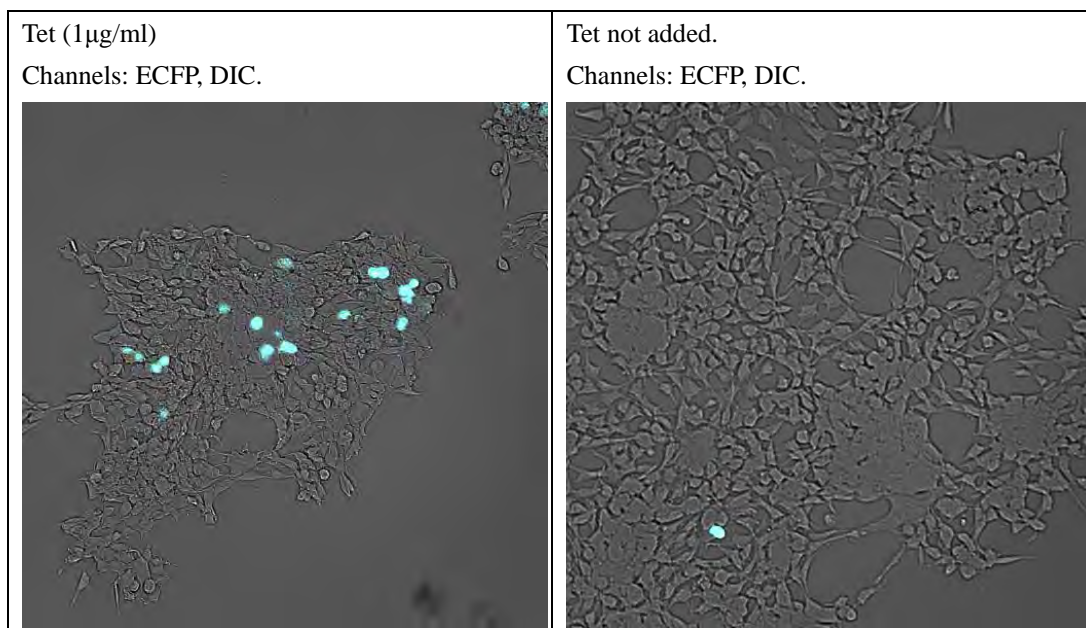
Figure 4.4-3 HH-3 was unable to specifically activate iECFP in HEK293 as CF was not significantly more intense than cells not transfected with HH-3 (Tet not added).
Cells induced with Tet (1 μg/ml) served as positive control for functional iECFP. HH-3 used here was 5' labelled with Cy5 (Red), as such increasing HH-3 transfection concentration increased the amount of HH-3 transfected into cells but did not increase iECFP activation. 1 μg (per well of 24 well plate) of Inducible system was transfected at 7:1 ratio of pTR:pPLG.

4.4.1.5 Switch to the use of HH.

To further focus our efforts to activate the inducible gene system in cell lines, we focused on the use of HH rather than Mz based sensors as HH were simpler constructs with high catalytic activities, and this reduced the need to activate the sensor with sNTS.

4.4.1.5.1 Use of tRNA expression system.

In previous experiments, transfection of HH into HEK293 apparently had no effect in activating the inducible gene system (iECFP). This could have been due to difficulties in maintaining appropriate secondary structures of HH (for e.g. if they form lasting complexes with transfection reagents) after transfection into cells.

As such, we decided to try to express HH-2 from within HEK293, by cloning HH-2 into an expression vector (at this stage, in-vitro experiments as described in Section 4.3.8 had confirmed the better specificity of Mz-2 over Mz-1 and Mz-3). Going through the literature, we discovered that a tRES had been successfully used to express Mz [Oshima et al, 2003] with several advantages compared to a pol II expression system, namely 1) high expression levels of one to three order of magnitude higher than pol II system [Sano & Taira, 2004], 2) ability to transport the expressed RNA into the cytoplasm (where they perform their function) via a tRNA nuclear export system [Kuwabara et al, 2001]. As such, we decided to employ the tRES to express HH-2.

Our tRES is an RNA expression plasmid adapted from [Sano & Taira, 2004], where the RNA to be expressed is placed under a pol III promoter. The version used here is that of a modified human tRNA for Valine (tRNA^{Val}) with the last seven bases removed and replaced with a short linker to prevent 3' end processing. The resulting expressed RNA has a ~ 91 nucleotide tRNA^{Val} attached to the 5' of the RNA. This extra 5' domain should fold like a tRNA with a cloverleaf structure, which facilitates the transport of the entire RNA into the cytoplasm.

4.4.1.5.2 **Design and construction of HH-2-1_tRES.**

We designed a tRES sequence for HH-2, which would express HH-2 with a partial tRNA^{Val} sequence 5' of the HH-2 sequence for the expression benefits as explained above. The final expressed RNA was denoted as HH-2-1_tRES.

The design of HH-2-1_tRES was accomplished as instructed in [Sano & Taira, 2004]. This essentially involved concatenating the following three RNA segments (as shown in **Table 4.4-3**)

together, in 5' to 3' order:

1) tRES 5' sequence, which included the required pol III promoter sequence.

2) Existing HH-2 sequence with one initial U on the Rz binding arm removed as a continuous run of four Us would terminate a pol III transcript.

3) tRES 3' termination sequence, which is four repeats of Us.

Table 4.4-3 Sequences for the design of HH-2-1_tRES.

Constructs	Sequence (5' to 3')	Length (nt)
HH-2-1_tRES	ACCGUUGGUUUCCGUAGUGUAGUGGUUAUCACGUUCGCCUAACAC GCGAA AGGUCCCCGGUUCGAAACCGGGCACUACAAAAACCAACAAAUUUC GACUG AUGAGGCCGAAAGGCCGAAAGUUUCAUAUUUU	132
tRES(5')	ACCGUUGGUUUCCGUAGUGUAGUGGUUAUCACGUUCGCCUAACAC GCGAA AGGUCCCCGGUUCGAAACCGGGCACUACAAAAACCAACAAA	91
HH-2-1	UUUCGACUGAUGAGGCCGAAAGGCCGAAAGUUUCAUA	37
tRES(3')	UUUU	4

As recommended in [Sano & Taira, 2004], we predicted the secondary structure of HH-2-1_tRES to ensure that the critical cloverleaf structure is formed (without which the entire RNA is likely to accumulate in the nucleus, rather than get exported to the cytoplasm where they are supposed to function). As shown in Figure 4.4-4, our HH-2-1_tRES was predicted to form the cloverleaf structure (nucleotides 4 to 88).

In addition, we wanted to further ensure that our HH-2 binding arms could bind correctly and with the right HH Rz structure to the substrate, without being interfered by, nor interfering with the cloverleaf structure. By adding the substrate (RTS-2) sequence 3' to the HH-2-1_tRES sequence and performing a secondary structure prediction for the entire

construct, as shown in Figure 4.4-5, we were assured that this was the case (HH Rz structure

was defined from nucleotides 92 to 128).
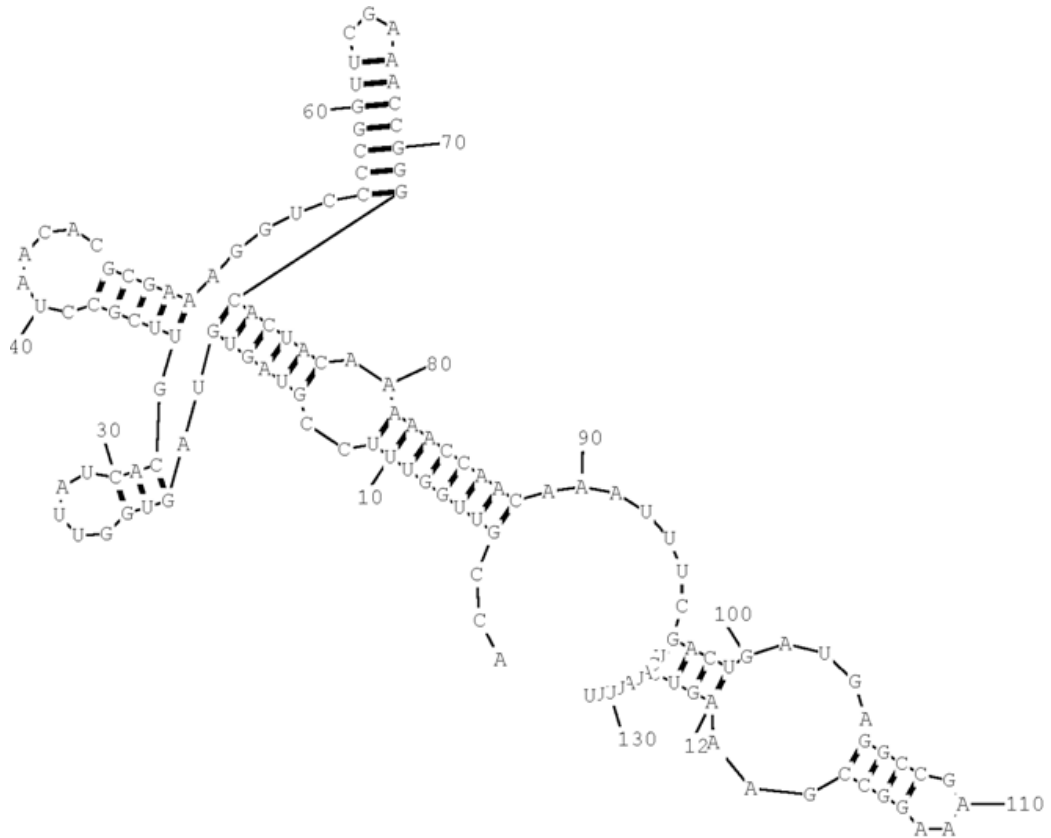
Figure 4.4-4 HH-2-1_tRES was predicted to form the cloverleaf structure (nucleotides 4 to 88).
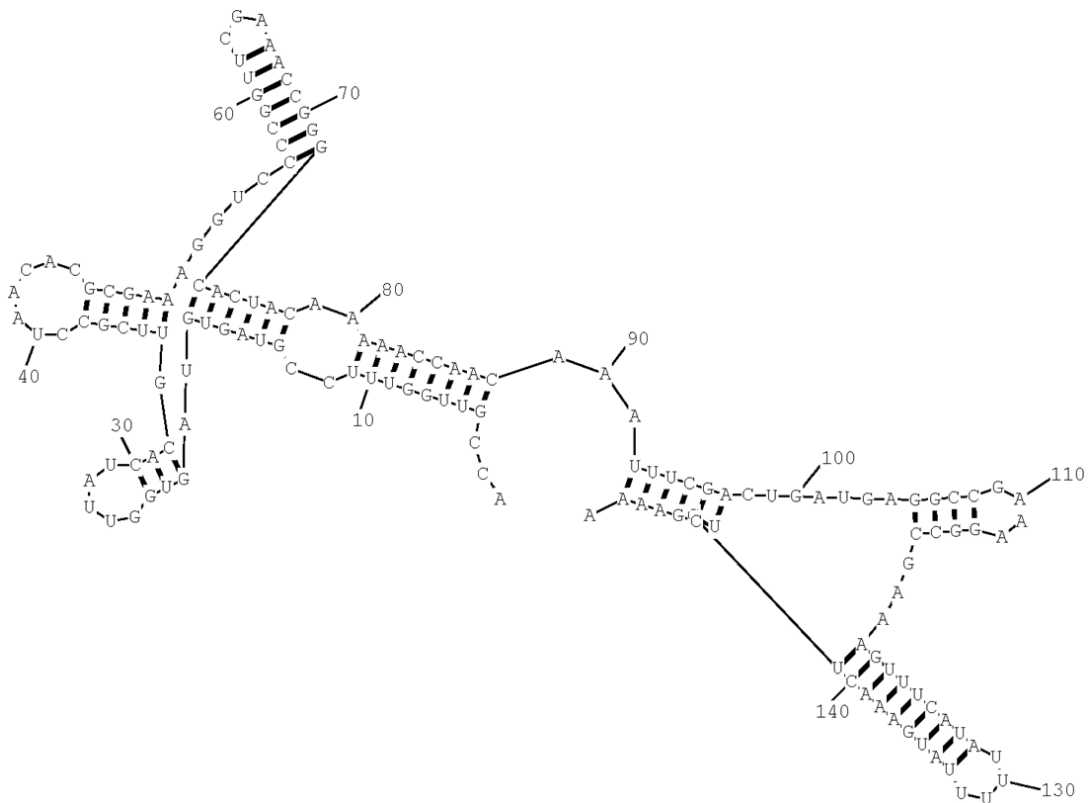
Figure 4.4-5 HH-2-1_tRES was predicted to be able to bind correctly and with the right HH Rz structure (nucleotides 92 to 128) to the substrate.

There was no predicted interference by, nor interference to the cloverleaf structure.

We constructed the double stranded DNA of HH-2-1_tRES with protruding 3' A de novo (using gene synthesis via oligonucleotide ligation, as described in Section 3.1.9) and then inserted it via TA-cloning into pGEM-T Easy vectors to created the plasmid pHH-2-1_tRES. Ligation oligos and their sequences for HH-2-1_tRES were as shown in Figure 4.4-6 and Table 4.4-4.
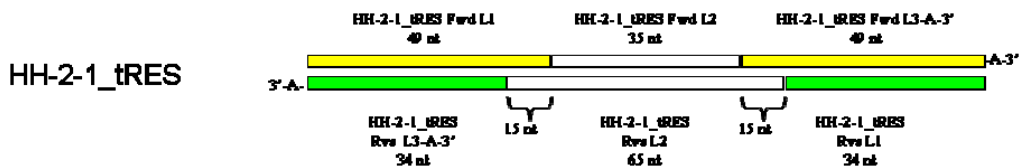
Figure 4.4-6 Ligation oligos configurations for HH-2-1_tRES.

Table 4.4-4 Ligation oligos sequences for HH-2-1_tRES.

Ligation Oligo	Sequence (5' to 3')
HH-2-1_tRES Fwd L1	ACCGTTGGTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCGA
HH-2-1_tRES Fwd L2	AAGGTCCCCGGTTCGAAACCGGGCACTACAAAAAC
HH-2-1_tRES Fwd L3-A-3'	CAACAAATTTCGACTGATGAGGCCGAAAGGCCGAAAGTTTCATATTTTA
HH-2-1_tRES Rvs L1	AAAATATGAAACTTTCGGCCTTTCGGCCTCATCA
HH-2-1_tRES Rvs L2	GTCGAAATTTGTTGGTTTTTGTAGTGCCCGGTTTCGAACCGGGGACCTTTCGCGTGTTAGGCGAA
HH-2-1_tRES Rvs L3-A-3'	CGTGATAACCACTACACTACGGAAACCAACGGTA

4.4.1.5.3 **Test of HH-2-1_tRES to activate iEGFP in HEK293.**

Using a similar experimental protocol as the previous experiment, we tested the ability of pHH-2-1_tRES to activate the inducible system, this time using iEGFP instead of iECFP. We created iEGFP similarly to the way we created iECFP, by cloning an EGFP sequence into pcDNA4/TO/myc-His A.

Briefly, on Day -1, HEK293 cells were seeded at a density equivalent to about 1×10^5 cells / well in 24 well plate. On Day 0, the inducible gene system (with pTR to iEGFP at ratios of 7:1, 3:1, 1:1) was transfected. On Day 1, instead of RNA transfection, pHH-2-1_RES was transfected at 1 and 2 μg, and tet (1μg/ml) induction were carried out. On Day 2 cells in each well were replated into 3 wells (with 1 coverslips) in a 24 wells plate. On Day 3 and Day 4, after formaldehyde fixation, cells on coverslips were mounted on slides using Fluorsave. Fluorescence microscopy was used to evaluate the slides for iEGFP activation, and pictures

were taken on at least one day.

Despite the testing of many conditions, we were unable to observe significant activation of iEGFP in pHH-2-1_tRES transfected wells compared to wells with no Tet induction (some representative pictures were as shown in Figure 4.4-7). iEGFP was inducible for cells induced with Tet (1 µg/ml).

HH-2-1_tRES 1 µg 7:1 1 µg iEGFP.
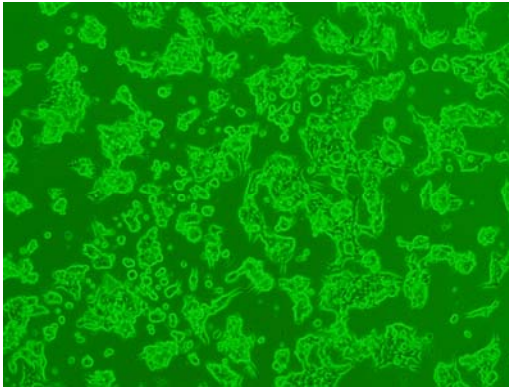
Channels: DIC.   Channels: EGFP.

HH-2-1_tRES 2 µg 7:1 1 µg iEGFP.
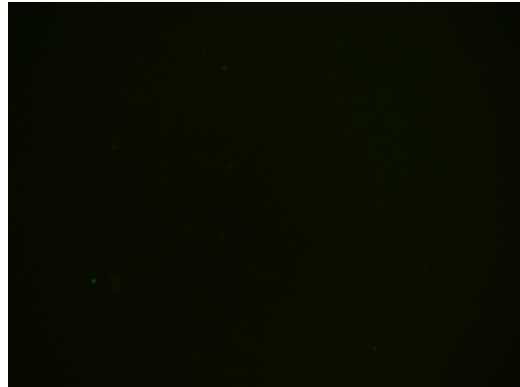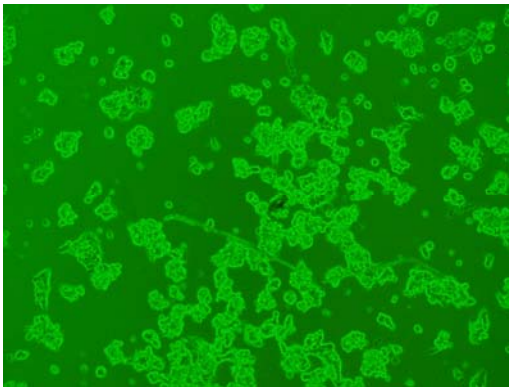
Channels: DIC.   Channels: EGFP.
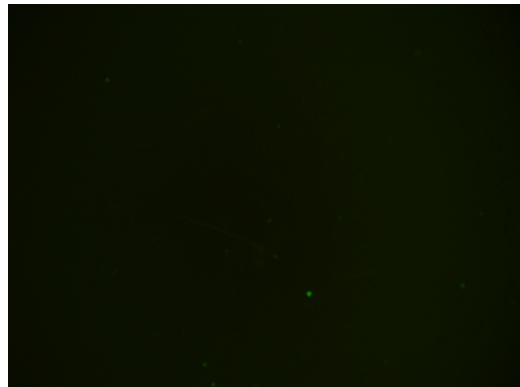
Tet not added.

Channels: DIC.   Channels: EGFP.

Tet (1µg/ml).
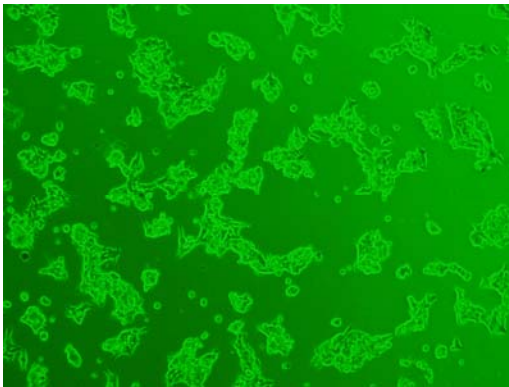
Channels: DIC.   Channels: EGFP.
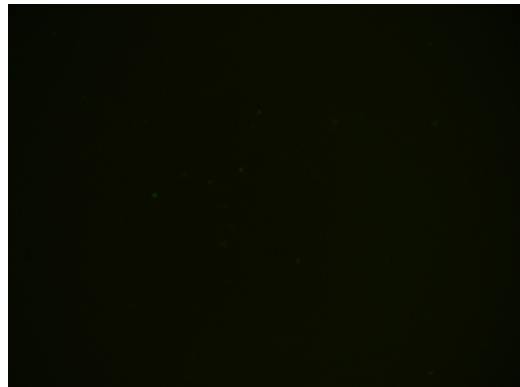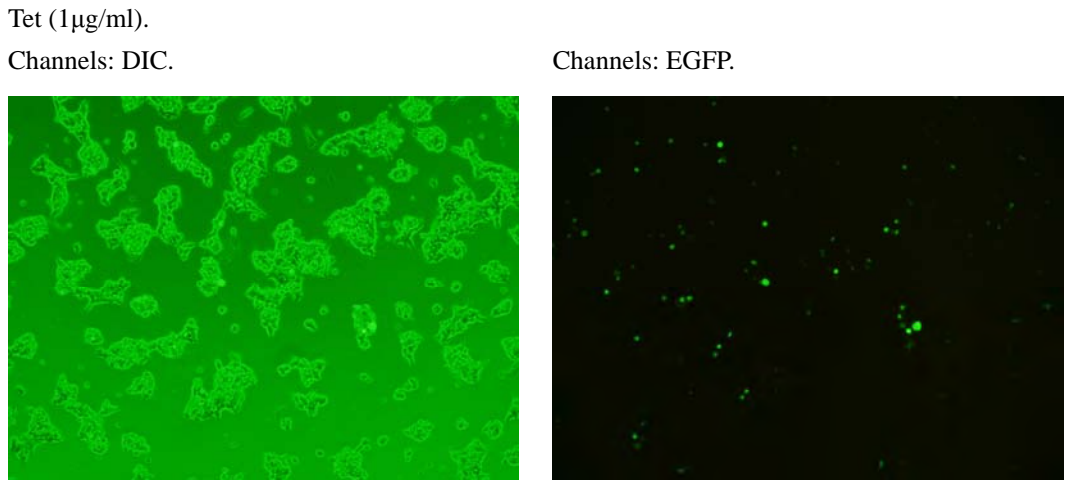
Figure 4.4-7 HH-2-1_tRES was unable to specifically activate iEGFP in HEK293.
This could be seen as GF was not significantly different from cells not transfected with
HH-2-1_tRES and with Tet not added. Cells induced with Tet (1 µg/ml) served as positive
control for functional iEGFP. Pictures shown here were from Day 3, for wells with 1 µg inducible
system transfected at 7:1 ratio of pTR:pPLG. Wells at other ratios did not have better results.

4.4.1.5.4 **RT-PCR of HH-2-1_tRES in transfected HEK293 cells.**

We decided to check if the lack of activation of iGFP by pHH-2-1_tRES could have been due

to the HH-2-1_tRES RNA not being expressed in cells. To investigate this possibility, we

scaled up the transfection of pHH-2-1_tRES in order to harvest sufficient cells for RNA

isolation and subsequent RT-PCR to determine level of HH-2-1_tRES expression, but

otherwise maintained the same or similar transfection protocol, such as DNA mass to cells

transfected.

Briefly, on Day -1, HEK293 cells were seeded at 7.5×10^5 cells / well for 6 wells plate (a

density equivalent to about 1.5×10^5 cells / well in 24 well plate). On Day 0, pHH-2-1_RES

was transfected at 5 µg / well (equivalent to 1 µg / well for 24 well plate).

On Day 1, Day 2, Day 3, Day 4 transfected cells were harvested for RNA isolation. For cells

harvested on Day 3 and Day 4, they were replated on Day 1 and Day 2 respectively, in order

to maintain similar phase of growth and cell density on day of harvest. Cells on Day 2 had 2 days of growth after transfection of pHH-2-1_tRES, similar to the Day 3 cells in the previous experiment. On Day 2, untransfected cells were also harvested for HEK293 RNA isolation. Harvested cells were pelleted and kept at -80 °C

RNA was isolated from all the harvested cells on the same day. RT-PCR was performed for HH-2-1_tRES and β-Actin (βA) as a positive control. Primer information and RT-PCR conditions were as detailed in Methodology. For each gene, a "-RT" control was done where the same procedure was used except that an equal volume of H_2O was added in place of the Reverse Transcriptase (RT) enzyme. This control would indicate whether (and the approximate amount if so) the amplified products were due to contaminating DNA sequences (plasmid or genomic DNA) in the isolated RNA. A "H_2O" or water control was performed for each gene where an equal volume of H_2O was added in place of the RNA sample to ensure that no contaminants in the reagents gave rise to amplified products.

HH-2-1_tRES primers were designed to give a 134 nt product while the β-Actin primers were designed to give a 643 nt product. RT-PCR products were analysed on an agarose gel stained with EtBr, as shown in Figure 4.4-8.
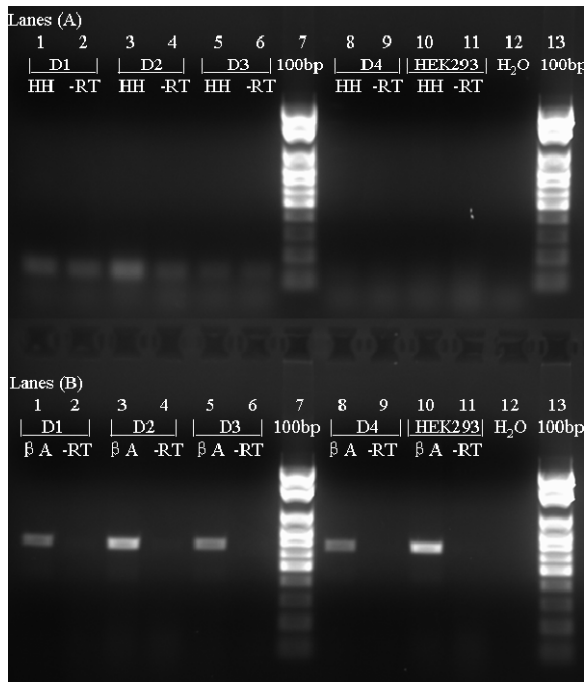
Figure 4.4-8 Agarose gel analysis of RT-PCR products for HEK293 cells transfected with pHH-2-1_tRES on Day 0.

For HH-2-1_tRES, Day 1, Day 2, Day 3 RNA all showed the expected product bands. However, these products were also shown in the respective –RT lanes, which indicated that there were likely some DNA amplification from transfected plasmids present in the isolated RNA.

For Day 2, the RT lane product seemed much more intense than that of the –RT lane, compared to those of Day 1 and Day 3. The β-Actin band was also more intense on Day 2 than Day 1 and Day 3, indicating that more RNA was likely to have been harvested on Day 2, as expected with normal cell proliferation. This indicated that HH-2-1_tRES RNA was a able to accumulate from Day 1 to Day 2, in tandem with normal cell proliferation. Since there was an increase in the ratio of RT lane band intensity to that of –RT lane, some of the RT lane product in Day 2 was likely amplified from expressed RNA, and that the amount of

HH-2-1_tRES RNA per well was higher on Day 2, compared to Day 1 and Day 3.

Day 4 RNA showed no significant HH-2-1_tRES products, although all β-actin RT lanes had

products, indicating that RNA were isolated from all samples and that HH-2-1_tRES RNA

and its plasmid were no longer present in Day 4 cells.

RNA from untransfected HEK293 cells did not give any HH-2-1_tRES product, indicating

that the Day 1 to Day 3 products were not from some native genes in HEK293 cells.

There were no products in H_2O lane, indicating that the products obtained in all lanes were not

due to contamination in the reagents.

From these results, we concluded that HH-2-1_tRES RNA was likely expressed in

pHH-2-1_tRES transfected HEK293 cells, and continued to be expressed on Day 2 (i.e. 2nd

day after transfection – the day of transfection was Day 0). However, it could be possible that

the level of expression was insufficient to activate the iEGFP system. This could also be due

to rapid removal of the expressed RNA as Day 3 cells already showed no significant presence

of HH-2-1_tRES RNA. As such, we decided to try to improve expression levels of

HH-2-1_tRES by including a T7 RNA Pol III promoter 5' of the HH-2-1_tRES sequence in

the pHH-2-1_tRES plasmid.

4.4.1.5.5 Design and construction of HH-2-2_tRES

We created the HH-2-2_tRES construct by adding a T7 RNA Pol III promoter 5' of the

HH-2-1_tRES sequence in the pHH-2-1_tRES plasmid, thus creating the plasmid,

pHH-2-2_tRES.

This essentially involved concatenating the following two RNA segments (as shown in **Table**

4.3-7) together, in 5' to 3' order:

1) T7 RNA Pol III promoter sequence (T7 Pol), from -17 to +2.

2) The original HH-2-1_tRES sequence.

T7 Pol was the minimum sequence required for efficient transcription. RNA Transcription with T7 Pol started at +1, hence the expressed RNA, HH-2-2_tRES would have an extra GG at 5' compared to HH-2-1_tRES. Like in the case of HH-2-1_tRES, transcription termination would occur after the four runs of U at the end of the sequence.

Table 4.4-5 Sequences for the design of HH-2-2_tRES.

Constructs	Sequence (5' to 3')	Length (nt)
HH-2-2_tRES	TAATACGACTCACTATAGGACCGUUGGUUUCCGUAGUGUAGUGGU UAUCACGUUCGCCUAACACGCGAAAGGUCCCCGGUUCGAAACCGG GCACUACAAAAACCAACAAAUUUCGACUGAUGAGGCCGAAAGGCC GAAAGUUUCAUAUUUU	151
T7 Pol	TAATACGACTCACTATAGG	19
HH-2-1_tRES	ACCGUUGGUUUCCGUAGUGUAGUGGUUAUCACGUUCGCCUAACAC GCGAAAGGUCCCCGGUUCGAAACCGGGCACUACAAAAACCAACAA AUUUCGACUGAUGAGGCCGAAAGGCCGAAAGUUUCAUAUUUU	132

As in the case for HH-2-1_tRES we predicted the secondary structure (**Figure 4.4-9**) of HH-2-2_tRES to ensure that even with the additional two bases at 5', the critical cloverleaf structure (nucleotides 6 to 90) was formed, and that the HH-2 binding arms could bind correctly and with the right HH Rz structure (defined from nucleotides 94 to 130) to the substrate, without being interfered by, nor interfering with the cloverleaf structure.
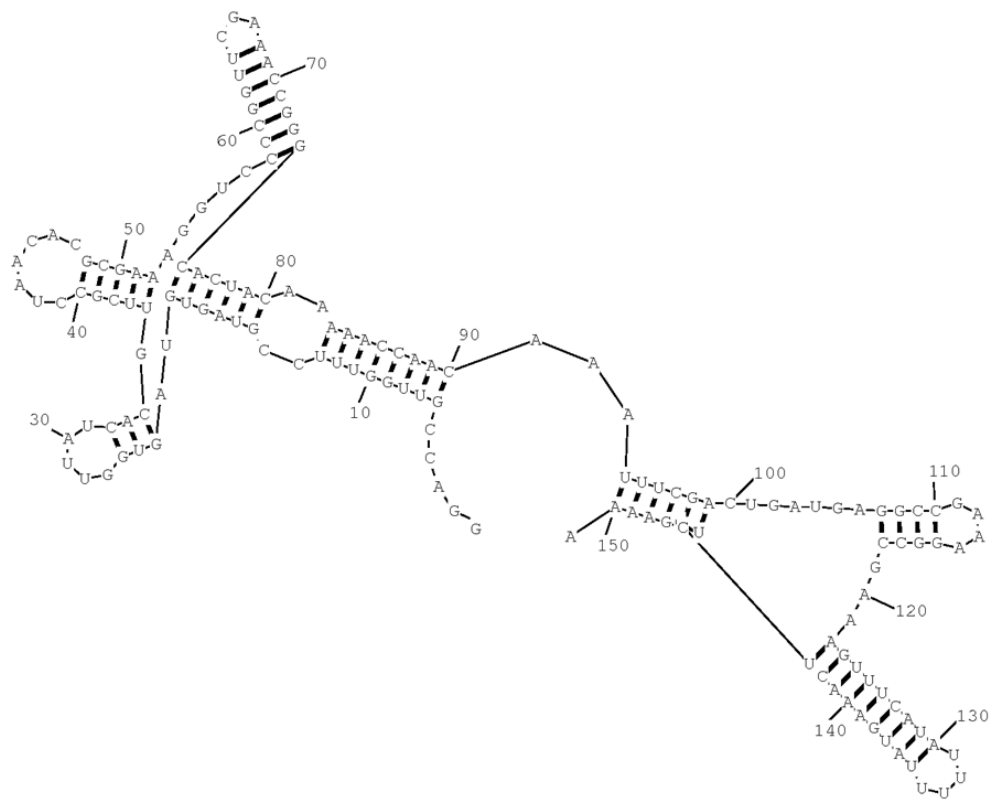
Figure 4.4-9 HH-2-2_tRES was predicted to be able to bind correctly and with the right HH Rz structure (nucleotides 94 to 130) to the substrate.
There was no predicted interference by, nor interference to the cloverleaf structure.

We synthesised the double stranded DNA of HH-2-2_tRES by adding via PCR a T7 Pol sequence (19 nt) to the 5' end of the PCR template, which was the double stranded DNA sequence of HH-2-1_tRES (obtained by PCR from pHH-2-1_tRES, using primers HH-2-1_tRES Fwd L1 and HH-2-1_tRES Rvs L1). The forward primer contained the T7 Pol sequence with a short 3' sequence matching the 5' start of the HH-2-1_tRES sense sequence, and the reverse primer (HH-2-1_tRES Rvs L1) matched the 5' end of the HH-2-1_tRES antisense sequence. The double stranded HH-2-2_tRES was then gel purified and inserted via TA-cloning into pGEM-T Easy vectors to create the plasmid pHH-2-2_tRES. Primer and template sequences were as shown in Table 4.4-6.

