Automated Design and Empirical Validation of Hammerhead Ribozymes

Nawwaf Kharma

A Thesis in the Department of Biology

Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science (Biology) at Concordia University Montreal, Quebec, Canada

November, 2015

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Nawwaf Kharma

Entitled: Automated Design and Empirical Validation of Hammerhead Ribozymes

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
Dr. Selvadurai Dayanandan

_____ External Examiner
Dr. David Walsh

_____ Examiner
Dr. Reginald Storms

_____ Examiner
Dr. Vincent Martin

_____ Supervisor
Dr. Luc Varin

Approved by _____
Dr. Selvadurai Dayanandan, Graduate Program Director

January 11, 2016 _____
Dr. André Roy, Dean of Faculty

ABSTRACT

Automated Design and Empirical Validation of Hammerhead Ribozymes

Nawwaf Kharma

Ribozymes are catalytic RNA molecules. Hammerhead ribozymes are one of a set of ribozymes capable of cleaving RNA molecules, *in cis* and *in trans*, without the help of other molecules, such as proteins. A trans-acting hammerhead ribozyme has a few related forms, which can be customized to target specific sites within RNA strands, including the transcripts of genes. As such, hammerhead ribozymes can be used to down-regulate or even silence any gene with valid cut-sites. All, except very short transcripts, will have multiple valid cut-sites. However, the efficiency of silencing by a hammerhead ribozyme depends on multiple often conflicting factors. Also, the efficiency of cleavage of any one ribozyme is normally low. Hence, it is useful to automate the process of design of hammerhead ribozymes to efficiently and without error explore the large space of possible designs. Computers are simply better than humans in doing a large amount of repetitive work without error.

This thesis describes an original computational algorithm that allows for the automated design of hammerhead ribozymes; the algorithm was implemented as a web service, by our collaborators, and is now freely available via the Internet. The algorithm takes into account all the relevant mathematically-modeled factors, influencing cleavage efficiency & target specificity, including cut-site availability, cut-site accessibility, ribozyme secondary structure, annealing temperature and off-target effects. Given an input sequence or gene, the algorithm proposes a list of potential hammerhead ribozymes that can, in theory, cleave the given RNA sequence. The latter part of this thesis describes the wet lab work, which involved the *in vitro* and hence, *in vivo* testing of several ribozymes targeting the transcript of the PABPN1 gene. This gene is the primary cause of a human disease, OPMD (Oculopharyngeal muscular dystrophy).

We describe the experimental methods used to test the cleavage efficacy and reaction kinetics of a small number of hammerhead ribozymes generated by the computational algorithm. We measured the *in vitro* transcript cleavage efficiency and enzyme kinetics, as well the *in vivo* gene knockdown effect of individual ribozymes. We also measured the enhanced effect of using combinations of two or more ribozymes all targeting the same transcript. Finally, we design a mutant PABPN1 gene that gives the same protein as the wild type gene, but one that generates a transcript supposedly immune to the catalytic activity of hammerhead ribozymes. We test this hypothesis, *in vivo*, and provide the results.

The results show that (a) every one of the ribozymes generated by the computational algorithm is a functional ribozyme; (b) the use of two or more ribozymes increases the overall cleavage efficiency; (c) these ribozymes are effective both *in vitro* and *in vivo*; (d) the immune transcript is not affected by any ribozyme. Hence, the computational algorithm is an effective design tool. The use of at least two hammerhead ribozymes is clearly more effective than one. And finally, it is possible to generate a hammerhead ribozyme (or more) that would cleave the transcript of a given gene (say PABPN1), while not cleaving the transcript of a mutated version of the gene, one that codes for the same protein as the original gene. This opens the door to potential therapeutic applications of our approach, perhaps where other gene editing approaches cannot be employed.

To Aida & our children, blessed with big hearts and infinite patience;

To my parents, who gave me life, love and the strength to be myself;

To Jonathan, whose energy & intelligence are truly infectious;

To Luc, without whom none of this would have been possible.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

hhRz	Hammerhead Ribozyme
TSM	Tertiary Stabilizing Motif
RNAi	RNA Interference
CTE	Constitutive Transport Element
E. coli	*Escherichia coli*
OPMD	Oculopharyngeal Muscular Dystrophy
PABPN1	Polyadenylate-Binding Protein Nuclear 1

Chapter 1: Introduction

RNA is the acronym for ribonucleic acid. RNA, DNA and proteins are believed to be the three macromolecules that are essential for all known forms of life. The central dogma of modern molecular biology states that the flow of genetic information goes from DNA through RNA to proteins (with each molecule acting as a template for the generation of the next one). The chemical structure of RNA is quite similar to that of DNA, with significant differences: (a) RNA contains the sugar ribose while DNA contains the slightly different sugar deoxyribose, (b) RNA has the nucleobase uracil while DNA contains thymine, and (c) within living cells, RNA usually exists as a single strand, while DNA usually exists in double stranded form. Since RNA usually exists as single strands, it can and does fold back on itself, forming complex spatial structures that may imbue it with functional capabilities, such as cleavage and splicing of RNA strands.

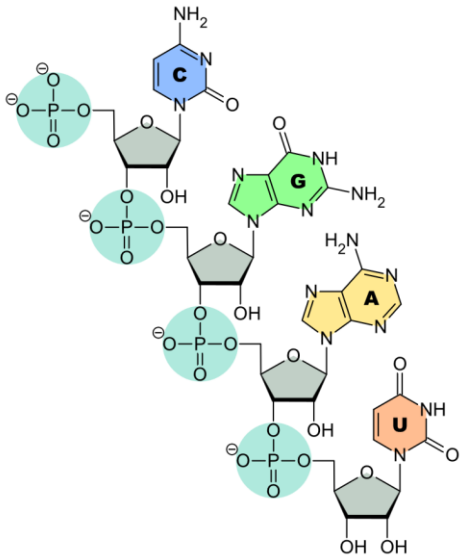


Figure 1. The Chemical Structure of a Strand of RNA [1]

RIBOZYMES

Ribozymes are RNA enzymes: RNA molecules with catalytic capabilities. The original discovery of ribozymes was twofold: RNA segments that cut themselves out of larger RNAs (self-splicing

introns) and a protein-assisted RNA enzyme (ribonuclease P) that cuts the leader sequences off all transfer RNAs [2]. Currently, ribozymes are divided into two groups: (a) self- and trans-cleaving ribozymes; and (b) splicing ribozymes, which are involved in the excision of introns from precursor RNAs. Hammerhead ribozymes are cleaving ribozymes and hence, we will present a brief review of such molecules.

Cleaving ribozymes include: the hammerhead (HH), hairpin (HP), hepatitis δ virus (HDV) and Varkud satellite (VS) ribozymes, which are predominantly found in satellite RNAs of plant origin; Ribozymes that reside within eukaryotic pre-mRNAs (*CPEB3*) and co-transcriptional cleavage (CoTC) ribozymes; a bacterial mRNA (*glmS*); RNase P, which is the only natural ribozyme that performs a multiple-turnover RNA cleavage reaction *in trans.*

Table 1. Naturally Occurring Cleaving Ribozymes [3]

Ribozyme Class	Function	Ligand	Size	Organisms
Hammerhead	Processing of multi-meric transcripts during rolling-circle replication	none	65	Plant viral satellite RNA, eukaryotes
Varkud Satellite		none	155	Satellite RNA of Neurospora spp. mitochondria
Hairpin		none	75	Plant viral satellite RNA
HDV		none	85	Human satellite virus
CoTC	Transcription termination	GTP	190	Primates
CPEB3	Splicing regulation	none	70	Mammals
glmS	Gene control	GlcN6P	170	Gram+ bacteria
RNase P	tRNA processing	none	140-500	Prokaryotes, eukaryotes

To our knowledge, the only ribozymes that have been modified to cleave *in trans* and without need of non-RNA co-enzymes are: the hhRz [4], HP [5], HDV [6] and the recently discovered Twister

ribozymes [7]. The smallest, easiest to adapt, and most widely used of these trans-acting ribozymes is the hammered.

HAMMERHEAD RIBOZYMES

Natural hammerhead ribozymes are all self-cleaving ribozymes. They have well-characterized structure and functionality. Structurally, the hammerhead ribozyme's full-length form includes tertiary structures (e.g., loops and bumps) far from the actual catalytic core. In contrast, the smallest (or minimal) hhRz capable of catalytic activity excludes all distal structures. The cleavage reaction is a phosphodiester isomerization reaction, where nucleophilic attack of the 20-OH on the adjacent phosphate causes backbone cleavage, leaving 2',3'-cyclic phosphate and 5'-OH termini. A minimal hammerhead ribozyme has a typical K_m value of 10 mm, and turnover rates of about 1 molecule/minute. A full-length hammerhead ribozyme has a similar K_m but may be ~1,000-fold faster.

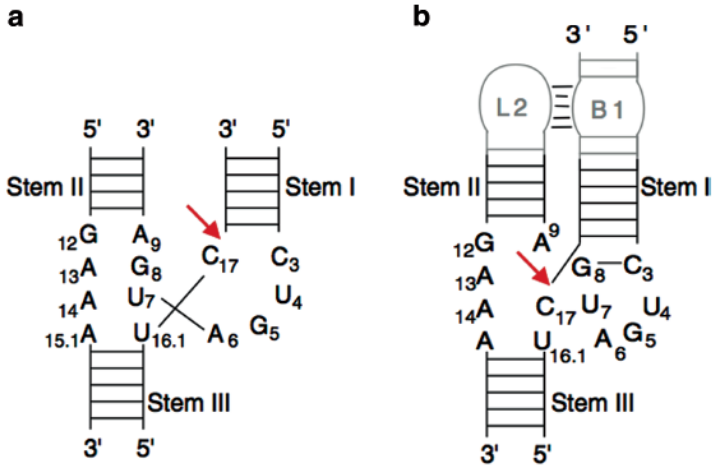


Figure 2. A Schematic Representation of the Secondary Structure of (a) the minimal and (b) the full-length hammerhead ribozyme. The conserved residues in the catalytic core are shown explicitly in each case, and cleavage sites are indicated with arrows. The tertiary contact in (b) between bulge B1 and loop L2 is indicated via horizontal lines [8].

Engineered *trans*-acting minimal hammerhead ribozymes consist of three domains: (a) a *substrate binding* domain, which is simply complementary to the sequence of the substrate; (b) a *catalytic* domain, which is necessary for cleavage to occur and finally (c) a *structural* domain that connects

both parts of the catalytic domain together and ensures that the ribozyme takes the right form in space. A schematic representation of a minimal hhRz is shown below.

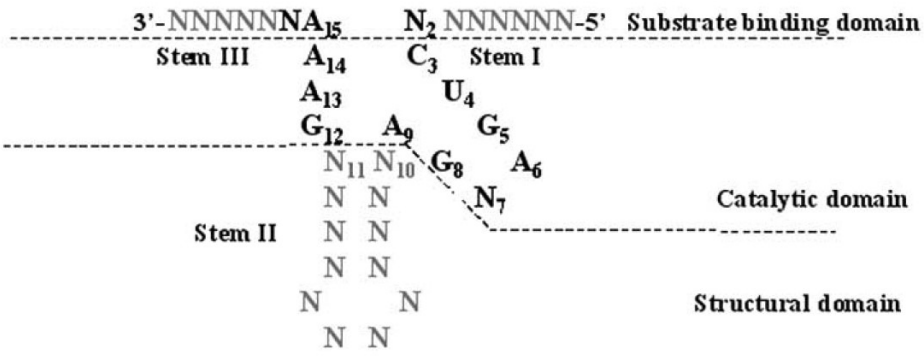


Figure 3. A Schematic Representation of the Secondary Structure of the minimal hammerhead ribozyme. The conserved residues in the catalytic core are shown explicitly. N stands for any nucleotide [9].

A distinguishing advantage of hammerhead ribozymes is the limited number of restrictions imposed on potential RNA targets. In essence, any (a) sufficiently *long* RNA strand with (b) an *accessible* region in the immediate neighborhood of (c) a GUC *triplet* (upstream of the intended cut-site) constitutes a potential target for cleavage by hammerhead ribozymes.

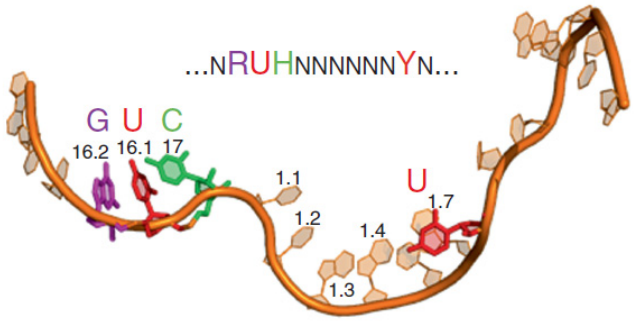


Figure 4. The Optimal Sequence of RNA Substrates Targetable by Hammerhead Ribozymes. R is preferably G, though A is possible. H is preferably C, though all residues are acceptable. Y at position 1.7 is preferably U, but C is possible [8].

