Examination of the structural and functional versatility of \textit{glmS} ribozymes by using \textit{in vitro} selection

Kristian H. Link\textsuperscript{1,2}, Lixia Guo\textsuperscript{1} and Ronald R. Breaker\textsuperscript{1,2,3,*}

\textsuperscript{1}Department of Molecular, Cellular and Developmental Biology and \textsuperscript{2}Howard Hughes Medical Institute and \textsuperscript{3}Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

Received April 28, 2006; Revised August 2, 2006; Accepted August 3, 2006

ABSTRACT

Self-cleaving ribozymes associated with the \textit{glmS} genes of many Gram-positive bacteria are activated by binding to glucosamine-6-phosphate (GlcN6P). Representatives of the \textit{glmS} ribozyme class function as metabolite-sensing riboswitches whose self-cleavage activities down-regulate the expression of GlmS enzymes that synthesizes GlcN6P. As with other riboswitches, natural \textit{glmS} ribozyme isolates are highly specific for their target metabolite. Other small molecules closely related to GlcN6P, such as glucose-6-phosphate, cannot activate self-cleavage. We applied \textit{in vitro} selection methods in an attempt to identify variants of a \textit{Bacillus cereus} \textit{glmS} ribozyme that expand the range of compounds that induce self-cleavage. In addition, we sought to increase the number of variant ribozymes of this class to further examine the proposed secondary structure model. Although numerous variant ribozymes were obtained that efficiently self-cleave, none exhibited changes in target specificity. These findings are consistent with the hypothesis that GlcN6P is used by the ribozyme as a coenzyme for RNA cleavage, rather than an allosteric effector.

INTRODUCTION

Riboswitches are structured RNA elements that most commonly reside within the 5'-untranslated regions (UTRs) of bacterial mRNAs where they control gene expression via direct binding of small molecules (1–3). A variety of natural ligands or effector molecules including coenzymes, amino acids and nucleobases are selectively sensed by natural aptamer domains that distinguish each riboswitch class (4). Each aptamer is usually closely associated with an expression platform (1,5), which interfaces with factors that are responsible for expressing genes. Ligand binding typically stabilizes one structural state of a riboswitch aptamer, and this alters the folding pathway taken by the expression platform located immediately downstream of (or sometimes overlapping) the aptamer.

The vast majority of riboswitches in bacteria appear to control gene expression by modulating transcription termination or translation initiation. Computer-aided nucleotide sequence analyses (6–11) reveal that many riboswitch aptamers reside upstream of a strong hairpin loop followed by a short repeat of U residues. This architecture conforms to the structure of known intrinsic transcription terminators (12,13), and some riboswitches have been experimentally demonstrated to modulate transcription termination upon ligand binding \textit{in vitro} (5,14–21). Likewise, bioinformatic studies (6–9) and biochemical studies (22,23) have revealed that riboswitches also use ligand-dependent differential base pairing to control accessibility of ribosome binding sites within some mRNAs.

As with most other riboswitches, \textit{glmS} ribozymes (24,25) (Figure 1A) are located in mRNAs that encode proteins involved in the metabolism or transport of the compound that is being detected. The protein encoded by \textit{glmS} mRNAs, \textit{L-glutamine:fructose 6-phosphate amidotransferase}, is an enzyme that produces GlcN6P (Figure 1B), which is selectively recognized by \textit{glmS} ribozymes from \textit{Bacillus subtilis} (24) and \textit{Bacillus cereus} (26,27). GlcN6P production is the first step in the biosynthesis of UDP-GlcNAc. This nucleotide-modified glucosamine compound is used in bacterial cell wall biosynthesis and is essential for bacterial cell survival (28).

Several characteristics of \textit{glmS} ribozymes are unique among riboswitches, which suggest they control gene expression using a distinct regulatory mechanism. The most prominent difference is that \textit{glmS} ribozymes undergo efficient site-specific self-cleavage in the presence of GlcN6P. When saturated with GlcN6P, the rate constant for self-cleavage is \(\text{~3 min}^{-1} \) (24,27). This corresponds to a rate acceleration of at least 10-million-fold over the uncatalyzed rate for RNA transesterification (29). Self-cleavage occurs via a phosphoester transfer reaction in which the 2'-oxygen of the labile linkage functions as a nucleophile and attacks the phosphate at the cleavage site. The products of the cleavage reaction possess a 5'-hydroxyl group and a 2',3'-cyclic phosphate. The cleavage site resides within the 5'-UTR of each \textit{glmS} mRNA, and therefore the riboswitch does not directly deactivate translation by disrupting the continuity of the open reading frame (ORF).
Although the precise mechanism for gene control has yet to be established, ribozyme activity is known to be essential for gene repression (24).