Table 4.4-6 Primer and template sequences for PCR synthesising HH-2-2_tRES.

Construct	Sequence (5' to 3')
HH-2-1_tRES (Sense)	ACCGTTGGTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTAACACGCGAA AGGTCCCCGGTTCGAAACCGGGCACTACAAAAACCAACAAATTTCGACTG ATGAGGCCGAAAGGCCGAAAGTTTCATATTTTA
HH-2-1_tRES (Antisense)	AAAATATGAAACTTTCGGCCTTTCGGCCTCATCAGTCGAAATTTGTTGGT TTTTGTAGTGCCCGGTTTCGAACCGGGGACCTTTCGCGTGTTAGGCGAAC GTGATAACCACTACACTACGGAAACCAACGGTA
HH-2-2_tRES Fwd	TAATACGACTCACTATAGGACCGTTGGTTT
HH-2-1_tRES Rvs L1	AAAATATGAAACTTTCGGCCTTTCGGCCTCATCA

4.4.1.5.6 Test of HH-2-2_tRES to activate iEGFP in HEK293.

Using a similar experimental protocol as that for HH-2-1_tRES, we tested the ability of pHH-2-2_tRES to activate the iEGFP inducible system.

Briefly, on Day -1, HEK293 cells were seeded at a density of about 1 x 10^5 cells / well in 24 well plate. On Day 0, 1 µg of the iEGFP inducible gene system (with pTR to iEGFP at ratios of 7:1) was transfected. On Day 1, pHH-2-2_RES was transfected at 1 and 2 µg. On Day 2, fresh media was introduced and tet (1µg/ml) induction was carried out. On Day 2, Day 3 and Day 4, cells were examined for iEGFP activation via Fluorescence microscopy, and pictures were taken on at least one day.

As for the case of HH-2-1_tRES, we were unable to observe significant activation of iEGFP in both 1 µg and 2 µg pHH-2-2_tRES transfected wells compared to wells with no Tet induction (some representative pictures were as shown in Figure 4.4-10). iEGFP was inducible for cells induced with Tet (1 µg/ml).
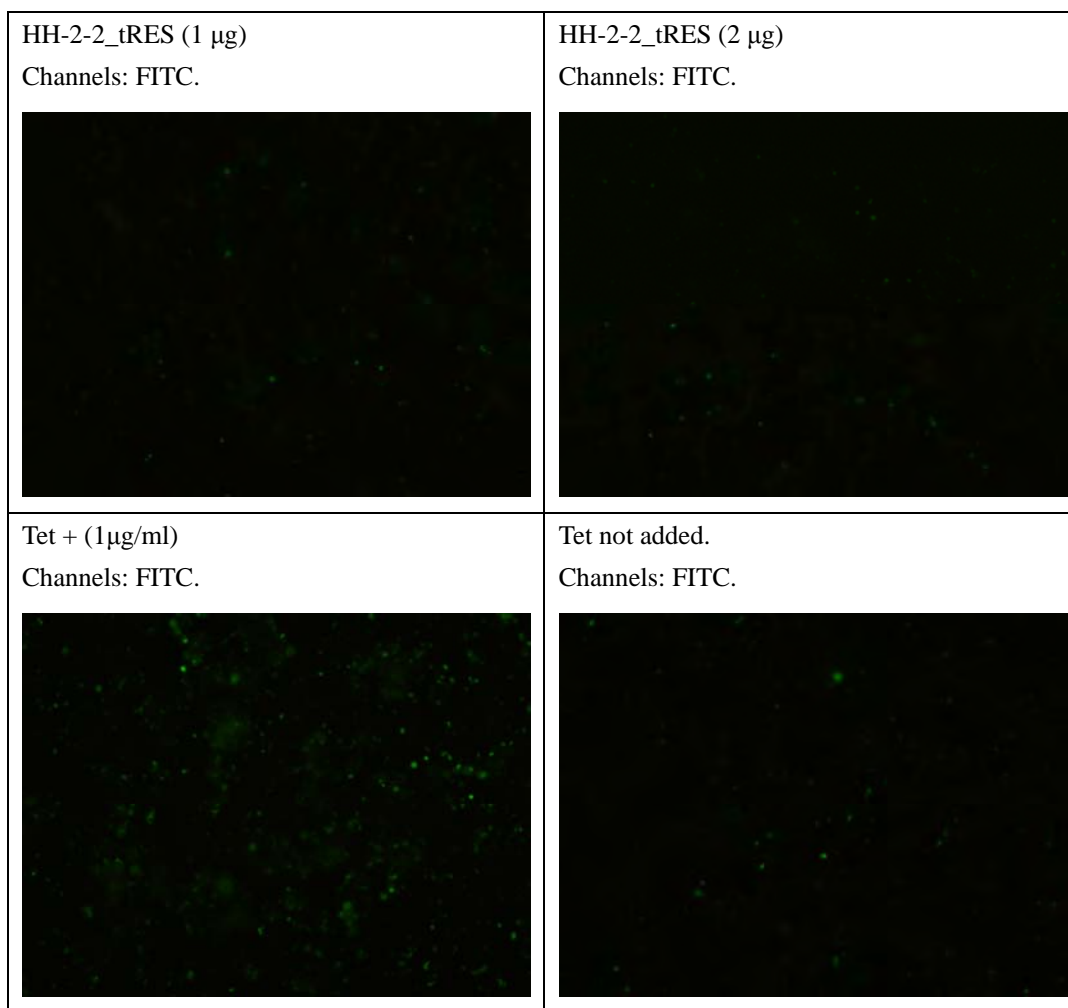
Figure 4.4-10 HH-2-2_tRES was unable to specifically activate iEGFP in HEK293 as GF was not significantly different from cells not transfected with HH-2-2_tRES and with Tet not added.

Cells induced with Tet (1 μg/ml) served as positive control for functional iEGFP. Pictures shown here were from Day 2. Observations on other days did yield better results.

4.4.2 Test of RS in place of inducible gene system.

For paradigm 2 of e-SRS (i.e. to activate a RS if and only if NTS is present), we attempted to transfect both e-SRS sensors and RSs into cells to allow detection of specific RNA within cells. However, the RSs degraded when transfected into culture vessels with cells grown and were thus not usable. The problem was not resolved by the use of nuclease resistant modifications, such as phosphorothioate bonds and addition of 2'-O- methyl groups.

4.4.2.1 Test of Mz-1,2,3 to activate RS in HEK293.

We tested the ability of Mz-1, Mz-2, Mz-3 to activate their respective RS within HEK293 cells. The concepts of Mz and RS were explained in Section 4.3, particularly Sections 4.3.1 to 4.3.3, and 4.3.7. Sequences were given in **Table 4.3-2** for RS and **Table 4.3-8** for Mz. For experiments in this Section, the RS used (RTS-1, RTS-2, RTS-3, and RTS-b1a2) were labelled at 5' with 6-FAM, and at 3' with BHQ1. When viewed with a fluorescence microscope, the cleavage of such RS would give rise to green fluorescence (GF) due to the conjugated 6-FAM. We tested the ability of Mz and HH to activate their respective RS (5' 6-FAM, 3' BHQ1) using:

1) Mz-1, Mz-2 (40 pmol) with

   a. sNTS 37a (40 pmol)

   b. sNNS 37b (40 pmol)

   c. RS

      i. 40 pmol

      ii. 400 pmol

2) Mz-3, Mz-b1a2 (40 pmol) with

   a. sNTS 37a (40 pmol)

   b. sNNS 37b (40 pmol)

   c. RS

      i. 400 pmol

3) HH-1, HH-2 (40 pmol) with

a. RS

    i. 40 pmol

    ii. 400 pmol

4) HH-3, HH-b1a2 (40 pmol) with

    a. RS (400 pmol)

Briefly, on Day -1, HEK293 cells were seeded at a density of about 1.5×10^5 cells / well in 24 well plate. On Day 0, RNA were transfected. On Day 1, fresh media was introduced. Cells were examined for RS activation via Fluorescence microscopy. Visual observations were made on both days and important results were recorded.

On Day 0, in the 1 hour immediately after transfection, all wells transfected with RS started showing some percentage of cells with GF, which increased with time until about 10 % of cells (for those with 400 pmol RS) were GF (percentage and GF intensity were less for samples with 40 pmol RTS). Since this included cells transfected with only RS (no Mz), the observation suggested that the GF observed were not due to activated Mz cleavage of RS. Control wells not transfected with RS were not GF.

On Day 1, after the change of media, it appeared that the major GF difference between wells seemed to be the amount of RS transfected, i.e., whether or not Mz (with sNTS or sNNS) was present did not seem to make a difference to GF detected. In general, for cells transfected with:

1) 40 pmol RS, ~ 20% of cells were GF, with weak intensity.

2) 400 pmol RS, greater than 90% of cells were GF, with medium intensity.

These observations strongly suggested that the GF produced were due to cleavage of RS

without the need for any activated Mz based sensors.

It was concluded that the RS transfected (synthesised from standard ribonucleotide bases) were most likely degraded, probably by nucleases present in the cell cultures.

4.4.2.2 Test of Mz-2, HH-2, and HH-2-2_tRES to activate RS (nuclease resistant) in HEK293.

In vitro testing of the various Mz based sensor had at this point revealed that Mz-2 had the best ability to specifically activate its RS in the presence of NTS-37a. As such, we focused our RS based cell line experiments on Mz-2, HH-2, and HH-2-2_tRES, all designed to activate specifically RTS-2.

In addition, as the findings from the previous section showed that RS made of standard ribonucleotide bases were easily degraded after transfection, we decided to confer nuclease resistant properties to the RS by using phosphorothioate bonds and 2'-O-methyl RNA bases where we believed they would not significantly interfere with Rz cleavage of the RS. As a result, we created a new RS with such modifications, named RTS-2_M-P2, based on the sequence of RTS-2. Sequence of RTS-2_M-P2 was as shown in Table 4.4-7. RTS-2_M-P2 was 17 nt, was labelled at 5' with Cy3, and at 3' with BHQ2. When viewed with a fluorescence microscope, its cleavage would give rise to red fluorescence (RF) due to the conjugated Cy3.

Table 4.4-7 Sequence of RTS-2_M-P2.

Sequence (5' to 3')
(Cy3)-mU*mA*mU*mG*mA*mA*rA*rC*rU*rCrU*rC*mG*mA*mA*mA*mA-(BHQ2)

Legend: *, Phosphorothioate Bond (PS); mN, 2'-O-methyl RNA base; rN - RNA base.

In case the RNA sNTS transfected were in some way inactivated (e.g. by degradation by

nuclease, or clumping to proteins), we also transfected some wells with NTS or NNS in the form of pECFP-Cav 2.1EFa or pECFP-Cav 2.1EFb, which would express the full length NTS or NNS. These plasmids were similar to the pEGFP-Cav2.1EFa and pmCherry-Cav2.1EFb used previously (Section 4.4.1.4.1) except that the EGFP and mCherry sequences were removed by restriction digest and the ECFP sequence from pECFP inserted in their places.

We tested the ability of HH-2-2_tRES, Mz-2, and HH-2 to activate their RTS-2_M-P2 (5' Cy3, 3' BHQ-2) using:

1) HH-2-2_tRES at

    a.   1, 2 μg

          i.   RS 40 pmol

2) Mz-2 (40 pmol) at

    a.   pECFP-Cav 2.1EFa / pECFP-Cav 2.1EFb (1 μg), Nothing

          i.   RS 4 pmol

          ii.   RS 40 pmol

          iii.   RS 400 pmol

    b.   sNTS / sNNS (40 pmol).

          i.   RS 4 pmol

          ii.   RS 400 pmol

3) HH-2 (40 pmol) at

    a.   RS 4 pmol

    b.   RS 400 pmol

4) Nothing at

    a. RS 4 pmol

    b. RS 40 pmol

    c. RS 400 pmol

Briefly, on Day -1, HEK293 cells were seeded at a density of about 1.0×10^5 cells / well in 24 well plate. On Day 0, HH-2-1_tRES (at 1 and 2 μg), pECFP-Cav 2.1EFa (1 μg), and pECFP-Cav 2.1EFb (1 μg) were transfected for specific wells. On Day 1, RNA (including RS) were transfected. On Day 2, fresh media was introduced. On Day 2, Day 3, and Day 4, cells were examined for RS activation via Fluorescence microscopy, and pictures were taken on at least one day (some representative pictures were as shown in Figure 4.4-11, Figure 4.4-12, and Figure 4.4-14).

HH-2-2_tRES (1 μg), RTS-2_M-P2 (40 pmol). Channels: WIY.	HH-2-2_tRES (2 μg), RTS-2_M-P2 (40 pmol). Channels: WIY.
RTS-2_M-P2 (40 pmol). Channels: WIY.	
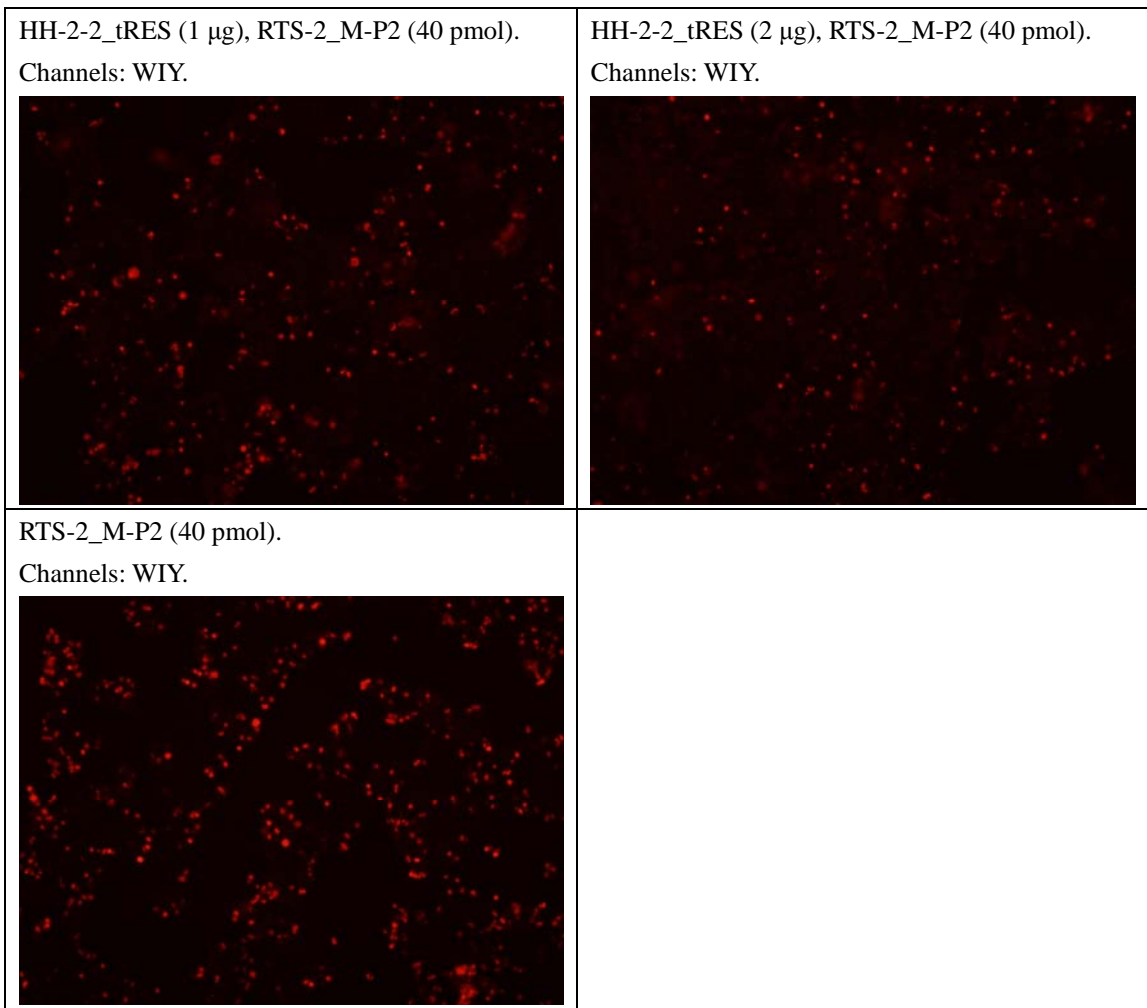
Figure 4.4-11 Test of HH-2-2_tRES to activate RTS-2_M-P2.

Similar activation of RTS-2_M-P2 (RF) was observed in all wells. The amount of HH-2-2_tRES transfected (0, 1 or 2 μg) did not appear to change activation of RS. Pictures shown here were from Day 2.

Mz-2 (40 pmol), NTS (1 µg), RTS-2_M-P2 (400 pmol). Channels: WIY.	Mz-2 (40 pmol), NNS (1 µg), RTS-2_M-P2 (400 pmol). Channels: WIY.
Mz-2 (40 pmol), RTS-2_M-P2 (400 pmol). Channels: WIY.	RTS-2_M-P2 (400 pmol). Channels: WIY.
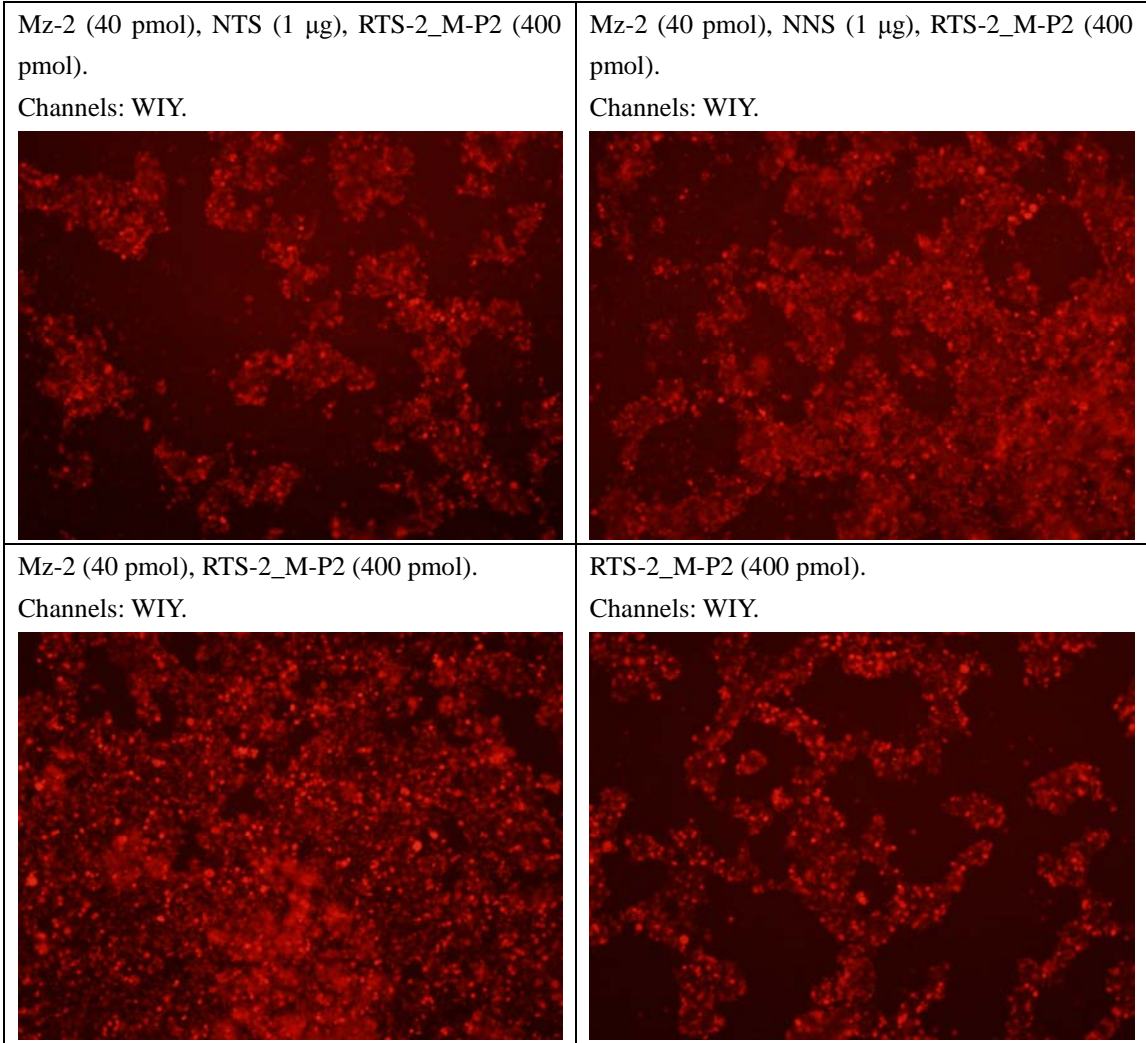
Figure 4.4-12 Test of Mz-2 with NTS-37a or NTS-37b to activate RTS-2_M-P2.
Similar activation of RTS-2_M-P2 (RF) was observed in all wells. The presence of Mz-2, NTS, and NNS did not appear to affect the activation of RTS-2_M-P2. All concentrations of RTS-2_M-P2 showed similar trends. Pictures shown here were from Day 2.

Mz-2 (40 pmol), sNTS (40 pmol), RTS-2_M-P2 (400 pmol). Channels: WIY.	Mz-2 (40 pmol), sNNS (40 pmol), RTS-2_M-P2 (400 pmol). Channels: WIY.
Mz-2 (40 pmol), RTS-2_M-P2 (400 pmol). Channels: WIY.	RTS-2_M-P2 (400 pmol). Channels: WIY.
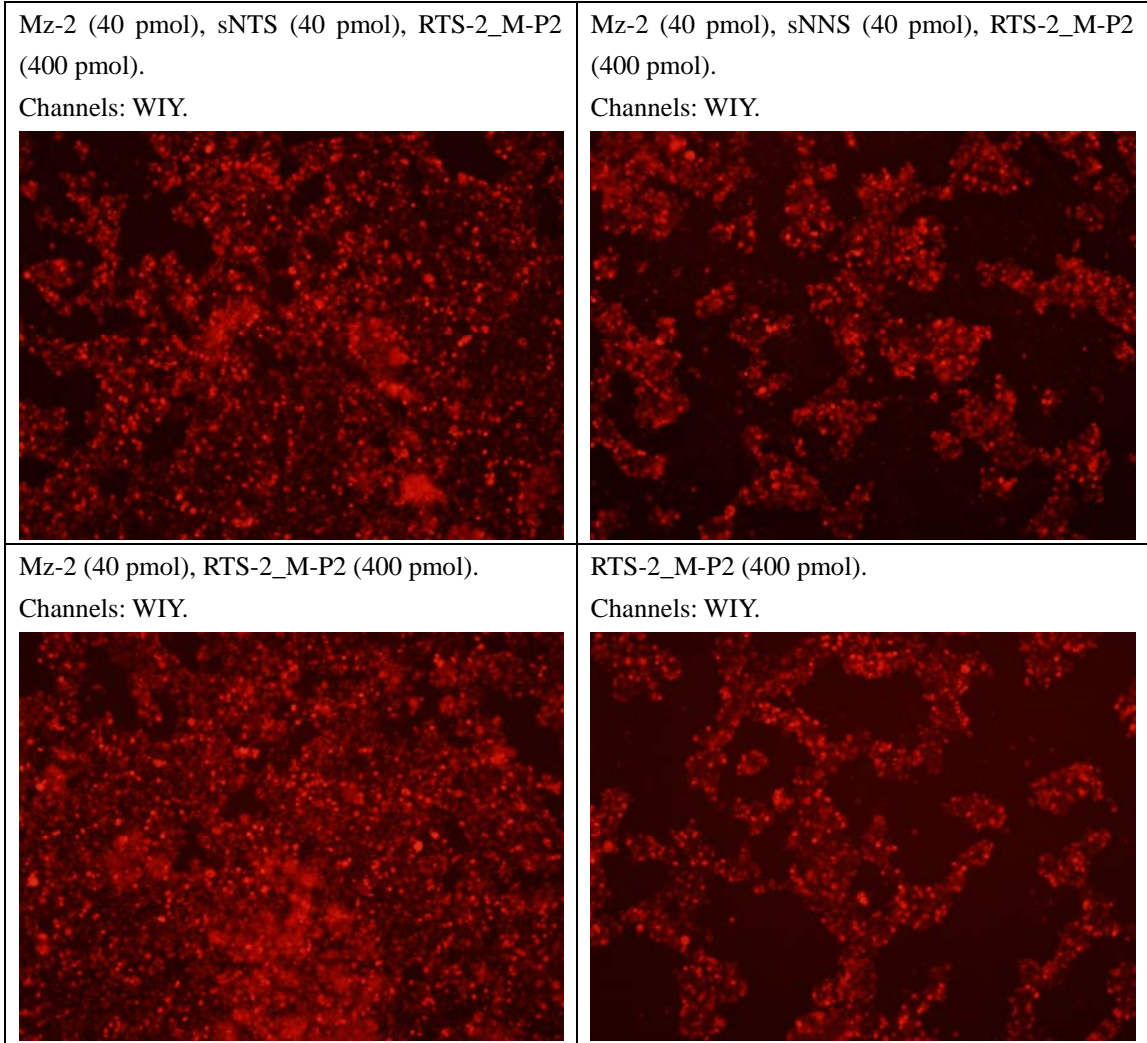
Figure 4.4-13 Test of Mz-2 with sNTS-37a or sNTS-37b to activate RTS-2_M-P2.
Similar activation of RTS-2_M-P2 (RF) was observed in all wells. The presence of Mz-2, sNTS, and sNNS did not appear to affect the activation of RTS-2_M-P2. All concentrations of RTS-2_M-P2 showed similar trends. Pictures shown here were from Day 2.

HH-2 (40 pmol), RTS-2_M-P2 (4 pmol). Channels: WIY.	HH-2 (40 pmol), RTS-2_M-P2 (400 pmol). Channels: WIY.
RTS-2_M-P2 (4 pmol). Channels: WIY.	RTS-2_M-P2 (400 pmol). Channels: WIY.
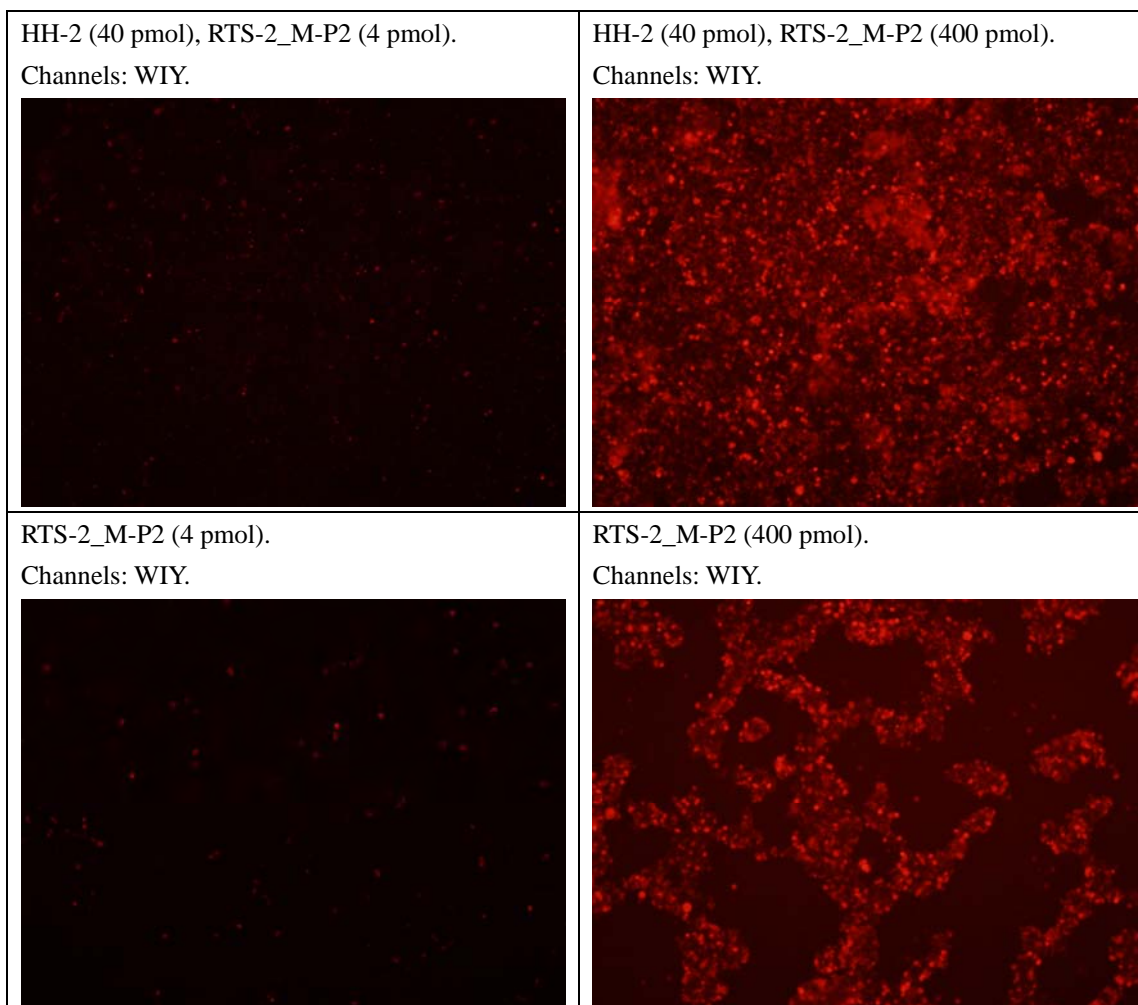
Figure 4.4-14 Test of HH-2 to activate RTS-2_M-P2.

Similar activation of RTS-2_M-P2 (RF) was observed in all wells. The presence of HH-2 did not appear to affect the activation of RTS-2_M-P2. All concentrations of RTS-2_M-P2 showed similar trends. Pictures shown here were from Day 2.

These observations strongly suggested that the RF produced were due to cleavage of RS

(RTS-2_M-P2) without the need for any activated Mz based sensors, HH-2, or HH-2-2_tRES.

It was concluded that even with significant nuclease resistant protection, the RS RTS-2_M-P2

was still degraded after transfection, probably by nucleases present in the cell cultures.

4.4.2.3 Test of transfection components to activate RTS-2_M-P2 and RTS-2 (RTA).

If we were able to limit the agent(s) responsible for degrading the RS to certain components of the transfection process, then it might still be possible to alter the transfection process so that such agent(s) were removed from contact with the RS until the RS were mostly transfected into cells. Hence, we proceeded to test transfection elements for their ability to degrade RTS-2_M-P2 and RTS-2.

Essentially, both RS (RTS-2_M-P2 and RTS-2) were added to agent(s) involved in transfection in a 96 well plate, and viewed with a fluorescence microscope. Activation (i.e. degradation) of the RS would be indicated by red fluorescence (RF) or green fluorescence (GF) for RTS-2_M-P2 and RTS-2 respectively. Pictures were taken as shown in Figure 4.4-15. RTS-2_M-P2 (5' Cy3, 3' BHQ-2) and RTS-2 (5' 6-FAM, 3' BHQ-1) were tested for activation when the following were added:

1) DMEM

2) Opti-MEM

3) Lipoffectamine 2000

4) Opti-MEM with Lipoffectamine 2000

All RS were added at 5 pmol (1000 nM). Where Lipoffectamine 2000 was added, it was added at a ratio (as for RNA transfection before) of 5 pmol RNA to 0.25 μl LF2000. Opti-MEM was added at 3.5 μl per well where needed. All wells were top up to 5 μl with DEPC treated water where needed.

RTS-2_M-P2 (5 pmol). Channels: WIY.			
DMEM	Opti-MEM	Lipoffectamine 2000	Opti-MEM with Lipoffectamine 2000

RTS-2 (5 pmol). Channels: FITC.			
DMEM	Opti-MEM	Lipoffectamine 2000	Opti-MEM with Lipoffectamine 2000
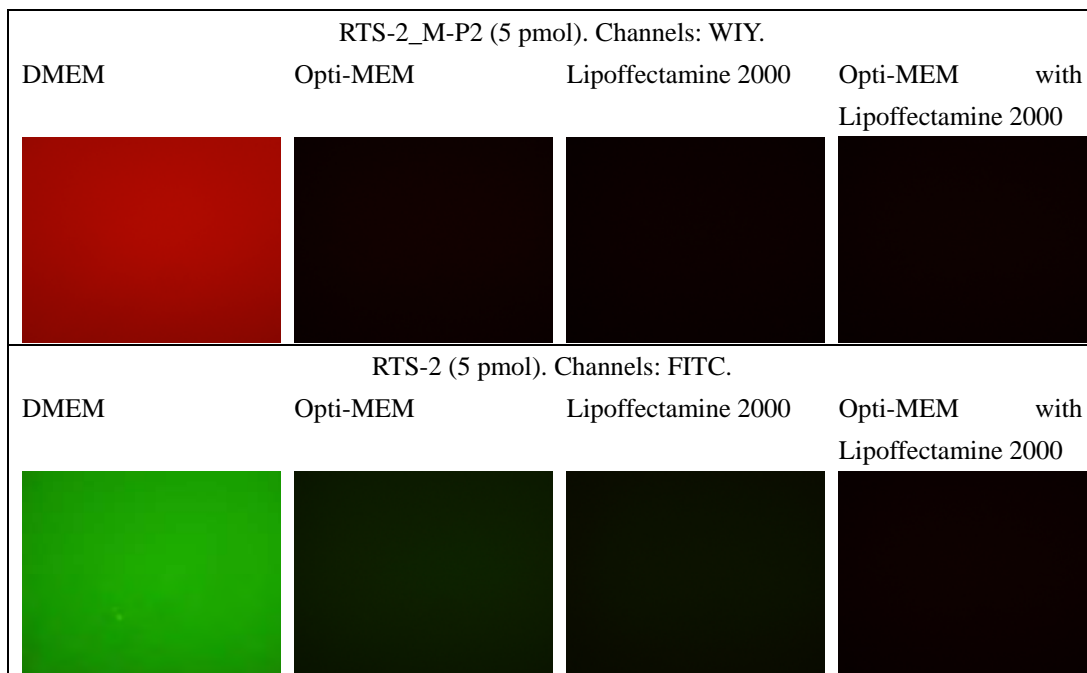
Figure 4.4-15 DMEM was identified as the agent that degraded both RS, while Opti-MEM and Lipoffectamine 2000, or their combinations did not activate RS.