However, the use of the minimal hammerhead ribozyme as a gene silencing agent for *in vivo* applications was impeded by the apparent need for unrealistically high (10 mM) concentrations of Mg^{2+} and their slow (~1 molecule/min) turnover rates. It was later discovered, however, that

hammerhead ribozymes do not strictly require divalent cations for catalysis, as long as a sufficiently high concentration of even non-metallic monovalent salt is present, permitting the RNA to fold correctly [10]. This meant that the full-length structure of naturally occurring (self-cleaving) ribozymes held the secret to their catalytic activity.

In 2003, it was shown that interactions between non-conserved elements (e.g., loops and bulges) outside the conserved catalytic core stabilize the ribozyme in a catalytically active conformation [11]. This allows ribozymes to cleave at low physiological relevant concentrations of Mg^{2+}. Still, these tertiary stabilizing motifs (or TSMs) imposed additional restrictions on the sequence of potential cleavage targets of trans-cleaving ribozymes. To remove these restrictions, SELEX was used to find new trans-cleaving hammerhead ribozymes, with TSMs, which are active at physiological concentrations of Mg^{2+} [12].

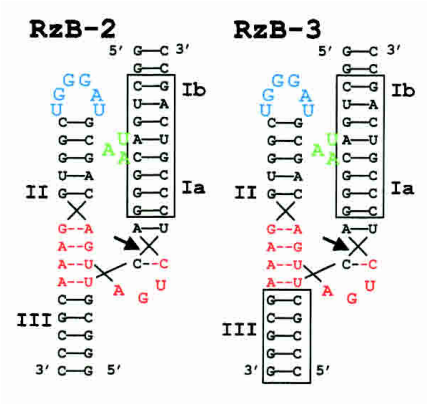


Figure 5. Examples of two full-length ribozymes (RzB-2 and RzB-3) annealed to RNA substrates [12]. Both ribozymes are distinguished by interaction between a UAA bump in helix I and a lager UGGGAU loop in helix II; this interaction stabilizes the ribozyme, without imposing additional conditions on the sequence of the substrate.

It is mainly these trans-acting full-length hammered ribozymes that will form the core of the augmented hammered ribozymes devised in the next section.

AUGMENTED HAMMERHEAD RIBOZYMES

For gene silencing applications, ribozymes target gene transcripts within a cellular environment very different from that of *in vitro* experiments. The average length of a mammalian mRNA is ~2.2 K bases. Such long mRNAs are highly structured, resulting in many double-stranded regions. This makes it much harder (often impossible) for an unaided small RNA to access a region with an intended cut-site.

In addition, in mammalian cells, mRNAs form in the cytoplasm and hence, it is important that hhRZs be co-localized to the cytoplasm, where they can attack mRNAs. The cytoplasm will have a temperature and certain ionic concentrations that must be taken into account when (artificially) folding RNAs (hhRz and target) to, for example, gain an *un*reliable idea about their secondary structure. It is critical that each of the substrate-binding part of the two 'arms' that make-up the substrate binding domain of a hhRz do not exceed 19 bases long each, to avoid the RNA interference (RNAi) mechanism [13]. On the other hand, *in vitro* experiments showed that short, 4+5 self-biding ends of a minimal ribozyme have virtually no effect on the kinetics of cleavage, while resulting in up to a 20-fold increase in the ribozyme's discrimination towards mismatching targets; the lengths of the substrate-binding part of the ribozyme's 'arms' were 9+9 [14]. Hence, unless a hammerhead ribozyme is going to be augmented with additional elements (our case), it appears possible to enhance the specificity of a hhRz (and maybe its resistance to exonucleases) by giving it short self-binding ends.

It was found that the entire structure of the transcript containing a hammerhead ribozyme determined not only the ribozyme's cleavage activity but also its intracellular half-life [15]. Further, all the tested ribozymes were chimeric $tRNA^{Val}$ ribozymes, which were transcribed in the cell nucleus, recognized by the nuclear protein exportin-t (Xpo-t) and hence, exported efficiently from the nucleus to the cytoplasm. This ensures that both the ribozymes and their target transcript are present in the same compartment. The cytoplasmic localization of $tRNA^{Val}$ ribozymes and the intracellular half-life and steady-state level of each $tRNA^{Val}$ ribozyme were the major determinants of their *in vivo* functional activity [15].

Since accessibility to the target cut-site is a major determinant of ribozyme activity, it would be useful to find a way to 'unzip' double-stranded portions of a structured RNA target; all RNA strands are structured to some degree. In [16], it was reasoned that part of a ribozyme's extended

structure could be an element that recruits a protein, which would open-up all secondary structures, thereby making all cut-sites on a substrate accessible to catalysis by the ribozyme. To create this hhRz, the hhRz was linked to an RNA motif, a constitutive transport element (CTE), which appears to interact with RNA helicase. An RNA helicase is a member of a class of helicase proteins with non-specific RNA binding, sliding and unwinding activities.

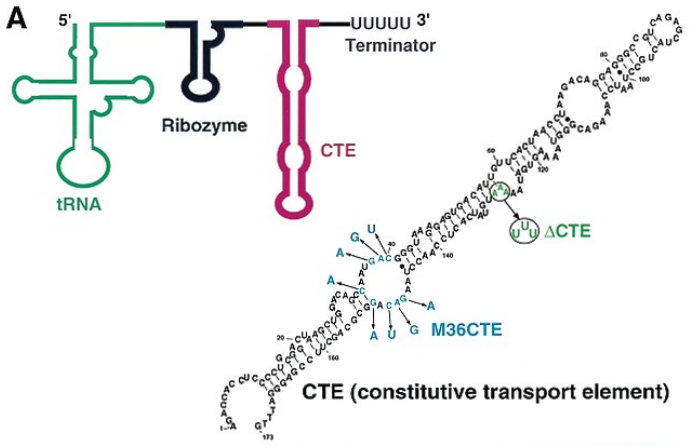


Figure 6. The schematic of a minimal hammerhead ribozyme, with two additional elements [16]. The 5'-end element is a $tRNA^{Val}$ for directing the whole transcript towards the cytoplasm, while the 3'-end element is a constitutive transport element (CTE) to recruit an RNA helicase, which in turn would unwind all double-stranded segments of the target RNA substrate.

HAMMERHEAD RIBOZYME RESISTANT TRANSCRIPTS

The genetic code is the mapping between the nucleotide triplets (or codons) on an mRNA strand and the respective peptides of a poly-peptide chain (or protein), a mapping which is almost invariant across all living organisms. Every codon maps to a peptide, including the start codon, which codes for Methionine, except for three codons, codenamed amber, ochre and opal. Given that there are 4 possible nucleotides, there are 64 codons, but only 20 standard amino acids. Hence, many codons map to the same amino acid, making the genetic code 'degenerate'. This, however, allows for silent mutations, where a change of a nucleotide in a codon does not lead to a change in the resulting amino acid.

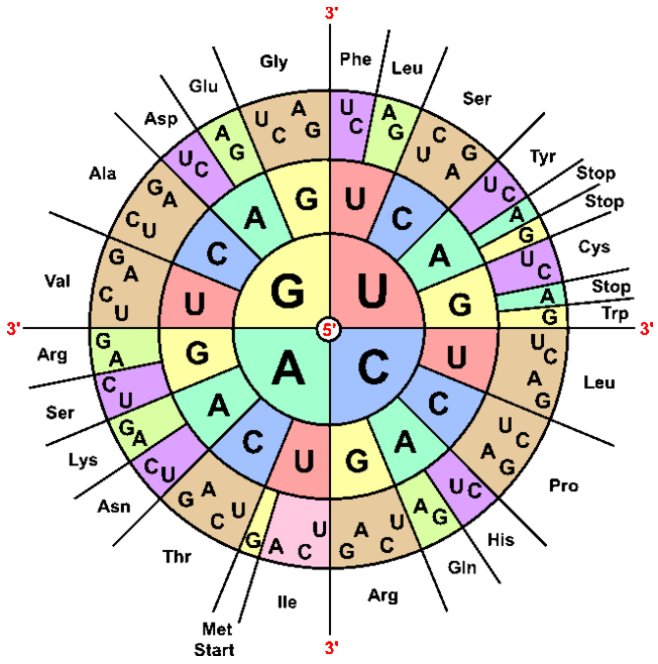


Figure 7. The genetic code [17].

Since ribozymes act on the RNA transcript of a gene and not on the corresponding protein, it is possible to generate a hhRz (or even several hhRzs) that cleave the transcript of a mutant gene, and simultaneously generate a hhRz-resistant transcript of a healthy version of the same gene. This strategy is probably applicable to all trans-acting ribozymes, but in the specific case of hammerheads, all that is need is to either (a) change the RUH cut-site to something else, preferably replacing the U with another nucleotide or (b) alter the transcript's sequence where the hhRz is supposed to anneal, to reduce the complementarity between that sequence and the ribozyme's arms, while ensuring that the codons of the new transcript map to the required poly-peptide chain. In fact, it may be appropriate to use this occasion to also optimize the codon sequence for maximal expression within a given organism.

Though it was well understood that many factors affect the rate of protein expression from an mRNA, such as translation initiation regions and RNA structural elements (throughout the transcript), there has been an 'obsession' with codon optimization of genes when expressed in a different host, mainly on the basis of codon usage frequencies. Common codon optimization strategies include (a) employing the *most* used synonymous codons within the host organism; (b) *randomly* selecting a synonymous codon after discarding very rare codons; (c) replacing highly/rarely used codon in the native species with *equally* probable codons in the target species.

Recent studies, however, have reminded researchers that such popular approaches rest on a bed of empirically unsubstantiated assumptions.

However, a recent thought-provoking study [18], did examine the assumptions that underlie common approaches to codon optimization, namely: (a) rare codons are rate-limiting for protein synthesis; (b) synonymous codons are interchangeable without affecting protein structure and function; (c) replacing rare codons with frequently used ones increases protein production. The study found or argued that (a) codons designated as rare may be incorrectly categorized and may not limit the rate of translation; (b) functional properties of expressed proteins (e.g., fluorescence) can be altered by synonymous codon replacement due to differences in protein folding; (c) though codon usage in mammals has not yet been studied systematically as in *E. coli* (see [19]), there is little reason to suggest that traditional codon optimization would lead to enhanced protein expression levels.

In the case of *E. coli*, a library of 154 genes was constructed and expressed in that organism, with all the genes expressing the same green fluorescent protein (GFP), but bearing random variations at synonymous sites. GFP protein levels varied 250-fold across the library, but codon bias did *not* correlate with gene expression. In contrast, the stability of mRNA folding near the ribosomal binding site explained more than half the variation in protein levels [19].

RIBOZYMES FOR GENE SILENCING

Several hammerhead ribozymes for the treatment of autosomal dominant mutations in the human rod opsin (RHO) gene, were tested for knockdown effect on human RHO gene expression. Computational simulation was used to select target mRNA regions likely to be single stranded and hence, accessible to hhRz annealing and cleavage. The maximum opsin protein level knockdown in this study was ~30%, over a 48 hour period [20].

Constitutive activation of nuclear factor-κB (NF-κB) is implicated in the tumorigenesis of several cancers including melanoma; inhibitor of κB kinase (IKK) functions as a major mediator of NF-κB activation. In [21], the authors constructed and tested a double hhRz expression system to silence IKKβ. Their *in vivo* data showed that the knockdown of IKKβ significantly reduced the growth of melanoma lesions in mice [21].

The chemokine receptors CXCR4 and CCR5 are required for HIV-1 virus to enter cells. In [22], it was shown that hhRz activity can lead to the inhibition of the cell surface expression of both CCR5 and CXCR4. This resulted in a significant inhibition of HIV-1 replication, when peripheral blood mononuclear cells were presented with the virus.

For a comprehensive summary of successful applications of hammerhead ribozymes, see [23], with targets including (a) Growth factors, receptors, transduction elements (e.g., PDGF-A, VEGF, TGF-b, HGF, Oestrogen receptor and G Protein); (b) Oncogenes, protoncogenes, fusion genes (e.g., H-ras, K-ras, N-ras, Her2/neu, bcr/abl, fos and PML/RAR); (c) Apoptosis, survival factors, drug resistance (e.g., Bcl-2, Survivin, fos and Telomerase RNA and MDR-1); (d) Transcription factors (e.g., HIF-1, E2F1, p300 and NFkB); (e) Extracellular matrix, matrix modulating factors (e.g., Osteopontin, Tenascin C, MMP-9 and Type-1 collagen); (f) Circulating factors (e.g., Apolipoprotein-A, Apolipoprotein-B, Antiboby k-chains and Pleiotropin); (g) Viral genome, viral genes (e.g., Hepatitis HCV, Hepatitis HBV, Herpes HSV, Immunodeficiency HIV, Tat, Rev and Env).

To our knowledge, however, no hammerhead ribozyme has been tested for its ability to knockdown PABPN1 expression *in vitro* or in a model organism.