The results of several previous studies indicate that the glmS ribozyme forms a series of base-paired structures and unpaired but conserved nucleotides that are important for ribozyme folding and catalysis (24,25,27,30). These results also suggest that the ribozyme is highly selective for GlcN6P (24,26,31) and that this ligand might function as a coenzyme for RNA transesterification (26,27,31,32). One possibility is that the 2-amino group of GlcN6P (Figure 1B) is responsible for proton transfer during cleavage of the phosphodiester linkage. In the current study, we have used in vitro selection (33,34) to determine whether variant glmS ribozymes could be isolated that exhibit altered ligand specificity, and to further examine the structural model for this ribozyme class. We speculated that if the glmS ribozyme uses GlcN6P as a coenzyme, then it might not be possible to isolate close variants of the parent ribozyme that use GlcN6P analogs that lack a 2-amino group of the natural ligand.

**MATERIALS AND METHODS**

**Chemicals and oligonucleotides**

GlcN6P and various analogs were purchased from Sigma, except for GlcNol6P, α-MeOGlcN6P and β-MeOGlcN6P, which were synthesized as reported elsewhere (31). Synthetic DNAs were prepared by standard solid-phase methods (W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University) and purified as described previously (35). Mutagenized DNA templates (primer 1: 5′-TAATACGACTCACTATAGTAAATTATAGAAGCGCCAGAACTACAAGTGTAGTTGACGAGGTGGGGTTTATCGAGATTTCGGCGGATGGCTCCCGGTTG and primer 2: 5′-CCGTGCCTCTTCTCTCATTACACTTTCACCTTTGTCCACTAAGTCAGCTTAATGATTTAAGTAAAAGCTTGCGGTTGTGATGTACAACCGGGAGCCATCCGG) were synthesized with a degeneracy (36) of 0.09 per position as indicated by the underlined nucleotides. Primer 3 (5′-TAATACGACTCACTATAGTAAATTATAGAAGCGCCAGAACTACAAG), which contains a T7
promoter for in vitro transcription, and primer 4 (5'-CCGGTGCTACTTCGCTCT) were used for reverse transcription and polymerase chain reaction (RT–PCR) amplification of selected RNAs during each round of in vitro selection.

Double-stranded template DNAs used to prepare the initial population (generation zero or G0) of RNAs by in vitro transcription were generated by primer extension in a reaction containing 200 pmol each of primers 1 and 2, 200 μM of the four deoxyribonucleotide 5'-triphosphates (dNTPs), 50 mM Tris–HCl (pH 8.3 at 23°C), 75 mM KCl, 3 mM MgCl2, 10 μM DTT and 8 U/μl SuperScript™II Reverse Transcriptase (SSII RT, Invitrogen). RNAs were generated by in vitro transcription using T7 RNA polymerase (T7 RNAP). A typical in vitro transcription contained 80 pmol of template DNA and was incubated in the presence of 50 mM HEPES (pH 7.5 at 23°C), 15 mM MgCl2, 2 mM spermidine, 5 mM DTT, 2.5 mM of the four ribonucleotide 5'-triphosphates (NTPs), 20 μCi of [α-32P]UTP and 40 U/μl T7 RNAP. The resulting RNA products were purified by denaturing 6% polyacrylamide gel electrophoresis (PAGE). RNAs were recovered from each gel by crush/soaking with 10 mM HEPES (pH 7.5 at 23°C), 200 mM NaCl and 1 mM EDTA followed by precipitation with ethanol. RNAs were resuspended in deionized water and stored at −20°C until used. The G0 population was prepared using 200 pmol of template and 50 μCi of [α-32P]UTP.

In vitro selection

RNAs at each round of in vitro selection were incubated in the presence of 50 mM HEPES (pH 7.5 at 23°C), 10 mM MgCl2, 200 mM KCl and 200 μM each (or as otherwise noted) of GlcN6P, glucosamine-6-sulfate (GlcN6S), glucose-6-phosphate (Glc6P), glucosamine (GlcN), glucose (Glc), 2-amino-2-deoxy-D-glucitol-6-phosphate (GlcNol6P), α-O-methyl glucosamine-6-phosphate (α-MeGlcN6P) and β-O-methyl glucosamine-6-phosphate (β-MeGlcN6P). In vitro selection reactions were initiated by the addition of ligand and terminated after incubation for times defined for each reaction by the addition of 1 vol of an aqueous gel loading solution containing 16% (w/v) sucrose, 0.08% (w/v) sodium dodecyl sulfate, 0.04% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanol, 7.25 M urea, 72 mM Tris–HCl, 72 mM borate and 100 mM EDTA. Reaction products were separated by 6% denaturing PAGE, and were imaged and quantitated by using a PhosphorImager (Amersham Biosciences).