As the results showed in Figure 4.4-15, the presence of DMEM quickly activated (degraded) both RS to release the respective fluorescence, while the other transfection agents had no significant effect on the RS.

Upon discovering that DMEM was apparently able to activate the RS, we repeated the experiments in Section 4.4.2.2 with the modification that transfection of RNA was done without DMEM in the media.

This was accomplished by washing cells 3 times with 500 μl of OPTI-MEM, with the final wash remaining as new media, just before the transfection of RNA. DMEM (10% FBS, 1% P/S) was returned as the media only 6 hours after transfection of RNA, by when most transfection of RNA into cells should have occurred.

However, similar results were obtained as before, that the RF produced were due to cleavage of RS (RTS-2_M-P2) without the need for any activated Mz based sensors, HH-2, or

HH-2-2_tRES.

It was concluded that even with significant nuclease resistant protection and the removal of DMEM during transfection, the RS, RTS-2_M-P2 was still degraded after transfection, probably by nucleases present within the cells.

At this stage, it was decided that based on the findings, time and resources available, it was best to develop e-SRS for in vitro applications.

4.5 Applications in Nucleic Acid *in vitro* detection.

Our experience from the previous section suggested that the e-SRS technology was not yet applicable in the environment of living cells. However, positive data in application of e-SRS in test tube environment lead us to redirect development towards in vitro diagnostic purposes, in the specific detection of Malaria infection (Plasmodium falciparum, Chloroquine resistance strains – CVIET and SVMNT haplotypes for the pfcrt gene, and the Chloroquine sensitive strain – CVMNK haplotype for pfcrt gene) and Dengue viral infection (Serotypes of DEN-1, DEN-2, DEN-3, DEN-4).

Specifically, we would try to develop in vitro diagnostic tests for specific detection of:

1) Malaria infection of Plasmodium falciparum

    a. Chloroquine sensitive strain – CVMNK haplotype for pfcrt gene, denoted here as Mfs.

    b. Chloroquine resistance strains, with pfcrt gene haplotypes:

        i. SVMNT (predominant in South America, but also found in South

East Asia), denoted here as Mfr1.

    ii. CVIET (predominant in Asia) , denoted here as Mfr2.

2) Dengue viral infection

    a. All four common serotypes (DEN-1, 2, 3, 4), denoted here as D1, D2, D3, D4.

4.5.1 sNTS

The sNTS sequences used for both Dengue and Malaria were as shown in Table 4.5-1.

Table 4.5-1 sNTS for Dengue and Malaria.

sNTS	Strain	Sequence (5' to 3')	Lengths (nt)
sNTS-D1	DEN-1	CGTCTCAGTGATCCGGGGG	19
sNTS-D2	DEN-2	CGCCACAAGGGCCATGAACAG	21
sNTS-D3	DEN-3	TAACATCATCATGAGACAGAGC	22
sNTS-D4	DEN-4	CTCTGTTGTCTTAAACAAGAGA	22
sNTS-Mfr1	S99	TGTAATGAATACAATTTTTGCTAA	24
sNTS-Mfr2	Dd2	TGTAATTGAAACAATTTTTGCTAA	24
sNTS-Mfs	3D7	TGTAATGAATAAAATTTTTGCTAA	24

The column "Strain" indicates the viral or parasitic strains from which DNA sequence have been derived. Highlighted bases for Malaria NTS indicate differences compared to sNTS-Mfs (the wild type Chloroquine sensitive sequence).

4.5.2 Computational design of Mz based sensors

In the previous section, with only one category in which one NTS was paired with one NNS, the design process was performed manually, which although rather time consuming, was still feasible. With current targeting of seven new NTS, with each NTS being paired with two or three NNS, there was not just a need to design significantly more (seven) new specific Mz based sensors, but each new sensor had a more complex level of specificity requirement

during the design phase. This prompted us to consider automating the Mz based sensor design process, for the three main reasons discussed below.

1) The most immediately obvious reason was to handle the dramatically increased design workload and time required. There were a lot more secondary structures to generate and assess. In fact, within each category (i.e. Dengue or Malaria), a linear increase in the number of NTS would lead to non-linear increase in design work and time as each new NTS meant not just a new specific Mz based sensor but also an additional NNS for the other sensors in the category. For a category with n number of NTS, each potential Mz design had to be evaluated for $(n + 1)$ secondary structures (*1* of Mz+NTS+STS, *(n-1)* of Mz+NNS+STS, and *1* of Mz+STS). This had to be done for every NTS. Hence, even if only one potential Mz design was assessed and accepted for each NTS (typically, many different designs had to be assessed before an appropriately specific design could be obtained), for a category with n NTS, the amount of secondary structures to be generated and assessed would be *n x (n + 1)*, or $(n^2 + n)$. In other words, the amount of work and time required in Mz based sensor design increased exponentially with the number of NTS targeted in each category. In addition, the increase in secondary structures to be assessed would increase the likelihood of error if these assessment continued to be done manually i.e. by visual examination of secondary structure figures. Before creating a computational tool for Mz based sensor design, we made a detailed estimate of the amount of design work required for our new NTS, as presented in Section 4.5.2.1.

2) Besides the increase in workload faced, the difficulty of design was also increased when there were more NTS in each category. This was because each modification of Mz design, say in order to address a particular problem in one RNA combination (e.g. Mz+NTS+STS), would likely also affect secondary structures of other RNA combinations. As such, it could become very difficult to make any corrections of inadequate Mz based sensor designs, via manual design.

3) Last but not least, automated Mz based sensor design would allow us to easily evaluate all possible configurations of SAS (Sensor Arm Split) designs in order to select the optimised Mz based sensor design for a given NTS, its associated NNS, and STS. While experience generated some strategies in the "trial and error" based manual design, it is likely that such an approach would not produce the optimised Mz design, particularly as the number of NTS in each category increased.

4.5.2.1 Estimation of the number of Mz based sensor designs to be examined for Dengue and Malaria detection.

In order to better determine whether there was a need to create a computational tool for Mz based sensor design for our new development in Dengue and Malaria detection, we made a detailed estimate of the amount of design work required for our new NTS. The finding that over 640 secondary structures had to be generated and assessed for correctness (active or inactive catalytic structure) supported our decision to automate the Mz based sensor design.

4.5.2.1.1 **RNA regions to adjust in designing a Mz based sensor.**

While we recognise 4 regions (sensor arm, Stem II, catalytic core, and catalytic arm) as important in deciding a Mz based sensor design (Section 4.3.6.2), the catalytic core sequence is conserved and hence need not be adjusted. This leaves 3 main regions of adjustment in designing a specific Mz based sensor, as follows:

1) Sensor arm

    a. Sequence and length can be adjusted depending on NTS.

    b. SAS can be adjusted, from 1 to one less than the length of the sensor arm.

2) Stem II

    a. Both MzL and MzR Stem II can be adjusted, without need for perfect matches or equal lengths.

3) Catalytic arm

    a. Sequence and length can be adjusted depending on STS.

    b. Catalytic Arm Split can be adjusted, from 1 to one less than the length of the catalytic arm, as long as the Rz recognition site on the STS is also adjusted accordingly.

In the current case, in order to have the greatest chance of success, we kept unchanged as many aspects of our functionally successful Mz-2 design as possible. Hence, we limited changes to only those aspects that must necessarily change, i.e., the sensor arms only (points 1(a) and (b) above).

4.5.2.1.2 **Number of Mz based sensor designs for Dengue and Malaria detection.**

The 7 NTS involved here have lengths of ranging from 19 to 24, (only one was less than 21 nt), meaning that the maximum SAS range for each NTS was 20 or more for six NTS and 19 for one. Assuming conservatively that the SAS range involved for each Mz based sensor was 20, point 1(b) above meant that there were 20 different variants of Mz designs that we could test for appropriate folding via RNAstructure 4.5.

As described in Section 4.3.6.1, for each Mz variant, there was a need to test for appropriate secondary structures for three *types* of RNA combinations (Mz+NTS+STS, Mz+NNS+STS, and Mz+STS), according to table **Table 4.5-2**.

Table 4.5-2 Number of RNA combinations to be assessed for each NTS.

Category	No. of NTS	No. of RNA Combinations			Total	Sub-Total (No. of NTS x Total no. of RNA Combinations)
		Mz+NTS+STS	Mz+NNS+STS	Mz+STS		
Malaria	3	1	2	1	4	12 (= 3x4)
Dengue	4	1	3	1	5	20 (= 4x5)
Total						32

Based on the above estimates, there were about 20 (variants per Mz) x 32 (number of RNA combinations per variant) $\approx$ 640 secondary structures to generate and assess for correctness (active or inactive).

In order to obtain the best possible designs (within the constraints of following most of Mz-2 design) we faced a conservative estimate of about 640 RNA secondary structures to generate and evaluate for correctness (active or inactive). The actual figure was 676 (4x(23+23+23) + 5x(18+20+21+21)).

4.5.2.2 Computational algorithm for optimising designs of Mz based sensor

In order to automate the design of Mz based e-SRS sensor, we developed in Perl 5.8.8 a computational algorithm that using RNAstructure 4.5 for secondary structure prediction, generated sequences for the optimal Mz based sensor to specifically recognise a given RNA target (NTS) and not any other given RNA sequences (NNS). This software tool (a Perl script named "eSRS.pl") greatly automated the 3 processes of: 1) designing sequences for specific e-SRS Mz-based sensor(s) for each NTS, 2) generating their resultant secondary structures, and 3) assessing these structures for appropriate specificity (presence of absence of catalytic conformation in various situations).

The entire source code of eSRS.pl was given in the Section Appendices.

4.5.2.2.1 Inputs required by algorithm

There are essentially 3 kinds of inputs that are required by the Perl script, eSRS.pl.

1) Sensor regions to be entered into Perl script eSRS.pl. As mentioned in Section 4.5.2.1.1, we did not want to change elements of our successful Mz-2 designs that were not required to be changed when targeting new NTS. As such, these sequences were hard-coded into the Perl scripts, as shown in **Table 4.5-3**. Should changes to these regions (e.g. Stem II sequences) be needed, the user could do a text search for the Declaration lines and substitute the existing values between the quotations marks for the desired values.

2) System information such as the Categories of NTS involved, and path information in

order for eSRS.pl to access required information, such as where the text file

containing NTS/NNS information is located, where RNAstructure 4.5 was installed,

etc. These inputs are shown in

Table 4.5-4. As above, these information can be edited. The difference in entering path information (except where stated), is that every back slash has to be replaced by two back slashes, for e.g. "C:\eSRS\" would become "C:\\eSRS\\". Other details for editing values were given in the table.

3) NTS sequences for each category, which are not hard-coded as they are likely to change with each use of the tool. The sequences for each category have to be provided in a tab delimited text file, named and placed in a folder as specified in

Table 4.5-4 for System information. In each file, there must be only two columns. The first column contains the NTS name, and the second column contains the NTS sequence in uppercases. An example for the Dengue category is shown in **Table 4.5-5**.

Table 4.5-3 Sensor regions hard-coded into eSRS.pl.

Region	Variable Name	Default Value	Declaration line
STS	$RTS2	UAUGAAACUCUCGAAAA	my $RTS2 = "UAUGAAACUCUCGAAAA";
MzL catalytic arm	$MzL_RTSrc5	UUUUCGA	my $MzL_RTSrc5 = "UUUUCGA";
MzL Catalytic core	$MzL_Loop	CUGAUGAG	my $MzL_Loop = "CUGAUGAG";
MzL Stem II	$MzL_stem2	UGA	my $MzL_stem2 = "UGA";
MzR Stem II	$MzR_stem2	AGCA	my $MzR_stem2 = "AGCA";
MzR Catalytic core	$MzR_Loop	CGAA	my $MzR_Loop = "CGAA";
MzL catalytic arm	$MzR_RTSrc3	AGUUUCAUA	my $MzR_RTSrc3 = "AGUUUCAUA";
Linker	$Loop	gaaa	my $Loop = "gaaa";

Table 4.5-4 System information required by eSRS.pl.

Variable type / Path for	Variable Name	Default Value	Declaration line
Categories.	@cat	Malaria Dengue	my @cat = qw(Malaria Dengue); To edit, insert new values between "qw(" and ");". Multiple values should be separated by a space. Unwanted values should be deleted.
NTS folder.	@NTS_path	C:\eSRS\NTS\NTS_Mal.xls C:\eSRS\NTS\NTS_Den.xls	@NTS_path = qw(C:\eSRS\NTS\NTS_Mal.xls C:\eSRS\NTS\NTS_Den.xls); Sequence of path info must follow that of categories above (i.e. the category listed first must also have the path info listed first, and so on). Use single back slash for path. Editing rules for Categories also applies here.
Containing folder for outputs.	$eSRS_path	C:\ eSRS \	my $eSRS_path = "C:\\eSRS\\";
Folder for sensor sequence and information output.	$eSRS_seq_path	C:\eSRS\Sensor_Seq\	my $eSRS_seq_path = "C:\\eSRS\\Sensor_Seq\\";
Folder for RNAstructure 4.5 installation.	$RNAst_path	C:\Program Files\RNAstructure 4.5\	my $RNAst_path = "C:\\Program Files\\RNAstructure 4.5\\";
Folder for combined RNA sequences.	$fold_seq_path	C:\eSRS\Fold_seq\$cat\	my $fold_seq_path = "C:\\eSRS\\Fold_seq\\$cat\\":

Variable type / Path for	Variable Name	Default Value	Declaration line
Folder for combined RNA sequences of functional Mz designs.	$fold_seq_func_path	C:\eSRS\Fold_seq\Functional\$cat\	my $fold_seq_func_path = "C:\\eSRS\\Fold_seq\\Functional\\$cat\\":
Folder for CT files.	$ct_path	C:\eSRS\CT\$cat\	my $ct_path = "C:\\eSRS\\CT\\$cat\\";
Folder for CT files of functional Mz designs.	$ct_func_path	C:\eSRS\CT\Functional\$cat\	my $ct_func_path = "C:\\eSRS\\CT\\Functional\\$cat\\";

Note: "$cat" in the last 4 rows should not be edited. It gives the last part of the path information that is specific to the category.

Table 4.5-5 An example of Dengue NTS file for use with eSRS.pl.

NTS name	Sequence (5' to 3')
sNTS-D1	CGUCUCAGUGAUCCGGGGG
sNTS-D2	CGCCACAAGGGCCAUGAACAG
sNTS-D3	UAACAUCAUCAUGAGACAGAGC
sNTS-D4	CUCUGUUGUCUUAAACAAGAGA

Note: The header row shown here is given for information only and should not be present in the NTS file. Columns should be tab delimited, meaning there should be a tab character between the NTS name and its sequence on each line. There should be only 1 tab character on each line. The NTS file should be saved as a tab delimited text file with a file name and extension as given in the "@NTS_path" variable.

4.5.2.2.2 **Outline of algorithm**

For each Category,

- Get list of NTS names and sequences into 2D (2 dimensional) array, @NTS.
    - 1st D array contains references to 2nd D arrays, one for each NTS.
    - 2nd D array has 2 elements. 1st element is the NTS name. 2nd element is the NTS sequence.
- Get possible MzL & MzR pairs of Stem II into 2D array, @stem2.
    - 1st D array contains references to 2nd D arrays, one for each possible Stem II. In the existing version, only one Stem II (that of Mz-2) was used.
    - 2nd D array has 3 elements. 1st element is the Stem II Name. 2nd and 3rd elements are the MzL and MzR Stem II sequence respectively.
        - Stem II Name was given as "<MzL StemII sequence>,<MzR StemII sequence>", for e.g. "UGA,AGCA". Each MzL and MzR sequence was made up to a minimum of 3 characters per stem with extra "-" for those with less than 3 characters.
- For each NTS,
    - Get possible MzL & MzR pairs of sensor arm into 3D array, @NTSrc.
        - 1st D array contains references to 2nd D arrays, one for each NTS.
        - 2nd D array has 2 or more elements. 1st element is the NTS Name. Remaining elements are 3rd D arrays, one for each possible sensor arm, generated by setting values of SAS from 1 to (length of NTS-1).
        - 3rd D array has 3 elements. 1st element is the SAS (possible range is 01 to 99) for the particular sensor arm. 2nd and 3rd elements are the MzL and MzR sensor arm sequences respectively.
- Get possible MzL & MzR pairs into 4D array, @mz.
    - 1st D array contains references to 2nd D arrays, one for each NTS.
    - 2nd D array has 2 or more elements. 1st element is the NTS Name. Remaining elements are references to 3rd D arrays, one for each possible Stem II.
    - 3rd D array has 2 or more elements. 1st element is the StemII Name. Remaining elements are references to 4th D arrays, one for each possible sensor arm for the current NTS.
    - 4th D array has 3 elements. 1st element is the Mz Name for the current sensor arm. 2nd and 3rd elements are the MzL and MzR sequences respectively.
        - Mz Name was given as "Mz_<NTS Name, less "sNTS">_<StemII Name>_<SAS (01-99)>", for e.g. Mz_D1_UGA,AGCA_05 and Mz_Mfr1_UGA,AGCA_13.
        - MzL and MzR sequences for a particular SAS were obtained by joining 4 RNA segments each, as explained in Sections 4.3.6.3.1 and 4.3.6.3.2 respectively. Briefly:

- MzL - 5' catalytic arm; MzL catalytic core; MzL StemII; 3' sensor arm.
- MzR - 5' sensor arm; MzR StemII; MzR catalytic core; 3' catalytic arm.

- For each NTS,
  - Get list of NNS names and sequences into 2D (2 dimensional) array, @NNS.
    - Structure is the same as that of @NTS.
    - Created by copying all elements of @NTS except the element containing current NTS.
  - Assess each Mz based sensor design.
    - Get MzL and MzR from @mz.
    - Assemble RNA combination(s) into .seq text files of (RNAstructure 4.5 format), write to an output folder, and send to RNAstructure 4.5 to assess activity, in the following order. Should a combination be assessed as unsuitable, remaining combinations are not assessed.
      - Mz+NTS+STS
      - Mz+STS
      - Mz+NNS+STS
    - .seq file Name: <Mz Name>_<T/N/A>_<NTS or NNS Name, less "sNTS">.<seq>.   E.g.: Mz_D1_UGA,AGCA_05_T_D1.seq
      - <T/N/A> referred to the RNA combination involved. T for Mz+NTS+STS, N for Mz+NNS+STS, A (stands for Mz alone) for Mz+STS.
    - Each .seq file would be assessed by RNAstructure 4.5, which would produce a .ct (Connectivity Table) text file describing its predicted secondary structure, and which would be stored in an output folder.
    - .ct file Name: <Mz Name>_<T/N/A>_<NTS or NNS Name, less "sNTS">.<ct>.   E.g.: Mz_D1_UGA,AGCA_05_T_D1.ct
      - <T/N/A> referred to the RNA combination involved. T for Mz+NTS+STS, N for Mz+NNS+STS, A (stands for Mz alone) for Mz+STS.
    - Validity of an RNA combination is assessed by examining the .ct file.
      - The most stable structure(s) are assessed based on .ct file information.
      - Based on the combined length of the structure, the minimum requirement for an active structure is calculated (as described in Section 4.3.6.4.1) and stored in a 2D array, @ss_active.
        - 1st D array contains references to 2nd D arrays, one for the pairing condition of each of the 15 bases defined as part of the sufficient conditions for an

active secondary structure.

- o 2nd D array has 2 elements. 1st element is one of the 15 bases, the 2nd element is its pairing partner.

- Each of the most stable structure of each .ct file is scanned and compared to @ss_active. For each RNA combination:
    - o Mz+NTS+STS

      If any of the most stable structure(s) is inactive, design is considered invalid. This is a conservative selection criteria favouring more active sensors.

    - o Mz+STS

      If any of the most stable structure(s) is active, design is considered invalid. This is a conservative selection criteria favouring more specific sensors.

    - o Mz+NNS+STS

      If any of the most stable structure(s) is active, design is considered invalid. This is a conservative selection criteria favouring more specific sensors.

    - ▪ For each sensor design, details will be written to an output text file (see next Section).

- o For valid sensor designs, .seq and .ct files would be moved to a "Functional" folder, for ease of reference (amongst the many files generated altogether).

4.5.2.2.3 **Outputs produced by algorithm**

There are essentially 3 outputs produced by the Perl script, eSRS.pl.

1) .seq files. These are eSRS.pl generated text files containing the sequences of a particular RNA combinations (Mz+NTS+STS, Mz+STS, or Mz+NNS+STS) for a particular Mz based sensor merged into a single RNA sequence with lowercase linker sequences "gaaa".

    a. Not all RNA combinations of a particular sensor design may be present, as explained in Section 4.5.2.2.2.

    b. Format follows the RNAstructure 4.5 .seq sequence file format.

c. .seq file Name: <Mz Name>_<T/N/A>_<NTS or NNS Name, less "sNTS">.<seq>. E.g.: Mz_D1_UGA,AGCA_05_T_D1.seq

   Details were explained in Section 4.5.2.2.1.

d. Output folder is C:\eSRS\Fold_seq\$cat\, $cat represents the name of the category involved, as explained in Section 4.5.2.2.1.

e. .seq files belonging to valid sensor designs would be moved to C:\eSRS\Fold_seq\Functional\$cat\

2) .ct files. These are RNAstructure 4.5 generated text files containing the predicted secondary structure(s) of a particular RNA combinations (Mz+NTS+STS, Mz+STS, or Mz+NNS+STS) for a particular Mz based sensor.

   a. Not all RNA combinations of a particular sensor design may be present, as explained in Section 4.5.2.2.2.

   b. Format follows the RNAstructure 4.5 CT (Connectivity Table) file format.

   c. .ct file Name: <Mz Name>_<T/N/A>_<NTS or NNS Name, less "sNTS">.<ct>. E.g.: Mz_D1_UGA,AGCA_05_T_D1.ct

      Details were explained in Section 4.5.2.2.1.

   d. Output folder is C:\eSRS\CT\$cat\, $cat represents the name of the category involved, as explained in Section 4.5.2.2.1.

   e. .ct files belonging to valid sensor designs would be moved to C:\eSRS\CT\Functional\$cat\

3) e-SRS Sensor sequence files. These are eSRS.pl generated tab delimited text files

containing the sequences and other information of all designed Mz based sensor for each category (one file per category). They can be easily viewed with Microsoft Excel.

    a. The format and information contained are illustrated in **Table 4.5-6**.

        i. Each row represents the values for one Mz based sensor design, of the Category given in the "Category" column.

        ii. MzL or MzR Name were given as "Mz<L or R>_<NTS Name, less "sNTS">_<StemII Name>_<SAS (01-99)>", for e.g. MzL_D1_UGA,AGCA_05 and MzR_D1_UGA,AGCA_05.

        iii. "MzL" and "MzR" Columns give the MzL and MzR RNA sequences (5' to 3') respectively.

        iv. After the "MzR" column follows a number of pairs of "NTS" and "Energy" Columns, one pair for each RNA combination in the Category. This would mean ($n+1$) pairs where n is the number of NTS in the Category. For e.g., Dengue (four NTS) with five RNA combinations would have five such pairs.

            1. Each pair indicates for a particular RNA combination, in the "NTS" column, the NTS Name for the NTS (or NNS) present, and in the "Energy" column, the free energy (in kcal/mol) predicted by RNAstructure 4.5 for its most stable structure.

2. The first pair indicates the values for Mz+NTS+STS.

3. The second pair indicates the values for Mz+STS. The NTS Name would be given as "A" for Mz <u>a</u>lone (i.e., without NTS or NNS).

4. The remaining pairs would be values for Mz+NNS+STS, one for each NNS.

v. The RNA combinations would be assessed in an order as explained in Section 4.5.2.2.2.

1. The first RNA combination that failed to meet design requirements would have an "!" indicated for its "Energy" column.

2. Following combinations would not be assessed and would have values of "-" for both "NTS" (except in the case of Mz+STS, which would still have "A") and "Energy" Columns.

3. An Mz based sensor design would be considered valid if it has numerical values for all "Energy" columns designs. It should have no "!" or "-" values.

b. Output folder is C:\eSRS\Sensor_Seq\, $cat represents the name of the category involved, as explained in Section 4.5.2.2.1.

Table 4.5-6 Format of e-SRS Sensor sequence files.

Category	MzL Name	MzL	MzR Name	MzR	NTS	Energy	NTS	Energy

Note: Header names are approximate only and actually output may differ slightly. Two pairs of "NTS" and "Energy" Columns are shown here. Actual number of such pairs would be (*n+1*) where *n* is the number of NTS in the Category.

4.5.2.2.4 Strategies and guidelines for selecting from valid sensors.

Our algorithm outputs sensor designs that are predicted to be valid for their given NTS and NNS. The valid sensor designs for Dengue and Malaria are shown in **Table A 5** and **Table A 6** respectively. As these results demonstrate, each NTS may have more than one valid design, indeed, for the Dengue category, there were between 4 to 13 valid designs for each NTS.

With limited resources, we employed the following selection strategies and guidelines to select one sensor per NTS for in-vitro testing. If upon in-vitro testing, any sensor was found unsatisfactory, by being 1) of low activity in detecting NTS, or 2) of exhibiting non-specific activity in the presence of NNS, or both, corresponding strategies (i.e. to increase specific activity or decrease non specific activity) from the following were used to select a new sensor.

To predict the relative activities of different RNA combinations, which we associate with their stability, we created an index called the Stability Buffer (SB). This indicated how much "buffer" the most stable structure had for it to stay in its conformation, and thus how stable that particular structure was likely to be. To appreciate this concept, one should bear in mind the dynamic nature of most RNA secondary structures. As explained in Section 4.3.6.1, each RNA combination could have more than one predicted secondary structure, each with a particular free energy that indicates its stability. Thus although all sensor designs considered at

this stage fit the validity criteria in having most stable Mz+NTS+STS structures that are active and most stable Mz+NNS+STS and Mz+STS structures that inactive, there could be undesirable active or inactive structures (as befit each combination) of lower stability. Such less stable structures might still have a chance of forming, which leads to either lower specific activity (for Mz+NTS+STS) or higher non-specific activity (for Mz+NTS+STS and Mz+STS), and thus should be kept to a minimum.

SB then measures the magnitude of the free energy difference between the most stable structure and the most stable undesired structure. It can be measured for all three RNA combinations:

1) For Mz+NTS+STS, it is the difference between the most stable structure (active) and the most stable inactive (MSI) structure.

2) For Mz+NNS+STS and Mz+STS, it is the difference between the most stable structure (inactive) and the most stable active (MSA) structure.

$$SB = Energy \, (most \, stable \, structure) - Energy \, (MSI \, or \, MSA)$$

Note: if MSI or MSA structures are not predicted, their free energies were assumed to be zero.

The larger the SB, the more likely the most stable (and desired) structure will be formed.

SB% is SB calculated as a percentage of the free energy value of the most stable structure.

$$SB\% = 100\% \times \frac{SB}{Energy \, (most \, stable \, structure)}$$

The larger the SB%, the greater the proportion of structures formed that is in the most stable (and desired) conformation. SB% allows the relative stability of different sensor designs to be compared to each other.

To increase rate of specific reaction, we select for sensor designs with greater SB% and SB of Mz+NTS+STS structure. This is so that more active Mz+NTS+STS structures would be formed, giving a greater reaction rate for STS cleavage, given that other factors are not limiting.

To increase specificity (i.e. to reduce non-specific activity against a particular NNS), we select for sensor designs with:

1) Absence of active conformation in the Mz+NNS+STS or Mz+STS structures predicted. As explained above, such active structures could contribute to non-specific activity (i.e. activity in the presence of NNS) and thus should be kept to a minimum. If such structures are present, then:

   a. The SB% should be as high as possible, so as to keep the proportion of inactive structures as high as possible.

   b. Such structures being ranked in stability as many ranks behind the most stable structure as possible.

   c. Having as little such structures as possible.

2) The most stable inactive Mz+NNS+STS for NNS concerned. If an Mz+NNS+STS structure is highly stable in an inactive conformation, the structure would be "locked" (i.e. unlikely to change out of such structure) in an inactive conformation and thus not contribute to non-specific activity.

4.5.3 Sequences of Mz based sensors

Mz based sensors were designed for Dengue and Malaria targets using above software.

For each NTS of the Dengue and Malaria category, valid sensor designs for Dengue and Malaria were obtained as shown in **Table A 5** and **Table A 6** respectively. Following the strategies and guidelines above, specific Mz based sensor were manually selected from the pool of valid designs and sent for synthesis. The final list of MzL and MzR sequences used and their respective NTS are as shown in **Table 4.3-8**. All RNA oligonucleotides (including STS, MzL, MzR, NTS, NNS) were synthesised commercially with HPLC purification.

Table 4.5-7 Final selection of MzL and MzR sequences.

NTS	Sensor Name	Constructs	Sequence (5' to 3')	Length (nt)
D1	Mz-D1	MzL_D1_UGA,AGCA_05	UUUUCGACUGAUGAGUGAGGAUCACUGAGACG	32
	SB: -4.60 SB%:8.5%	MzR_D1_UGA,AGCA_05	CCCCCAGCACGAAAGUUUCAUA	22
D2	Mz-D2	MzL_D2_UGA,AGCA_06	UUUUCGACUGAUGAGUGAAUGGCCCUUGUGGCG	33
	SB: -4.10 SB%: 7.3%	MzR_D2_UGA,AGCA_06	CUGUUCAGCACGAAAGUUUCAUA	23
D3	Mz-D3	MzL_D3_UGA,AGCA_05	UUUUCGACUGAUGAGUGAGUCUCAUGAUGAUGUUA	35
	SB: -3.80 SB%: 7.8%	MzR_D3_UGA,AGCA_05	GCUCUAGCACGAAAGUUUCAUA	22
D4	Mz-D4	MzL_D4_UGA,AGCA_13	UUUUCGACUGAUGAGUGAACAACAGAG	27
	SB: -4.00 SB%: 8.8%	MzR_D4_UGA,AGCA_13	UCUCUUGUUUAAGAGCACGAAAGUUUCAUA	30
Mfs	Mz-Mfs	MzL_Mfs_UGA,AGCA_11	UUUUCGACUGAUGAGUGAUUUAUUCAUUACA	31
	SB: -3.40 SB%: 8.8%	MzR_Mfs_UGA,AGCA_11	UUAGCAAAAAUAGCACGAAAGUUUCAUA	28
Mfr1	Mz-Mfr1	MzL_Mfr1_UGA,AGCA_13	UUUUCGACUGAUGAGUGAUAUUCAUUACA	29
	SB: -4.10 SB%: 10.0%	MzR_Mfr1_UGA,AGCA_13	UUAGCAAAAAUUGAGCACGAAAGUUUCAUA	30
Mfr2	Mz-Mfr2	MzL_Mfr2_UGA,AGCA_15	UUUUCGACUGAUGAGUGAUCAAUUACA	27
	SB: -3.70 SB%: 9.1%	MzR_Mfr2_UGA,AGCA_15	UUAGCAAAAAUUGUUAGCACGAAAGUUUCAUA	32

STS used were all RTS-2. The unit of SB is kcal/mol

4.5.4 Detection of Dengue Serotypes D1, D2, D3, D4 sNTS.

We were able to distinguish between sNTS of D1, D2, D3, and D4 (sNTS-D1, sNTS-D2, sNTS-D3, and sNTS-D4) using their specific Mz, Mz-D1, Mz-D2, Mz-D3, and Mz-D4 respectively. As shown in

Figure **4.5-1**, RTA assays were specific for each Mz. These results are representative of at

least 3 repeats.