RIBOZYMES FOR PABPN1 GENE SILENCING

Oculopharyngeal muscular dystrophy (OPMD) is a lateonset muscle disease associated with progressive ptosis of the eyelids, dysphagia, and unique tubulofilamentous intranuclear inclusions (INIs) in skeletal muscle. In 1998, the common dominant and rarer recessive forms of OPMD both were found to be caused by short $(GCG)_{11–17}$/polyalanine expansions in the polyadenylate-binding protein nuclear 1 gene (PABPN1) [24]. Dominant and recessive OPMD are caused by mitotically and meiotically stable short triplet repeat expansions of a $(GCN)_{10}$ and, more rarely, point mutations, leading to a lengthening of a polyalanine stretch in the protein [25].

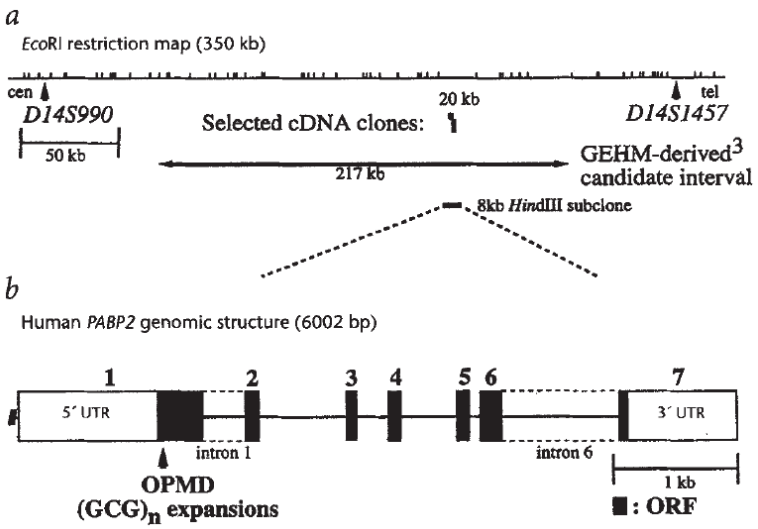


Figure 8. Positional cloning of the PABPN1 gene; (a) positions of the PABPN1 selected cDNA clones in relation to the EcoRI restriction map; (b) genomic structure of the PABPN1 gene and the location of the $(GCG)_n$ expansions- all taken from [24].

THE SEQUEL

The rest of this thesis is divided into a number of chapters. The following chapter 2 presents the *computational algorithm* of the bioinformatics web service (RiboSoft) used to design the hammerhead ribozymes that were indeed tested, *in vitro* as well as *in vivo,* for their effect on the PABPN1 transcript. Chapter 3 presents the experimental *methods* & *results* of the wet lab experiments. Chapter 4 presents a brief *conclusion* and a suggestion for *future* work- some of which is in progress, at the Montreal Neurological Institute & Hospital.

Chapter 2: Computational Algorithm

This chapter describes a new computational algorithm, which was implemented as a web application called RiboSoft by Gabriel Belmonte & Anas Ambri. The algorithm employs published declarative and procedural knowledge, but it does so as part of a holistic design & optimization procedure. The procedure emulates, to a large extent, the way in which an expert human would go about designing an efficient hammerhead ribozyme. The algorithm automatically generates a list of ranked hammerhead ribozymes to target an mRNA or arbitrary sequence of nucleotides, inputted by a user. There are other programs that share some of the attributes of RiboSoft; these programs are reviewed below and the advantages of our program are identified. This chapter provides details of the algorithm in parallel with a user perspective of the implementing program. This is the programs used to design the hammerhead ribozymes, which in chapter 3 are tested for their efficacy in cleaving the PABPN1 transcript.

BIOINFORMATICS' REVIEW

RiboSubstrates [26] is a web accessible bioinformatic tool that allows the user to search a cDNA database for all potential substrates for a given ribozyme (hammerhead ribozymes as well as SOFA-HDV ribozymes). However, it is not a proper design tool. A design tool takes into consideration all of the objectives & constraints that are of concern to the user, so its output meets the requirements of the user and does not require further manual processing- especially if that processing could have been carried out by a computer.

Aladdin (seArch computing tooL for hAmmerheaD ribozyme DesIgN) [27] is the only web service that offers hammerhead design capability similar to RiboSoft's. This tool, which evolved from an earlier computational method [28], starts by mapping all canonical NUH cleavage sites in the RNA sequence to be cleaved, and then identifies all the target sites including the flanking elements. It uses the UNAFold package to predict the accessibility of each target according to calculated secondary structures [29]. Finally, it outputs a set of suitable hammerhead ribozyme candidates with optimal structural folding.

The web tool is quite handy and scientifically sound. However, it suffers from several disadvantages. First, Aladdin only allows the user to provide a target sequence (to be cleaved). Other significant specifications that include (but are not limited to) the nature of the environment, the type and size of the ribozyme and the nature of the application. For an 'automated' tool, the user will end-up searching through and manually optimizing the 'optimized' output of Aladdin, before he can in fact start the process of synthesizing it.

Second, Aladdin provides the following information to the user on the proposed ribozymes: the exact nature (e.g., GUC) and numerical position, on the substrate, of the targeted cut-site; the binding energy of the two arms of the proposed ribozyme; accessibility of the two arms of the ribozyme to the cut-site; the uniqueness of the ribozyme and, naturally, its actual sequence. The output also includes a link that can be used to view the secondary structure of the ribozyme and a numerical score (in [0,1]) that is purely based on structural features. As such, Aladdin provides a lot of useful information, but (a) it does not provide any off-target specificity analysis; (b) does not utilize the latest research on target-site accessibility [30]; (c) does not present the user with a single *overall* rank that can be used to pick multi-objectively optimal ribozymes; (d) it does not offer the user the ability to sort the proposed ribozymes according to one of the evaluated criteria.

Finally, Aladdin does not adopt a multi-objective optimization approach to the evaluation and ranking of the possible hammerhead ribozymes, but rather goes through filters that successively eliminate candidate ribozyme that do not satisfy one criterion at a time. All in all, Aladdin – or rather the computational algorithm behind it – is a solid contribution to automated ribozyme design tools, but it is a touch out-of-date, incomplete and not as easy to use or interpret as RiboSoft.

COMPUTATIONAL ALGORITHM

A riboswitch is an RNA that turns genes on or off, when it binds to a ligand input or a combination of ligands inputs [31]. Riboswitches can be combined with ribozymes so that the binding of their ligands trigger RNA cleavage [32][33]. Engineered examples of riboswitch-ribozymes implementing Boolean functions include some that responded to small-molecule ligands [34]. However, the ligands were found empirically in two steps. First, via SELEX [35], a set of RNA sequences were found that bind to a targeted molecule, then many of those RNA sequences were tested as part of an extended arm of a hammerhead ribozyme, until one was found that could trigger a conformational change in the ribozyme, leading to its activation (or inactivation).

In [36], a purely computational approach was developed that found complementary RNA ligands that altered the conformation and hence activity of a hammerhead ribozyme. In fact, they tested *in vitro* computationally designed ribozymes that implemented four Boolean logic functions: AND, OR, NOT and YES. It is worth noting that only a precisely delineated part of the ribozyme need to be altered to give the whole ribozyme this unique allosteric capability; the rest of the ribozyme stays fixed.

New Hammerhead Ribozyme Automated Design Algorithm

Before the design algorithm can run, valid input from the user must be entered. The required input takes the form of: (1) a target nucleotide sequence, which may be provided by: direct entry into a text box, uploading a pure text (.txt) or FASTA format file, or typing in a GenBank sequence number; and (2) a set of design parameter values.

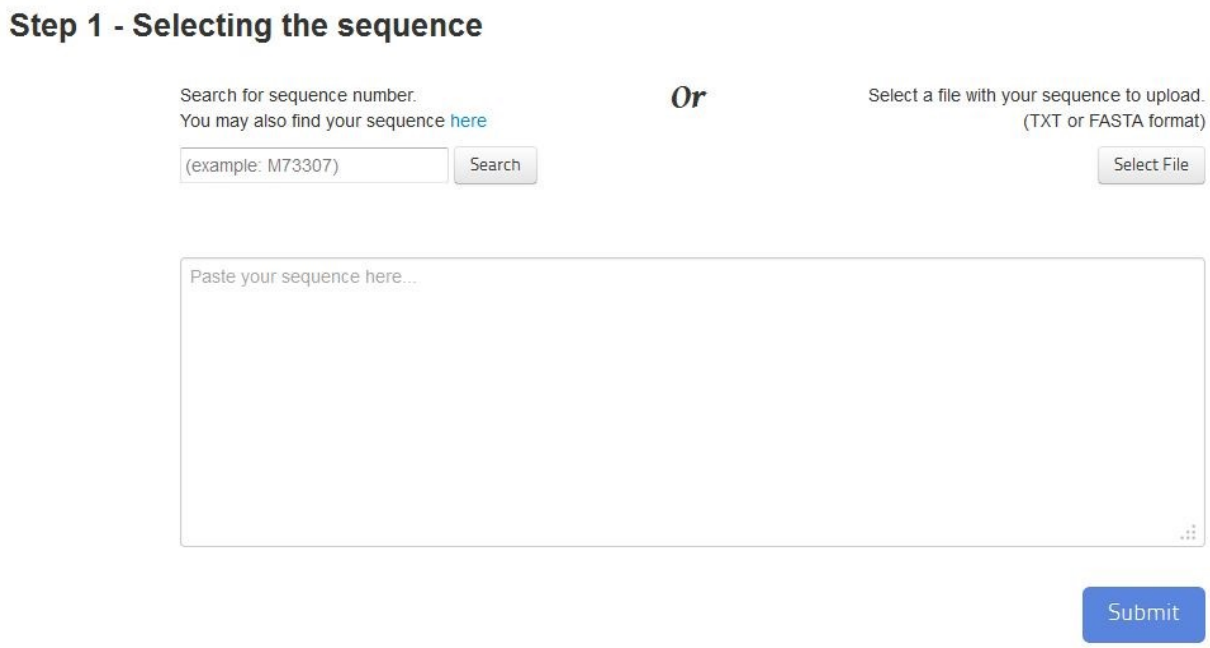


Figure. 9. A Snapshot of the Input Screen (step 1) of RiboSoft: Asking the User for a Sequence, a File or a Sequence Code.

The design parameters are: (a) the target environment: whether it is *in vitro* or *in vivo*, and if it is *in vivo*, then the name or code of the intended organism; (b) environment properties: temperature as well as oligomer, sodium and magnesium ion concentrations; (c) potential cut-sites (e.g., GUC); (d) ribozyme structure: minimal or extended (i.e., involving loop-loop interaction) [11]; (d)

advanced options: the minimum and maximum lengths of helices I and III (the targeting arms of the ribozyme), the presence or absence of a T7 promoter (usually used for *in vitro* experiments) and whether specificity should take into consideration off-target sites that allow annealing only or annealing and cleavage by the designed ribozyme.

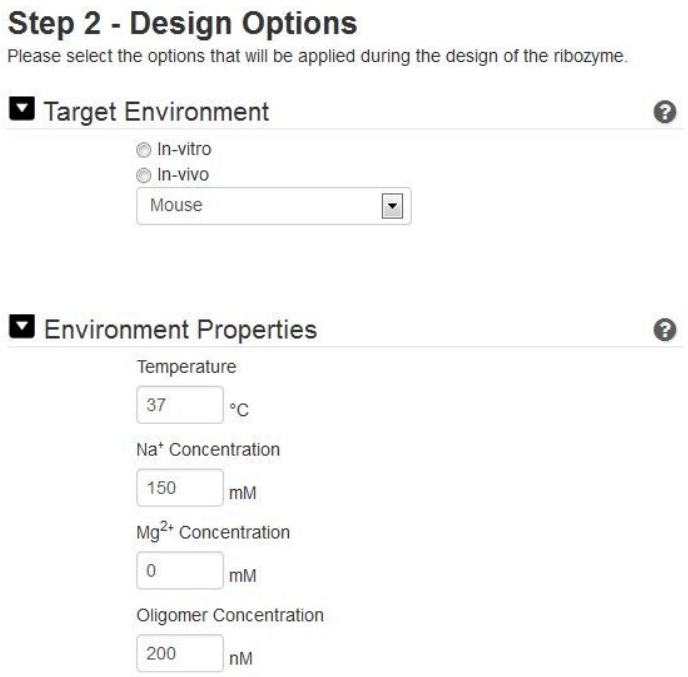


Figure. 10. A Partial Snapshot of the Input Screen (step 2) for RiboSoft: Allowing for User Specification of Various Parameters.

A target sequence must be provided, but most of the design parameters have default values, which are acceptable for most applications. The design algorithm then initiates the actual design and optimization procedure. This procedure has six steps, which starts with the generation of a large number of candidate ribozymes, assessing them using various measures of quality, including (but not limited to) target cut-site accessibility and specificity to the target sequence as well as proper folding of the ribozyme structure. The procedure concludes with the multi-objective assessment and ranking of resulting candidate ribozymes. The algorithms for each part of the procedure are explained below.

Step 3 - Summary

Here are the design options that will be applied to the sequence.

Design Option	Value
Sequence Length	3 nucleotides
Target Environment	In-vitro
Temperature	37
Na^+ Concentration	150
Mg^{2+} Concentration	0
Oligomer Concentration	200
Cut-sites	GUC
Fold Shape	Basic
Helix I Length	Between 3 and 8
Helix III Length	Between 3 and 8
Use Promoter T7	No
Specificity	Cleavage only

Figure. 11. A Snapshot of the Input Screen (step 3) for RiboSoft: Confirmation of User Input.

Candidate Generation

In principle, a hammerhead ribozyme (hhRz) can cleave, with different efficiency levels (inconsistently measured by different studies [37][38]), at locations on RNA strands with an NUH motif. An RNA strand may have any number of different cut-sites within it, repeated any number of times. The most widely targeted cut-site is GUC, as it is thought to be the most susceptible to hhRz cleavage.

Since the language of RNA has 4 letters in its alphabet: A, U, C and G, the probability of any 3-letter sequence is $(¼)^3 = 0.015625$, or one occurrence every 64 bases, assuming a uniformly random distribution of bases over a strand- an assumption often violated. Nevertheless, since there are empirically established databases of the genomes of many model organisms, including humans, it is usually possible to identify the exact number and position of any cut-site on the mRNA of any gene.

For all cut-sites, all possible candidates with perfectly hybridizing arms of lengths, between the minimum and maximum values specified by the user, are generated. The total number of

candidates is calculable using the formula:

$$\approx (\textit{arm-length range})^2 \cdot (\textit{number of cut-sites}) \quad (1)$$

This implies that the number of candidates is also proportional to the target sequence length, since more cut-sites will be found. With a conservative length of 3-10 for the hybridizing arms and 2 cut-site types, a sequence of 3000 nucleotides yields about 100 cut-sites for 2 cut-site types. This results in a total of 4900 ribozyme candidates. This provides a rough estimate of the number of candidates that are typically generated by RiboSoft.

Annealing Temperature

As a refresher, a trans-cleaving hammerhead ribozyme may be viewed as a three-part construct: a fixed largely double-stranded central segment ('hammerhead'), plus two single-stranded arms (stems I and III) up- and down-stream of the hammerhead. It is those two arms that provide the selectivity that permit the ribozyme to strongly anneal to a particular site on a target RNA molecule amongst myriad RNA strands in the medium and hence, cleave that target.

The shorter an arm is, the lower the degree of selectivity and strength of annealing (annealing temperature) between the arm and the target RNA. In converse, the longer an arm is the more selective the annealing and the more heat is required to de-anneal the arm from the target RNA. Thus, it may appear that the solution is to automatically give every hhRz very long arms, but the catch is that longer single-stranded arms make it more likely that an arm would anneal to itself and/or with the other arm (partially or fully). Also, ribozymes with very long arms will anneal so strongly to their target that they will continue to anneal to fragments resulting from the cleavage event and hence, not turn over to anneal to new targets.

In brief, the annealing temperature of an hhRz may be approximated by the sum of the annealing temperatures of the two arms of the ribozyme. All environmental conditions (e.g., ionic concentrations) being equal, the annealing temperature is a function of the nucleotide make-up and lengths of the arms of a ribozyme. A number of formulas have been used and compared [39] to calculate the annealing temperature of RNA (and DNA) molecules, including the fairly standard base-stacking approach employed by our algorithm (adopted from [40]).

Very low and very high annealing temperatures are not desirable. Since calculation of the annealing temperature is computationally simple (relative to RNA folding) we use this calculation

in a discriminatory manner to weed out all candidate hhRz with arms that are too short or too long. In fact, the range of the annealing temperature we use is from 0 to 65°C above the temperature of the intended environment (e.g., 37°C for mammalian cells). The upper bound of this range might appear too high, but this value is for both arms, and so the annealing temperature for one arm might in some cases be only a few degrees above the lowest reasonable temperature for effective annealing.

For a given cut-site on a target RNA strand, the sequence of any arm of an hhRz is dictated by the RNA sequence around the cut-site, as an hhRz should anneal perfectly to its target. Hence, every hhRz with target-matching arms and an overall annealing temperature within the pre-specified range is an acceptable hhRz candidate generated by this phase of the design process.

Ribozyme Structure

Every hhRz candidate surviving the annealing temperature phase is evaluated for its structure. This is a measure of how open (single-stranded) the arms are, which makes it easier for a ribozyme to find and anneal to its intended target sequence and hence, perform its trans-cleavage function. The Sfold program [41] is used on the server side to fold all the candidates; it uses a statistical Boltzmann ensemble model to generate possible secondary structures for the candidate RNAs.

This means that Sfold generates more than one secondary structure for each candidate, groups them by similarity and assigns them a frequency (or probability) value. In fact, 10 secondary structures are reported, each with its own probability. This probability is necessarily less than 1, and the sum of all probabilities is 1. The probability is used as a weight when considering the structure of a ribozyme. All of the temperature values used are in Kelvin, as cancelation may erroneously occur with negative temperatures.

The quality of a given secondary structure (of many for one hhRz) is given by an algorithm described by the pseudo-code:

```
sum = 0;
for each candidateStructure
  ContinuousPairs = FindContinuousBasePairs();
```

```
  for each ContinuousPair in ContinuousPairs
    sum += MeltingTemperature(ContinuousPair);
  end for each;
end for each;
```

A candidateStructure is a possible secondary structure of an hhRz. FindContinousBasePairs will find all double-stranded sequences in the structure, such that there is at most one mismatch in each sequence. This list will be stored in ContinousPairs. For each of these double stranded segments (or ContinuousPair), MeltingTemperature will be computed and added to sum. This sum represents the melting temperature of the whole candidateStructure.

Hence, the algorithm uses the sum of the melting temperatures of all continuous stretches of double-stranded RNA as an (inverse) measure of the openness of the two arms of the ribozyme. It is worth noting that we allow up to one mismatch in each 'continuous' stretch of double-stranded RNA.

For each secondary structure of the ten secondary structures associated with a given hhRz, we compute one overall melting temperature. Naturally, continuous pairs in the hammerhead core (Helix-II) will not be counted, since they are meant to be there. However, a penalty of 50 degrees is added to any structure lacking the hammerhead core. Since the resultant temperature here is associated with only one of ten possible structures, a weighted sum (Q_f) of all ten temperatures is calculated, for each hhRz candidate, using the formula:

$$Q_f = \sum_{i \in \mathbf{Q}} f_s \cdot Q_s \qquad (2)$$

Q_s is the melting temperature of one secondary structure; f_s is the probability associated with that secondary structure; **Q** is the set of all ten structures. Naturally, lower values for Q_f are preferred and as such, this measure of quality should be minimized.

Target Accessibility

Accessibility is a measure of how easy it is to access the target. Targets that fold unto themselves strongly in the region around and including a cut-site will reduce the efficiency of a ribozyme designed for that cut-site.

Two accessibility terms are computed here. One term, *cut-site inaccessibility*, inversely measures the openness of a target cut-site, to a particular hhRz candidate, with specific arms' lengths. This is reflected by the annealing temperature of the region of the target cut-site, the region to which the arms of an hhRz must anneal. The other term, *disruption energy* ($\Delta G_{\text{disruption}}$) is based on research done by [30], and is in fact: the free energy cost to melt any local secondary structure at the target site, making it accessible to hhRz binding.

Both calculations are described succinctly by the pseudo-code:

Cut-site inaccessibility

```
sum = 0;
for each fold in targetFolds
  ContinuousPairs = FindContinuousBasePairs();
  for each ContinuousPair in ContinuousPairs
    sum += MeltingTemperature(ContinuousPair);
  end for each;
end for each;
```

The algorithm is identical to the one used for Ribozyme Structure explained earlier. Only the regions have changed. The regions are now the cut-site regions (or folds) in targetFolds.

Disruption energy

```
for each cut-site
  constrainedFolds =
   FoldTargetWithOpenCutsite(target, cut-site);
  cut-site.disruptionE =
```

```
    targetFolds.LFE-constrainedFolds.LFE;
end for each;
```

The cut-site is that region of the target substrate that is complementary to the arms of the hhRz. The constrainedFolds represent secondary structures of the target such that the target is forced open at the cut-site. These secondary structures are obtained through the FoldTargetWithOpenCutsites operation. The disruptionE energy is the difference between two energies: targetFolds.LFE and constrainedFolds.LFE. targetFolds.LFE. The former is the free energy of the unconstrained folding of the target substrate, while the latter is the free energy of the target substrate, with the cut-site region forced open.

The computation of cut-site inaccessibility is done per candidate hhRz, as different ribozymes targeting the same cut-site will have different arms' lengths. The substrate is folded using Sfold, which returns 10 secondary structures with 10 potentially different secondary structures for the cut-site region. Hence, 10 annealing temperature values are computed for that region, and these are summed using the probabilities of the corresponding secondary structures as weights. The result of this calculation is the raw cut-site accessibility value.

The calculation of disruption energy is done per cut-site. The reason for this is the significant computational cost associated with the computation of the secondary structure for a long RNA strand; typical gene transcripts are one to several thousand bases long. Playing it safe, we assume the maximum allowable lengths for the arms of the hhRz, and hence compute $\Delta G_{\text{disruption}}$. Since Sfold returns ten possible secondary structures for the target transcript, we will also have ten $\Delta G_{\text{disruption}}$ values. These values are used to calculate a weighted linear sum in a fashion identical to what was done for cut-site accessibility.

As with ribozyme structure, cut-site inaccessibility and disruption energy should be minimized.

Specificity Assessment

It is often important, especially for *in vivo* applications, that a given ribozyme affects the targeted transcript and no other. Typically, an hhRz would affect a transcript by cleaving it. It is possible, however, for an hhRz to anneal to a transcript and hence, lead to the transcript's destruction or to a reduction in its translational efficiency, without directly (or indirectly) cleaving it. This is part of what is known as off-target effects. Hence, the advanced options' section of RiboSoft's input pages

allow the user to check for either off-target hits that would result in cleavage or *all* off-target hits, including those that do not lead to cleavage.

The NCBI BLAST web-service is queried with a high threshold (sensitivity) using the region around the targeted cut-site corresponding to the maximal possible lengths of hhRz arms. The following pseudo-code describes the process:

```
Specificity = 0;
for each cut-site
  results = Queryblast(cut-site);
  for each result in results
    Weight(result);
    if result is XM or XR
      result.Weight = 0;
    end if;
    Specificity += result.Weight;
  end for each;
end for each;
```

Specificity is initialized to zero. For each cut-site, results will hold information obtained from querying BLAST on that cut-site. A cut-site is defined here as the region of the target that is complementary to the arms of the hhRz. Each BLAST result will be dealt with individually, even if two or more results are attributed to a single transcript. Weight will weigh each result based on the perfection of its match to the hhRz arms. These weighted results will be summed into Specificity. An XM or XR Blast hit will not be used in the specificity calculation.

The results are weighted based on the number of matched pairs, thus a perfect match yields a weight of one (the targeted transcript). It should be noted that speculative RNA (XM and XR) are ignored in the count, but are still reported in the results' page.

The program will only return 0 if either no organism was found or specified or the RNA sequence does not exist, since there is not necessarily *a priori* knowledge of the transcript being targeted (users can enter arbitrary sequences directly).

Optimally, the value of specificity is 1, entailing a perfect match to only one cut-site on the targeted transcript. Higher specificity values are detrimental to the quality of the hhRz candidate, and as such this measure of quality should be minimized.

Fitness Evaluation

The quality of a candidate ribozyme is reflected in all five terms described in the sections above. It should be noted that the raw values returned by some of those quality measures, are not intuitively comprehensible. For instance, greater values for ribozyme structure indicate worse structures. In addition, it is often not clear what the lower and upper bounds are, for a given measure.