Optimised reactions conditions for all Dengue Mz based sensors were 200 nM Mz, 200 nM

sNTS, 1000 nM RTS-2, 50 mM Tris (pH 8.0), 10 mM MgCl$_2$, and incubation at 37 °C.
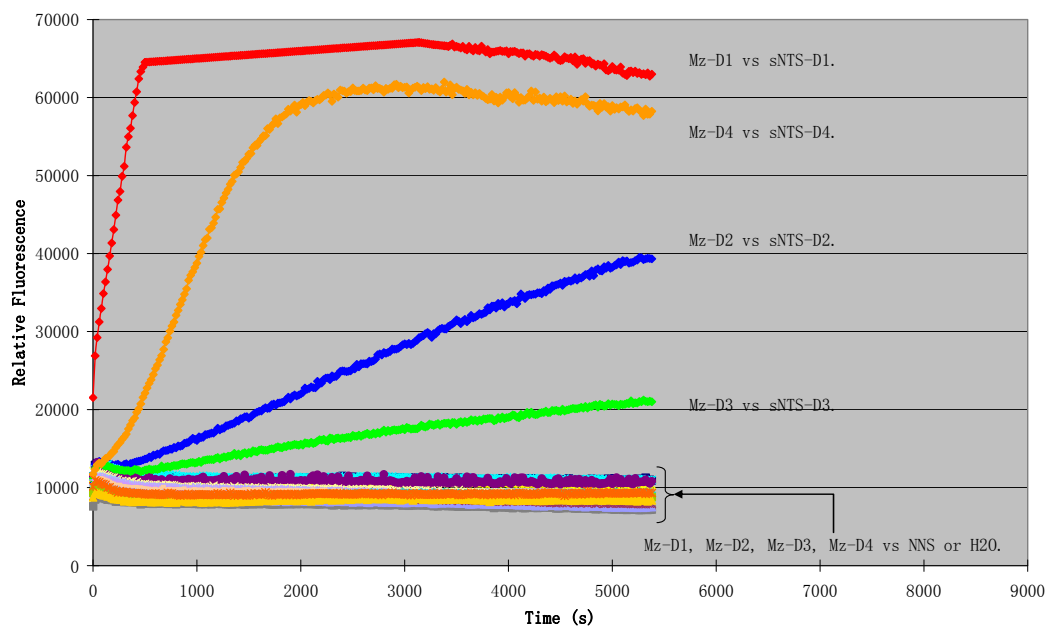
Figure 4.5-1 RTA results for sNTS using Mz based sensors for Dengue.
The RTA result for each Mz-based sensor was colour coded: Red for Mz-D1; Blue for Mz-D2;
Green for Mz-D3; Orange for Mz-D4. Each sensor was tested against all conditions, i.e. vs NTS;
vs NNS; vs H$_2$O (i.e. alone). Separations of the fluorescence curves for sensors vs NTS and
those vs NNS or H$_2$O were clear. The RTA responses of all sensors vs NNS or H$_2$O were
essentially at background levels. Hollow points indicate averaged replacement values for
values that reached saturation (for Mz-D1 vs sNTS-D1).

4.5.5 Detection of Malaria Strains Mfs, Mfr1, Mfr2 sNTS.

We were able to distinguish between sNTS of Mfs, Mfr1, and Mfr2 (sNTS-Mfs, sNTS-Mfr1,

and sNTS-Mfr2) using their specific Mz, Mz-Mfs, Mz-Mfr1, and Mz-Mfr2 respectively. As

shown in

Figure **4.5-2**, RTA assays were specific for each Mz. These results were representative of at least 3 repeats.

In the cases of Mz-Mfs and Mz-Mfr1, we initially had difficulty in differentiating between sNTS-Mfr1 and sNTS-Mfs. We attempted to improve specificity by optimising Mg^{2+} concentration (from 2mM to 10 mM in steps of 1mM, and then from 1 to 2 mM in steps of 0.2 mM) and incubation temperature (37, 39.5, 42 °C). While changing incubation temperature did not seem to improve specificity, reducing Mg^{2+} improved specificity for both. However, only Mz-Mfs became totally specific at 2 mM Mg^{2+}. Mz-Mfr1 seemed most specific at 3 mM Mg^{2+} but still showed too much activation against sNTS-Mfs.

Subsequently, we discovered that when sNTS-Mfr2 was added into the RTA mix for Mz-Mfr1, it reduced the signal for reactions against sNTS-Mfs (i.e., the NNS), while not significantly reducing that against sNTS-Mfr1 (i.e., the NTS). sNTS-Mfr2 seemed to act as a "competitor nucleotide" (CN) against both NTS and NNS for their binding for Mz, but with a greater reduction of the NNS (sNTS-Mfs) affinity for Mz-Mfr1, than that of the NTS affinity for Mz-Mfr1. The critical issue was thus to add just the right amount of CN that would derive the maximum affinity reduction for the NNS with the least impact on NTS affinity to the Mz.

Hence, to improve the specificity of Mz-Mfr1 to acceptable levels, we optimised the concentration of the competitor nucleotide sNTS-Mfr2 added to the RTA (testing 0, 20, 100, 200, 400, 1000, 2000, 4000, 10000, 20000 nM). We discovered that 2000 nM of sNTS-Mfr2 added to RTA for Mz-Mfr1 made it specific enough to distinguish the NTS (sNTS-Mfr1) from NNS (sNTS-Mfr2).

Optimised reactions conditions were as follows:

For sNTS-Mfs:

200 nM Mz, 200 nM sNTS, 1000 nM RTS-2, 50 mM Tris (pH 8.0), 2 mM $MgCl_2$, and

incubation at 37 °C.

For sNTS-Mfr1:

200 nM Mz, 200 nM sNTS, 1000 nM RTS-2, 50 mM Tris (pH 8.0), 3 mM $MgCl_2$, 2000 nM

sNTS-Mfr2 (as CN to improve specificity), and incubation at 37 °C.

For sNTS-Mfr2:

200 nM Mz, 200 nM sNTS, 1000 nM RTS-2, 50 mM Tris (pH 8.0), 3 mM $MgCl_2$, and

incubation at 37 °C.
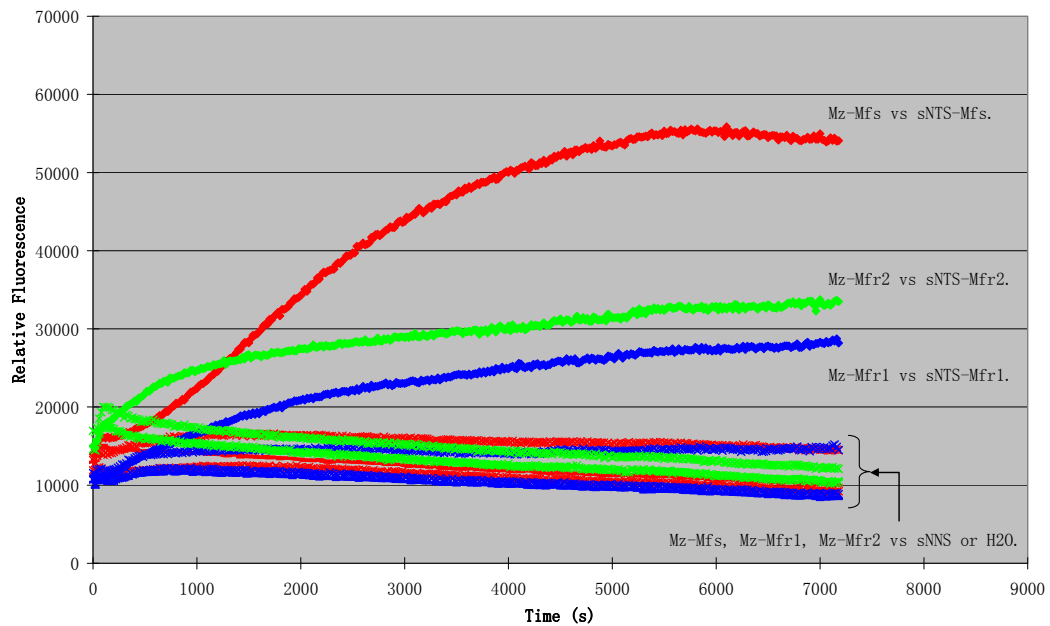
Figure 4.5-2 RTA results for sNTS using Mz based sensors for Malaria.
The RTA result for each Mz-based sensor was colour coded: Red for Mz-Mfs; Blue for Mz-Mfr1; Green for Mz-Mfr2. Each sensor was tested against all conditions, i.e. vs sNTS ; vs sNNS; vs H_2O (i.e. alone). The points for sNTS, sNNS, and H_2O were marked with diamonds, crosses, and dashes (-) respectively. Separations of the fluorescence curves for sensors vs sNTS and those vs sNNS or H_2O were clear. The RTA responses of all sensors vs sNNS or H_2O were essentially at background levels, and were labelled on the graph together as "Mz-Mfs, Mz-Mfr1, Mz-Mfr2 vs sNNS or H2O.".

4.5.6 Detection of Malaria Strains Mfs, Mfr1, Mfr2 NASBA NTS.

4.5.6.1 Use of NASBA to detection DNA NTS.

One way to distinguish between genomic DNA of Mfs, Mfr1, and Mfr2 was to amplify desired genomic segments into RNA via a process called NASBA. NASBA is a 41 °C isothermal reaction that uses a forward (with T7 promoter attached) and reverse primer to produce from the NASBA template, a DNA amplicon (with T7 promoter attached) that can then transcribe many RNA copies of the amplicon (less the T7 promoter region).

The starting NASBA template can be an RNA, in which case the included Avian Myeloblastosis Virus Reverse Transcriptase (AMV-RT) creates a DNA copy using the forward primer. The RNA from this DNA-RNA duplex is then digested by the included RNase H. The AMV-RT then reverse transcribe this DNA strand into a double stranded version (using the reverse primer), which serves as an in-vitro transcription (IVT) template. If the starting template is genomic or other double stranded DNA, an initial denaturation is required, after which AMV-RT would also create a suitable IVT template using the forward primer. From the IVT template, many copies of the RNA amplicon would be transcribed by the included T7 RNA polymerase. More IVT templates would be produced from the RNA amplicons, thereby generating an amplification cycle with an amplification factor of 10^6 to 10^9 in 90 minutes.

After NASBA is performed, the RNA in the NASBA product can then be subjected to e-SRS detection in a NASBA RTA reaction.

4.5.6.2 Cloning of NTS segment from genome into plasmids.

To test this approach, we first cloned a segment of the 3 Plasmodium falciparum strains' genomes containing the NTS into plasmids. We managed to PCR out approximately 200 nt amplicons using primers Mal Fwd 02 and Mal Rvs 01 for Mfs and Mfr1, and Mal Fwd 01 and Mal Rvs 01 for Mfr2. Mal Fwd 02 added an additional T7 promoter sequence used in NASBA to the amplicon. These amplicons were gel purified and cloned into pGEM-T Easy vectors. The plasmids were then sequenced with both forward and reverse vector primers to ensure amplification were what we expected and thus contained original sequences of NTS.

Oligonucleotide information were as shown in **Table A 7**. Using these plasmids, we performed

NASBA template synthesis via PCR (Elongase Enzyme Mix) with primers Mal Fwd 02 and

Mal Rvs 01. PCR products of the right sizes (223 nt) were gel purified and served as NASBA

templates.

4.5.6.3 Initial tests of NASBA RTA.

We performed 90 minutes of NASBA using $6x10^5$, $6x10^7$, and $6x10^9$ copies of NASBA

template per μl of NASBA reaction. 1 μl of these NASBA products were used in 50 μl of

NASBA RTA reaction. As NASBA was supposed to provide amplification of 10^6 to 10^9 in 90

minutes, only about 6.022×10^3 to 6.022×10^6 of NASBA template per μl NASBA reaction

was required to produce the standard 200 nM of NTS that was routinely detectable in sNTS

RTA. However, NASBA RTA produced no specific signals for all three NTS (Mfs, Mfr1,

Mfr2).

We tried to improve NASBA efficiency via reducing the size of the NASBA amplicons by

selecting another forward primer, Mal Fwd 03. Using Mal Fwd 03 and Mal Rvs 01, shorter

DNA and RNA amplicons were generated (142 and 120 nt respectively, as compared to 223

and 201 nt previously). Oligonucleotide information were as shown in Table 4.5-8.

Using the new primers, we repeated the entire process from NASBA template synthesis to

NASBA RTA, but still did not obtain any specific NASBA RTA signals.

Table 4.5-8 Oligonucleotides for NASBA RTA of Mfs, Mfr1, Mfr2.

Oligonucleotides	Role	Sequence (5' to 3')	Length (nt)

Oligonucleo tides	Role	Sequence (5' to 3')	Length (nt)
Mal Fwd 03	NASBA forward primer.	AATTCTAATACGACTCACTATAGGGAGAA GGTATTTTAAGTATTATTTATTTAAGTGT A	59
Mal Rvs 01	NASBA reverse primer.	AAGTTGTGAGTTTCGGATGTTAC	23
DNA Amplicon in NASBA.	NASBA template and NASBA DNA amplicon (i.e., IVT template) using primers: Mal Fwd 03 Mal Rvs 01	AATTCTAATACGACTCACTATAGGGAGAA GGTATTTTAAGTATTATTTATTTAAGTGT ATGTGTAATGAATAAAATTTTTGCTAAAA GAACTTTAAACAAAATTGGTAACTATAGT TTTGTAACATCCGAAACTCACAACTT	142
RNA Amplicon in NASBA.	NASBA RNA amplicon using primers: Mal Fwd 03 Mal Rvs 01	GGGAGAAGGTATTTTAAGTATTATTTATT TAAGTGTATG<mark>TGTAATGAATAAAATTTTT GCTAA</mark>AAGAACTTTAAACAAAATTGGTAA CTATAGTTTTGTAACATCCGAAACTCACA ACTT	120

NASBA DNA and RNA Amplicons shown were those for Mfs. Those for Mfr1 and Mfr2 were of the same length, with a few different bases at the NTS region. The grey highlight indicated the NTS detection region.

In order to further ascertain that the lack of RTA signal was not due to insufficient NTS concentration, we tried to further increase the concentration of NASBA NTS (nNTS) in RTA by increasing the amount of NASBA product used in RTA. From the previous 1 μl NASBA product in 50 μl of RTA reaction, we tested 1, 5, 10, and 15 μl NASBA products, using the previous maximum of 6×10^9 copies of NASBA template per μl of NASBA reaction. If the NASBA process performed as claimed by the kit manufacturer, these parameters would provided estimated concentrations of NTS far in excess of what worked well for sNTS RTA (200 nM). The estimated nNTS concentrations were as shown in Table 4.5-9. To obtain the NTS concentration for RTA, the following formula was used:

$$\frac{\text{NASBA product (L) x starting template per L x Amplification factor}}{\text{RTA volume (L) x Avogadro number } (6.022 \times 10^{23})}$$

Table 4.5-9 Estimated nNTS concentrations achieved in RTA.

Volume of NASBA product (ul)	RTA NTS Concentration (mM) with amplification factor of	
	10^6	10^9
1	0.2	199.3
5	1.0	996.3
10	2.0	1992.7
15	3.0	2989.0

NASBA was done with starting template of 6×10^9 Copies / µl. RTA reaction volume was 50 µl.

However, the RTA results showed that no specific signals were obtained, although the RTA

signals appeared to increase somewhat for greater volumes of NASBA products used,

including NASBA water (i.e. NASBA reaction with no starting template), which showed that

NASBA reagents might somehow contribute to signals in RTA.

4.5.6.4 Use of Antisense oligonucleotides to activate detection of long NTS from NASBA.

As this stage, we believed that insufficient NTS concentration was unlikely to be the cause of

the lack of RTA activation, and suspected that the problem was likely due to secondary

structures obstructing the Mz binding site (i.e. the NTS), which could be a common issue

when attempting to detect long NTS. We decided to try to improve access to the NTS by

adding antisense oligonucleotides (AS) that could bind to sequences close to and flanking the

NTS. Upon binding of the AS, the secondary structure that previously obscured the NTS

would be at least partially denatured, thus improving access to the NTS by the e-SRS. In

addition, as RNase H was provided by in NASBA, RNase H digestion of the RNA strand at

the DNA/RNA duplex regions flanking the NTS (where the AS bind) would release the NTS

as a sNTS, thus allowing RTA to proceed as if sNTS were supplied (instead of long NTS). However, if the RNase H digest occurred during NASBA, it could very well reduce the amplification factor during NASBA, as additional IVT template would not be generated from the digested RNA.

4.5.6.4.1 **NASBA with AS added during or after NASBA.**

As such, we decided to conduct a series of NASBA reactions with AS present during the NASBA reaction, and another series of NASBA reactions where the AS were added after the NASBA reaction.

5' and 3' AS, Mal Rvs 194-215, Mal Rvs 242-271 were designed and used for this experiment (AS information was as shown in **Table 3.4-5**).

For AS present during NASBA, 7 μl reactions with 4.2×10^{10} gel purified DNA template using primers F3R1 were performed. Final AS concentrations were 100 μM, 10 μM, 100 nM, and 1 nM. NASBA was conducted at 41°C for about 1.5 hours.

For AS added after NASBA, 10 μl reactions with 6.0×10^{10} GP DNA (previously performed) using primers F3R1 were performed. NASBA was conducted at 41°C for about 1.5 hours. Subsequently, 5.6 μl of NASBA product was added to 1.4 μl of AS and incubated at 41°C for 2 hours for denaturation with AS. Final AS concentrations were 10 μM, 100 nM, and 1 nM.

4.5.6.4.2 **NASBA RTA with AS added during or after NASBA.**

9 NASBA RTA was conducted as before with the exception of using 5 μl of NASBA product, some of which had AS added (either during or after NASBA, as describe above). The

reactions were as follows:

For AS added during NASBA: 4 Mz+NTS+STS reactions with 100 μM, 10 μM, 100 nM, and 1 nM of AS.

For AS added after NASBA: 3 Mz+NTS+STS reactions with 10 μM, 100 nM, and 1 nM of AS.

In addition, there was 1 Mz+NTS+STS reaction with no AS added, and 1 Mz+STS reaction with 10 μM AS.

In parallel, 4 sNTS Mfs RTA was also conducted to observe the possible effects of NASBA on the previously optimised sNTS Mfs RTA. The sNTS RTA were conducted as before, with the exception of adding various amount of AS. The reactions were as follows:

2 Mz+NTS+STS reactions with 100 μM, and 10 μM of AS.

1 Mz+NTS+STS reactions with no AS.

1 Mz+STS reaction with no AS.

As shown in Figure **4.5-3** and Figure **4.5-4**, for 10 μM and 100 μM AS NASBA RTA reactions, specific RTA signals were obtained, indicating that such concentration of AS could indeed help activate RTA, regardless of whether they were added during or after NASBA reactions (for 10 μM AS).

However, the remaining concentrations of AS (100, 1, 0 nM) did not help to activate NASBA RTA. About 8 minutes into the RTA, these reactions still did not show any specific signals, whereas those reactions that did show specific signals activated as soon as observation started.

To further verify the ability of AS to help activate NASBA RTA, the RTA measurement was

temporarily suspended, the plate extracted, and an additional 100 μM AS was quickly added

(by addition of 5 μl 1000 μM AS to existing reaction) to the following reactions that had not

activated:

At 839.0 s, to the 2 NASBA RTA 1 nM AS reactions.

At 1210 s, to sNTS RTA water reaction (Mz+STS) with no AS.

At 2199 s, to NASBA RTA with no AS, and NASBA water (Mz+STS) with 10 μM AS.

Upon addition of AS, the RTA plate was returned to the fluorometer and measurements

resumed.

For the NASBA RTA Mz+NTS+STS with 1 nM and 0 nM AS reactions, there were no signs

of any specific signals until the additional 100 μM AS were added (about 14 and 36 minutes

respectively into the RTA). After addition of AS, specific signals were detected immediately

upon replacement of the RTA plate into the fluorometer and resumption of fluorescence

measurement.

The fact that similar addition of AS into sNTS RTA (Figure **4.5-5**) and NASBA RTA water

reactions about 20 and 36 minutes respectively into the RTA, had no activation effect on RTA

indicated that the AS themselves did not activate the cleavage of RS. In addition, as shown in

Figure **4.5-5**, the presence of 100 μM and 10 μM of AS in sNTS RTA of Mfs did not

significantly affect the RTA results as compared to sNTS RTA with no AS.

The observed immediate effect of AS on inducing RTA signal implied that the denaturation

action of AS were almost immediate and suggested that AS functioned via simple AS pairing

with nNTS (resulting in denaturation of secondary structures around the NTS) rather than via

RNase H action.
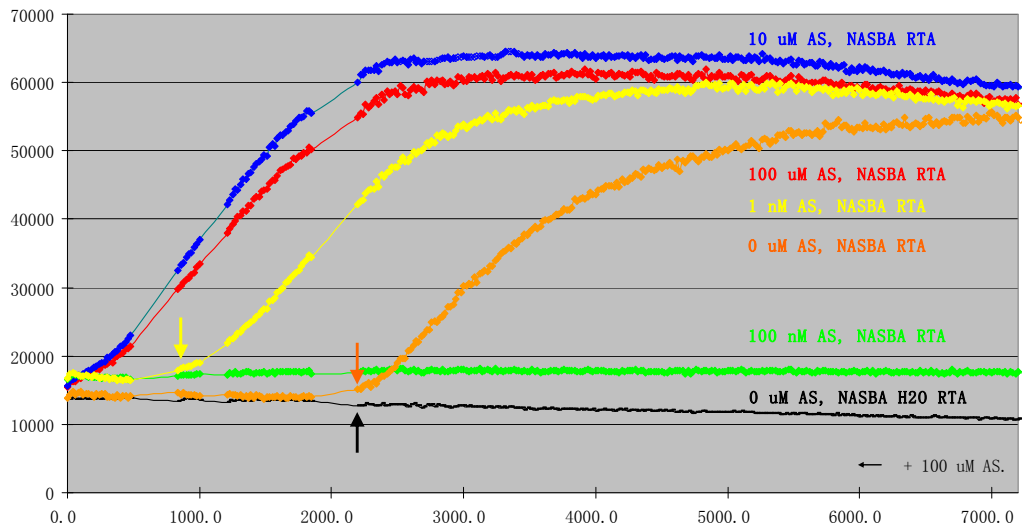
Figure 4.5-3 RTA results for NASBA RTA (with AS added during NASBA)

using Mz based sensors for Mfs.

Arrows indicated the addition of 100 μM AS. Labels and arrows were the same colours as the lines they applied to. Hollow points indicate averaged replacement values for values that reached saturation. X-Axis was in seconds. Y-Axis was in relative fluorescence units.
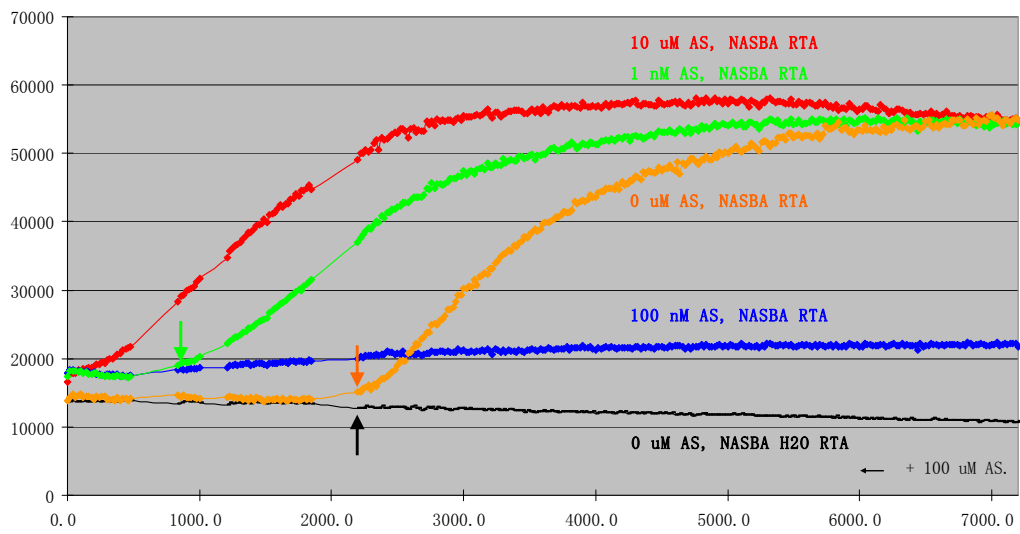
Figure 4.5-4 RTA results for NASBA RTA (with AS added after NASBA) using Mz

based sensors for Mfs.

Arrows indicated the addition of 100 μM AS. Labels, arrows and lines were given the same colour for the results of the same Mz based sensor. X-Axis was in seconds. Y-Axis was in relative fluorescence units.
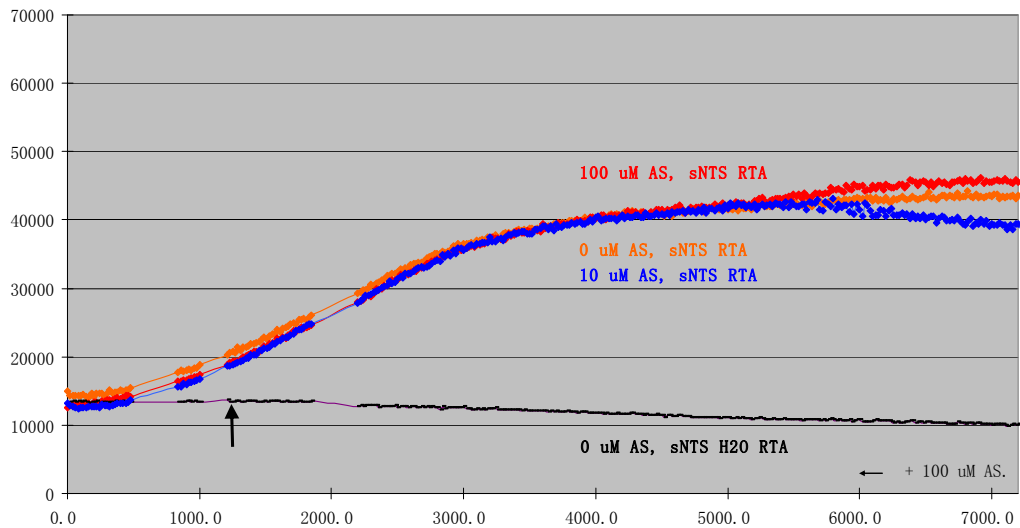
Figure 4.5-5 RTA results for sNTS RTA (with some AS added in RTA) using Mz based

sensors for Mfs.

Arrows indicated the addition of 100 µM AS. Labels and arrows were the same colours as the lines they applied to. X-Axis was in seconds. Y-Axis was in relative fluorescence units.

The results of this experiment indicated that denaturation via addition of AS appeared to be a good method for overcoming issues of secondary structure obscuring the NTS when dealing with long NTS RTA.

4.5.6.5 Optimised conditions for NASBA RTA (AS added at RTA).

Since AS denaturation appeared effective even when added after NASBA, we decided to perform AS denaturation by adding AS to the NASBA products, with incubation at 70 °C for 2 minutes, followed by immediate incubation on ice before use. This process was effective. After testing of various concentrations and sequences to use as AS, we eventually used a 5 µl denaturation mix with 3 µl NASBA product, 0.75 µl each of 100 µM 5' and 3' AS (final concentration of 15 µM), and 0.5 µl of DEPC treated water. 3 different AS sequences were used as shown in Table 4.5-10.

Table 4.5-10 Oligonucleotides in NASBA RTA.

Oligonucleotides	Role	Sequence (5' to 3')	Length (nt)
Mal Rvs 194-215	5' AS for Mfr2.	CATACACTTAAATAAATAATAC	22
Mal Rvs 206-214	5' AS for Mfs, Mfr1.	ATACACTTA	09
Mal Rvs 242-271	3' AS for Mfs, Mfr1, Mfr2.	AACTATAGTTACCAATTTTGTTTAAAGTTC	30
Mal Fwd 03	NASBA forward primer.	AATTCTAATACGACTCACTATAGGGAGAAGGTATTTTAAGTATTATTTATTTAAGTGTA	59
Mal Rvs 01	NASBA reverse primer.	AAGTTGTGAGTTTCGGATGTTAC	23

Oligonucleotides	Role	Sequence (5' to 3')	Length (nt)
NASBA F3R1 DNA	NASBA DNA Amplicon	AATTCTAATACGACTCACTATAGGGAGAA GGTATTTTAAGTATTATTTATTTAAGTGT ATGTGTAATGAATAAAATTTTTGCTAAAA GAACTTTAAACAAAATTGGTAACTATAGT TTTGTAACATCCGAAACTCACAACTT	142
NASBA F3R1 Mfs RNA	NASBA RNA Amplicon	GGGAGAAGGUAUUUUAAGUAUUAUUUAUU UAAGUGUAUGUGUAAUGAAUAAAAUUUUU GCUAAAAGAACUUUAAACAAAAUUGGUAA CUAUAGUUUUGUAACAUCCGAAACUCACA ACUU	120
NASBA F3R1 Mfr1 RNA	NASBA RNA Amplicon	GGGAGAAGGUAUUUUAAGUAUUAUUUAUU UAAGUGUAUGUGUAAUGAAUACAAUUUUU GCUAAAAGAACUUUAAACAAAAUUGGUAA CUAUAGUUUUGUAACAUCCGAAACUCACA ACUU	120
NASBA F3R1 Mfr2 RNA	NASBA RNA Amplicon	GGGAGAAGGUAUUUUAAGUAUUAUUUAUU UAAGUGUAUGUGUAAUUGAAACAAUUUUU GCUAAAAGAACUUUAAACAAAAUUGGUAA CUAUAGUUUUGUAACAUCCGAAACUCACA ACUU	120

NASBA DNA Amplicon shown was that for Mfs. Those for Mfr1 and Mfr2 were of the same length, with a few different bases at the NTS region. The grey highlight indicated the NTS detection region, which are bases 40 to 63 of the NASBA RNA amplicon. Underlined sequences were those that differed amongst the 3 NTS. The yellow and green highlights indicated the regions where the 5' and 3' AS bind respectively.

Eventually, we were able to distinguish between NASBA amplified NTS (in the form of DNA amplified from regions of genomic DNA) of Mfs, Mfr1, and Mfr2 (N-Mfs, N-Mfr1, and N-Mfr2) by their using specific Mz based e-SRS sensors, for Mz-Mfs and Mz-Mfr2. As shown in

Figure **4.5-6**, RTA assays were specific for each Mz. These results were representative of at least 3 repeats.

To distinguish between NASBA amplified NTS of Mfs and Mfr1 (N-Mfs and N-Mfr1) using

Mz-Mfr1, we were only able to perform one repeat, due to exhaustion of reagents. If more reagents were available, we would extent the incubation during to 6 hours and for another 3 repeats.

Optimised reactions conditions were as follows:

For NASBA NTS-Mfs:

200 nM Mz, 3 µl NASBA NTS, 1000 nM RTS-2, 50 mM Tris (pH 8.0), 2 mM $MgCl_2$, and incubation at 37 °C.

For NASBA NTS-Mfr1:

200 nM Mz, 3 µl NASBA NTS, 1000 nM RTS-2, 50 mM Tris (pH 8.0), 3 mM $MgCl_2$, 2000 nM sNTS-Mfr2 (to improve specificity), and incubation at 37 °C for at least 6 hours.

For NASBA NTS-Mfr2:

200 nM Mz, 3 µl NASBA NTS, 1000 nM RTS-2, 50 mM Tris (pH 8.0), 3 mM $MgCl_2$, and incubation at 37 °C.

Each of the NASBA NTS should be denatured at 70 °C for 2 mins with 10 µM of AS (Mal Rvs 194-215 / Mal Rvs 206-214) followed by quick chill on ice.
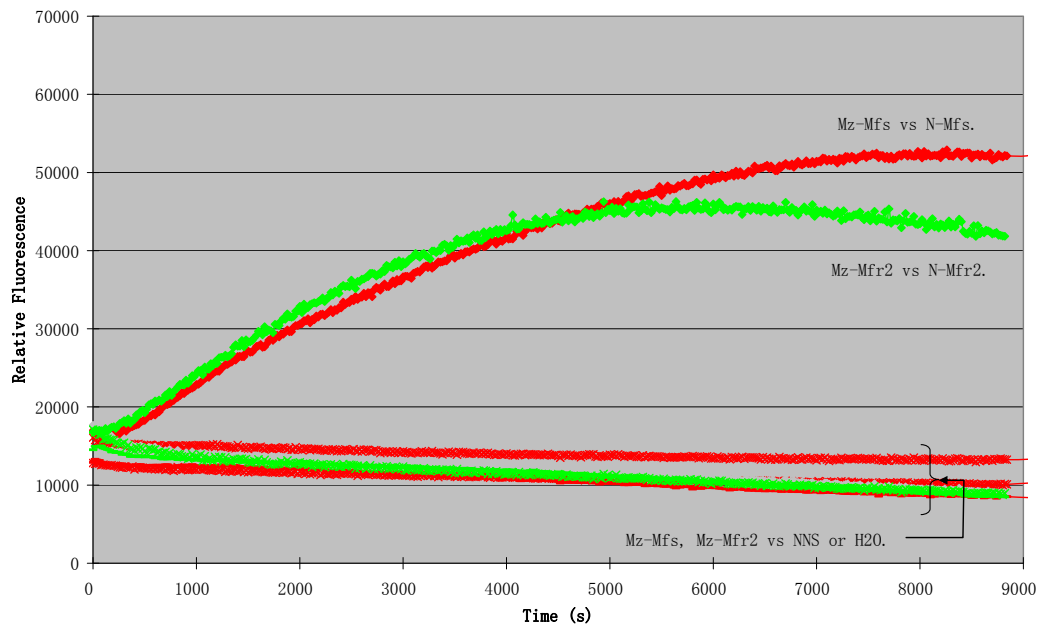
Figure 4.5-6 RTA results for NASBA amplified NTS using Mz based sensors for Mfs and Mfr2.

Separations of the fluorescence curves for NTS and those for NNS or H_2O were clear.
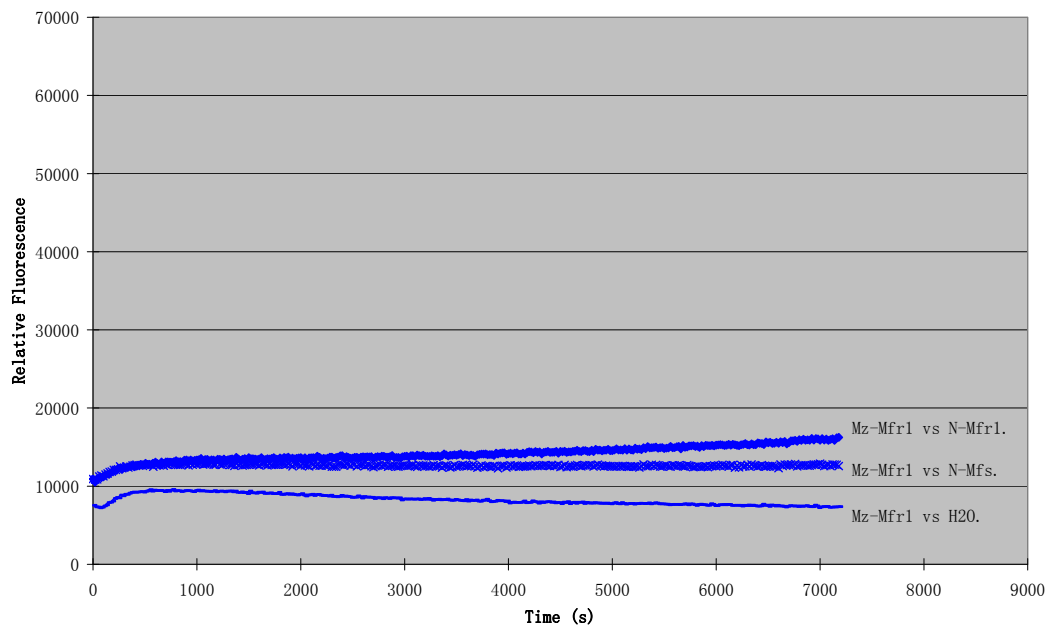
Figure 4.5-7 RTA results for NASBA amplified NTS using Mz based sensors for Mfr1.

Separation of the signal between Mfr1 and Mfs is expected to increase with increased incubation time, as the fluorescence curve for Mfs remained flat (at 7000 s) while that for Mfr1 was rising steadily.

CHAPTER 5   DISCUSSIONS

5.1 Overview of project.

As summarised in **Figure 5.1-1**, e-SRS development can be broadly viewed as 3 major phases,

1) the initial design and testing of an intelligent gene expression platform known as e-SIGE, 2)

the design and testing of the present e-SRS, and 3) application of e-SRS in the detection of

Nucleic Acids in test tubes using RSs.

The following sections will discuss what we learnt from some of the important problems

encountered, and their solutions. The relevance of e-SRS as a molecular gene detection

method, particularly as it pertains to Malaria diagnosis would be discussed. Finally, future

directions for adaptation of e-SRS for other detection formats would be covered.

e-SIGE - Initial design for gene expression platform that activate upon sensing specific nucleic acid sequences.	Designed e-SIGE system [4.1] with sequences for a basic design, Integrated RNA silencing Construct (IRC). Aim 1a and 1b designed to test key elements of IRC mechanisms [4.2]. Synthesised Aim 1a and 1b contructs [4.2.2]. Problem: Assembly PCR yielded many sequence errors. Action: "Gene synthesis via oligonucleotide ligation". Tested IRC via Ribozyme Cis-cleavage Assays (RCA) [4.2.3]. RCA results analysed via Denaturing PAGE were mostly unexpected. Conclusions: 1) Initial e-SIGE sensor was unworkable. 2) New assay needed for analysing RCA that gave clear and immediate signals, and did not involve gel running. 3) Therefore, a new sensor and system had to be designed.

e-SRS with Maxizyme (Mz) based sensors to replace e-SIGE.	Designed e-SRS system [4.3] using modification of allosteric ribozyme, the Maxizyme (Mz). Created RS [4.3.2] to assay Ribozyme Trans-cleavage Assays (RTA).
	
	Successfully tested e-SRS RTA in vitro with RS [4.3.8].
	
	Tested e-SRS in cell lines to:
	    1. Activate inducible gene system [4.4.1].
	Problem: Mz-based sensors did not activate e-SRS.
	Action: Focused on using the simpler HH-2 ribozyme instead.
	Problem: Transfected HH-2 RNA did not activate e-SRS.
	Action: Expressed HH-2 with tRNA promoter, T7 RNA Pol promoter.
	    2. Activate RS [4.4.2].
	Problem: RS degraded when transfected.
	Action: Added nuclease protection to RS.
	Problem: RS degraded by DMEM despite nuclease protection.
	Action: Removed DMEM during transfection.
	
	Conclusions: 1) RS likely degraded in cells. 2) Therefore, it was best to develop e-SRS for in vitro applications.

e-SRS for in vitro nucleic acid detection using RS [4.5].	Designed Mz-based e-SRS sensors for 3 Malaria & 4 Dengue targets [4.5.2]. Problem: Too many Mz-based sensor designs to manually generate & assess. Action: Computational algorithm to greatly automate the design and assessment of Mz-based sensors. Tested e-SRS RTA [4.5.4, 4.5.5]. Problem: Mz-Mfr1 unable to distinguish 1 nt difference. Action: Added "competitor nucleotide" to improve specificity. Tested Malaria e-SRS via NASBA RTA [4.5.6]. Problem: NASBA products did not activate RTA. Action: Added antisense DNA to remove interfering secondary structures. Conclusions: 1) Malaria & Dengue RTA works. 2) Malaria NASBA RTA specific for 2 strains.

Figure 5.1-1 Overview of e-SRS development (relevant Sections in square brackets).

5.2 Gene synthesis via oligonucleotide ligation .

In Section 4.2.2, we attempted to synthesise double stranded DNA templates of our test constructs via assemebly PCR. Surprisingly, this long established technique, which had been used to synthesise genes as long as 3 kb in a single step [Stemmer et al, 1995] failed to produce a single sequenced clone with correct DNA sequence for our short constructs. Many of these clones had stretches of missing segments. For the 89 nt DNA template of HP-LR-CS, we were able to extend a 70 nt middle segment by end primers, using 2 rounds of PCR. However, an attempt to repeat this method for the 184 nt DNA template of T1a (even with PAGE purification of synthesized oligonucleotide) produced similar sequence errors as in assembly PCR.

The problem was unlikely to be due to routine DNA polymerase insertion error as our amplicons were relatively short, and we also eventually used high fidelity DNA polymerases (Elongase Enzyme Mix from Invitrogen and PfuTurbo DNA polymerase from Stratagene). Rather, we suspected that it was the large degree of secondary structures present in our test constructs that caused DNA polymerases to skip certain areas of tight secondary structures and thus resulted in the missing segments.

In contrast, the DNA ligation method, Gene Synthesis via Oligonucleotide Ligation that we used to synthesise the remaining constructs easily and quickly produced clones with correct sequences. The execution of Gene Synthesis via Oligonucleotide Ligation (as described in Sections 3.1.9 and 4.2.2.3) was possibly even simpler than that of assembly PCR, which

required two rounds of PCR. Gene Synthesis via Oligonucleotide Ligation required only one step to anneal (in about 1 hour) all required oligonucleotides, and then an overnight 16 °C ligation of an aliquot of annealed oligonucleotides. The generation of ligation oligonucleotides was also fairly simple, with lengths between about 40 to 70 nts and overlaps of 15 nt. We did not take extra precautions to avoid secondary structures at overlap regions.

A potential feature of Gene Synthesis via Oligonucleotide Ligation that makes it particularly suitable for templates with regions of sequence similarity (as is the case for highly structured templates) is that the annealing process is performed from an intial (denaturating) temperature of 95 °C to an end point of 25 °C at a gradual pace that drops 0.02 °C per second (which can be done in many thermal cycler). This gradual annealing should allow better annealing accuracy than the typical fast annealing process in each amplification cycle of a typical PCR.

The initial stage of assembly PCR involves assembly of an incomplete template, i.e. a semi-double stranded DNA template that has single stranded regions. These single stranded areas are more likely to form secondary structures and thus any PCR based methods utilising incomplete templates are particularly unsuitable for highly structured constructs.

In addition, as PCR tend to favour shorter amplicons (such as those with skipped regions), incorrect sequences with skipped regions tend to be selectively amplified. There is no such selective generation of incorrect sequences for Gene Synthesis via Oligonucleotide Ligation.

In this work, Gene Synthesis via Oligonucleotide Ligation successfully synthesised up to 200 nt long DNA templates with high degree of secondary structures using commercially synthesised unpurified oligonucleotides. This technique also yielded a large percentage of

constructs with correct sequences. We thus believe that Gene Synthesis via Oligonucleotide Ligation can serve as a simple and cost effective way for gene synthesis, particularly those sequences with significant secondary structures.

5.3 Computational algorithm to optimise & assess Maxizyme designs.

Traditionally, the instruction for Maxizyme (Mz) design (given a specific NTS and STS) in the literature [Tanabe et al, 2000; Kuwabara et al, 1998 (Figure 1b)] provides only a fundamental principle that for a valid design, in the presence of NTS, an active secondary structure should be most stable (i.e. lowest free energy), and that in the absence of NTS or presence of NNS, an inactive secondary structure should be most stable. This fundamental principle was elaborated upon in Section 4.3.6.1. As to how such a valid design could be obtained, after deciding on the NTS, NNS, and STS, the existing literature did not seem to elaborate on, and it seemed mainly a process of trial and error, where Mz sequences were adjusted until some valid design(s) emerged in the secondary structure prediction programme. We initially also designed our Mz-based sensors via some trial and error based adjustment of invalid designs until we obtained valid designs (Section 4.3.6.4.2). In the process, we came up with some general strategies & guidelines for achieving desired secondary structures. However, these guidelines were merely that, and as expected, they quite often do not immediately lead to a valid design the first time. There were still a significant amount of trial and error adjustments of Mz-based sensor designs. As such, when the number of categories

and NTS increased, the amount of design work increased greatly (as explained in Section 4.5.2), and we decided to create a computational algorithm to automate much of the design work.

The first step involved making a definition of active and inactive secondary structure that could be computationally assessed (Section 4.3.6.4.1). To do that, we had to first analyse the anatomy of the Mz to delineate the functions of various regions, particularly those responsible for catalytic activity (Section 4.3.6.2). We placed priority on specificity over sensitivity as the strength of e-SRS lies in its ability to distinguish between closely related nucleic acid sequences. This meant that emphasis was placed on avoiding false positives of specific Mz designs, i.e. where RNA combinations without NTS (intended and predicted to be inactive) were actually active. In other words, any structure that had any chance at all of being active should be declared as active in our algorithm. As such, we established the definition of an active structure as the bare minimum required for a Mz to cleave STS. All other structures were considered inactive.

With a computationally assessable definition of an active structure, all that was needed was to create (and then assess) all possible Mz designs (given the NTS and STS). As described in Section 4.5.2.1.1, this could be done by varying three factors: 1) Sensor arm, particularly the SAS, 2) Catalytic arm, particularly the Catalytic Arm Split, and 3) Stem II sequences. However, in order to have the greatest chance of success, we kept unchanged as many aspects of our functionally successful Mz-2 design as possible. Therefore, our algorithm only tested each possible SAS to gather valid designs.

Our computational algorithm did not only reduce the manual design work. As it performs an exhaustive scan of all the possible designs by using all possible SAS, it also produces all the valid Mz designs possible within the given boundaries. If more than one valid design was available, one could select all or the preferred ones for in vitro testing. This should significantly improve the chances of finding the optimal Mz compared to manual trial and error based Mz design. Since there were often more than one valid design per NTS, we also came up with some strategies and guidelines for selecting amongst the valid designs (Section 4.5.2.2.4).

Since this was our first computational algorithm for designing Mz-based e-SRS sensor, there are certainly many areas for improvement. An improved version of the algorithm could include:

1) The ability to insert desired STS. Currently, the STS is fixed.

2) The ability to allow scanning of all possible Catalytic Arm Split by varying the STS sequence around all potential cleavage sites on STS.

3) The ability to insert the desired Stem II sequences.

4) The ability to exhaustively scan all possible Stem II sequences limited by the length of each strand of the stem.

5) The ability to sort and rank valid designs according to various strategies for their selection.

6) A Graphical User Interface (GUI) such that data entry and general usage is easy, even for users without computing background.

In conclusion, our computational algorithm is a prototype tool that paves the way for automatic and systematic Mz design. It allows anyone wishing to design Mz, not necessarily for e-SRS sensors, to save time and labour, and gain the ability to select the optimal designs amongst valid ones.

5.4 Detection of single nucleotide difference.

We had no difficulties in distinguishing the 4 Dengue sNTS, which were greatly different in sequences and were of length between 19 and 22 nt. However, sNTS-Mfs and sNTS-Mfr1 were hardly distinguishable by their respective Mz-based sensors, when we started with standard conditions – both Mz-Mfs and Mz-Mfr1 were strongly activated by both sNTS. The 3 Malaria sNTS were 24 nt long and between Mfs and Mfr1, there was only a Single Nucleotide Difference (SND).

To improve the specificity of the two Mz-based sensors for their own NTS, we tried to optimise RTA temperature and Mg^{2+} concentration. Apparently, only the latter was somewhat effective. This is likely due to the two roles of Mg^{2+} in RTA, which are 1) to participate in catalytic ribozyme cleavage reaction when recruited by the catalytic core of the Mz, and 2) to help anneal the negatively charged RNA strands. As such, with the reduction of RTA Mg^{2+} concentration, the annealing condition becomes more stringent (specific), and the requirement of proper secondary structure (brought about by correct binding of NTS and STS) for the recruitment of Mg^{2+} (so as to activate the Mz) possibly increases.

Mz RTA protocol in general uses 10 mM Mg^{2+}. However, our titration downwards of Mg^{2+}

concentration showed that while RTA activity may decrease, RTA specificity may improve with less than 10 mM Mg^{2+}. In fact, at 2 mM Mg^{2+}, Mz-Mfs became specific, while Mz-Mfr1was the most specific at 3 mM Mg^{2+} (although still not totally specific). This also indicates that for all new Mz designed, RTA could potentially be optimised by varying the Mg^{2+} concentration, although 10 mM might be a good starting point when selecting amongst several valid starting designs.

As it was, while testing the ability of each Mz-based sensor to activate in the presence of both NTS and NNS, we discovered that the presence of NNS invariably reduced the activation produced by the NTS. This provided the inspiration to use a "competitor nucleotide" (CN), which could be a NNS that had been found not to activate the Mz-based sensor (such as sNTS-Mfr2 for the case of Mz-Mfr1) in order to improve specificity. After optimisation, we discovered that adding 2000 nM of sNTS-Mfr2 as CN made Mz-Mfr1 completely specific.

CN seemed to improve specificity by reducing the activity of Mz-based sensor against both NTS and NNS, but doing so more for NNS than for NTS. Too much CN (e.g. 10000 nM and above) can reduce activity for both to non-existant levels. Hence, the key was to find a level of CN that allowed the NTS to activate the Mz-based sensor but reduced activation by NNS to non-significant levels.

Judging by such behaviour of CN, we speculate that CN could bind to Mz-based sensors (at the sensor arm) but due to the significant mismatches with the sensor arm, holds the Mz-based sensor in an inactive conformation. In other words, CN "competes" (hence the name "competitor nucleotide") with both NTS and NNS for binding with the Mz-based sensors, and

holds the sensor in an inactive conformation. As such, CN sequence should be close enough to the NTS sequence to bind the Mz-based sensor, yet different enough that it does not result in sensor activation. In our case, the CN sNTS-Mfr2 had 3 nucleotides different from that of sNTS-Mfr1 (see **Table 4.5-1**) while the NNS sNTS-Mfs had only 1 nucleotide different from sNTS-Mfr1. Since the NTS (sNTS-Mfr1) had complete complementary match to its Mz-based sensor's sensor arm, while the NNS (sNTS-Mfs) had a single complementary mismatch with the same sensor arm, it seems logical that the CN would compete more strongly with the NNS than with the NTS. As a result, the CN reduced the ability to activate the Mz-based sensor more strongly for the NNS than for the NTS.

We therefore believe that the design and use of an appropriate CN could be a useful strategy to help improve Mz specificity that is still lacking after optimisation of Mg^{2+} concentration, especially when there are no other valid designs that could be tested (Mz-Mfr1 had only 1 valid design, as shown in **Table A 6**).

Molecular Beacons (MB) have been established to show good SND distinguishing ability for probe region of 15 nt [Bonnet et al, 1999]. Compared to this, Mz-based e-SRS seems at least comparable in the detection of sNTS, as we could distinguish between 1 SND in a much longer probe region (24 nt). While more SND should be tested for proper comparisons of specificity between Mz-based e-SRS and MB, this initial result is promising. In fact, SND specificity will likely improve if we were to shorten the detection region (to say 15 nt, like that of the MB), as the disruptive effect of the single base mismatch would be even greater on the secondary structure of the Mz-based sensor, thereby reducing the chances of 1) sensor arm

binding to NNS, and 2) an appropriate secondary structure of the Mz catalytic core being formed even if the NNS did bind with the sensor arm.

In testing NASBA RTA, we again had trouble distinguishing SND for Mz-Mfs and Mz-Mfr1, despite using the optimised conditions derived from the sNTS RTA. One possibility is the carry over of Mg^{2+} from the NASBA reaction. However, the NASBA reaction is not stated by the manufacturers to contain Mg^{2+} and our use at 3μl of NASBA product for 50 μl of RTA would likely significantly dilute any Mg^{2+} carried over from NASBA. NASBA reaction does contain K^+, which would add 0.6 mM K^+ eventually to RTA. If this small amount of K^+ had any effect, it would likely be limited to slightly reducing the specificity of RNA annealing, and not serve to increase the ease of Mz catalytic core activation, which requires the recruitment of Mg^{2+}.

We suspect that the major issue was the increased length of NTS provided by NASBA (120 nt) compared to that of sNTS (24 nt). The additional length may result in secondary structures that partially obscure the NTS region where the SND exist and therefore reduce the specificity of Mz-based sensor annealing to the NTS. The additional secondary structures may also interfere with the predicted secondary structure of the Mz-based sensor. The use of AS (antisense oligonucleotides) could be improved on to help in this case, for instance, by having more units of AS that span most of the long NTS such that essentially only the NTS region is single stranded and available for e-SRS sensor detection. This would greatly reduce the amount of single stranded region and thus secondary structures that the long NTS could form and therefore reduce their interference with the Mz-based sensor.

5.5 Use of AS to facilitate the detection of long NTS.

Our earliest attempts at NASBA RTAs produced no significant fluorescence due to the secondary structures of the detected RNA. We initially suspected that a likely factor was that of insufficient NASBA NTS (nNTS) for detection (Section 4.5.6.3). However, despite increasing the amount of NASBA template as well as NASBA products used in RTA, the problem was not resolved. As such, we suspected that the problem was due to secondary structures of the relatively long NTS obscuring the 24 nt detection region, thus denying the Mz-based sensor access to it.

In order to open up the RNA surrounding the detection region, we added AS (antisense oligonucleotides) that would bind to regions flanking and close by to the NTS detection region. As our results showed, AS indeed resolved the non-activation issue in NASBA RTA, and we optimised a set of AS sequences and concentration for the three Malaria NTS.

To our delight, the denaturation action of AS were almost immediate upon their addition to the RTA mix, which suggested that AS functioned via simple AS pairing with nNTS that denatured secondary structures around the NTS rather than via RNase H action that released sNTS from their original nNTS. Indeed, this was corroborated when similar long NTS generated via in vitro transcription (with no RNase H added) could be detected by NASBA RTA when AS were added, but not without addition of AS (data not shown).

We further analysed the functions of AS by predicting secondary structures (using RNAstructure 4.5) of Mz-based sensors activated by nNTS, both in the presence and absence of AS. To predict nNTS structure in the presence of AS (denoted as nNTS+AS) , we combined

the various RNA (e.g. AS, sNTS, Mz, STS) sequences into a single sequence, with "gaaa"

separating each segment, just like how we predicted the binding of Mz-based sensors with

NTS and STS (Section 4.3.6.3). The sequences were joined as given in Table 5.5-1. Our

secondary structure predictions showed that AS bind to their designated regions and prevented

the NTS detection regions from forming secondary structures with external regions.

Table 5.5-1 Strand sequences used in Mz based sensor design.

RNA combinations	Strand sequences (5' to 3')
nNTS+AS.	5' AS, nNTS, 3' AS.
nNTS with Mz and STS.	STS, MzL, nNTS, MzR.
nNTS+AS with Mz and STS.	STS, MzL, 5' AS, nNTS, 3' AS, MzR.

Note that for strand sequences, each comma indicated the addition of a linker "gaaa".

One way to determine whether AS were able to improve the chances of Mz-based sensor

activation was to see if the free energy change of formation for the activated complex (i.e.,

nNTS correctly binding to the Mz-based sensor and STS) was larger with AS added. The

larger (i.e. more negative) the free energy change of formation, the more stable (and thus

likely to occur) the activated complex is compared to the structures before binding. We

therefore compared the free energy change of binding in two cases: 1) nNTS alone

binding with the Mz and STS to form activated complex, 2) nNTS+AS (i.e. nNTS already

in the presence of AS) binding with the Mz and STS to form activated complex. As **Table**

5.5-2 showed, the formation of the actived complex from nNTS resulted in significantly

greater stabilisation (larger free energy change, note last two columns in Table) when AS

was present as compared to when no AS was present. This clearly showed that the

presence of AS promoted the formation of the activated complex, compared to having

only the nNTS alone.

Table 5.5-2 Comparison of the free energy values of nNTS e-SRS activation with and without AS.

NTS	AS		Free Energy Change of formation/process (kcal/mol)						Stability increase for activation in presence of AS	
	5'	3'	nNTS	Activated Complex	e-SRS activation	nNTS+AS	Activated Complex with AS	e-SRS activation	kcal/mol	As a % of the value without AS
Mfs	Mal Rvs 206-214	Mal Rvs 242-271	-15.8	-41.3	-25.5	-45.1	-84	-38.9	-13.4	52.5%
Mfr1	Mal Rvs 206-214	Mal Rvs 242-271	-15.7	-43.4	-27.7	-46.9	-86.1	-39.2	-11.5	41.5%
Mfr2	Mal Rvs 194-215	Mal Rvs 242-271	-14.2	-43.2	-29	-64.3	-104.3	-40	-11	37.9%

Note that a negative value for "Stability increase for activation in presence of AS in kcal/mol" indicates an increase in stability. Free energy change of e-SRS activation was calculated as the free energy of the activated complex subtracted from that of the sNTS (either with or without AS).

It should be noted that the increases in free energy change (in the presence of AS) were not due to the free energy change of AS binding the nNTS as such bindings were already present at the initial stage (nNTS+AS). One factor that may account for this increase in free energy change (when AS is present), is that when AS prevent the detection region from binding to other segments of the nNTS, the detection region would either remain single stranded, or only form bonds within itself, which by probability are usually less stable than those that could have been formed with outside regions (since there are more sequences outside the detection region for long NTS, and binding with outside regions would likely produce less conformational strain). As such, when forming the activated complex with Mz-based sensor and STS, less energy would be needed to break bonds that the detection region had already formed.

Based on the above understanding, AS may be used to facilitate in general all DNA or RNA hybridisation where the probe and the target have difficulty annealing due to secondary structures of the target nucleic acid. This may include facilitating RNA knock down in cells, whether by siRNA, microRNA, or ribozyme, as well probe hybridisation whether in vivo or in vitro. The general principle would be to synthesise AS flanking the detection region. If secondary structure prediction shows the detection region to pair with far off regions, AS could also be targeted to those far off regions. Multiple (i.e. more than 2) sites could in principle be targeted by AS. If choices between several AS had to be made, a comparison of the relative benefits of the AS could be made by comparing which AS enables a greater change in free energy for the intended probe-target hybridisation, as was done above. Using

such a methodology, a computational algortithm could be written to suggest the best AS (of defined number and length) to use for any given probe-target hybridisation.

5.6 Use of e-SRS in cell lines.

e-SRS was not successfully developed in cell lines. The RS system was shown to be unfeasible due to degradation of the RS (Section 4.4.2.3), despite the use of significant nuclease resistant protection. This could be due to the presence of a single unprotected normal phosphodiester RNA bond in the RS, which was required if the Mz-based sensor was to be able to cleave it.

The other e-SRS cellular paradigm was to activate an inducible gene expression system in and only in the presence of NTS. This was also not successfully developed. The problem did not appear to be simply due to the activation of Mz-based sensor, as even the use of the simpler HH-2 (a Hammerhead ribozyme) in an attempt to knock down the TetR mRNA did not allow even the slightest activation of the PLG (Section 4.4.1.5). This was despite various ways of expressing HH-2 in cell lines, where it was affirmed by RT-PCR that HH-2 was indeed expressed (Section 4.4.1.5.4).

A simple explanation could be that the TetR mRNA was too highly expressed to knock down. However, after our subsequent experience of how the optimised conditions of sNTS RTA could fail to work when NASBA NTS, which were long NTS, were used, and how the addition of appropriate AS (antisense oligonucleotides) could almost instantly activate NASBA RTA, we believe that a similar problem of secondary structure was responsible in the

non-activation of PLG. In other words, secondary structures of the TetR mRNA and/or long NTS obscuring the STS cleavage region or NTS detection region could be the main factor preventing the activation of PLG.

As such, a possible solution would be to design and add AS around the STS cleavage region (or elsewhere on the TetR mRNA, as described in Section 5.5) to help activate PLG using HH-2. Once this is successful, transfected Mz-based sensor and sNTS could be tested to activate PLG. Eventually, long NTS as expressed in cells could be tested with Mz-based sensors to activate PLG, and if activation difficulties arise, AS could be designed for the long NTS as well. Note that such AS should be RNA oligonucleotides in order to avoid undesired RNase H degradation of the NTS. In addition, the length of AS RNA should be kept significantly shorter than the typical 19 to 21 nt siRNA length. This is in order to avoid having the duplexed RNA region from being subject to siRNA effects.

5.7 Comparison of e-SRS to other molecular gene detection methods.

As covered in the Introduction, conventional methods of detecting Nucleic Acid sequences (NA) based on hybridisation of complementary oligonucleotides probes to target NA have two intrinsic weaknesses: 1) They are often limited in their ability to detect small differences in sequence, especially single nucleotide differences (SND), and 2) They usually have poor Signal to Noise Ratio (SNR) as typically they have each only 1 signal per hybridisation of probe and target, and that unhybridised probes contribute to a significant background.

In contrast, the existing e-SRS sensor, the Maxizyme (Mz) has been shown to be highly specific in recognition, being able to distinguish SND both in the literature (SND in 16 nt) [Kong et al, 2003; Kong et al, 2005] as well as in our own work (SND in 24 nt). The overall secondary structure of the Mz must be correct before catalytic structure can be achieved for signal generation. The overall secondary structure of the Mz is in turn highly dependent on the correct pairing of the NTS and sensor arm of the Mz. As such, any mismatch in the pairing has a greater chance to disrupt the complex structural requirement of an active complex, thus leading to abolishment of catalytic activity and hence signal generation. In comparison, conventional probes rely merely on hybridisation of 2 short strands of NA, wherein successful hybridisation, whether perfect or with mismatch(es) would lead to signal generation.

In addition, as the Mz-based sensors are designed specifically to distinguish between NTS and NNS via RNA secondary structure prediction software, sensors can be computationally screened to be amongst the better ones (out of many possibilities) to distinguish NTS from NNS.

While having high specificity in detection, the e-SRS sensor is not so specialised as to be limited to detecting only SND, as in the case for some systems, such as QUAL (see Section 5.7.2). Such systems would likely face problems in differentiating between 3 or more NA species with multiple mismatches. This is not a problem for e-SRS as correct secondary structure along the entire Mz sensor arm (not just a specific point along the sensor arm), contributes towards inducing the appropriate conformational change into an active Mz.

Currently, e-SRS utilises fluorescence based RSs that are quenched when non-activated,

resulting in low background signal. In addition, e-SRS has signal amplification power as each activated Mz can cleave many RS as long as the NTS is present, resulting in multiple fluorescent reporter per copy of NTS detected. Both factors contribute to improve SNR for e-SRS.

e-SRS also separates NTS detection and signal generation (RS cleavage) into 2 distinct modules and hence is more amenable to turnover optimisation, where required.

e-SRS therefore has many advantages compared to conventional hybridisation probes. It is hence of greater interest to compare e-SRS with more advanced gene detection methods, such as the Molecule Beacon and some of its derivatives and other methods that incorporate signal amplification.

5.7.1 Comparison with Molecular Beacons and derivatives.

The relatively recent introduction of Molecular Beacons (MB) has significantly improved specificity and SNR over conventional hybridisation probes [Tyagi & Kramer, 1996]. A MB comprises of an oligonucleotide conjugated on opposite ends with a fluorescence reporter and a quencher that are constitutively close by, and which separate upon binding of target NA and thus generate significant fluorescent signal only when hybridised. However a limitation of MB is that hybridisation to target NA is required to generate fluorescence, hence at most 1 fluorescent reporter is generated by each copy of target NA. In addition, specificity is still determined primarily by hybridisation of NA, hence there is still room for improvement in resolution of closely related sequences.

There have been attempts to improve specificity and/or sensitivity of the MB, such as the Hairpin Inversion Probes that reduces interaction of stem sequences with targeted Nucleic Acid (NA) [Browne, 2005]; Binary DNA Probes that separates hybridisation sequences into 2 portions to improve specificity [Kolpashchikov, 2006]; Tentacle probes that provide additional sensing surface to improve sensitivity and specificity [Satterfield et al, 2007]; Nuclease S1 to improve sensitivity by digesting DNA templates of non-perfect matches [Ye et al, 2007]; Single Wall Carbon Nanotube (SWNT) quenches unhybridised MB to reduce background and increases specificity [Yang et al, 2008]. These methods however, have not become mainstream methods, and none of them have amplification of signal.

Activation of the MB is based primarily on hybridisation to create conformational strain that can unwind the MB stem region. Hence, successful binding with mismatch(es) of a wrong sequence will still activate MB, since binding, whether fully complementary or with mismatches will unwind the MB stem, thus releasing quenching of the reporter.

Unlike the MB, for the e-SRS Mz-based sensors to activate, binding of the target NA is required to provide the necessary energetic stabilisation so that correct conformation change can occur for the Mz catalytic core to capture an Mg^{2+} and become catalytically active. As such, activation of the Mz-based sensor upon recognition of target NA does not depend only on hybridisation of the sensor arm with target NA, but also on the induction of appropriate secondary structure changes (set into motion by the hybridisation) that allows an appropriate catalytic core conformation to be adopted. If there was a mismatch(es) in the hybridised target NA, the distortion in secondary structure could impact negatively on the adoption of the active

conformation and thus successful binding of a wrong sequence will not necessarily activate e-SRS.

For applications of MBs with signal amplification, there are two approaches. One of these is an enzyme based method called Nicking Enzyme Signal Amplification (NESA) [Li et al, 2008]. In the basic version, NESA uses a DNA nicking enzyme (N.BstNB I) to cleave an activated MB that has hybridised to its target, which must contains a nicking enzyme recognition sequence (GACTC). After cleaving, the two fragments that remain of the MB dissociate (with the fragment containing the fluorescent reporter permanently unquenched) and allow more MBs to bind and be cleaved, thus creating amplification of signal. An extended version of NESA requires the additional elements of DNA padlock probe, DNA ligase, DNA polymerase, and primer that together allow Rolling Circle Amplification of the target NA, which are then detected like in basic NESA. Rolling Circle Amplification creates many more copies of the target NA with the nicking enzyme recognition sequence added if it was not orginally present in the target NA. This allows greater amplification as well as removed the restriction of only targeting NA with the nicking enzyme recognition sequence.

Compared to e-SRS, NESA, especially the extended version requires the provision of additional protein enzymes and reaction buffers, which tends to make a system more complicated and fragile. Basic NESA requires the DNA nicking enzyme recognition site to be present on both the MB and the sequence of detection, resulting in a very restricted choice of NTS. Rolling Circle Amplification allows additional amplification, but is really a "generic" amplification technique that could be used by conventional hybridisation probes as well

e-SRS. The critical difference it made to NESA is that it allowed NA without the nicking enzyme recognition sequence to be targeted. Ultimately, NESA whether with or without RCA still depends on MB for specificity, and like other MB based sensor systems, NESA is unlikely to be as specific as the e-SRS Mz-based sensors.

The other approach to signal amplification with MB is via the use of catalytic DNA molecules called deoxyribozymes (Dz) that are fused with a MB domain. One group created constructs they called catalytic molecular beacons, whereby they modified a hammerhead-type Dz (able to cleave RNA) with an added MB domain such that one of the MB stem is also the binding arm of the Dz [Stojanovic et al, 2001]. When the MB domain is not activated, the other stem anneals with the binding arm and prevents full binding of a separate DNA substrate sequence. This substrate sequence, like our RS, has a constitutively quenched reporter. Upon activation of the MB domain, the binding arm of the Dz is no longer blocked and allows binding and cleavage of the substrate sequence, resulting in fluorescence. Since apparently the Dz domain remains in a catalytically active conformation even without addition of target NA, activation of such catalytic molecular beacons, like the MB, would be based primarily on simple hybridisation (albeit with competition with a blocking MB stem), and thus is unlikely to match Mz for specificity.