For these reasons, the following equation is used to normalize the values returned by most of the quality measures:

$$M_{normalized} = \frac{(M_{original} - M_{\min})}{(M_{max} - M_{min})} \qquad (3)$$

Where M_{min} is the smallest value for that measure among all candidates, and vice versa for M_{max}. $M_{original}$ is the original (raw) value for that measure, while $M_{normalized}$ is that value after normalization. Also, if for a given measure, larger values are detrimental, the following equation is applied to the output, so that 1 represents the best possible value and 0 the worst:

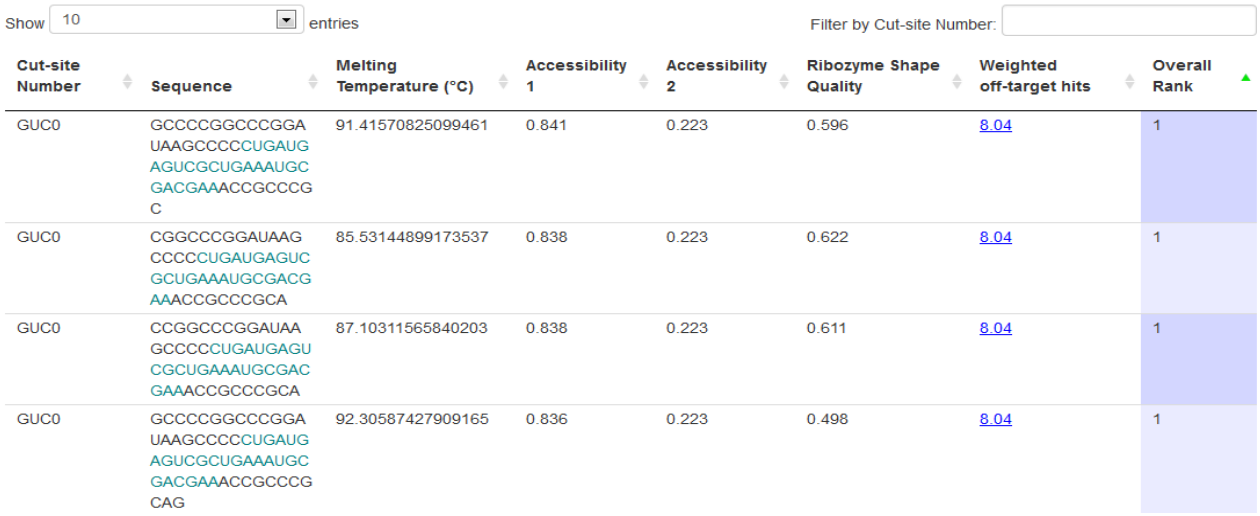
$$M_{adjusted} = 1 - M_{normalized} \qquad (4)$$

$M_{normalized}$ is the normalized value for that measure, while $M_{adjsuted}$ is that value inverted, and is the final value used by the multi-objective optimization procedure outlined below. Since melting temperature and specificity can be accurately judged by the user, they are not normalized or adjusted.

Depending on the application, the designer will have a greater interest in certain parameters versus others. For example, specificity or lack of off-target effects, may be of great value to some *in vivo* applications, but would probably be of no concern to most *in vitro* applications. Hence, it is not

possible to identify, beforehand, and for all possible applications of an hhRz, what the relative importance of the various terms should be. Even for a particular well-defined application, it is sometimes impossible for a user to spell-out numerically the relative importance of the various quality terms. As such, it was our decision to adopt a multi-objective optimization approach to the ranking of the ribozymes.

This adopted approach (described in full in [42]) allows us to present the various candidates, to users, as members of various Pareto optimal fronts. All ribozymes belonging to the first front are awarded the same rank (1); all ribozymes belonging to the second front (once the first front has been removed) are awarded rank 2, and so on, until all candidates are ranked. In order for the user to make a final decision as to the best hhRz for his application: both rank and the various quality measures are included in a table of all hhRz candidates. The table can be sorted, by the user, according to any of the individual measures as well as overall rank. This table constitutes the output of the web service (RiboSoft) implementing the algorithm.


Show 10 entries

Filter by Cut-site Number:

Cut-site Number	Sequence	Melting Temperature (°C)	Accessibility 1	Accessibility 2	Ribozyme Shape Quality	Weighted off-target hits	Overall Rank
GUC0	GCCCCGGCCCGGAUAAGCCCCCUGAUGAGUCGCUGAAAUGCGACGAAACCGCCCGC	91.41570825099461	0.841	0.223	0.596	8.04	1
GUC0	CGGCCCGGAUAAGCCCCCUGAUGAGUCGCUGAAAUGCGACGAAACCGCCCGCA	85.53144899173537	0.838	0.223	0.622	8.04	1
GUC0	CCGGCCCGGAUAAGCCCCCUGAUGAGUCGCUGAAAUGCGACGAAACCGCCCGCA	87.10311565840203	0.838	0.223	0.611	8.04	1
GUC0	GCCCCGGCCCGGAUAAGCCCCCUGAUGAGUCGCUGAAAUGCGACGAAACCGCCCGCAG	92.30587427909165	0.836	0.223	0.498	8.04	1

Figure. 12. Sample Output from RiboSoft

A brief representation of the algorithm used for fitness ranking is presented below.

```
ParetoRank (candidates, rank)
{
  for each pivot in candidates
    for each candidate in candidates
      if (Dominates (candidate, pivot))
        exit inner-loop,
      continue outer-loop -> try new pivot;
    end for each
    pivot.rank = rank;
  end for each
  remove all x in candidates with pivot.rank = rank;
  ParetoRank(cadidates, rank - 1);
}
```

The first time this operation is called, it is called using a list of all the candidates called candidates and a rank equal to 1. A pivot is a candidate that will be compared to every other candidate. Candidate A dominates candidate B if it is at least as good as B in all measures of quality and better than B in at least one quality. If pivot is not dominated by any other as determined by Dominates, then its rank is set to rank. When all candidates have been evaluated, all those that have been assigned the current rank value are removed from the candidates list in this function. The function will be re-invoked with the ParetoRank call, with a decreased set of candidates and a decremented rank value.

Chapter 3: Biological Application

In addition to specifying and devising a comprehensive hammerhead ribozyme design algorithm, then implementing it as web service, we have also used the service for a particular gene silencing application (the PABPN1 gene), whose mutation results in the muscular dystrophy known as OPMD. Before proceeding to *in vivo* trials in various model organisms, which is a costly affair, we proceeded to test the validity of ribozymes *in vitro* and *in vivo*.

In this chapter, we describe the wet lab experimental methods used to test the cleavage efficacy and reaction kinetics of a small number (4) of hammerhead ribozymes generated by the computational algorithm. We measured the *in vitro* transcript cleavage efficiency & enzyme kinetics, as well the *in vivo* gene knockdown effect of individual ribozymes. We also measured the enhanced effect of using combinations of two or more ribozymes all targeting the same transcript. Finally, we design a mutant PABPN1 gene that gives the same protein as the wild type gene, but one that generates a transcript supposedly immune to the catalytic activity of hammerhead ribozymes. We test this hypothesis, *in vivo*, and provide the results.

The following sections describe in detail the wet lab experimental methods followed to test the ribozymes and the results of these tests. The results were positive and provide preliminary evidence of the value of the new algorithm.

EXPERIMENTAL METHODS

Hammerhead Ribozymes

The four hammerhead DNA templates, recommended by Ribosoft for PABPN1, were prepared by 5 cycles of PCR using Taq DNA polymerase and 1 μM of the T7 promoter TAATACGACTCACTATAGCG with 1 μM of each oligonucleotide (sequences complementary to the T7 promoter are underlined):

5’GGCCTGGAGTTTCGTCGCATTTCAGCGACTCATCAGTGAGGTTAAACTGGAGCCTGCGCTATAGTGAGTCGTATTA3’,

5’TATCAAAGCTCGAGTTTCGTCGCATTTCAGCGACTCATCAGAGGGATTAGATGGAGGAAGACGCTATAGTGAGTCGTATTA3’,

5’AGATGAATATGAGTTTCGTCGCATTTCAGCGACTCATCAGCACCTTTACCAGGCAATGCGCTATAGTGAGTCGTATTA3’,

5’GGGTCGCGTTTCGTCGCATTTCAGCGACTCATCAGTACAGTTAGGGCCGGCGCTATAGTGAGTCGTATTA3’

corresponding respectively to the constructions for Yz144, Yz363, Yz437 and Yz867, with an annealing temperature of 55°C with 200 µM dNTP, 2 mM $MgCl_2$, 120 mM Tris–HCl pH 8.8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$ and 0.1% Triton X-100. The sizes of the DNA templates were confirmed on 1.5% agarose gel. The sequences of these four ribozymes are:

Yz144: 5' CAGGCTCCAGTTTAACCTCACTGATGAGTCGCTGAAATGCGACGAAACTCCAGGCC 3'

Yz363: 5' TCTTCCTCCATCTAATCCCTCTGATGAGTCGCTGAAATGCGACGAAACTCGAGCTTTGATA 3'

Yz437: 5' GCATTGCCTGGTAAAGGTGCTGATGAGTCGCTGAAATGCGACGAAACTCATATTCATCT 3'

Yz867: 5' CCGGCCCTAACTGTACTGATGAGTCGCTGAAATGCGACGAAACGCGACCC 3'

PABPN1 Gene

The PABPN1 gene used to create by PCR the DNA template for transcription was the D323 clone (kindly provided by Guy Rouleau’s Lab) used in [24]. The first PCR product was made with the oligonucleotide 5’GACTACGGGAACGGCCTGGAGTC3’ starting 129 nucleotides downstream of the start codon and with the oligonucleotide 5’AAAGGGAACAAAAGCTGGAG3’ 147 nucleotides downstream of the stop codon (as reverse complement); covering 936 nucleotides. To generate a full-length PCR product, 20% betaine was incorporated to the PCR reaction (same reaction conditions as above) for 30 cycles at an annealing temperature of 51°C. A second PCR reaction was performed on the first PCR product with identical conditions, except for the first oligonucleotide being replaced by the oligonucleotide 5’TAATACGACTCACTATAGCGACTACGGGAACGGCCTGGAGTC3’ to incorporate the T7 promoter to the DNA template. The size of both DNA templates were confirmed on 0.7% agarose gel. The PCR product was purified with an EZ-10 Spin Column PCR Products Purification Kit (Bio Basic Inc.) and precipitated in ethanol with NaOAc.

RNA Synthesis of Hammerhead Ribozymes

Ribozymes were synthesized by *in vitro* run-off transcription in 100 µl reactions using 20 pmol of double-stranded DNA template, 80 mM HEPES–KOH pH 7.5, 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 3.0 mM of each NTP, 0.01 U of yeast pyrophosphatase and 2 µg of purified T7 RNA polymerase at 37°C for 3 h [43]. The reactions were stopped by adding 5 U of DNase (RNase

free) and incubating at 37°C for 30 min. The mixtures were extracted twice with both phenol and chloroform, and the nucleic acid precipitated with ethanol. After dissolution in equal volumes of water and formamide dye buffer (95% formamide, 10 mM EDTA, 0.025% bromophenol blue and 0.025% xylene cyanol) the ribozymes were fractionated by 10% denaturing (8 M urea) polyacrylamide gel electrophoresis (PAGE; 19:1 ratio of acrylamide to bisacrylamide) in buffer containing 45 mM Tris-borate pH 7.5 and 1 mM EDTA. The reaction products were visualized by UV shadowing and the bands corresponding to the ribozymes cut out. The transcripts were eluted from the gel slices overnight at room temperature in a solution containing 0.1% SDS, 10 mM EDTA and 0.5 M ammonium acetate. The purified RNA was precipitated by the addition of 0.1 volume of 2 M sodium acetate pH 4.5 and 2.5 volumes of ethanol. After washing in 70% ethanol and drying, the pellets were resuspended in ultrapure water, and the quantity of RNA determined by spectrophotometry at 260 nm.

RNA Synthesis of PABPN1 Transcript

The full-length PABPN1 gene could not be transcribed efficiently. Various conditions and constructions were tested until a construction with a truncated 5′-end was successfully transcribed at a reasonable amount. The ‘standard’ transcription protocol described for the ribozymes was however inefficient for this gene. Transcripts were obtained only with the HiScribe™ T7 *in Vitro* Transcription Kit (New England Biolabs) where 50 pmol of the column-purified and precipitated DNA template was in presence of 20% betaine and the kit’s supplied buffer, High Molecular Weight solution, rNTPs and T7 RNA polymerase enzyme. The transcription reaction proceeded for 4h30 at 42°C and was stopped by the addition of 5 U of DNase (RNase free) and incubating at 37°C for 45 min. Followed by phenol-chloroform extraction and precipitation, the RNA was purified on 5% 8 M urea PAGE. Elution, precipitation and concentration determination were performed as stated above.

Generation of Immune WT PABPN1

Codon optimized-resistant PABPN1 cDNA sequence of human PABPN1 was generated by Geneart (Regensburg, Germany). cDNA sequences were modified to include a consensus Kozak sequence. Codon usage was optimized based on transfer RNA frequencies in humans. Most critically, all GTC triplets (which do not necessarily map to codons) were altered to be as close as possible to N*not*-TG (where N is any nucleotide) *and* at least 3 nucleotides on either side of the

original GTC were changed to reduce the hybridization efficiency of any hammerhead ribozyme targeting a GTC cut-site. Optimization of gene transcript alters transgene sequences yet retain the translated amino acid sequence, and potentially increase the amount of gene product. The full sequence of the Immune WT PABPN1 follows.

5′ GTG GCT GCT GCA GCC GCT GCC GCC GCA GCA GCT GGC GCT GCT GGC GGA AGA GGA TCT
GGA CCT GGA AGG CGG AGA CAC CTG GTG CCT GGC GCT GGC GGA GAA GCT GGC GAA GGC GCT
CCA GGC GGA GCT GGC GAT TAT GGC AAT GGC CTC GAA AGC GAA GAG CTG GAA CCT GAG GAA
CTG CTG CTG GAA CCC GAG CCT GAG CCC GAG CCA GAG GAA GAA CCC CCT AGA CCT AGA GCC
CCT CCT GGC GCT CCT GGA CCA GGA CCT GGA TCA GGC GCA CCT GGC AGC CAG GAA GAA GAA
GAG GAA CCT GGC CTG GTG GAA GGC GAT CCT GGC GAC GGC GCC ATT GAG GAC CCT GAG CTG
GAA GCC ATC AAG GCC AGA GTG CGC GAG ATG GAA GAG GAA GCC GAG AAG CTG AAA GAA CTG
CAG AAC GAG GTG GAA AAG CAA ATG AAC ATG AGC CCC CCA CCC GGA AAC GCC GGC CCT GTG
ATC ATG AGC ATC GAG GAA AAA ATG GAA GCC GAC GCC CGG TCC ATC TAC GTG GGC AAT GTG
GAT TAC GGC GCC ACC GCC GAA GAA CTG GAA GCC CAC TTT CAC GGC TGC GGC TCC GTG AAC
AGA GTG ACC ATC CTG TGC GAC AAG TTC AGC GGC CAC CCC AAG GGC TTC GCC TAC ATC GAG
TTC AGC GAC AAA GAA AGC GTG CGG ACC AGC CTG GCC CTG GAC GAG TCT CTG TTC CGG GGC
AGA CAG ATC AAA GTG ATC CCC AAG CGG ACC AAC AGA CCC GGC ATC AGC ACC ACC GAC AGA
GGC TTC CCC AGA GCC CGG TAC AGA GCC AGA ACC ACC AAC TAC AAC AGC AGC AGA AGC CGG
TTC TAC AGC GGC TTC AAC AGC CGG CCC AGA GGC AGA GTG TAC AGA GGA AGG GCC AGA GCC
ACA AGC TGG TAC AGC CCC TAC TGA 3′

RNA Labeling

Purified PABPN1 RNA (6 pmol) was 5′ dephosphorylated with 5 U antarctic phosphatase in 50 mM Bis-Tris-Propane-HCl pH 6.0, 1 mM $MgCl_2$, and 0.1 mM $ZnCl_2$ for 30 min at 37°C followed by heat inactivation 5 min at 65°C. The RNA was radioactively 5′-^{32}P-labeled using 10 U of T4 polynucleotide kinase (NEB) with 20 μCi of [γ-^{32}P]-ATP (3000 Ci/mmol; Perkin-Elmer) for 1 h at 37°C in 70 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$ and 5 mM DTT. The labeled RNA was gel purified on 5% 8 M urea PAGE, eluted, precipitated as stated above except for the RNA bands being identified by radioactivity rather than UV shadowing.

Kinetic Measurements of Ribozyme Cleavage

Trace amount of [5′-^{32}P]-radioactively labeled PABPN1 RNA and 10 nM of ribozyme (10 nM of each ribozymes when in combination) were separately pre-incubated 10 min at 37°C in the presence of cleavage buffer mimicking physiological conditions (50 mM Tris pH 7.5, 100 μM

$MgCl_2$, 100 mM NaCl and 25 mM KCl). The cleavage reaction was initiated by adding an equal volume of target to the ribozyme. Aliquots (1.5 μL) were removed at intervals and the reaction terminated by addition to a 5 μL chilled mixture containing 95% (v/v) formamide, 10 mM EDTA and electrophoresis dyes. Ribozyme and cleavage products were separated by electrophoresis in a dual concentration 8 M urea 19:1 polyacrylamide gel with a 4% top two-third and 20% bottom third. Gels were exposed to storage phosphor screens, and quantified by phosphorimaging. Progress curves were fitted by nonlinear regression analysis to single exponential function.

Construction of Plasmids for Ribozyme Expression in Mammalian Cells

The same four hhRz against PABPN1 were used, namely Yz144, Yz363, Yz437, and Yz867. Templates containing these hammerhead catalytic sequences were synthesized chemically.

PCR products were digested with KpnI and Csp45I restriction endonucleases and were ligated with pUC-KE-tRNA-CTE plasmids that had been digested with both KpnI and Csp45I obtained kindly from Nawrot [44]. The DNA sequence of every plasmid was verified using Sanger sequencing with the PCR primer P7 (5'-CGCCAGGGTTTTCCCAGTCACGAC-3').

Cell Culture and Transfection I

Human transformed primary embryonal kidney cell lines HEK293 were cultured in Dulbecco's modification of Eagle's medium (Invitrogen), supplemented with 10% foetal bovine serum, G418 (200 μg/ml) was used as selection antibiotic for ribozyme-transfected HEK293 cells. Cells were grown in monolayer, at 37 ^{0}C in cell culture incubator with 5% CO_2 in 6-well plates and transfected at 70-80% confluence. Transfection with 2 μg of the ribozyme containing plasmid was carried out using the Jet prime reagent (Polyplus-transfection Inc., France) according to the manufacturer's protocol. One week later, selected G418 cells were collected for RNA and protein extractions in parallel.

Cell Culture and Transfection II (for immune WT Experiments)

Human transformed primary embryonal kidney cell lines HEK293 and mouse C2C12 muscle cell lines were cultured in Dulbecco's modification of Eagle's medium (Invitrogen), supplemented with 10% foetal bovine serum, G418 (200 ug/ml) was used as the selection antibiotic for ribozymes transfected HEK293 cells. Cells were grown in monolayer, at 37 ^{0}C in cell culture incubator of 5% CO_2 in 6-well plates and transfected at 70-80% confluence. Transfection with 2 ug of the ribozyme

containing plasmid was carried out using the Jet prime reagent (Polyplus-transfection Inc., France) according to the manufacturer's protocol. One week later, selected G418 resistant cells were collected for RNA and protein extractions in parallel. To assess the impact of codon optimization on PABPN1 expression, plasmids expressing non-codon-optimized human PABPN1, codon-optimized resistant human PABPN1, or codon-optimized resistant human PABPN1 with ribozyme were transfected into HEK293T cells.

Western Blotting I

Total proteins were extracted in SUB buffer (containing 8M urea, 2% mercaptoethanol and 0.5% SDS) at different time points, and electrophoresed in 12% SDS-PAGE gels. Proteins were blotted into nitrocellulose membrane and probed with specific antibodies against PABPN1 (1:2000) (Abcam, USA). Parallel samples were probed using monoclonal actin antibody (Chemicon International, USA) to confirm the equal loading of lysates between lanes. After incubation with specific secondary HRP-conjugated antibodies, the membranes were revealed using the western blot chemiluminescence reagent plus kit (NEN Life Sciences Products, Boston, MA, USA).

Quantification of the band intensity was performed using the ImageJ software. Values were normalized by β-actin. Quantification of PABPN1 protein level was determined by western blot using the ImageJ software, normalized to actin and presented as fold change relative to the level of PABPN1 in untransfected HEK293T cells.

Western Blotting II (for Immune WT Experiments)

Total proteins were collected 5 days after transfection and selection with G418. Total proteins were extracted in SUB buffer at different time points (containing 8M urea, 2% mercaptoethanol and 0.5% SDS), and electrophoresed in 12% SDS-PAGE gels. Proteins were blotted into nitrocellulose membrane and probed with specific antibodies against PABPN1 (1:2000) (Abcam, USA). Parallel samples were probed using monoclonal actin antibody (1:10000) (Chemicon International, USA) to confirm the equal loading of lysates between lanes. After incubation with specific secondary HRP-conjugated antibodies, the membranes were revealed using the western blot chemiluminescence reagent plus kit (NEN Life Sciences Products, Boston, MA, USA).

EXPERIMENTAL RESULTS

Single Ribozyme Cleavage and Additive Effect of Combinations

The *in vitro* efficiency of cleavage by the designed hammerhead ribozymes of the PABPN1 transcript is shown in figure 13. The experiment measured the efficiency of cleavage of all single, double, triple and quadruple combinations of four ribozymes. We tested the efficiency of the ribozymes in conditions similar to those that are the norm within mammalian cells (37°C, 100 mM Na^+ and 100 uM Mg^{2+}). The only exception being the very first lane in figure 13, where a combination of all four ribozymes was used at zero magnesium ion concentration.

The information shown in figure 13 show the RNA after 40 minutes from the initiation of the cleavage reaction. The efficiency of cleavage, represented as a percentage of the original un-cleaved transcript, is shown at the bottom of the figure. With the exception of the first lane, the results are arranged, from left to right to reflect an increasing number of participating ribozymes. Naturally, whatever remains of the original transcript is represented by the bands at the top of the figure, while the various fragments resulting from cleavage are spread all over the gel below. The expected lengths of the transcript fragments are listed to the left of the gel, while the actual lengths of the detected fragments are listed on the right at the proper location.

The results indicate several important facts: (a) all single ribozymes cleave the transcript, with an efficiency of cleavage ranging between 39.3% and 81.5%. (b) A greater number of ribozymes invariably cleaves more efficiently than a smaller number of ribozymes; the average cleavage percentage for 1, 2, 3 and 4 ribozymes is: 66.15%, 87.65%, 95.575%, and 97.7%. (c) Oddly enough, a combination of all four ribozymes in the absence of magnesium ions also cleaved very efficiently (96.6%). Magnesium is thought to be essential for efficient cleavage by hammerhead ribozymes. (d) The greatest increase in the efficiency of cleavage resulted from using two instead of one ribozyme.

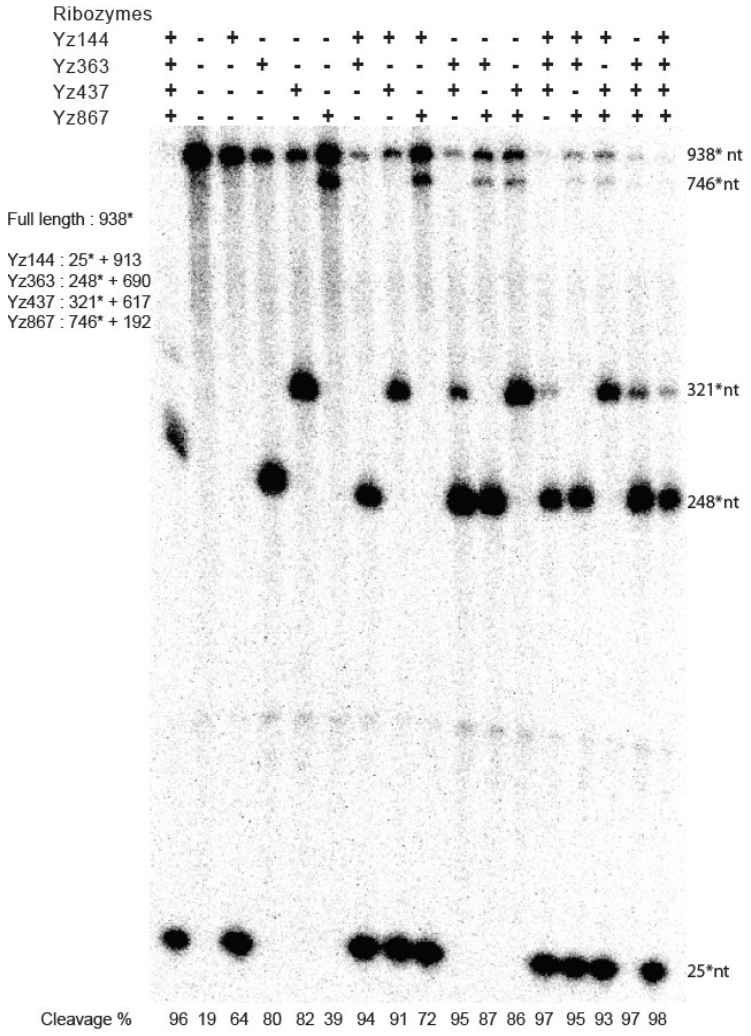


Fig. 13. Single and combined ribozyme cleavage. The 938 bases substrate labeled in 5′ was incubated 40 min. at 37°C, with 100 nM of the indicated ribozymes, in 50 mM Tris, pH 7.5, 100 mM NaCl and 100 μM $MgCl_2$, except for lane 1, which was in 0 $MgCl_2$. Reactions were loaded on a 4% (top)/ 20% (bottom) denaturing PAGE. Numbers with asterisks correspond to the labeled fragment.

This result confirms that the design of the ribozymes, amongst many generated by RiboSoft, were correct, in that they led to functional ribozymes. However, this does not prove their effectiveness *in vivo*, as many factors play a role in that environment, including known and unknown RNA binding proteins, known and unknown RNA cleaving enzymes, and the rates of production and breakdown of both the ribozymes and target mRNAs.

Single & Combination Ribozyme Kinetic Reaction Results

The progress of the cleavage reaction over time, for all combinations of the four ribozymes, is presented in figure 14. The purpose of this data is to see and mathematically model the rate of the cleavage reaction. This allows us to identify the time it takes the reaction to reach its steady-state. Also, the plots provide a more complete picture of the dynamic relationship between the number of ribozymes (participating in a reaction) and the change of the rate of cleavage over time.

The gel images in figure 14 have nine columns, each representing a different time point in the ribozyme cleavage reaction and, hence, the varying concentrations of the un-cleaved transcript and the cleaved fragments as the reaction continues. A greater degree of cleavage is indicated by a lighter band at the top of a column. The time steps are: 0s, 20s, 40s, 1m, 2m, 5m, 10m, 30m and 1h for columns 1-9, from left to right. The efficiency of cleavage at those time steps can be seen in corresponding figure 14(E). This figure also provides the constants of the cleavage rate (k_{obs}). In 14(A), exactly one ribozyme was used in this reaction (Yz144), while in 14(B, C and D) two, three and four ribozymes were used, respectively.