Another group created the Catalytic Beacons [Xiao et al, 2004]. Starting with a Dz that is able to intercalate Hemin (an iron-containing porphyrin) to exhibit peroxidase like activity, they extended the sequence of the Dz with a sequence that can bind the target NA. In the absence of target NA, the additional sequences would form secondary structures with the Dz, thus

disrupting its active structure and rendering it inactive. Upon binding of the targeted NA, secondary structure disruption to the Dz domain is removed and the activated Dz can then generate chemiluminescence in the presence of Hemin, H_2O_2 and luminol. Like the e-SRS Mz-based sensor, the activation of Catalytic Beacon requires conformation change initiated by the binding of target NA. However, for the Mz-based sensor, the target NA helps to structure the binding arm domain while for the catalytic beacon, the target NA serves to remove a complementary segment that disrupted the active conformation. Since the degree of disruption and the ability of target NA to remove this disruption defers for different target NA sequence, it is not clear that such conformational changes can be easily generalised to any target NA sequence. In addition, when the target NA was provided at 200 nM (the concentration of sNTS used in e-SRS sNTS RTA), the activity of the catalytic beacon seemed only slightly above that of when no target NA was provided. This implied that the degree to which the target NA could control the activation of the catalytic beacon was not as high as that for the e-SRS Mz-based sensor.

5.7.2 Comparison with methods with signal amplification.

QUenched Auto-Ligation (QUAL) probes consist of a pair of probes (typically 7 and 17 nt each) designed to anneal on adjacent location on target NA, and which have between them a reporter and a quencher. Upon being positioned adjacent to each other, auto-ligation occurs between the two probes via a 3' phosphorothioate group on the 5' probe, joining them into one long probe, with the concurrent displacement of the reporter or quencher. This results in the

release of fluorescence signal [Xu et al, 2001; Sando et al, 2002]. While QUAL may potentially generate multiple reporters per copy of target NA, this process requires ligated probes to dissociate from the target NA strand. This would be less efficient than in the case of e-SRS Mz as the ligated probe would be around 20 nt, while each cleaved fragment of RS would be about 7-10 nt only.

Further modification was made to QUAL via addition of a short "universal linker" to destabilise the ligated product (which would otherwise be too stable to dissociate significantly for multiple auto-ligation on each copy of target NA) so as to allow some amplification of signal [Abe & Kool, 2004]. However, dissociation rates may not be as high as that of 2 separate 7-10 nt strands (as in e-SRS), as the dissociation of regions near the "universal linker" (without total dissociation of ligated probe) may be sufficient to greatly reduce its destabilising effect, particularly if the combined probe is long.

Pertaining to the optimisation of substrate turnover, QUAL has an intrinsic design limitation as the reporter sequence is the reverse complement of the sequence of detection (i.e. NTS), and hence is dictated by and dependent on the specific NTS of interest. In e-SRS, the RS binding region of the Mz-based sensor is a distinct entity and entirely independent of the sequence of the NTS binding region (responsible for target NA specificity). As such, substrate turnover and NTS binding can be fairly independently optimised, and substrate turnover is much less dependent on the specific NTS of interest. For instance, while the current e-SRS RS is 17 nt long (before cleavage), it can be shortened significantly (to improve dissociation after cleavage) without significantly affecting specificity.

The auto-ligation step limits the specificity of QUAL as this reaction (which results in the activation of QUAL) will occur so long as the two QUAL probes are place adjacent in the right 5' to 3' orientation. This means that the segment of NTS where the sequence can be reliably determined is restricted to the bases close to the junction of the two probes. The system would likely face problem in differentiating between 3 or more NA species with multiple mismatches, where there is no Single Nucleotide Difference (SND) that is unique for a particular NTS. Hence QUAL appears more suited for SND detection (and which must be unique amongst species present) rather than all-purpose NA detection, like for e-SRS. In addition, while QUAL typically distinguishes SND in ~17 nt sequences, we have shown e-SRS capable of distinguishing SND in up to 24 nt sequences.

5.8 Potential advantages of e-SRS compared to PCR based diagnosis of Malaria.

The current gold standard for Malaria diagnosis is the microscopic examination of Giemsa-stained thick and thin blood films [Murray et al, 2008]. Immunochromatographic antigen detection assays have also been developed as Malaria Rapid Diagnostic Tests (MRDTs), which are fast and simple to use [Murray et al, 2008]. However, as discussed in the introduction, both such methods have significant disadvantages.

The key advantage of PCR based methods is their superior sensitivity achieved via the power of exponential amplification of PCR. Unfortunately, the sensitivity of e-SRS NASBA RTA was not determined in this work. Before it was realised that initial NASBA RTAs were

unsuccessful due to the secondary structures of the detected RNA, it was suspected that a likely factor was that of insufficient NTS for detection. As such, increasingly high amounts of NASBA template as well as NASBA products were used in NASBA RTA.This lead to the eventual high concentration of estimated NTS provided in NASBA RTA, even though we eventually realised that the problem was due to NTS secondary structure rather than NTS concentration. Unfortunately, by the end of the project, there were insufficient resources to establish the sensitivity limits of NASBA RTA.

The sensitivity of e-SRS should be further established in future work by titrating downwards the amount of NASBA template used. While the sensitivity of e-SRS NASBA RTA has not yet been established, the process may hold several potential advantages when compared to PCR based methods for diagnosis of diseases.

5.8.1 Specificity.

The amplification power of PCR based methods can be a double edged sword, as any non-specific priming can easily and quickly lead to amplification of non-specific amplicons, resulting in false positives and hence reduced specificity.

PCR based assays exist that provide specificity at the level of Malaria as well as individual human infecting Plasmodium species. There are Malaria consensus quantitative PCR (qPCR) primers that can amplify all 4 human infecting Plasmodium species, and each of these could be distinguished via melting curve analysis [Mangold et al, 2005]. For P. falciparum strains, qPCR assays were developed that distinguishes between Chloroquine sensitive (CQS)

CVMNK haplotype and resistant (CQR) SVMNT and CVIET haplotypes [Farcas et al, 2006].

Consensus proprietary primers were used to amplify the region around the P.f pfcrt gene codon 76, which differs between the three strains. Sequence specific FRET probes for each strain were used. Subsequently, melting curve analysis was used to distinguish between the 3 strains.

If PCR products are assayed directly via non-sequence specific probes, such as ethidum bromide in agarose gel electrophoresis, the reduced specificity of amplification applies directly to the assay specificity if the non-specific amplicons have similar sizes as the specific amplicons. However, if sequence specific probes are used, the reduced specificity of amplification can be modulated by the specificity of the probing process.

This is the case for e-SRS NASBA RTA. While NASBA also amplifies, e-SRS NASBA RTA has an additional layer of specificity as detection requires specific activation of the e-SRS Mz-based sensor. Even if NASBA amplifies a lot of non-specific products, an optimised e-SRS would still not activate unless the intended target was present. This was clearly shown (Section 4.5.6.5) when the optimised e-SRS Mz-Mfs and Mz-Mfr2 only activated when their intended targets were provided, and not when similar sequences were provided. As discussed in Sections 5.4 and 5.7, e-SRS sensor specificity is better than conventional hybridisation based nucleic acid probes and at least comparable to the Molecular Beacon.

In distinguishing between species, the use of consensus amplification primers have the advantage of reducing amplification bias due to primer-target affinity differences [McNamara et al, 2004]. As shown in **Table 4.5-10**, e-SRS NASBA RTA used consensus NASBA primers

for all 3 strains detected. This would be an advantage in quantification accuracy between different strains when compared to PCR based assays that use different primers to distinguish between species with closely related sequences.

5.8.2 Ease of use and flexibility in application.

The e-SRS sensor as applied in NASBA RTA provide great ease of use. Unlike a PCR-based assay, no expensive and sophisticated equipment such as the thermal cycler is required. All that is needed are a constant temperature (41 °C and 37 °C) water bath and a fluorometer. There is also no need for any significant expertise in the analysis of results, unlike that for quantitative PCR melting curve analysis. In addition, fast real time results could be obtained, with no need to run gels or apply probes in a second step.

The PCR as applied in molecular diagnostic is in essence only an amplification method. Only in the case of real-time PCR (without probes) can it be considered to be a detection as well as amplification method (however, the equipment and skills required are sophisticated). e-SRS (employed with the RS) however, is intrinsically both an amplification and a detection method.

It would be desirable to further improve simplicity in future development. This could be done by developing the use of minimally processed samples instead of extracted genetic material. For instance, it would be highly desirable if mosquito tissues or blood, or patients' blood could be used as input. As far as e-SRS is concerned, the major concern is probably ensuring that RNase activities are inhibited.

In addition, as discussed in Section 5.3, our current software tool for e-SRS Mz-based sensor design could be further developed into a user friendly version that allows a typical wet-lab biologist to easily use e-SRS for detection of their nucleic acids of interest.

Due to the modular nature of the e-SRS, meaning the existence of the sensor and the RS as two physically separate entities, there is great flexibility in coupling e-SRS with other amplification (e.g. NASBA) or detection methods. As the situation requires, the existing fluorescence based detection of the e-SRS RTA can be converted into various powerful detection formats, as shown in the following sections.

5.9 Application of e-SRS in other formats of detection.

5.9.1 Coloured dye based detection

To reduce the need for complex equipment, a coloured dye based assay could be developed as a field suitable assay kit. There are two variations of such a colour based assay that could be developed as a small hand held kit with some transparent chambers:

1) RS attached to surface in chamber.

   In this "RS attached" version, the RS is immobilised on a surface in chamber 1 with the free end conjugated to a coloured dye. RTA mixture (less the RS) including the sample Nucleic Acid (NA) is then introduced to chamber 1.

   RTA then occurs in chamber 1, and the cleavage of RS enables its free end with the coloured dye to be released into solution. Upon completion of RTA, the solution is allowed to flow (or is washed) into chamber 2, which was previously colourless or

coloured in contrast to the dye attached to the RS.

Comparison of colours between the 2 chambers can indicate the degree of e-SRS activation.

2) RS not attached (i.e. in solution).

In this "RS not attached" version, the RS is free in a buffer with one end biotinylated and the other end conjugated to a coloured dye. RTA mixture including the RS and the sample NA is prepared and introduced into chamber 1, an inert chamber.

RTA then occurs in chamber 1. Upon completion of RTA, the solution is allowed to flow (or is washed) into chamber 2, with a layer of immobilised streptavidin. The dye conjugated end of the cleaved RS will not attach to the streptavidin coated chamber 2, while those of the uncleaved RS will attach via the conjugated biotin (at the other end of the RS) and remain in chamber 2.

The solution is allowed to flow (or washed) into chamber 3, which was previously colourless or coloured in contrast to the dye attached to the RS.

Comparison of colours between chambers 2 and 3 will indicate the degree of e-SRS activation.

Both methods have their advantages. The "RS attached" method is more convenient, utilises less material, and is thus suitable for mass production of kits that are unlikely to change in reagents (in particular the RS sequence). The "RS not attached" method is more flexible, since the RS required can be ordered and synthesised separately, according to the needs of individual applications. The kit itself (less the RS) can therefore be manufactured in large bulk,

and be offered for different applications.

5.9.2 Silicon Nanowire based electrical detection

For greater sensitivity and digital readout, the e-SRS sensor could be coupled to the Silicon

Nanowires (SiNW) electrical detection method to combine the selectivity and amplification of

the e-SRS sensor with the sensitivity and the simple, rapid and electrical read out of the

SiNW.

SiNW as an electrical biomolecular detection method has in recent been shown to hold great

potential [Gao et al, 2007]. Sensitivity had been reported to reach 1 fM of RNA (signal

strength increases with DNA concentration) [Zhang et al, 2009] while selectivity for single

base mismatch had been reported [Zhang et al, 2008a; Zhang et al, 2009]. The electrical

detection output is easily digitised and the whole setup can be minaturised into tiny hand held

SiNW chips that provide results without the use of labels (e.g. fluorescent probes or coloured

dyes) [Gao et al, 2007].

SiNW based biomolecular detection is based on the synthesis of a SiNW chip, which typically

contains an array of many SiNW, each with molecular probes preattached. A small volume of

the sample nucleic acid (NA) is then introduced in an appropriate buffer solution onto the

SiNW chip. Each SiNW will change significantly in electrical resistance when target

molecules bind to the attached probes, as these target molecules bring significant amount of

charges into close proximity of the SiNW [Zhang et al, 2008b]. In the case of NA,

complementary Peptide Nucleic Acid (PNA) probes could be conjugated onto SiNW. When

target NA bind to their probe on the SiNW, the negative charges on the target NA sugar phosphate backbones would cause an increase in the resistance of the SiNW. PNA probes are commonly used as the lack of charges on their backbone enable very low contribution to the background electrical resistance, and allow high hybridisation efficiency [Zhang et al, 2008a]. The changes in resistance of the SiNW can be easily and accurately determined, leading to easy realtime quantification of target-probe binding.

There are two variations by which the e-SRS sensor could be coupled to the SiNW chip to obtain an e-SRS SiNW chip for detection of NA:

1) Direct RS conjugation to SiNW.

   In this "direct RS conjugation" method, instead of having PNA probes conjugated to SiNW, the RSs are conjugated in their place.

   The RTA mixture with an appropriate e-SRS sensor (less the RS) including the sample Nucleic Acid (NA) is prepared with appropriate buffer. A small volume of the whole mixture is added to the SiNW chip, where RTA then occurs. e-SRS sensors activated by the target NA (if present) would then cleave the RS conjugated on the SiNW.

   Since the RS were RNA and thus contained negative charges on their sugar phosphate backbones, their cleavage would reduce the negative charges concentration on their attached SiNW, and therefore result in a reduction of electrical resistance of the SiNW. The reduction of resistance of the SiNW would therefore indicate the amount of activation of e-SRS sensors, which in turn indicate the amount of target NA binding

by the e-SRS sensoss.

2) Indirect RS conjugation via binding to short PNA linkers on SiNW.

In this "indirect RS conjugation" method, instead of directly conjugating the RS to SiNW, RS are annealed to short complimentary PNA linkers that are conjugated on the SiNW. These short PNA linkers (with short defined sequence, such as "GGGGG") are conjugated in place of the PNA probes on existing SiNW chips. The complementary linker sequence (e.g. "CCCCC") is added to the appropriate end of the RS sequence, such that the RS can anneal to the PNA linkers (with the remainder of the RS pointing away from the SiNW).

The RS can then be added in two ways:

a. The RS is applied first to the SiNW chip, and allowed to bind onto the PNA linkers. The SiNW resistance is monitored to ensure sufficient RS has bound to the SiNW via the PNA linkers. When the resistance value stabalise, excess RS are washed off. The e-SRS SiNW chip is then ready for use as in the "direct RS conjugation" method.

b. The same RS as above (with linkers attached) is used, but the RTA mixture is prepared in whole (i.e. with all components, including RS, e-SRS sensor, and sample NA), and a small volume is added to the SiNW chip (with PNA linker conjugated). The RS, whether cleaved or uncleaved will then bind to the PNA linkers on the SiNW. The results would be measured and analysed just like in the "direct RS conjugation" method.

While the SiNW based detection is in itself a very promising biomolecular detection method, there are significant advantages that combining e-SRS with SiNW can bring to existing SiNW based NA detection.

1) Improved sensitivity via amplification by e-SRS.

   The coupling of e-SRS sensors with SiNW detection would allow for amplification of signal due to the ability of each activated e-SRS sensor to cleave many RS. This would therefore further increase the sensitivity of the assay.

2) Improved selectivity conferred by the e-SRS sensor.

   The existing SiNW NA detection is still based on simple hybridisation of target NA with their probes conjugated on SiNW, just like the conventional hybridisation probes described in Section 1.1. As such the e-SRS sensors could offer similar improvements on selectivity for SiNW as they could for the conventional hybridisation probes.

   As discussed in [Zhang et al, 2008b], non-fully complementary target nucleic acid (NA) seemed to be able to hybridise almost as effectively as fully complementary target NA. In diagnosing between highly similar NA, this would lead to some signal being generated even when the NA bound to probe is not the target NA, since the non-targeted NA bound would also bring similar amount of negative charge close to the SiNW, leading to resistance changes. As explained in Section 5.7.1, successful binding of a non-targeted (i.e. non-fully complementary) NA would not necessarily activate e-SRS, meaning that such binding would not activate the e-SRS sensor to cleave the RS (on the SiNW), and therefore not lead to resistance changes of the SiNW (which are indicative of NA binding).

In [Zhang et al, 2008a], for a 22 nt DNA at 1 nM, target NA hybridisation resulted in a signal about 6.375 times that for a Single Nucleotide Difference (SND) hybridisation. In the case of e-SRS RTA, as described in Section 4.5.5 and as shown in **Figure 4.5-2** for the sNTS-Mfs (a 24 nt RNA) at 200 nM (with 2000 nM competitor nucleotide), at about 95 minutes, the e-SRS sensor registered relative fluorescence units of 55500, 15375, and 10890 respectively for the target NA, SND NA, and water control. After normalising with respect to the water control, target NA hybridisation resulted in signal about ~9.946 times that of SND hybridisation. As such, the e-SRS sensor could potentially add to the specificity of the SiNW chip.

3) Simpler fabrication of SiNW chips.

A major challenge in SiNW based detection is the consistent and precise large scale fabrication of the SiNW chip, which includes the attachment of various probes on the SiNW [Gao et al, 2007]. The conjugation of probe of different sequences and lengths on each new SiNW chip could add to variation of fabrication conditions and thus increase the difficulty in ensuring consistency in SiNW chip fabrication.

When coupled to e-SRS using the "direct RS conjugation", the same RS could be conjugated on all chips, with the application specific e-SRS sensor added to the sample rather than pre-attached on the SiNW chip. Alternatively, if the "indirect RS conjugation" is used, fabrication is even easier as only a standard short PNA linker need be conjugated to the SiNW. This configuration could be mass produced for different applications which could have their own specific RS easily synthesised as

short RNAs. Hence, fabrication of e-SRS SiNW chips would be simpler than traditional SiNW chips and result in greater ease of production and consistency of signal.

4) Improvement in detection format.

Typical SiNW based detection for NA measures the increase in SiNW resistance when NA bind to PNA probes [Gao et al, 2007]. The SiNW resistance is actually determined by monitoring the current following through the SiNW, when a known voltage is applied. The higher the resistance becomes, the smaller the current monitored, which has a minimum of zero ampere (for a positive current). Due to the inverse relationship between current and resistance, the larger the resistance increase, the smaller the current decrease. As such, this detection format becomes less sensitive as more target NA bind to their probe. In e-SRS SiNW based detections where the RS were attached before the addition of sample NA, the resistance of SiNW actually reduces as more RS are cleaved from the SiNW. This means that the monitored current actually increases, and increases more rapidly with more RS being cleaved. Therefore incorporating the e-SRS sensor in SiNW detection may actually improve the existing detection format.

Bibliography

1. Abe, H. and Kool, E.T. (2004) Destabilizing universal linkers for signal amplification in self-ligating probes for RNA. J. Am. Chem. Soc. 126, 13980–13986.

2. Ananvoranich S, Perreault JP. (1998) Substrate specificity of delta ribozyme cleavage. J Biol Chem. 1998 May 22;273(21):13182-8.

3. Bayer TS, Smolke CD. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat Biotechnol. 2005 Mar;23(3):337-43. Epub 2005 Feb 20.

4. Bergeron LJ, Perreault JP. (2005) Target-dependent on/off switch increases ribozyme fidelity. Nucleic Acids Res. 2005 Feb 24;33(4):1240-8. Print 2005.

5. Berzal-Herranz A, Joseph S, Burke JM. (1992) In vitro selection of active hairpin ribozymes by sequential RNA-catalyzed cleavage and ligation reactions. Genes Dev. 1992 Jan;6(1):129-34.

6. Bonnet G, Tyagi S, Libchaber A, Kramer FR. (1999) Thermodynamic basis of the enhanced specificity of structured DNA probes. Proc Natl Acad Sci U S A. 1999 May 25;96(11):6171-6.

7. Boyd SD. (2008) Everything you wanted to know about small RNA but were afraid to ask. Lab Invest. 2008 Jun;88(6):569-78. Epub 2008 Apr 21. Review.

8. Browne KA. (2004) Sequence-specific, self-reporting hairpin inversion probes. J Am Chem Soc. 2005 Feb 16;127(6):1989-94.

9. Breaker RR. (2004) Natural and engineered nucleic acids as tools to explore biology. Nature. 2004 Dec 16;432(7019):838-45. Review.

10. Buck AH, Campbell CJ, Dickinson P, Mountford CP, Stoquert HC, Terry JG, Evans SA, Keane LM, Su TJ, Mount AR, Walton AJ, Beattie JS, Crain J, Ghazal P. (2007) DNA nanoswitch as a biosensor. Anal Chem. 2007 Jun 15;79(12):4724-8. Epub 2007 May 18.

11. Chowrira BM, Berzal-Herranz A, Burke JM. (1991) Novel guanosine requirement for catalysis by the hairpin ribozyme. Nature. 1991 Nov 28;354(6351):320-2.

12. Chowrira BM, Pavco PA, McSwiggen JA. (1994) In vitro and in vivo comparison of hammerhead, hairpin, and hepatitis delta virus self-processing ribozyme cassettes. J Biol Chem. 1994 Oct 14;269(41):25856-64.

13. Deen JL, Harris E, Wills B, Balmaseda A, Hammond SN, Rocha C, Dung NM, Hung NT, Hien TT, Farrar JJ. (2006) The WHO dengue classification and case definitions: time for a reassessment. Lancet. 2006 Jul 8;368(9530):170-3.

14. Dirks RW, Molenaar C, Tanke HJ. (2003) Visualizing RNA molecules inside the nucleus of living cells. Methods, vol. 29, no. 1, pp. 51-57, Jan 2003.

15. Du Quan (1999) http://academic.brooklyn.cuny.edu/chem/zhuang/QD/toppage1.htm

16. Gao Z, Agarwal A, Trigg AD, Singh N, Fang C, Tung CH, Fan Y, Buddharaju KD, Kong J. (2007) Silicon nanowire arrays for label-free detection of DNA. Anal Chem. 2007 May 1;79(9):3291-7. Epub 2007 Apr 4.

17. Griffith KS, Lewis LS, Mali S, Parise ME. (2007) Treatment of malaria in the United States: a systematic review. JAMA. 2007 May 23;297(20):2264-77. Review.

18. Farcas GA, Soeller R, Zhong K, Zahirieh A, Kain KC. (2006) Real-time polymerase chain reaction assay for the rapid detection and characterization of chloroquine-resistant Plasmodium falciparum malaria in returned travelers. Clin Infect Dis. 2006 Mar 1;42(5):622-7. Epub 2006 Jan 25.

19. Fujitani K, Sasaki-Tozawa N, Kikuchi Y. (1993) Different target-site specificities of the hairpin ribozyme in cis and trans cleavages. FEBS Lett. 1993 Sep 27;331(1-2):155-8.

20. Guzmán MG, Kourí G. (2004) Dengue diagnosis, advances and challenges. Int J Infect Dis. 2004 Mar;8(2):69-80.

21. Hänscheid T. (2003) Current strategies to avoid misdiagnosis of malaria. Clin Microbiol Infect. 2003 Jun;9(6):497-504.

22. Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. Science 287, 820−825, 2000.

23. Hertel KJ, Herschlag D, Uhlenbeck OC. (1996) Specificity of hammerhead ribozyme cleavage. EMBO J. 1996 Jul 15;15(14):3751-7.

24. Invention Disclosure, ETPL (2005) Invention Disclosure (ID) titled "ENVIRONMENT-SENSING INDUCED GENE EXPRESSION (E-SIGE)" submitted to Exploit Technologies Pte Ltd (ETPL) on 17th May 2005.

25. Jonkman A, Chibwe RA, Khoromana CO, Liabunya UL, Chaponda ME, Kandiero GE, Molyneux ME, Taylor TE. (1995) Cost-saving through microscopy-based versus presumptive diagnosis of malaria in adult outpatients in Malawi. Bull World Health Organ. 1995;73(2):223-7.

26. Kilian AH, Metzger WG, Mutschelknauss EJ, Kabagambe G, Langi P, Korte R, von Sonnenburg F. (2000) Reliability of malaria microscopy in epidemiological studies: results of quality control. Trop Med Int Health. 2000 Jan;5(1):3-8.

27. Kong XJ, Song YH, Lin JS, Huang HJ, Wang NX, Liu NZ, Li B, Jin YX. (2003) Maxizyme-mediated specific inhibition on mutant-type p53 in vitro. World J Gastroenterol. 2003 Jul;9(7):1571-5.

28. Kong XJ, Lin JS, Song YH, Jin YX. (2005) Inhibition of mutant-type p53 by a chimeric U6 maxizyme in hepatocellular carcinoma cell lines. Zhonghua Gan Zang Bing Za Zhi. 2005 Oct;13(10):759-62. Chinese.

29. Kolpashchikov DM. (2006) A binary DNA probe for highly specific nucleic Acid recognition. J Am Chem Soc. 2006 Aug 16;128(32):10625-8.

30. Kubota T, Ikeda S, Okamoto A. (2007) RNA analysis with a novel fluorescent oligonucleotide. Nucleic Acids Symp Ser (Oxf). 2007;(51):291-2.

31. Kuwabara T, Warashina M, Tanabe T, Tani K, Asano S, Taira K. (1998) A novel allosterically trans-activated ribozyme, the maxizyme, with exceptional specificity in vitro and in vivo. Mol Cell. 1998 Nov;2(5):617-27.

32. Kuwabara T, Warashina M, Sano M, Tang H, Wong-Staal F, Munekata E, Taira K. (2001) Recognition of engineered tRNAs with an extended 3' end by Exportin-t

(Xpo-t) and transport of tRNA-attached ribozymes to the cytoplasm in somatic cells. Biomacromolecules. 2001 Winter;2(4):1229-42.

33. Lanciotti RS, Calisher CH, Gubler DJ, Chang GJ, Vorndam AV. (1992) Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. J Clin Microbiol. 1992 Mar;30(3):545-51.

34. Li JJ, Chu Y, Lee BY, Xie XS. (2008) Enzymatic signal amplification of molecular beacons for sensitive DNA detection. Nucleic Acids Res. 2008 Apr;36(6):e36. Epub 2008 Feb 27.

35. Lukhtanov EA, Lokhov SG, Gorn VV, Podyminogin MA, Mahoney W. (2007) Novel DNA probes with low background and high hybridization-triggered fluorescence. Nucleic Acids Res. 2007;35(5):e30. Epub 2007 Jan 26.

36. McNamara DT, Thomson JM, Kasehagen LJ, Zimmerman PA. (2004) Development of a multiplex PCR-ligase detection reaction assay for diagnosis of infection by the four parasite species causing malaria in humans. J Clin Microbiol. 2004 Jun;42(6):2403-10.

37. Mangold KA, Manson RU, Koay ES, Stephens L, Regner M, Thomson RB Jr, Peterson LR, Kaul KL. (2005) Real-time PCR for detection and identification of Plasmodium spp. J Clin Microbiol. 2005 May;43(5):2435-40.

38. Mens P, Spieker N, Omar S, Heijnen M, Schallig H, Kager PA. (2007) Is molecular biology the best alternative for diagnosis of malaria to microscopy? A comparison between microscopy, antigen detection and molecular tests in rural Kenya and urban Tanzania. Trop Med Int Health. 2007 Feb;12(2):238-44.

39. Ministry of Health. (2004) A guide on infectious diseases of public health importance in Singapore (6th Edition).

40. Ministry of Health. (2006) Communicable Disease Surveillance in Singapore 2006.

41. Ministry of Health. (2009) Weekly Infectious Disease Bulletin, for Epidemiological Week 37. (http://www.moh.gov.sg/mohcorp/statisticsweeklybulletins.aspx)

42. Murray CK, Gasser RA Jr, Magill AJ, Miller RS. (2008) Update on rapid diagnostic testing for malaria. Clin Microbiol Rev. 2008 Jan;21(1):97-110. Review.

43. Nimjee SM, Rusconi CP, Sullenger BA. (2005) Aptamers: an emerging class of therapeutics. Annu Rev Med. 2005;56:555-83. Review.

44. Oishi K, Saito M, Mapua CA, Natividad FF. (2007) Dengue illness: clinical features and pathogenesis. J Infect Chemother. 2007 Jun;13(3):125-33. Epub 2007 Jun 21. Review.

45. Oshima K, Kawasaki H, Soda Y, Tani K, Asano S, Taira K. (2003) Maxizymes and small hairpin-type RNAs that are driven by a tRNA promoter specifically cleave a chimeric gene associated with leukemia in vitro and in vivo. Cancer Res. 2003 Oct 15;63(20):6809-14.

46. Pérez-Ruiz M, Barroso-DelJesus A, Berzal-Herranz A. (1999) Specificity of the hairpin ribozyme. Sequence requirements surrounding the cleavage site. J Biol Chem. 1999 Oct 8;274(41):29376-80.

47. Rusconi CP, Roberts JD, Pitoc GA, Nimjee SM, White RR, Quick G Jr, Scardino E,

Fay WP, Sullenger BA. (2004) Antidote-mediated control of an anticoagulant aptamer in vivo. Nat Biotechnol. 2004 Nov;22(11):1423-8.

48. Rydzanicz R, Zhao XS, Johnson PE. (2005) Assembly PCR oligo maker: a tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W521-5.

49. Philip Samuel P, Tyagi BK. (2006) Diagnostic methods for detection & isolation of dengue viruses from vector mosquitoes. Indian J Med Res. 2006 May;123(5):615-28. Review.

50. Sando, S. and Kool, E.T. (2002) Quencher as leaving group: efficient detection of DNA-joining reactions. J. Am. Chem. Soc. 124, 2096–2097

51. Sano M, Taira K. (2004) Ribozyme expression systems. Methods Mol Biol. 2004;252:195-207.

52. Satterfield BC, West JA, Caplan MR. (2007) Tentacle probes: eliminating false positives without sacrificing sensitivity. Nucleic Acids Res. 2007;35(10):e76. Epub 2007 May 21.

53. SBIC Innovative Grant (2005) 1st SBIC Innovative Grant (2005) application. The grant was awarded to develop e-SIGE as a bioimaging tool. Project Title: A Novel E-SIGE Technology to Bioimage Alternative Splicing Activity in Neuron and Muscle.

54. Schneider P, Wolters L, Schoone G, Schallig H, Sillekens P, Hermsen R, Sauerwein R. (2005) Real-time nucleic acid sequence-based amplification is more convenient than real-time PCR for quantification of Plasmodium falciparum. J Clin Microbiol. 2005 Jan;43(1):402-5.

55. Simmons CP, Farrar J. (2009) Changing patterns of dengue epidemiology and implications for clinical management and vaccines. PLoS Med. 2009 Sep;6(9):e1000129. Epub 2009 Sep 1.

56. Soda Y, Tani K, Bai Y, Saiki M, Chen M, Izawa K, Kobayashi S, Takahashi S, Uchimaru K, Kuwabara T, Warashina M, Tanabe T, Miyoshi H, Sugita K, Nakazawa S, Tojo A, Taira K, Asano S. (2004) A novel maxizyme vector targeting a bcr-abl fusion gene induced specific cell death in Philadelphia chromosome-positive acute lymphoblastic leukemia. Blood. 2004 Jul 15;104(2):356-63. Epub 2004 Mar 23.

57. Soukup JK, Soukup GA. (2004) Riboswitches exert genetic control through metabolite-induced conformational change. Curr Opin Struct Biol. 2004 Jun;14(3):344-9. Review.

58. Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. Gene. 1995 Oct 16;164(1):49-53.

59. Stojanovic, M.N., de Prada, P. and Landry, D.W. (2001) Catalytic molecular beacons. Chembiochem, 2, 411–415.

60. Stoppacher R, Adams SP. (2003) Malaria deaths in the United States: case report and review of deaths, 1979-1998. J Forensic Sci. 2003 Mar;48(2):404-8. Review.

61. Sun LQ, Cairns MJ, Saravolac EG, Baker A, Gerlach WL. (2000) Catalytic nucleic

acids: from lab to applications. Pharmacol Rev. 2000 Sep;52(3):325-47. Review.

62. Taira K, Warashina T, Warashina M, Kawasaki H, Hara T, Nozawa I. (2004) NOVEL MAXIZYME. United States Patent Application Publication No.: US 2004/0248114 A1. Publication Date: Dec. 9, 2004.

63. Tanabe T, Takata I, Kuwabara T, Warashina M, Kawasaki H, Tani K, Ohta S, Asano S, Taira K. (2000) Maxizymes, novel allosterically controllable ribozymes, can be designed to cleave various substrates. Biomacromolecules. 2000 Spring;1(1):108-17.

64. Tuerk C, Gold L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science. 1990 Aug 3;249(4968):505-10.

65. Tyagi, S. and Kramer, F.R. (1996) Molecular beacons: probes that fluoresce upon hybridization. Nat. Biotechnol. 14, 303–308.

66. Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. Nature. 2004 Mar 18;428(6980):281-6.