It can be seen from the plots that both the rate of the cleavage reaction and the steady-state value of cleavage efficiency have increased due to the presence of an additional ribozyme. The two-ribozyme combination was able to cleave about 43% of the transcript in 10 minutes, while the single ribozyme cleaved about 15% in the same time period. Also, the two ribozyme combination had a steady-state cleavage efficiency of ~90% a clear improvement over the steady-state of the single ribozyme of ~ 30%. Similarly, the combination of three ribozymes increased further the cleavage efficiency of the RNA substrate. Indeed, only five minutes was required to cleave 44% of the target RNA, half as much time as with two ribozymes, which achieved 43% cleavage in ten minutes. In spite of this, three ribozymes presented only a minor improvement over the already highly efficient combination of two ribozymes, 90% cleavage. No measurable improvement was obtained with four ribozymes compared to three.

It is recommended that one uses two proven hammerhead ribozymes in gene silencing applications; three or more hhRz is not advisable, as that increases the cost of construction and delivery without considerably enhancing the steady-state value of cleavage efficiency. Also, expressing many ribozymes increases the likelihood of unwanted side-effects such as off-target hits and toxicity for *in vivo* applications.

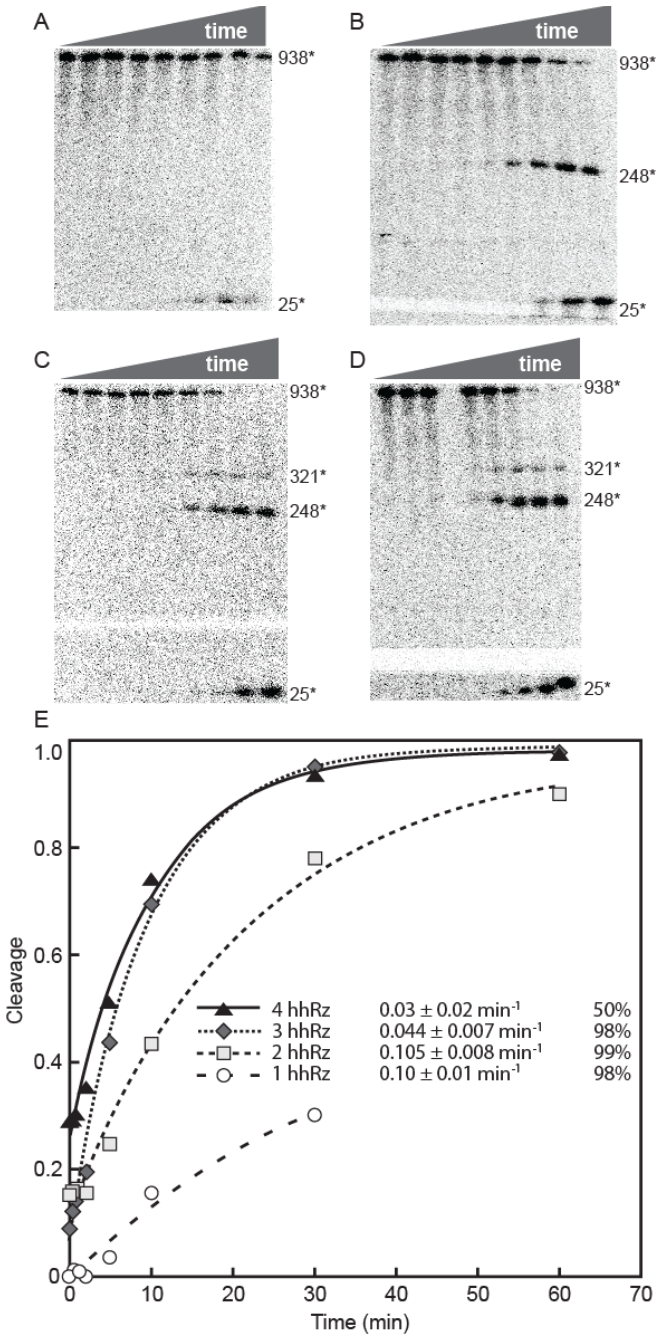


Fig. 14. Single & combined ribozyme kinetic. The rate of cleavage of 1, 2, 3 and 4 combined ribozymes is illustrated. (A) Gel of RNA substrate cleaved by a single ribozyme (Yz144), (B) by two ribozymes (Yz144 and Yz363), (C) three ribozymes (Yz144, Yz363, Yz437) and (D) with all ribozymes. In all cases, 9 time points are taken, from 0s to 1h. The size of observable cleaved products is indicated as in Fig. 13. (E) Cleavage was quantified and plotted to determine rate constants, which are indicated in the graph along with final percentage of cleavage.

Specific hhRzs are Able to Inhibit PABPN1 Expression Level

To test the hhRzs silencing effect, we used the human cell line HEK293T and transfected the cells with all different hhRzs individually. Stable clones of hhRzs were maintained in G418 selective media all the time during manipulations. We then performed Western blot on cell extracts to test the effectiveness of hhRzs in knocking down the endogenous PABPN1 level in this cell line. Western blot results revealed hhRzs sequences: Yz144, and Yz437 as optimal candidates, as they showed high efficiency in knock-down PABPN1 (Figure 15A). We then transfected combinations of two hhRzs simultaneously in HEK293T cells and performed Western blot to test if they are more effective in knocking down PABPN1 protein level. A combination of two hhRzs showed more reduction in PABPN1 protein level (Figure 15B) than individual hhRzs (Figure 15A), reducing expression down to 5% for the combinations of Yz363 with either Yz437 or Yz144.

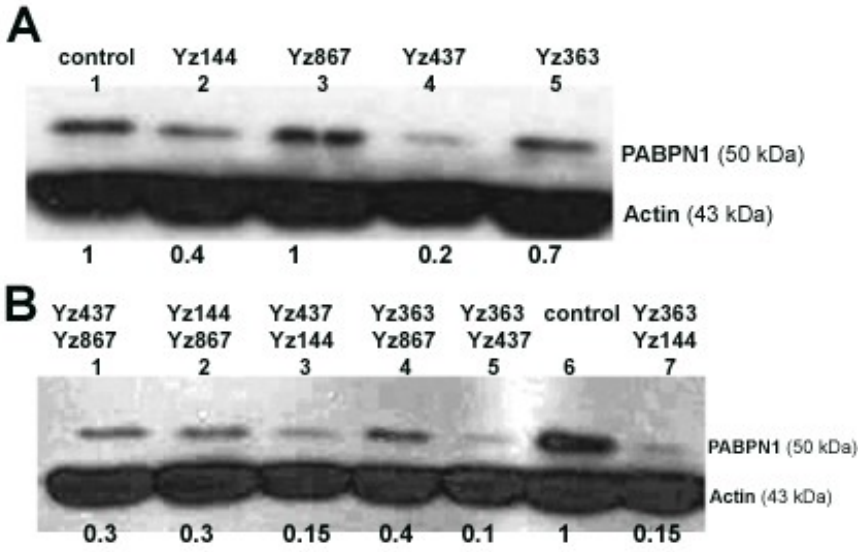


Figure 15. Ribozymes are effective in knocking down PABPN1. (A) Different individual ribozymes targeting PABPN1 transcript are able to knock down the level of PABPN1 protein. The endogenous expression of PABPN1 was analyzed by Western blot. HEK 293T cells were transfected with the corresponding individual ribozymes, and selected with G418 for 5 days before protein extraction. (B) Combination of two ribozymes targeting PABPN1 transcript are more efficient in knocking down the level of PABPN1 protein. The endogenous expression of PABPN1 was analyzed by Western blot. HEK 293 cells were transfected with the combined two indicated hhRzs. For (A) and (B), the "control" lanes correspond to plasmid without ribozymes. The values below the gel indicate PABPN1 protein signal intensities (quantified using ImageJ) after normalization to the control signal intensities (lane 1 in 15A and lane 6 in 15B).

Immune WT PABPN1 Expression Level is Resistant to Knockdown by hhRzs

Different individual ribozymes targeting PABPN1 transcript are able to knock down the level of PABPN1 protein (lane 1-4). Combination of two ribozymes targeting PABPN1 transcript are more efficient in knocking down the level of PABPN1 protein (lane 5). Immune PABPN1 increases protein level of PABPN1 (lane 8). Immune PABPN1 is resistant to knock down by ribozyme (lane 9-10). The control lane (7) corresponds to plasmid without immune PABPN1. NT (lane 6) refers to non-transfected cells. The endogenous expression of PABPN1 was analyzed by Western blot. HEK 293T cells were transfected with the corresponding individual ribozymes, or the combined indicated ribozymes, and selected with G418 for 5 days before protein extraction. Actin antibody was used to correct for differences in protein loading. The results are shown in Figure 16.

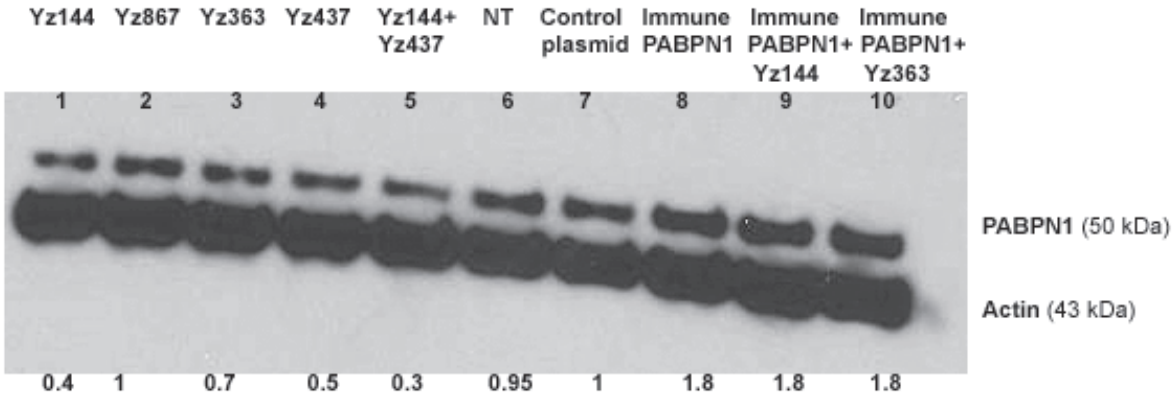


Figure 16. Except for Yz867, every one of the individual hhRz (lanes 1-4) as well as the binary combination of lane 5 was able to knockdown the expression of PABPN1. NT (of lane 6) stands for cells that are not transfected by either ribozymes or the Immune WT. In lane 7, the cells were transfected by the same plasmid used to express the Immune WT in lanes 8-10, except that the control plasmid expressed nothing. Critically, lane 8 shows a clear increase in the level of expression of PABPN1, while lanes 9 & 10 show no noticeable decrease in the level of expression of PANBPN1 (relative to lane 8), due to the potential activity of the co-transfected hammerhead ribozymes. The values below the gel indicate PABPN1 protein signal intensities (quantified using ImageJ) after normalization to the control plasmid signal intensity.

Chapter 4: Conclusion & Future Work

This thesis presents two contributions to the world of ribozyme design and ribozyme application. Chapter 2 describes the design of an algorithm (implemented as a web service, RiboSoft) for automated hammerhead ribozyme design. It can be used to design hammerhead ribozymes to target any RNA sequence, including the transcript of any gene. The program has clear and valuable advantages over existing hammerhead design services, related to usability, ribozyme quality assessment, and multi-objective optimization. It unifies in a single tool the functionalities of targeting ribozyme design softwares that have been done in the last decade: it finds target sites, computes target cut-site accessibility, ribozyme structure validity and potential off-target effects. More importantly for the future community of users, it provides a user friendly interface with the flexibility required for any application: from *in vitro* assays, to gene targeting in *E. coli*, in humans or in any desired species and summarizes all the complexity of the analysis in one simple ranking for the user to choose from.

The other contribution is the automated design of four hhRz targeting the transcript of the PABPN1 gene, and the testing these ribozymes, *in vitro* and *in vivo*, for their cleavage efficiency. A mutated PABPN1 is the root cause of OPMD (oculopharyngeal muscular dystrophy) in humans, a late onset muscle disease associated with progressive ptosis of the eyelids, and dysphagia. The experiments in chapter 3 validated the designs as they demonstrated cleavage by each of the four ribozymes. In spite of using single turnover conditions for in vitro assays, the results in vivo show that the ribozymes designed by RiboSoft lead to an efficient knockdown of PABPN1. This is likely due to the use of a strong promoter for ribozyme expression and higher stability of the structured ribozyme RNAs, as compared to the mRNA, which generally have low expression as compared to most non-coding RNAs. Overall, according to our results, we recommend using two hhRzs for researchers interested in targeting a gene for silencing.