67. Wrzesinski J, Legiewicz M, Smólska B, Ciesiolka J. (2001) Catalytic cleavage of cis- and trans-acting antigenomic delta ribozymes in the presence of various divalent metal ions. Nucleic Acids Res. 2001 Nov 1;29(21):4482-92.

68. Xiao Y, Pavlov V, Niazov T, Dishon A, Kotler M, Willner I. (2004) Catalytic beacons for the detection of DNA and telomerase activity. J Am Chem Soc. 2004 Jun 23;126(24):7430-1.

69. Xu Y, Karalkar NB, Kool ET. (2001) Nonenzymatic autoligation in direct three-color detection of RNA and DNA point mutations. Nat. Biotechnol. 19,148–152

70. Yang R, Jin J, Chen Y, Shao N, Kang H, Xiao Z, Tang Z, Wu Y, Zhu Z, Tan W. (2008) Carbon nanotube-quenched fluorescent oligonucleotides: probes that fluoresce upon hybridization. J Am Chem Soc. 2008 Jul 2;130(26):8351-8. Epub 2008 Jun 5.

71. Ye S, Miyajima Y, Ohnishi T, Yamamoto Y, Komiyama M. (2007) Combination of peptide nucleic acid beacon and nuclease S1 for clear-cut genotyping of single nucleotide polymorphisms. Anal Biochem. 2007 Apr 15;363(2):300-2. Epub 2006 Dec 28.

72. Zhang GJ, Chua JH, Chee RE, Agarwal A, Wong SM, Buddharaju KD, Balasubramanian N. (2008a) Highly sensitive measurements of PNA-DNA hybridization using oxide-etched silicon nanowire biosensors. Biosens Bioelectron. 2008 Jun 15;23(11):1701-7. Epub 2008 Feb 13.

73. Zhang GJ, Zhang G, Chua JH, Chee RE, Wong EH, Agarwal A, Buddharaju KD, Singh N, Gao Z, Balasubramanian N. (2008b) DNA sensing by silicon nanowire: charge layer distance dependence. Nano Lett. 2008 Apr;8(4):1066-70. Epub 2008 Mar 1.

74. Zhang GJ, Chua JH, Chee RE, Agarwal A, Wong SM. (2009) Label-free direct detection of MiRNAs with silicon nanowire biosensors. Biosens Bioelectron. 2009 Apr 15;24(8):2504-8. Epub 2009 Jan 4.

Appendices

Sequences

Sequences in PCR of PLG for adding short tags with restriction sites.

Table A 1 Sequences in PCR of PLG for adding short tags with restriction sites.

Role / Construct / Length (nt)	Sequence (5' to 3')
Forward primer GFP_XhoI Fwd_1 20	AATACTCGAGCGCCACCATG
Reverse primer GFP_XbaI Rvs_1 22	GCCCTCTAGACTTGTACAGCTC
EGFP amplicon XhoI-EGFP-XbaI 744	AATACTCGAGCGCCACCATGgtgagcaagggcgaggagctgttcaccgg ggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaag ttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctga ccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccac cctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctacccc gaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggct acgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagac ccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgag ctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagc tggagtacaactacaacagccacaacgtctatatcatggccgacaagca gaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggac ggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcg acggccccgtgctgctgcccgacaaccactacctgagcacccagtccgc cctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggag ttcgtgaccgccgccgggatcactctcggcatggacGAGCTGTACAAGT CTAGAGGGC

Role / Construct / Length (nt)	Sequence (5' to 3')
ECFP amplicon XhoI-ECFP-XbaI 744	AATA`CTCGAG`CGCCACCATGgtgagcaagggcgaggagctgttcaccgg ggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaag ttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctga ccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccac cctcgtgaccacccтgacctggggcgtgcagtgcttcagccgctacccc gaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggct acgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagac ccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgag ctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagc tggagtacaactacatcagccacaacgtctatatcaccgccgacaagca gaagaacggcatcaaggccaacttcaagatccgccacaacatcgaggac ggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcg acggcccccgtgctgctgcccgacaaccactacctgagcacccagtccgc cctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggag ttcgtgaccgccgccgggatcactctcggcatggac<u>GAGCTGTACAAGT</u> <u>CTAGA</u>GGGC
EYFP amplicon XhoI-EYFP-XbaI 744	AATA`CTCGAG`CGCCACCATGgtgagcaagggcgaggagctgttcaccgg ggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaag ttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctga ccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccac cctcgtgaccaccttcggctacggcctgcagtgcttcgcccgctacccc gaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggct acgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagac ccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgag ctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagc tggagtacaactacaacagccacaacgtctatatcatggccgacaagca gaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggac ggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcg acggcccccgtgctgctgcccgacaaccactacctgagctaccagtccgc cctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggag ttcgtgaccgccgccgggatcactctcggcatggac<u>GAGCTGTACAAGT</u> <u>CTAGA</u>GGGC

Shaded regions indicated restriction sites. In amplicon sequences, upper cases indicated primer sequences or their reverse complements, and underlined regions indicated segments on template that annealed to primers.

Possible sequences of e-SIGE IRC segments.

Table A 2 Possible sequences of e-SIGE IRC segments.

Segments	Sequences (5' to 3')
Hairpin ribozyme RC Variant 1	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
Hairpin ribozyme RC Variant 2	AACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
Hairpin ribozyme RC Variant 3	ACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
Hairpin ribozyme CS Variant 1	TGACa^gtcCTGTTT
Hairpin ribozyme CS Variant 2	TGACa^gtcCTGTT
Hairpin ribozyme CS Variant 3	TGACa^gtcCTGT
Hairpin ribozyme CS-1'_R	GTCACCAAATAGGA
Hairpin ribozyme RC-Linker-CS	RC- AGATCT -CS
Hairpin ribozyme CS_D	TGACttcgCTGTTT
Hammerhead ribozymes RC Variant 1	TCCGGTCTGATGAGTCCGTGAGGACGAAACAGG
Hammerhead ribozymes RC Variant 2	CCGGTCTGATGAGTCCGTGAGGACGAAACAGG
Hammerhead ribozymes RC Variant 3	CGGTCTGATGAGTCCGTGAGGACGAAACAGG

Segments	Sequences (5' to 3')
Hammerhead ribozymes CS Variant 1	CTgtc^aCCGG
Hammerhead ribozymes CS Variant 2	Tgtc^aCCGG
Hammerhead ribozymes CS Variant 3	Tgtc^aCCG
Hammerhead ribozymes CS_D	CTactgCCGG
Delta ribozyme RC Variant 1	GGGTCCACCTCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTAAGGGACCC
Delta ribozyme RC Variant 2	GGGTCCACCTCCTCGCGGTaacacatGGGCATCCGTTCGCGGATGGCTAAGGGACCC
Delta ribozyme RC Variant 3	GGGTCCACCTCCTCGCGGTtactgatGGGCATCCGTTCGCGGATGGCTAAGGGACCC
Delta ribozyme CS Variant 1	^GTGGTTT
Delta ribozyme CS Variant 2	^GTGTGTT
Delta ribozyme CS Variant 3	^GTCAGTA
Delta ribozyme CS|Linker|RC	CS|AT|RC
Delta ribozyme CS_D	TTCCTTT
A1 Variant 1	gttcggcagc
A2 Variant 1	tggatccc

Segments	Sequences (5' to 3')
A3 Variant 1	aa
A3A2A1 Variant 1	aatggatcccgttcggcagc
A3A2A1 Variant 2	tcgggtgcgcactatcacag
A3A2A1 Variant 3	tgtaaccgcgccaagctgtg
A3_2A2A1 Variant 1	aatggatcccgttcggcagc
B1 Variant 1	GCTGCCGAAC
B1 Variant 2	CTGTGATAGT
B1 Variant 3	CACAGCTTGG
B2 Variant 1	GGGATCCA
B2 Variant 2	GCGCACCC
B2 Variant 3	CGCGGTTA
B3 Variant 1	TT
B3 Variant 2	GA
B3 Variant 3	CA
C variant 1	gggatccatt
C variant 2	gcgcacccga
C variant 3	cgcggttaca
C-1_R	CTTGTACGGA
E variant 1 (37a)	ATTCACTATA
E variant 2 (37a)	GCCCCCT
E variant 3 (37b)	ATGCCTTACC
E variant 4 (37b)	CCCACCC

Segments	Sequences (5' to 3')
D2D1 variant 1a (37a)	tacatgtccttatagtgaat
D2D1 variant 1b (37b)	tacatgtccgggtaaggcat
D2D1 variant 2a (37a)	tgcctaagccgagaggggggc
D2D1 variant 2b (37b)	tccccagacccagggggtggg
F variant 1	ccaatacttacgcgtcg
	gtacggggtgactacta
	acagtcgcgcggaagcc
F variant 2	ccaatacttacgcgt
	gtacggggtgactac
	Acagtcgcgcggaag
F variant 3	ccaatacttacg
	gtacggggtgac
	acagtcgcgcgg
L1	TTCAAGAGA
L2	TTCAAGAGA
L3	TTCAAGAGA

Note: For RC and CS, lower cases indicated ribozyme cleavage active site sequence (on both substrates and ribozyme).

Aim 1a construct segments.

Table A 3 Aim 1a construct segments.

Construct	Segment	Variants	Sequence
Test 1a	RC1	HP-1	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
(T1a)	D2D1	1a	TACATGTCCTTATAGTGAAT
	CS2-1'	DT-1	AAACCAC
	A3_1	1	AA
	C_R	1	CTTGTACGGA
	CS1'_R	HP-1	GTCACCAAATAGGA
	L1	1	TTCAAGAGA
	CS1	HP-1	TGACagtcCTGTTT
	A3_2A2A1	1	AATGGATCCCGTTCGGCAGC
	L2	1	TTCAAGAGA
	B1	1	GCTGCCGAAC

Construct	Segment	Variants	Sequence
Control 1a	RC1	HP-1	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
(C1a)	D2D1	1a	TACATGTCCTTATAGTGAAT
	CS2-1'	DT-1	AAACCAC
	A3_1	1	AA
	C	1	GGGATCCATT
	CS1'	HP-1	AAACAGgactGTCA
	L1	1	TTCAAGAGA
	CS1	HP-1	TGACagtcCTGTTT
	A3_2A2A1	1	AATGGATCCCGTTCGGCAGC
	L2	1	TTCAAGAGA
	B1	1	GCTGCCGAAC
Control	RC1	HP-1	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
1a(-)	D2D1	1a	TACATGTCCTTATAGTGAAT
(C1a(-))	CS2-1'	DT-1	AAACCAC
	A3_1	1	AA
	C	1	GGGATCCATT
	CS1_D'	HP-1	AAACAGcgaaGTCA
	L1	1	TTCAAGAGA
	CS1_D	HP-1	TGACttcgCTGTTT
	A3_2A2A1	1	AATGGATCCCGTTCGGCAGC
	L2	1	TTCAAGAGA
	B1	1	GCTGCCGAAC

Construct	Segment	Variants	Sequence
HP-LR-CS	RC1	HP-1	AAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTA
	LR	HP-1	AGATCT
	CS1	HP-1	TGACagtcCTGTTT

Sequences of ligation oligos for Aim 1a and 1b constructs.

Table A 4 Sequences of ligation oligos for Aim 1a and 1b constructs.

Ligation Oligo	Sequence (5' to 3')
T/C1a Fwd L1	TAATACGACTCACTATAGGAAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTAtac
T1a Fwd L2	atgtccttatagtgaatAAACCACAACTTGTACGGAGTCA
T1a Fwd L3-A-3'	CCAAATAGGATTCAAGAGATGACagtcCTGTTTAATGGATCCCGTTCGGCAGCTTCAAGAGAGCTGCCGAACA
T1a Rvs L1	GTTCGGCAGCTCTCTTGAAGCTGCCGAACGGGATCCATTAAACAGgactGTCATCTC
T1a Rvs L2	TTGAATCCTATTTGGTGACTCCGTACAAGTTGTGGTTTattcactataaggacatgtaTACCAGGTAATA
T/C1a Rvs L3-A-3'	TACCACAACGTGTGTTTCTCTGGTTGACttctCTGTTTCCTATAGTGAGTCGTATTAA
C1a/C1a(-) Fwd L2	atgtccttatagtgaatAAACCACAAGGGATCCATTAAAC
C1a Fwd L3-A-3'	AGgactGTCATTCAAGAGATGACagtcCTGTTTAATGGATCCCGTTCGGCAGCTTCAAGAGAGCTGCCGAACA
C1a Rvs L1	GTTCGGCAGCTCTCTTGAAGCTGCCGAACGGGATCCATTAAACAGgactGTCATCTC
C1a Rvs L2	TTGAATGACagtcCTGTTTAATGGATCCCTTGTGGTTTattcactataaggacatgtaTACCAGGTAATA
C1a(-) Fwd L3-A-3'	AGcgaaGTCATTCAAGAGATGACttcgCTGTTTAATGGATCCCGTTCGGCAGCTTCAAGAGAGCTGCCGAACA
C1a(-) Rvs L1	GTTCGGCAGCTCTCTTGAAGCTGCCGAACGGGATCCATTAAACAGcgaaGTCATCTC
C1a(-) Rvs L2	TTGAATGACttcgCTGTTTAATGGATCCCTTGTGGTTTattcactataaggacatgtaTACCAGGTAATA
T1b/T1b(-) Fwd L1	TAATACGACTCACTATAGGAAACAGagaaGTCAACCAGAGAAACACACGTTGTGGTATATTACCTGGTAtacatgtc
T1b Fwd L2	cttatagtgaatAAACCACAATTCAAGAGATTGTGGTTTccaatac
T1b/T1b(-) Fwd L3-A-3'	ttacgcgtcgATTCACTATAGGGTCCACCTCCTCGCGGTaaaccatGGGCATCCGTTCGCGGATGGCTAAGGGACCCA
T1b/T1b(-) Rvs L1	GGGTCCCTTAGCCATCCGCGAACGGATGCCCatggtttACCGCGAGGAGGTGGACCCTATAG
T1b Rvs L2	TGAATcgacgcgtaagtattggAAACCACAATCTCTTGAATTGTGGTTTattcactataaggacatgtaTACCAGG
T1b/T1b(-) Rvs L3-A-3'	TAATATACCACAACGTGTGTTTCTCTGGTTGACttctCTGTTTCCTATAGTGAGTCGTATTAA

Ligation Oligo	Sequence (5' to 3')
T1b(-) Fwd L2	cttatagtgaatAAAGGAAAATTCAAGAGATTTTCCTTTccaatac
T1b(-) Rvs L2	TGAATcgacgcgtaagtattggAAAGGAAAATCTCTTGAATTTTCCTTTattcactataaggacatgtaTACCAGG
DT-LR-CS Fwd L1	TAATACGACTCACTATAGGTTCAAGAGATTGTGGTTTATGGGTCCACCTCCTCGCG
DT-LR-CS Fwd L2-A-3'	GTaaaccatGGGCATCCGTTCGCGGATGGCTAAGGGACCCA
DT-LR-CS Rvs L1	GGGTCCCTTAGCCATCCGCGAACGGATGCCCatggtttACCGCGAGGAGGTGGACC
DT-LR-CS Rvs L2-A-3'	CATAAACCACAATCTCTTGAACCTATAGTGAGTCGTATTAA
sNTS 37a Fwd-A-3'	TAATACGACTCACTATAGGATTCACTATAAGGACATGTAA
sNTS 37a Rvs-A-3'	TACATGTCCTTATAGTGAATCCTATAGTGAGTCGTATTAA
sNTS 37b Fwd-A-3'	TAATACGACTCACTATAGGATGCCTTACCCGGACATGTAA
sNTS 37b Rvs-A-3'	TACATGTCCGGGTAAGGCATCCTATAGTGAGTCGTATTAA

Valid e-SRS sensor designs for Dengue

Table A 5 Valid e-SRS sensor designs for Dengue.

Mz Name	MzL	MzR	NTS	Energy	NTS	Energy	NNS	Energy	NNS	Energy	NNS	Energy
Mz_D1_UGA,AGCA_02	UUUUCGACUGAUGAGUGACCCGGAUCACUGAGACG	CCAGCACGAAAGUUUCAUA	D1	-53.2	A	-17.9	D2	-21.9	D3	-23.6	D4	-24
Mz_D1_UGA,AGCA_04	UUUUCGACUGAUGAGUGACGGAUCACUGAGACG	CCCCAGCACGAAAGUUUCAUA	D1	-53.2	A	-17.5	D2	-21.7	D3	-23.2	D4	-23.8
Mz_D1_UGA,AGCA_05	UUUUCGACUGAUGAGUGAGGAUCACUGAGACG	CCCCCAGCACGAAAGUUUCAUA	D1	-54.1	A	-17	D2	-21.3	D3	-20.7	D4	-23.4
Mz_D1_UGA,AGCA_07	UUUUCGACUGAUGAGUGAAUCACUGAGACG	CCCCCGGAGCACGAAAGUUUCAUA	D1	-53	A	-18.1	D2	-23	D3	-20.1	D4	-24.1
Mz_D1_UGA,AGCA_08	UUUUCGACUGAUGAGUGAUCACUGAGACG	CCCCCGGAAGCACGAAAGUUUCAUA	D1	-53.5	A	-17.1	D2	-23	D3	-20.6	D4	-24.1
Mz_D1_UGA,AGCA_12	UUUUCGACUGAUGAGUGAUGAGACG	CCCCCGGAUCACAGCACGAAAGUUUCAUA	D1	-53.9	A	-20.3	D2	-21.3	D3	-24.6	D4	-23.9
Mz_D1_UGA,AGCA_13	UUUUCGACUGAUGAGUGAGAGACG	CCCCCGGAUCACUAGCACGAAAGUUUCAUA	D1	-52.9	A	-19.2	D2	-20.2	D3	-20.1	D4	-23.2
Mz_D1_UGA,AGCA_15	UUUUCGACUGAUGAGUGAGACG	CCCCCGGAUCACUGAAGCACGAAAGUUUCAUA	D1	-53	A	-22.3	D2	-23.3	D3	-20.5	D4	-25.9
Mz_D1_UGA,AGCA_16	UUUUCGACUGAUGAGUGAACG	CCCCCGGAUCACUGAGAGCACGAAAGUUUCAUA	D1	-53	A	-19.3	D2	-22.2	D3	-21.1	D4	-22.9

Mz Name	MzL	MzR	NTS	Energy	NTS	Energy	NNS	Energy	NNS	Energy	NNS	Energy
Mz_D1_UGA,AGCA_17	UUUUCGACUGAUGAGUGACG	CCCCCGGAUCACUGAGAAGCACGAAAGUUUCAUA	D1	-52.9	A	-18	D2	-23.4	D3	-20.6	D4	-21.6
Mz_D2_UGA,AGCA_04	UUUUCGACUGAUGAGUGAUCAUGGCCCUUGUGGCG	CUGUAGCACGAAAGUUUCAUA	D2	-55.6	A	-20.7	D1	-27.5	D3	-30.2	D4	-27
Mz_D2_UGA,AGCA_05	UUUUCGACUGAUGAGUGACAUGGCCCUUGUGGCG	CUGUUAGCACGAAAGUUUCAUA	D2	-54.6	A	-18	D1	-25.9	D3	-28.7	D4	-24.4
Mz_D2_UGA,AGCA_06	UUUUCGACUGAUGAGUGAAUGGCCCUUGUGGCG	CUGUUCAGCACGAAAGUUUCAUA	D2	-55.9	A	-20.6	D1	-27.6	D3	-28.5	D4	-25
Mz_D2_UGA,AGCA_07	UUUUCGACUGAUGAGUGAUGGCCCUUGUGGCG	CUGUUCAAGCACGAAAGUUUCAUA	D2	-55.5	A	-19.3	D1	-26.6	D3	-29.1	D4	-24.3
Mz_D2_UGA,AGCA_08	UUUUCGACUGAUGAGUGAGGCCCUUGUGGCG	CUGUUCAUAGCACGAAAGUUUCAUA	D2	-54.9	A	-18.6	D1	-25.9	D3	-27.5	D4	-24.3
Mz_D2_UGA,AGCA_11	UUUUCGACUGAUGAGUGACCUUGUGGCG	CUGUUCAUGGCAGCACGAAAGUUUCAUA	D2	-55.2	A	-18.9	D1	-26.1	D3	-27.2	D4	-25.7
Mz_D2_UGA,AGCA_16	UUUUCGACUGAUGAGUGAUGGCG	CUGUUCAUGGCCCUUGAGCACGAAAGUUUCAUA	D2	-55.2	A	-19	D1	-31.7	D3	-30.3	D4	-25.5
Mz_D2_UGA,AGCA_17	UUUUCGACUGAUGAGUGAGGCG	CUGUUCAUGGCCCUUGUAGCACGAAAGUUUCAUA	D2	-54.9	A	-18.1	D1	-35.6	D3	-26	D4	-23.7
Mz_D3_UGA,AGCA_02	UUUUCGACUGAUGAGUGAUCUGUCUCAUGAUGAUGUUA	GCAGCACGAAAGUUUCAUA	D3	-49.7	A	-19.1	D1	-23.9	D2	-23.4	D4	-24.9
Mz_D3_UGA,AGCA_03	UUUUCGACUGAUGAGUGACUGUCUCAUGAUGAUGUUA	GCUAGCACGAAAGUUUCAUA	D3	-48.4	A	-19.9	D1	-23.5	D2	-24.9	D4	-25.1

Mz Name	MzL	MzR	NTS	Energy	NTS	Energy	NNS	Energy	NNS	Energy	NNS	Energy
Mz_D3_UGA,AGCA_04	UUUUCGACUGAUGAGUGAUGUCUCAUGAUGAUGUUA	GCUCAGCACGAAAGUUUCAUA	D3	-49.7	A	-20.8	D1	-26.5	D2	-26.4	D4	-25.5
Mz_D3_UGA,AGCA_05	UUUUCGACUGAUGAGUGAGUCUCAUGAUGAUGUUA	GCUCUAGCACGAAAGUUUCAUA	D3	-48.7	A	-20.1	D1	-27.5	D2	-24.2	D4	-25.5
Mz_D3_UGA,AGCA_06	UUUUCGACUGAUGAGUGAUCUCAUGAUGAUGUUA	GCUCUGAGCACGAAAGUUUCAUA	D3	-49	A	-19.9	D1	-30.8	D2	-25.1	D4	-26.4
Mz_D3_UGA,AGCA_07	UUUUCGACUGAUGAGUGACUCAUGAUGAUGUUA	GCUCUGUAGCACGAAAGUUUCAUA	D3	-48.4	A	-18.8	D1	-29.5	D2	-25.3	D4	-25.3
Mz_D3_UGA,AGCA_08	UUUUCGACUGAUGAGUGAUCAUGAUGAUGUUA	GCUCUGUCAGCACGAAAGUUUCAUA	D3	-49.7	A	-21.7	D1	-30	D2	-25.7	D4	-27.4
Mz_D3_UGA,AGCA_09	UUUUCGACUGAUGAGUGACAUGAUGAUGUUA	GCUCUGUCUAGCACGAAAGUUUCAUA	D3	-48.4	A	-20.5	D1	-28.5	D2	-25	D4	-26.8
Mz_D3_UGA,AGCA_11	UUUUCGACUGAUGAGUGAUGAUGAUGUUA	GCUCUGUCUCAAGCACGAAAGUUUCAUA	D3	-49.3	A	-18.1	D1	-26.1	D2	-22.9	D4	-23.5
Mz_D3_UGA,AGCA_14	UUUUCGACUGAUGAGUGAUGAUGUUA	GCUCUGUCUCAUGAAGCACGAAAGUUUCAUA	D3	-49.3	A	-18.2	D1	-26.8	D2	-26.6	D4	-24
Mz_D3_UGA,AGCA_15	UUUUCGACUGAUGAGUGAGAUGUUA	GCUCUGUCUCAUGAUAGCACGAAAGUUUCAUA	D3	-48.7	A	-20.9	D1	-30.8	D2	-27.2	D4	-26.7
Mz_D3_UGA,AGCA_17	UUUUCGACUGAUGAGUGAUGUUA	GCUCUGUCUCAUGAUGAAGCACGAAAGUUUCAUA	D3	-49.3	A	-18.7	D1	-25.7	D2	-26.6	D4	-25.1
Mz_D3_UGA,AGCA_18	UUUUCGACUGAUGAGUGAGUUA	GCUCUGUCUCAUGAUGAUAGCACGAAAGUUUCAUA	D3	-48.7	A	-21.2	D1	-25.3	D2	-27.5	D4	-27

Mz Name	MzL	MzR	NTS	Energy	NTS	Energy	NNS	Energy	NNS	Energy	NNS	Energy
Mz_D4_UGA,AGCA_03	UUUUCGACUGAUGAGUGACUUGUUUAAGACAACAGAG	UCUAGCACGAAAGUUUCAUA	D4	-45.3	A	-16.8	D1	-22.2	D2	-20	D3	-22.7
Mz_D4_UGA,AGCA_13	UUUUCGACUGAUGAGUGAACAACAGAG	UCUCUUGUUUAAGAGCACGAAAGUUUCAUA	D4	-45.7	A	-20.5	D1	-26.6	D2	-23.9	D3	-24.6
Mz_D4_UGA,AGCA_14	UUUUCGACUGAUGAGUGACAACAGAG	UCUCUUGUUUAAGAAGCACGAAAGUUUCAUA	D4	-45.6	A	-19	D1	-25.6	D2	-23.6	D3	-24.7
Mz_D4_UGA,AGCA_16	UUUUCGACUGAUGAGUGAACAGAG	UCUCUUGUUUAAGACAAGCACGAAAGUUUCAUA	D4	-46.4	A	-20.5	D1	-24.2	D2	-22.8	D3	-25.5

Valid e-SRS sensor designs for Malaria.

Table A 6 Valid e-SRS sensor designs for Malaria.

Mz Name	MzL	MzR	NTS	Energy	NTS	Energy	NNS	Energy	NNS	Energy
Mz_Mfs_UGA,AGCA_11	UUUUCGACUGAUGAGUGAUUUAUUCAUUACA	UUAGCAAAAAUAGCACGAAAGUUUCAUA	Mfs	-38.8	A	-20	Mfr1	-36	Mfr2	-31
Mz_Mfs_UGA,AGCA_13	UUUUCGACUGAUGAGUGAUAUUCAUUACA	UUAGCAAAAAUUUAGCACGAAAGUUUCAUA	Mfs	-38.8	A	-18.6	Mfr1	-38	Mfr2	-30.8
Mz_Mfs_UGA,AGCA_12	UUUUCGACUGAUGAGUGAUUAUUCAUUACA	UUAGCAAAAAUUAGCACGAAAGUUUCAUA	Mfs	-38.8	A	-20.1	Mfr1	-38.2	Mfr2	-32.4

Mz Name	MzL	MzR	NTS	Energy	NTS	Energy	NNS	Energy	NNS	Energy
Mz_Mfs_UGA,AGCA_16	UUUUCGACUGAUGAGUGAUCAUUACA	UUAGCAAAAAUUUUAUAGCACGAAAGUUUCAUA	Mfs	-38.8	A	-17.8	Mfr1	-36.5	Mfr2	-31.4
Mz_Mfr1_UGA,AGCA_13	UUUUCGACUGAUGAGUGAUAUUCAUUACA	UUAGCAAAAAUUGAGCACGAAAGUUUCAUA	Mfr1	-40.9	A	-18.6	Mfr2	-35.3	Mfs	-38.5
Mz_Mfr2_UGA,AGCA_17	UUUUCGACUGAUGAGUGAAAUUACA	UUAGCAAAAAUUGUUUCAGCACGAAAGUUUCAUA	Mfr2	-41	A	-15.5	Mfr1	-35.7	Mfs	-31.7
Mz_Mfr2_UGA,AGCA_15	UUUUCGACUGAUGAGUGAUCAAUUACA	UUAGCAAAAAUUGUUAGCACGAAAGUUUCAUA	Mfr2	-40.7	A	-18.3	Mfr1	-36.9	Mfs	-32.7
Mz_Mfr2_UGA,AGCA_20	UUUUCGACUGAUGAGUGAUACA	UUAGCAAAAAUUGUUUCAAUAGCACGAAAGUUUCAUA	Mfr2	-40.7	A	-15	Mfr1	-35.2	Mfs	-31.2

Oligonucleotides in cloning of Mfs, Mfr1, Mfr2 NTS.

Table A 7 Oligonucleotides in cloning of Mfs, Mfr1, Mfr2 NTS.

Oligonucleotides	Role	Sequence (5' to 3')	Length (nt)
NASBA F3R1 DNA	NASBA DNA Amplicon	AATTCTAATACGACTCACTATAGGGAGAA GGTATTTTAAGTATTATTTATTTAAGTGT ATGTGTAATGAATAAAATTTTTGCTAAAA GAACTTTAAACAAAATTGGTAACTATAGT TTTGTAACATCCGAAACTCACAACTT	
DNA Amplicon in PCR.	Cloned sequence using primers: Mal Fwd 01 Mal Rvs 01	ACGTTTAGGTGGAGGTTCTTGTCTTGGTA AATGTGCTCATGTGTTTAAACTTATTTTT AAAGAGATTAAGGATAATATTTTTATTTA TATTTTAAGTATTATTTATTTAAGTGTAT GTGTAATGAATAAAATTTTTGCTAAAAGA ACTTTAAACAAAATTGGTAACTATAGTTT TGTAACATCCGAAACTCACAACTT	198
DNA Amplicon in PCR and NASBA.	Cloned sequence as well as IVT template produced in NASBA using primers: Mal Fwd 02 Mal Rvs 01	AATTCTAATACGACTCACTATAGGGACGT TTAGGTGGAGGTTCTTGTCTTGGTAAATG TGCTCATGTGTTTAAACTTATTTTTAAAG AGATTAAGGATAATATTTTTATTTATATT TTAAGTATTATTTATTTAAGTGTATGTGT AATGAATAAAATTTTTGCTAAAAGAACTT TAAACAAAATTGGTAACTATAGTTTTGTA ACATCCGAAACTCACAACTT	223
RNA Amplicon in NASBA.	NASBA using primers: Mal Fwd 02 Mal Rvs 01	GGGACGTTTAGGTGGAGGTTCTTGTCTTG GTAAATGTGCTCATGTGTTTAAACTTATT TTTAAAGAGATTAAGGATAATATTTTTAT TTATATTTTAAGTATTATTTATTTAAGTG TATGTGTAATGAATAAAATTTTTGCTAAA AGAACTTTAAACAAAATTGGTAACTATAG TTTTGTAACATCCGAAACTCACAACTT	201
Mal Fwd 02	Cloning and possible NASBA forward primers.	AATTCTAATACGACTCACTATAGGGACGT TTAGGTGGAGGTTCT	44
Mal Fwd 01	Cloning forward primer.	ACGTTTAGGTGGAGGTTCT	19
Mal Rvs 01	Cloning and possible NASBA forward primers.	AAGTTGTGAGTTTCGGATGTTAC	23

Oligonucleo tides	Role	Sequence (5' to 3')	Length (nt)
M13 (24nt) Fwd	Sequencing forward primer.	CGCCAGGGTTTTCCCAGTCACGAC	24
M13 (22nt) Rvs	Sequencing reverse primer.	TCACACAGGAAACAGCTATGAC	22

NASBA DNA and RNA Amplicons shown were those for Mfs. Those for Mfr1 and Mfr2 were of the same length, with a few different bases at the NTS region.

Oligonucleo tides	Role	Sequence (5' to 3')	Length (nt)
M13 (24nt) Fwd	Sequencing forward primer.		

Using the computational algorithm for e-SRS Mz-based sensor design.

Ensure that Perl 5.8.8 or later is installed in the computer running the algorithm. This can be done by downloading and installing ActivePerl by following the instuctions at http://www.activestate.com/activeperl/.

After proper installation of an appropriate version of Perl, copy the computer programme (written in Perl) for the algorithm, named "eSRS.pl" in the accompanying CD to any desired location on the computer with Perl installed. Alternatively, eSRS.pl can be obtained by copying the source code of eSRS.pl (provided below) onto a "Plain Text" file (with no additional formatting whatsoever) and saved as a "Plain Text" file named "eSRS.pl". Copying the eSRS.pl file from the CD is preferred for minimising the chances of errors in obtaining eSRS.pl.

Follow the instructions in Section 4.5.2.2.1 to edit the appropriate parameters in eSRS.pl and provide the required inputs. eSRS.pl is a text file that can be opened and edited with any text editor, such as Notepad or Wordpad. Word processors such as Microsoft Word, Abiword, or OpenOffice.org Writer can also be used, but take note to save the edited file as "Plain Text". Ensure that the file extension is ".pl" after saving (change it to ".pl" if it became ".txt").

After correctly saving the edited version of eSRS.pl, and ensuring that all the required preparations have been made (as described in Section 4.5.2.2.1), such as providing the appropriate NTS files in the appropriate folder, the algorithm in eSRS.pl can be executed. This can be done by double clicking the eSRS.pl file (if ActivePerl was properly installed), or the following command can be entered at the Command Promt, followed by the "Enter" key:

```
perl c:\eSRS.pl
```

Replace "c:\eSRS.pl" with the actual path information of the eSRS.pl file, which varies according to where the eSRS.pl file was saved.

When eSRS.pl is executing, a pop up window may appear and display various useful information about the execution process. Once eSRS.pl has finished executing, as indicated by the disappearance of the pop up window, or by the command prompt returning to await your next input, outputs of the algorithm can be found in the relevant folders as described in Seciton 4.5.2.2.3. Any remaining command prompt windows for eSRS.pl can be closed.