Any upcoming gene targeting projects can benefit from Ribosoft, which will easily provide multiple ribozyme designs against any transcript. This fact will allow anyone to select multiple ribozymes in order to attain the best effect possible. Indeed, as we have demonstrated, a dramatic

additive effect is obtained by combining two ribozymes, and in the few cases where it would be insufficient, three ribozymes are likely to be enough, making Ribosoft all the more valuable.

FUTURE WORK

There are two ways in which this work can be developed. One direction is that of expanding the computational design tool to include not just hammerhead ribozymes, as it does now, but other ribozymes, such as hairpin and HDV ribozymes. These ribozymes share the basic architecture of a hammerhead ribozyme, which is made of hybridizing arms responsible for the selective capability of the ribozyme and a largely conserved core. As such, many ribozymes can, in principle, be optimized using the same overall multi-objective approach employed by RiboSoft, which essentially takes into account the (a) structure of the ribozyme and exact sequence of its recognition arms; (b) structure of the region of the targeted cut-site and its exact sequence; (c) off-target effects, for in vivo applications. We cannot see any conceptual or technical impediments to the customization of the general approach to the various ribozymes whose functionality is a function of these three (a-c) sets of considerations.

Another direction for development involves the modification of the design of the resulting hammerhead ribozymes to test their gene silencing functionality, as well as unwanted effects, in cell models and animal models. We have already started the next phase of this work, which involves the measurement of the PABPN1 gene silencing efficiency, within human cells, of the same set of ribozymes presented in this thesis. However, the suppression of the PABPN1 gene using these ribozymes (or any other means) does not represent a complete therapeutic approach to OPMD. These ribozymes indiscriminately cleave both the wild type as well as the mutant (disease-causing) transcripts of PABPN1. And, the PABPN1 protein is required for efficient polymerization of the poly(A) tails of the 3'-ends of transcripts of eukaryotic genes and controls the size of these tails. An obvious solution to this problem is the provisioning of a PABPN1 gene, one that generates the same protein as the WT, but via a transcript that is immune to the catalytic effects of hammerhead ribozymes. A hhRz-ImmuneWT PABPN1 gene was in fact designed and tested. The initial results (presented at the end of chapter 3) prove the basic correctness of this approach.

In future work, we shall present the full test results of the effect of hammerhead ribozymes on both a mutant PABPN1 and the ImmuneWT PABPN1 in the human cells of OPMD afflicted individuals.

References

[1] T. Port and A. Cepaitis. (2015, September 10). Science Prof Online [Online]. Available: http://www.scienceprofonline.com/index.html
[2] N. G. Walter and D. R. Engelke, "Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs.," *Biologist (London).*, vol. 49, no. 5, pp. 199–203, 2002.
[3] A. Serganov and D. J. Patel, "Ribozymes, riboswitches and beyond: regulation of gene expression without proteins.," *Nat. Rev. Genet.*, vol. 8, no. 10, pp. 776–90, Oct. 2007.
[4] S. Hagiwara, K. Nakamura, H. Hamada, K. Sasaki, Y. Ito, K. Kuribayashi, T. Sato, Y. Sato, M. Takahashi, K. Kogawa, J. Kato, T. Terui, T. Takayama, T. Matsunaga, K. Taira, and Y. Niitsu, "Inhibition of type I procollagen production by tRNAVal CTE-HSP47 ribozyme," *J. Gene Med.*, vol. 5, no. 9, pp. 784–794, 2003.
[5] J. O. Ojwang, a Hampel, D. J. Looney, F. Wong-Staal, and J. Rappaport, "Inhibition of human immunodeficiency virus type 1 expression by a hairpin ribozyme.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 22, pp. 10802–10806, 1992.
[6] T. Perrotta and M. D. Been, "Cleavage of oligoribonucleotides by a ribozyme derived from the hepatitis delta virus RNA sequence.," *Biochemistry*, vol. 31, no. 1, pp. 16–21, 1992.
[7] A. Roth, Z. Weinberg, A. G. Y. Chen, P. B. Kim, T. D. Ames, and R. R. Breaker, "A widespread self-cleaving ribozyme class is revealed by bioinformatics.," *Nat. Chem. Biol.*, vol. 10, no. 1, pp. 56–60, 2014.
[8] J. Barciszewski, "RNA Technologies and Their Applications," *Gene Expr.*, pp. 393–418, 2010.
[9] L. Citti and G. Rainaldi, "Synthetic hammerhead ribozymes as therapeutic tools to control disease genes.," *Curr. Gene Ther.*, vol. 5, no. 1, pp. 11–24, Feb. 2005. [10] Murray, J.B., Seyhan, A.A., Walter, N.G., Burke, J.M. and Scott, W.G. (1998) The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem. Biol.*, **5**, 587–595.
[11] A. Khvorova, A. Lescoute, E. Westhof, and S. D. Jayasena, "Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity.," *Nat. Struct. Biol.*, vol. 10, no. 9, pp. 708–12, Sep. 2003.
[12] V. Saksmerprome, M. Roychowdhury-Saha, S. Jayasena, A. Khvorova, and D. H. Burke, "Artificial tertiary motifs stabilize trans-cleaving hammerhead ribozymes under conditions of submillimolar divalent ions and high temperatures.," *RNA*, vol. 10, no. 12, pp. 1916–1924, 2004.
[13] N. Agrawal, P. V. N. Dasaradhi, A. Mohmmed, P. Malhotra, R. K. Bhatnagar, and S. K. Mukherjee, "RNA Interference : Biology , Mechanism , and Applications," vol. 67, no. 4, pp. 657–685, 2003.
[14] T. Ohmichi and E. T. Kool, "The virtues of self-binding: high sequence specificity for RNA cleavage by self-processed hammerhead ribozymes.," *Nucleic Acids Res.*, vol. 28, no. 3, pp. 776–83, Feb. 2000.
[15] S. Koseki, T. Tanabe, K. Tani, S. Asano, T. Shioda, Y. Nagai, T. Shimada, J. Ohkawa, and K. Taira, "Factors governing the activity in vivo of ribozymes transcribed by RNA polymerase III.," *J. Virol.*, vol. 73, no. 3, pp. 1868–1877, 1999.

[16] M. Warashina, T. Kuwabara, Y. Kato, M. Sano, and K. Taira, “RNA-protein hybrid ribozymes that efficiently cleave any mRNA independently of the structure of the target RNA.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 10, pp. 5572–5577, 2001.
[17] R. Lewis. (2015, November, 20). PLOS Blogs [Online]. Available: http://blogs.plos.org/dnascience/2014/12/18/use-genetic-code-passwords/
[18] V. P. Mauro and S. a. Chappell, “A critical analysis of codon optimization in human therapeutics,” *Trends Mol. Med.*, vol. 20, no. 11, pp. 604–613, 2014.
[19] G. Kudla, A. W. Murray, D. Tollervey, and J. B. Plotkin, “Coding-sequence determinants of gene expression in Escherichia coli.,” *Science*, vol. 324, no. 5924, pp. 255–258, Apr. 2009.
[20] L. Qiaozhen, Z. Xiaoyang, T. McIntosh, H. Davis, J. F. Nemeth, C. Pendley, S. L. Wu, and W. S. Hancock, “Development of different analysis platforms with LC-MS for pharmacokinetic studies of protein drugs,” *Analytical Chemistry*, vol. 81, no. 21. pp. 8715–8723, 2009.
[21] J. Yang, W. Pan, G. a Clawson, and A. Richmond, “NIH Public Access,” vol. 67, no. 7, pp. 3127–3134, 2009.
[22] A. Qureshi, R. Zheng, T. Parlett, X. Shi, P. Balaraman, S. Cheloufi, B. Murphy, C. Guntermann, and P. Eagles, “Gene silencing of HIV chemokine receptors using ribozymes and single-stranded antisense RNA.,” *Biochem. J.*, vol. 394, no. Pt 2, pp. 511–518, 2006.
[23] L. Tedeschi, C. Lande, A. Cecchettini, and L. Citti, “Hammerhead ribozymes in therapeutic target discovery and validation.,” *Drug Discov. Today*, vol. 14, no. 15–16, pp. 776–83, Aug. 2009.
[24] B. Brais, J. P. Bouchard, Y. G. Xie, D. L. Rochefort, N. Chrétien, F. M. Tomé, R. G. Lafrenière, J. M. Rommens, E. Uyama, O. Nohira, S. Blumen, A. D. Korczyn, P. Heutink, J. Mathieu, A. Duranceau, F. Codère, M. Fardeau, and G. A. Rouleau, “Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy.,” *Nat. Genet.*, 1998.
[25] B. Brais, “Oculopharyngeal muscular dystrophy: a polyalanine myopathy.,” *Curr. Neurol. Neurosci. Rep.*, vol. 9, no. 1, pp. 76–82, Jan. 2009.
[26] J.-F. Lucier, L. J. Bergeron, F. P. Brière, R. Ouellette, S. A. Elela, and J.-P. Perreault, “RiboSubstrates: a web application addressing the cleavage specificities of ribozymes in designated genomes.,” *BMC Bioinformatics*, vol. 7, p. 480, Jan. 2006.
[27] A. Mercatanti, C. Lande and L. Citti. (2015, September 10). Aladin [Online]. Available: http://aladdin.ifc.cnr.it/aladdin.
[28] A. Mercatanti, G. Rainaldi, , L. Mariani, , R. Marangoni, and L. Citti, “A method for prediction of accessible sites on an mRNA sequence for target selection of hammerhead ribozymes.,” *Journal of Computational Biology*, vol. 9, pp. 641-653, July 2004.
[29] N. R. Markham and M. Zuker, “UNAFold: Software for nucleic acid folding and hybridization,” *Methods Mol. Biol.*, vol. 453, pp. 3–31, 2008.
[30] Y. Shao, S. Wu, C. Y. Chan, J. R. Klapper, E. Schneider, and Y. Ding, “A structural analysis of in vitro catalytic activities of hammerhead ribozymes.,” *BMC Bioinformatics*, vol. 8, p. 469, Jan. 2007.
[31] A.L. Edwards and R.T. Batey, “Riboswitches: A Common RNA Regulatory Element.,” *Nature Education,* vol. 3, p. 9, 2010.
[32] E. R. Lee, J. L. Baker, Z. Weinberg, N. Sudarsan and R. R. Breaker, “An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger.,” *Science*, vol. 329, pp. 845-848, August 2010.
[33] H. Gu, K. Furukawa, and R. R. Breaker, “Engineered allosteric ribozymes that sense the bacterial second messenger cyclic diguanosyl 5’-monophosphate.,” *Anal. Chem.*, vol. 84, no. 11, pp. 4935–41, Jun. 2012.

[34] M. N. Win and C. D. Smolke, “Higher-order cellular information processing with synthetic RNA devices.,” *Science*, vol. 322, no. 5900, pp. 456–60, Oct. 2008.
[35] D. Ellington and J. W. Szostak, “In vitro selection of RNA molecules that bind specific ligands.,” *Nature*, vol. 346, no. 6287, pp. 818–822, 1990.
[36] Penchovsky and R. R. Breaker, “Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes.,” *Nat. Biotechnol.*, vol. 23, no. 11, pp. 1424–33, Nov. 2005.
[37] M. Zoumadakis and M. Tabler, “Comparative analysis of cleavage rates after systematic permutation of the NUX consensus target motif for hammerhead ribozymes.,” *Nucleic Acids Res.*, vol. 23, no. 7, pp. 1192–6, Apr. 1995.
[38] T. Shimayama, S. Nishikawa, and K. Taira, “Generality of the NUX rule: kinetic analysis of the results of systematic mutations in the trinucleotide at the cleavage site of hammerhead ribozymes,” *Biochemistry*, vol. 34, no. 11, pp. 3649–3654, 1995.
[39] A. Panjkovich and F. Melo, “Comparison of different melting temperature calculation methods for short DNA sequences.,” *Bioinformatics*, vol. 21, no. 6, pp. 711–22, Mar. 2005.
[40] http://www.promega.com/techserv/tools/biomath/calc11.htm, last acceded July 4th, 2015.
[41] Y. Ding, C. Y. Chan, and C. E. Lawrence, “Sfold web server for statistical folding and rational design of nucleic acids.,” *Nucleic Acids Res.*, vol. 32, no. Web Server issue, pp. W135–41, Jul. 2004.
[42] K. Deb, K. *Multi-objective optimization using evolutionary algorithms*. New York, NY: John Wiley & Sons, 2001.
[43] J. F. Milligan, D. R. Groebe, G. W. Whherell, and O. C. Uhlenbeck, “Nucleic Acids Research © IR L Pnss Limited , Oxford , England . Nucleic Acids Research,” *Nucleic Acids Res.*, vol. 15, no. 21, pp. 8783–8798, 1987.
[44] B. Nawrot, S. Antoszczyk, M. Maszewska, T. Kuwabara, M. Warashina, K. Taira, and W. J. Stec, “Efficient inhibition of beta-secretase gene expression in HEK293 cells by tRNAVal-driven and CTE-helicase associated hammerhead ribozymes,” *Eur. J. Biochem.*, vol. 270, no. 19, pp. 3962–3970, Sep. 2003.