Contents of the accompanying CD.

The accompanying CD contains the following files:

1) eSRS.pl

   This is a copy of the computer programme that implements the algorithm for e-SRS

Mz-based sensor design.

2) NTS_Den.xls

    This is the NTS file used for the Dengue category, and can serve as a sample NTS file

    to test eSRS.pl.

3) NTS_Mal.xls

    This is the NTS file used for the Malaria category, and can serve as a sample NTS file

    to test eSRS.pl.

Source code of eSRS.pl.

```perl
#!/usr/local/bin/perl
# No input from <STDI> or files.


my @cat = qw(Malaria Dengue);
my @NTS_path = qw(C:\eSRS\NTS\NTS_Mal.xls C:\eSRS\NTS\NTS_Den.xls);
my $eSRS_path = "C:\\eSRS\\";
my $eSRS_seq_path = "C:\\eSRS\\Sensor_Seq\\";
my $RNAst_path = "C:\\Program Files\\RNAstructure 4.5\\";


my $RTS2 = "UAUGAAACUCUCGAAAA"; #STS   RTS-2   UAUGAAACUCUCGAAAA
my $MzL_RTSrc5 = "UUUUCGA";     #MzL   5' catalytic arm UUUUCGA
my $MzL_Loop = "CUGAUGAG";      #      catalytic core   CUGAUGAG
my $MzL_stem2 = "UGA";          #   Stem II UGA
my $MzL_NTSrc3;                 #   3' sensor arm
my $MzR_NTSrc5;                 #MzR  5' sensor arm
my $MzR_stem2 = "AGCA";         #        Stem II   AGCA
my $MzR_Loop = "CGAA";          #   catalytic core   CGAA
my $MzR_RTSrc3 = "AGUUUCAUA";   #      3' catalytic arm AGUUUCAUA
my $Loop = "gaaa";              #Linker  Unfolded loop gaaa


foreach $i (0..$#cat) {
  print     "\n***********************Designing    Mz    for
$cat[$i]...***********************\n";
  mz_design($cat[$i], $NTS_path[$i]);
  print    "\n***********************Mz   Designs   for   $cat[$i]
Done***********************\n";
}


# sub mz_design($cat, $NTS_path) to
 # $_[0]: category of NTS, e.g. Malaria, Dengue.
 # $_[1]: full path info of text file containing NTS names & seqs.
 # Function: Create & test all possible Mz combination with given STS
& Mz Stem2 seq.
 # Output: for all qualified Mz designs:
  # .seq files for Mz, NTS/NNS/Nothing, STS combined seq.
  # .ct files for above.
  # Summary files showing MzL & MzR seqs, & energy values for secondary
structures generated.
```

```perl
sub mz_design {
    my $cat = $_[0];
    my $NTS_path = $_[1];
    my $fold_seq_path = "C:\\eSRS\\Fold_seq\\$cat\\";
    my                        $fold_seq_func_path                        =
"C:\\eSRS\\Fold_seq\\Functional\\$cat\\";
    my $ct_path = "C:\\eSRS\\CT\\$cat\\";
    my $ct_func_path = "C:\\eSRS\\CT\\Functional\\$cat\\";
    system("mkdir \"$fold_seq_path\"");
    system("mkdir \"$fold_seq_func_path\"");
    system("mkdir \"$ct_path\"");
    system("mkdir \"$ct_func_path\"");
    system("mkdir \"$eSRS_seq_path\"");

    open NTS, "$NTS_path" or die "Can't read $NTS_path: $!\n";
    my @NTS;
    my $iter = 0;
    # Get list of NTS names and sequences into array, @NTS.
    while (<NTS>) {
       chomp $_;
    # Extract patterns from NTS_n\tNTS_s
       if ($_ =~ /^(.*)\t(.*)$/) {
          my $NTS_n = $1; # print $NTS_n;
          my $NTS_s = $2; # print ":\t$NTS_s\n";
    # Ensure $NTS_s is in RNA form (Convert all T to U).
          $NTS_s = rc(rc($NTS_s, R), R);
          $NTS[$iter] = [$NTS_n,$NTS_s];
          $iter++;
       }
    }
    close NTS;

    #   Possible MzL & MzR pairs of Stem II.
    #  @stem = [\@P1, \@P2, ...]
    #   @P1 = [StemII Name, StemII L, StemII R]
    #   Naming: <StemII L,StemII R (made up to 3 char per stem with "-")>
    #  e.g.: UGA,AGCA
    my @stem2;
    $stem2[0] = [$MzL_stem2.",".$MzR_stem2, $MzL_stem2, $MzR_stem2];

    #   Possible MzL & MzR pairs of sensor arm.
    #   @NTSrc = [\@NTS_1, \@NTS_2, ...]
```

```perl
    #   @NTS_1 = [NTS Name, \@P1, \@P2, ...]
    #    @P1 = [SAS, NTS' L, NTS' R]
    #   SAS: 01-99.
    my @NTSrc;
    foreach $i_NTS (0..$#NTS) {
        # Get NTS name.
        $NTSrc[$i_NTS][0] = $NTS[$i_NTS][0];
        my $NTS = $NTS[$i_NTS][1];
        my $NTS_length = length($NTS);
        my $NTSrc = rc($NTS, "R");
        foreach $i_pos (1..($NTS_length-1)) {
            $NTSrc[$i_NTS][$i_pos]   =   [pad($i_pos,   0,   2,   F),
seq_split($NTSrc, $i_pos, A)];
        }
    }


    #  Possible MzL & MzR pairs
    #  @mz = [\@NTS_1, \@NTS_2, ...]
    #   @NTS_1 = [NTS Name, \@StemII_P1, \@StemII_P2, ...]
    #    @StemII_P1 = [StemII Name, \@NTS_P1, \@NTS_P2, ...]
    #     @NTS_P1 = [Mz Name, MzL, MzR]
    #      Naming:
    #      "Mz_<NTS Name, less "sNTS">_<StemII Name>_<SAS (01-99)>"
    #      e.g.:  #Mz_Mfr1_UGA,AGCA_01
    #  MzL:
    #  5' part of STS' (UUUUCGA);MzL loop (CUGAUGAG);StemII L (UGA);3'
part of NTS'
    #  MzR:
    #  5' part of NTS';StemII R(GCA);MzR loop (CGAA);3' part of
STS'(AGUUUCAUA)
    my @Mz;
    my $i_NTS = 0;


    #  For each NTS:


    foreach $ele_NTS (@NTS) {
        my $NTS_name_l = $ele_NTS->[0];  # $NTS_name_l - Long form of
NTS name.
        $Mz[$i_NTS][0] = $NTS_name_l;
        $NTS_name_l =~ /^sNTS-(.*)$/;
        $NTS_name_s = $1;                    # $NTS_name_s - Short form
of NTS name (without "sNTS").
```

```perl
    my $i_stem = 1;

     #  For each StemII:

    foreach $ele_stem (@stem2) {
       my $stem2_name = $ele_stem->[0];
       $stem2L = $ele_stem->[1];
       $stem2R = $ele_stem->[2];
       $Mz[$i_NTS][$i_stem][0] = $stem2_name;

          #  For each SAS of current NTS (read from @NTSrc):

       foreach $ele_NTSrc_NTS (@NTSrc) {
          if ($ele_NTSrc_NTS->[0] eq $NTS_name_l) {
             my $i_NTSrc_pos = 1;
             foreach                         $ele_NTSrc_pos
(@{$ele_NTSrc_NTS}[1..$#{$ele_NTSrc_NTS}]) {
                my $NTSrc_pos = $ele_NTSrc_pos->[0];
                my $NTSrc5 = $ele_NTSrc_pos->[1];
                my $NTSrc3 = $ele_NTSrc_pos->[2];

                   #  Create Mz Name, MzL, MzR.
                $Mz[$i_NTS][$i_stem][$i_NTSrc_pos][0]          =
"Mz_".$NTS_name_s."_".$stem2_name."_"."$NTSrc_pos";
                $Mz[$i_NTS][$i_stem][$i_NTSrc_pos][1]          =
$MzL_RTSrc5.$MzL_Loop.$stem2L.$NTSrc3;
                $Mz[$i_NTS][$i_stem][$i_NTSrc_pos][2]          =
$NTSrc5.$stem2R.$MzR_Loop.$MzR_RTSrc3;
                $i_NTSrc_pos ++;
             }
          }
       }
       $i_stem ++;
    }
    $i_NTS ++;
 }

 print "\@Mz is:\n";
 print_a(\@Mz, 0);

 # Get functional Mz.
 my @Mz_func;
```

```perl
    push @Mz_func, [qw (Category MzL_Name MzL MzR_Name MzR NTS
Energy)];
    foreach $NTS (@NTS) {
      push @{$Mz_func[-1]}, ("NTS", "Energy");
    }


    foreach $i_NTS (0..$#NTS) {
       # Create list of NNS.
      my @NNS;
      if ($i_NTS == 0) {
        @NNS = @NTS[1..$#NTS];
      } elsif ($i_NTS != $#NTS) {
        @NNS = @NTS[0..($i_NTS-1),($i_NTS+1)..$#NTS];
      } elsif ($i_NTS == $#NTS) {
        @NNS = @NTS[0..($i_NTS-1)];
      }
      my $NTS_n_l = $NTS[$i_NTS]->[0];
      my $NTS = $NTS[$i_NTS]->[1];
      $NTS_n_l =~ /^sNTS-(.*)$/;
      my $NTS_n_s = $1;
      foreach $ele_mz_0 (@Mz) {                    #   $ele_mz_0   =
\@NTS_1
          # Only proceed to use Mz designs for current NTS.
        if ($ele_mz_0->[0] eq $NTS_n_l) { print "\n\nLT01";
          my @ele_mz_2 = @{$ele_mz_0->[1]}; #    @ele_mz_2    =
@StemII_P1
          shift @ele_mz_2;                         #          Remove
$StemII_P1[0], i.e. StemII Name.
          foreach $ele_mz_2 (@ele_mz_2) {      #   $ele_mz_2   =
\@NTS_P1...\@NTS_Pn.
            my $Mz_n = $ele_mz_2->[0];         # $ele_mz_2->[0] =
Mz Name
            $Mz_n =~ /^(Mz)(.*)$/;
            my $MzL_n = $1."L".$2;
            my $MzR_n = $1."R".$2;
            my $MzL = $ele_mz_2->[1];        # $ele_mz_2->[1] = MzL
            my $MzR = $ele_mz_2->[2];        # $ele_mz_2->[2] = MzR
            my $energy_T = "-"; my $energy_A = "-"; my @energy_N;
            my $active_T = 1; my $active_A = 0; my $active_N = 0;

                # Test fold_seq_T.
            my $fold_seq_T = assemble(T, $MzL, $MzR, $NTS);
```

```perl
                  print              "\nMzL:              $MzL_n:$MzL,\tMzR:
$MzR_n:$MzR,\tfold_seq_T:$fold_seq_T\n";
                    #Naming:  <Mz   Name>_<T/N/A>_<NTS  name  ,  less
"sNTS">.<seq/ct>. E.g.: Mz_Mfr1_UGA,AGCA_01_T_Mfr1.seq
          my $fold_seq_T_n = $Mz_n."_T_".$NTS_n_s.'.seq';
          my $fold_ct_T_n = $Mz_n."_T_".$NTS_n_s.'.ct';
          write_seq($fold_seq_path.$fold_seq_T_n, $fold_seq_T);
          RNAstructure($fold_seq_path.$fold_seq_T_n,
$ct_path.$fold_ct_T_n, "-w 3");
          my @line_s = @{(stable($ct_path.$fold_ct_T_n))[0]};
          my $len = (stable($ct_path.$fold_ct_T_n))[2];
          my @ss_active = ss_active($len);                    #
Determine active ss template based on total length.
                # If any of the most stable ss is inactive, Mz design
is considered invalid.
          foreach $line_s (@line_s) {
              if     (active($ct_path.$fold_ct_T_n,      $line_s,
\@ss_active) == 0) {
                  $active_T = 0;
                  $energy_T = "!";
                  foreach $i_energy_N (@energy_N[0..$#NNS]) {
                    @{$i_energy_N} = ("-", "-");
                  }
                  last;
              }
          }
          if ($active_T == 1) {
              $energy_T = (stable($ct_path.$fold_ct_T_n))[1];

                  # Test fold_seq_A.
              my $fold_seq_A = assemble(A, $MzL, $MzR, $NTS);
                  #Naming:  <Mz   Name>_<T/N/A>_<NTS  name  ,  less
"sNTS">.<seq/ct>. E.g.: Mz_Mfr1_UGA,AGCA_01_A_Mfr1.seq
              my $fold_seq_A_n = $Mz_n."_A_".$NTS_n_s.'.seq';
              my $fold_ct_A_n = $Mz_n."_A_".$NTS_n_s.'.ct';
              write_seq($fold_seq_path.$fold_seq_A_n,
$fold_seq_A);
              RNAstructure($fold_seq_path.$fold_seq_A_n,
$ct_path.$fold_ct_A_n, "-w 3");
              my @line_s = @{(stable($ct_path.$fold_ct_A_n))[0]};
              my $len = (stable($ct_path.$fold_ct_A_n))[2];
              print "$ct_path.$fold_ct_A_n \$len is $len.\n";
```

```perl
                my @ss_active = ss_active($len);
                    # If any of the most stable ss is active, Mz design
is considered invalid.
                foreach $line_s (@line_s) {
                    if    (active($ct_path.$fold_ct_A_n,    $line_s,
\@ss_active) == 1) {
                        $active_A = 1;
                        $energy_A = "!";
                        foreach $i_energy_N (@energy_N[0..$#NNS]) {
                            @{$i_energy_N} = ("-", "-");
                        }
                        last;
                    }
                }
                if ($active_A == 0) {
                    $energy_A = (stable($ct_path.$fold_ct_A_n))[1];

                        # Test fold_seq_N.
                        my $NNS_check = 0;
                    foreach $i_NNS (0..$#NNS) {
                            if ($NNS_check == 1) {
                            foreach                        $i_energy_N
(@energy_N[$i_NNS..$#NNS]) {
                                @{$i_energy_N} = ("-", "-");
                            }
                                last;
                            }
                        my $NNS_n_l = $NNS[$i_NNS]->[0];
                        my $NNS = $NNS[$i_NNS]->[1];
                        $NNS_n_l =~ /^sNTS-(.*)$/;
                        my $NNS_n_s = $1;
                            print "\$NNS_n_l = $NNS_n_l, \$NNS_n_s =
$NNS_n_s\n";
                        my $fold_seq_N = assemble(N, $MzL, $MzR, $NNS);
                            #Naming: <Mz Name>_<T/N/A>_<NTS name ,
less "sNTS">.<seq/ct>.  E.g.: Mz_Mfr1_UGA,AGCA_01_N_Mfr2.seq
                        my $fold_seq_N_n = $Mz_n."_N_".$NNS_n_s.'.seq';
                        my $fold_ct_N_n = $Mz_n."_N_".$NNS_n_s.'.ct';
                        write_seq($fold_seq_path.$fold_seq_N_n,
$fold_seq_N);
                        RNAstructure($fold_seq_path.$fold_seq_N_n,
$ct_path.$fold_ct_N_n, "-w 3");
```

```perl
                    my                    @line_s                    =
@{(stable($ct_path.$fold_ct_N_n))[0]};
                    my $len = (stable($ct_path.$fold_ct_N_n))[2];
                    my @ss_active = ss_active($len);
                            # If any of the most stable ss is active,
Mz design is considered invalid.
                    foreach $line_s (@line_s) {
                        $energy_N[$i_NNS][0] = $NNS_n_s;
                        if (active($ct_path.$fold_ct_N_n, $line_s,
\@ss_active) == 1) {
                            $active_N = 1; $NNS_check = 1;
                            $energy_N[$i_NNS][1] = "!";
                            last;
                        }
                        $energy_N[$i_NNS][1]                     =
(stable($ct_path.$fold_ct_N_n))[1];
                    }
                }
            }
        }
            # If Mz is suitable, move all seq, ct files into
"Functional" folder.
        if (($active_T == 1) and ($active_A == 0) and ($active_N
== 0)) {
            opendir DH, $fold_seq_path or die "Couldn't open the
Directory $fold_seq_path\n$!";
            while ($_ = readdir(DH)) {
                    # Extract patterns from:
                    #<Mz   Name>_<T/N/A>_<NTS   name   ,   less
"sNTS">.<seq/ct>
                    #e.g.: Mz_Mfr1_UGA,AGCA_01_T_Mfr1.seq
                if ($_ =~ /^$Mz_n.*$/) {
                    my $file_n = $_;
                        print "\nMoving .seq files\n";
                        print              "move              /Y
\"$fold_seq_path$file_n\" \"$fold_seq_func_path$file_n\"";
                    system("move    /Y    \"$fold_seq_path$file_n\"
\"$fold_seq_func_path$file_n\"");
                }
            }
            opendir DH, $ct_path or die "Couldn't open the
Directory $ct_path\n$!";
```

```perl
                while ($_ = readdir(DH)) {
                        # Extract patterns from:
                        #<Mz   Name>_<T/N/A>_<NTS   name   ,   less
"sNTS">.<seq/ct>
                        #e.g.: Mz_Mfr1_UGA,AGCA_01_T_Mfr1.ct
                if ($_ =~ /^$Mz_n.*$/) {
                    my $file_n = $_;
                            print "\nMoving .ct files\n";
                            print  "move   /Y   \"$ct_path$file_n\"
\"$ct_func_path$file_n\"";
                    system("move       /Y       \"$ct_path$file_n\"
\"$ct_func_path$file_n\"");
                }
            }
        }
                # Write @Mz_func;
                # (Category, MzL Name, MzL, MzR Name, MzR, Energies:
fold_seq_t, fold_seq_a, fold_seq_n1, fold_seq_n2, fold_seq_n3)
            push @Mz_func, [$cat, $MzL_n, $MzL, $MzR_n, $MzR,
$NTS_n_s, $energy_T, "A", $energy_A];
            foreach $i (@energy_N) {
                push @{$Mz_func[-1]}, @{$i};
            }
          }
        }
      }
    }
    # print @Mz_func.
    my $Sensor_seq = "Sensor_Seq_$cat\.txt";
    _2d_print(\@Mz_func, "\t", ">$eSRS_seq_path$Sensor_seq");
}


# sub _2d_print(\@chk, "\t", ">$chk_file") to
 # $_[0]: 2d array; $_[1]: delimiter Enclose with care!
  # '' only process \\ and \' All else are taken literally!!
  # E.g. '\t' is a string, equivalent with '\\t' and "\\t".
  # "\t" is a Tab char, and a Tab is still a Tab when using \Q.
 # $_[2]: Output file to print to,
  # include > to overwrite, or >> to append.
  # print to stdout if file not given;
 # $_[3]: n - no delimiter for last column,
```

```perl
   # by default the there will be delimiter for all coluNNS;
   # if printing to stdout, use "" for $_[2].
 # no output. Print 2d array separated by delimiter.
sub _2d_print {
   my @array_in = @{$_[0]};
   my $dlm = $_[1];
   my $file_out = $_[2];
   my $last_col = 1;
   $last_col = 0 if ($_[3] eq 'n');
   if ($file_out) {
      open FILE, "$file_out";
      if ($last_col == 1) {
         foreach $i_01 (@array_in) {
            print FILE $i_01 if !(@{$i_01});   # print entry that is not
an array reference.
            foreach $i_02 (@{$i_01}) {
               print FILE "$i_02$dlm";
            }
            print FILE "\n";
         }
      } elsif ($last_col == 0) {
         foreach $i_01 (@array_in) {
            print FILE $i_01 if !(@{$i_01});   # print entry that is not
an array reference.
            foreach $i_02 (@{$i_01}) {
               if ($i_02 eq $i_01->[-1]) {
                  print FILE "$i_02";
               } else {
                  print FILE "$i_02$dlm";
               }
            }
            print FILE "\n";
         }
      }
      close FILE;
   } else {
      if ($last_col == 1) {
         foreach $i_01 (@array_in) {
            print $i_01 if !(@{$i_01});   # print entry that is not an
array reference.
            foreach $i_02 (@{$i_01}) {
               print "$i_02$dlm";
```

```perl
                }
            print "\n";
        }
    } elsif ($last_col == 0) {
        foreach $i_01 (@array_in) {
            print $i_01 if !(@{$i_01});   # print entry that is not an
array reference.
            foreach $i_02 (@{$i_01}) {
                if ($i_02 eq $i_01->[-1]) {
                    print "$i_02";
                } else {
                    print "$i_02$dlm";
                }
            }
        }
        print "\n";
        }
    }
  }
}


# sub assemble(T/N/A, "$MzL", "$MzR", "$NTS") to
 # $_[0]: T, N, A to indicate type of NTS. A: "Alone", i.e. no NTS,
 # $_[1]: MzL seq.
 # $_[2]: MzR seq.
 # $_[3]: NTS seq.
   # Includes both NTS & NNS seqs.
 # Output: Assembled fold_seq.
sub assemble {
  my $type = $_[0];
  my $MzL = $_[1];
  my $MzR = $_[2];
  my $NTS = $_[3];
print "Starting assemble() for \$type = $type...\t";
  my $fold_seq;
  if ($type eq A) {
#fold_seq_a
#RTS, Loop, MzL, Loop, MzR.
     $fold_seq = $RTS2.$Loop.$MzL.$Loop.$MzR;
  } elsif (($type eq T)|($type eq N)) {
#fold_seq_t/n
#RTS, Loop, MzL, Loop, NNS, Loop, MzR.
     $fold_seq = $RTS2.$Loop.$MzL.$Loop.$NTS.$Loop.$MzR;
```

```
   }
print "Finishing assemble()\n";
   return $fold_seq;
}


#          sub          RNAstructure($fold_seq_path.$fold_seq_N_n,
$ct_path.$fold_ct_N_n, "-w 3") to
 # $_[0]: $fold_seq .seq file with full path info.
 # $_[1]: $fold_seq .ct file with full path info.
 # $_[2]: Parameters for RNAstructure 4.5.
 # Function: Create CT for given fold_seq.
   # Note that RNAstructure command will only work when given in working
Directory that has RNAstructure installed. A simple way to achieve this
is to use chdir EXPR to change working directory to EXPR. If EXPR is
omitted, changes to home directory. Returns T upon success, F otherwise.
 # Output: .ct text file.
sub RNAstructure {
print "Starting RNAstructure()...\n";
   my $fold_seq = $_[0];
   my $fold_seq_ct = $_[1];
   my $para = $_[2];
   my $sys = "RNAstructure /fold -s $fold_seq -c $fold_seq_ct $para";

   chdir $RNAst_path;
   print "system(\"$sys\");", "\n";
   system("$sys");
print "Finished RNAstructure()\n";

#-d - to use DNA folding parameters.
#-n xxxx to indicate the maximum number of structures (20 is assumed).
#-w xxxx to indicate the window size (5 is assumed, 3 is suggested ).
#-p xxxx to indicate the maximum percent difference (20% is assumed).
Specify percent as integer.
}


# sub write_seq($fold_seq_name, $fold_seq) to
 # $_[0]: $fold_seq Name with full path info.
 # $_[1]: fold_seq.
 # Function: write fold_seq as text file in RNAstructure 4.5 format
for .seq file.
 # Output: .seq text file.
sub write_seq {
```

```perl
   my $fold_seq_n = $_[0];
   my $fold_seq = $_[1];
print "Starting write_seq() for $fold_seq_n...\t";
# delete old files of the same name, if exist.
   system("del \"$fold_seq_n\"");
   open SEQ, ">$fold_seq_n" or die "$fold_seq_n: $!\n";
# Write to .seq file:
# ";\n" # ";" to indicate start of file.
   print SEQ ";\n";
# "fold_seq_name\n" # Title.
   print SEQ "$fold_seq_n\n";
# Sequence "\n"
   print SEQ "$fold_seq\n";
# "1" # Sequence must end with "1".
   print SEQ "1";
   close SEQ;
print "Finished write_seq()\n";
}


# sub stable($ct_path.$fold_ct_T_n) to
 # $_[0]: $fold_seq .ct file with full path info.
 # Function: Check for number of most stable structures, return starting
line of each, the energy, and fold_seq length ($len).
 # Output: (Array ref of starting lines; Lowest energy, $len).
sub stable {
   my $ct = $_[0];
    print "Starting stable() for $ct...\t";
   my @line_s; my @energy; my $line_c = 0; my $len;
   open CT, "$ct" or die "Can't read $ct: $!\n";

   while (<CT>) {
      $line_c ++;
      chomp;
        # Extract patterns from CT for:
        #                      111          ENERGY        =        -43.4
C:\eSRS\Fold_seq\Mz_Mfr1_UGA,AGCA_01_T_Mfr1.seq
      if ($_ =~ /^\s*(\d*)\s*ENERGY = (\S*)\s*.*$/) {
         $len = $1;   #print "\$1 is $1\n";
         push @energy, $2; #print "01: \$len: $len\tEnergy: $2\n";
         push @line_s, $line_c;
         if (scalar @energy > 1) {
               #   Check if the last entry has more positive energy than
```

previous entries, if so, reject it and end searh.

```perl
        if ($energy[-1] > $energy[-2]) {
            pop @line_s; pop @energy;
                print "Finishing stable()\n";
            close CT;
            return (\@line_s, $energy[0], $len);
        }
      }
    }
  }
   print "Finishing stable()\n";
  close CT;
  return(\@line_s, $energy[0], $len); # This line will only be reached
if 1) There was < 2 "ENERGY = " lines; or 2) All the energies have the
same value.
}


# sub ss_active($len) to
 # $_[0]: length of fold_seq.
 # Function: Determine active ss of fold_seq of such length.
 # Output: @ss_active.
sub ss_active {
  $len = $_[0];
  my @ss_active;
  $ss_active[0] = [8,$len-7];
  $ss_active[1] = [9,$len-8];
  $ss_active[2] = [10,0];
  $ss_active[3] = [11,28];
  $ss_active[4] = [29,0];
  $ss_active[5] = [30,0];
  $ss_active[6] = [31,0];
  $ss_active[7] = [32,0];
  $ss_active[8] = [33,0];
  $ss_active[9] = [34,0];
  $ss_active[10] = [35,0];
  $ss_active[11] = [36,$len-12];
  $ss_active[12] = [$len-11,0];
  $ss_active[13] = [$len-10,0];
  $ss_active[14] = [$len-9,0];
  return @ss_active;
}
```

```perl
# sub active($ct_path.$fold_ct_T_n, $line_s, \@ss_active) to
 # $_[0]: $fold_seq .ct file with full path info.
 # $_[1]: Starting line to check.
 # $_[2]: @ss_active.
 # Function: Check starting from $line_s against @ss_active for exact
match for active structure.
 # Output: 1 for active structure, 0 for inactive structure.
sub active {
   my $ct = $_[0];
   my $line_s = $_[1]; my $line_c = 0; my $pos_s = 0; my $pos = 0; my
$iter_ss = 0;
   my $n; my $pp; my $active = 0;
   my @ss_check = @{$_[2]};
print "Starting active() for $ct...\n";
   open CT, "$ct" or die "Can't read $ct: $!\n";
#                   111                ENERGY          =          -43.4
C:\eSRS\Fold_seq\Mz_Mfr1_UGA,AGCA_01_T_Mfr1.seq
#    1 U      0    2 111    1
# CT file: 2nd line onwards, each line has info about a given base,
from left to right:
# Base number, n;# Base.;# n-1.;# n+1.;# Number of the base to which
n is paired. No pairing is indicated by 0 (zero).;# Natural numbering.
RNAstructure ignores the actual value given in natural numbering, so
it is easiest to repeat n here.
   while (<CT>) {
      chomp;
      if (++$line_c == $line_s) {
         $pos_s = 1;    #print "\$pos_s = $pos_s\n'";
         next;
      }
      if ($pos_s == 1) {
         $pos ++;
# Extract patterns from CT for:
#    1 U      0    2 111    1
         $_ =~ /^\s*(\d+)\s*\w\s*\d+\s*\d+\s*(\d*)\s*\d*$/;
         $n = $1; $pp = $2;
         my @ss_check_c = @{$ss_check[$iter_ss]};
         if ($n != $pos) {
            print "Error: \$pos = $pos, \$n = $n\n";
            last;
         } elsif ($pos == $ss_check[$iter_ss]->[0]) {
             if ($pp == $ss_check[$iter_ss]->[1]) {
```

```perl
                print "At pos $pos, pairing partner is #$pp, same as
expected ($ss_check[$iter_ss]->[1])\.\n";
                $iter_ss ++;
                    # If we've checked the last of the 15 bases (as
active), end function and return structure as active.
                if ($iter_ss == scalar(@ss_check)) {
                    $active = 1;
                        print  "Finishing   active(),   \$active   =
$active.\n";
                    close CT;
                    return $active;
                }
                next;
            } else {
                $active = 0;
                print "Not active: At pos $pos, pairing partner is #$pp
instead of $ss_check[$iter_ss]->[1].\n";
                    print "Finishing active(), \$active = $active.\n";
                close CT;
                return $active;
            }
        }
    }
  }
# if file ends before reaching 1st element of @ss_check:
  if ($iter_ss != scalar(@ss_check)) {
     print "CT file is finished, but we've yet to check element $iter_ss
of \@ss_check!\n";
       print "Finishing active(), \$active = $active.\n";
  }
}


# sub com($seq, "D") to
 # $_[0]: DNA/RNA seq to complement; $_[1] D, R;
 # Function: Complement DNA/RNA seq.
 # if D given, output will have T instead of U.
 # if R given, output will have U instead of T.
 # Output: Complemented DNA/RNA seq.
sub com {
  my $seq = $_[0];
  my $seq_com = $seq;
  if ($_[1] eq "D") {
```

```perl
      $seq_com =~ tr/ATUGC/TAACG/;
   } elsif ($_[1] eq "R") {
      $seq_com =~ tr/AUTGC/UAACG/;
   }
   return $seq_com;
}


# sub pad($text, 0, 2, F) to
 # $_[0]: text to pad; $_[1]: Char used to pad; $_[2]: Number of char
after padding;
 # $_[3]: "F" to pad in front, "B" to pad at back;
 # Function: Pads text with char till given length.
 # Output: Padded text.
sub pad {
   my ($text, $char, $length, $FB) = @_;
   my $extra = $length-length($text); my $text_pad = $text;
   if ($extra > 0) {
      if ($FB eq "F") {
         $text_pad = ($char x ($extra)).$text;
      } elsif ($FB eq "B") {
         $text_pad = $text.($char x ($extra));
      }
   }
   return $text_pad;
}


# sub print_a($array, 0) to
 # $_[0]: Reference of array to print; $_[1]: 0 - Level of indent for
1st level of array;
 # print all elements of array.
 # All lower dimension (i.e., referenced) array to start on new line,
with 1 more indent.
 # Outputs reverse complemented DNA/RNA seq.
sub print_a {
   my $array = $_[0];
   my @array = @{$array};
   my $indent = $_[1];
   my $indent_c = 0;
   foreach $ele (@array) {
      if (ref($ele)) {
         $indent ++;
         print_a ($ele, $indent);
```

```perl
        $indent--;
        $indent_c = 0;
      } else {
        if ($indent_c > 0) {
          print "$ele\t";
        } else {
          print "\n","\t"x($indent), "L$indent: ", "$ele\t";
          $indent_c ++;
        }
      }
    }
  }
}
#my @array_1 = [1,2,3]; my @array_2 = (1,2,3,["4a", "4b", ["4c"]],5);
#print_a(\@array_2, 0);


# sub rc($Seq, D) to
 # $_[0]: DNA/RNA seq to reverse complement; $_[1] D, R;
 # Function: Reverse complement DNA/RNA seq.
 # Output: Reverse complemented DNA/RNA seq.
sub rc {
  my $seq = $_[0];
  my $type = $_[1];
  return rev(com($seq, $type));
}
#print "$RTS2\n";
#print rc(rc($RTS2, R), R), "\n";


# sub rev($text) to
 # $_[0]: text seq to reverse;
 # Function: Reverse text seq.
 # Output: Reversed text seq.
sub rev {
  my $seq = $_[0];
  my @seq_f; my @seq_r; my $seq_r;
# Extract individual letters from text seq, and put each in an array.
  while ($seq) {
    if ($seq =~ /^(\w)(.*)$/) {
      push @seq_f, $1; $seq = $2;
    }
  }
  @seq_r = reverse(@seq_f);
  foreach $i (@seq_r) {
```

```perl
      $seq_r = $seq_r . $i;
   }
   return $seq_r;
}


# sub seq_split($NTSrc, $i_pos, "5") to
 # $_[0]: text to split into two; $_[1]: Text position after which to
split;
 # $_[2]: 5 to return 1st half of split text, 3 to return 2nd half of
split text, "A" to return both fragments as a list;
 # Function: Splits text into two fragments.
 # Output: One or both fragments.
sub seq_split {
   my ($seq, $i_pos, $frag) = @_;
   my @frag;
# Extract split $seq, and put each in @frag.
   if ($seq =~ /^(\w{$i_pos})(.*)$/) {
      @frag = ($1, $2);
   }
   if ($frag eq "5") {
      return $frag[0];
   } elsif ($frag eq "3") {
      return $frag[1];
   } elsif ($frag eq "A") {
      return @frag;
   }
}
```