The 3′ cleavage products carrying the mutagenized region were recovered from each gel by elution as described above. Recovered sequences were amplified by RT–PCR. RT reactions were incubated for 1 h at 37°C and contained the 3′ cleavage product recovered from the selection reactions, 50 μM primer 2, 2.5 mM each of the four dNTPs, 50 mM Tris–HCl (pH 8.3 at 23°C), 75 mM KCl, 3 mM MgCl2, 10 μM DTT and 8 U/μl of SSII RT. PCR reactions contained 20 pmol each of primers 3 and 4, 200 μM each of the four dNTPs, 10 mM Tris–HCl (pH 8.3 at 23°C), 1.5 mM MgCl2, 40 mM KCl, 0.01% gelatin and 1 U/μl of Taq DNA polymerase. Reactions were subjected to thermocycling until near maximal PCR product was produced as determined by agarose gel electrophoresis and product detection by UV luminosity of ethidium bromide stained gels. PCR products were recovered from agarose gels by purification using QIAquick™ gel extraction kits (Qiagen). Individual variant DNAs were isolated by cloning (TA cloning kit™, Invitrogen) and analyzed by sequencing (W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University).

Ribozyme cleavage assays

Self-cleavage activities of wild-type B.cereus and variant glmS ribozymes were assessed by incubating trace amounts of internally 32P-labeled RNA under in vitro selection conditions as described above. Reactions were initiated by the addition of ligand(s) as noted for each assay, and incubated at 23°C. Reactions were quenched when aliquots were withdrawn and mixed with 1 vol of aqueous gel loading buffer as defined above.

Rate constants were determined by plotting the natural logarithm of the fraction of RNA remaining uncleaved versus time, and deriving the negative slope of the resulting line that corresponds to kobs. The data were established using incubation times that limited the extent of cleavage to <30%. The data were not adjusted for the amount of RNA remaining uncleaved after exhaustive incubation.

RESULTS AND DISCUSSION

Construct design and selection strategy

The conserved secondary structure of the glmS ribozyme from B.cereus contains eight base-paired elements (Figure 1A). These stems are distributed between the catalytic core domain of the ribozyme (P1, P2, P2.1 and P2.2) (24,37) and an accessory domain (P3, P3.1, P4 and P4.1) that permits the catalytic core to function when incubated with physiologically relevant concentrations of Mg2+ (27,30). Together these domains span 142 contiguous nucleotides, beginning 1 nt upstream of the cleavage site and ending at nt 141 (Figure 1A).

The starting pool of variant ribozymes was generated by introducing mutations at 122 of the 142 nt that encompass the glmS ribozyme from B.cereus (Figure 1A; see also Materials and Methods section). A degeneracy in nucleotide identity of 0.09 was introduced at each mutagenized position (36), to yield RNAs where these sites have a 91% chance of carrying the wild-type nucleotide and a 3% chance of carrying one of the other 3 nt. Approximately 200 pmol (1.2 × 1014 molecules) of DNA template was used to produce the starting RNA population (G0), providing complete coverage of all 6-error mutants. Subsequent populations included 10 pmol of RNA, each prepared by in vitro transcription in the absence of ligands. Full-length precursor RNAs were purified by denaturing 6% PAGE and subsequently incubated in selection buffer supplemented with 200 μM each of eight compounds (Figure 1B). The resulting 3′ cleavage products carrying the randomized ribozyme domains were purified by denaturing 6% PAGE and amplified by RT–PCR. One iteration of this selective-amplification process constitutes one round of in vitro selection.

Eight compounds (Figure 1B) were added to the initial in vitro selection reactions. Three of these compounds (GlcN6P, GlcN6S and GlcN) are known to activate the wild-type glmS ribozyme from B.cereus under conditions
used in this study for in vitro selection, although GlcN6S and GlcN likely are bound far less tightly than is GlcN6P (24,26,38). Since these three compounds activate the wild-type ribozyme, their inclusion in the in vitro selection reactions would allow us to isolate ribozyme variants that retained activity with the natural ligand. Moreover, any mutations that improved the affinity or activity of the ribozyme with the less effective ligands might be recovered. The remaining five ligands (Glc, Glc6P, GlcNol6P, α-MeO-GlcN6P and β-MeO-GlcN6P) do not exhibit significant activation of wild-type ribozyme self-cleavage even when present at a concentration of 1 mM (31). The addition of these compounds to the in vitro selection reactions should permit the isolation of ribozyme variants that could use these normally inactive compounds to promote cleavage if such variants were sufficiently close in sequence space to be present in the RNA population.

Ligand-binding properties of a glmS ribozyme resist alteration

The RNA population resulting from two rounds of in vitro selection (G2) was found to exhibit considerable activation of ribozyme function when a mixture of the eight compounds was present (Figure 2A). Systematic analysis of the effects of adding individual compounds revealed that the G3 RNA population mostly contained ribozymes that respond to GlcN6P (data not shown). The G3 population was split into two lineages (I and II) to permit separate attempts to optimize activity with GlcN6P and to determine if variant ribozymes could be isolated that broaden ligand specificity.

Lineage I was extended by an additional 13 rounds of in vitro selection using only GlcN6P as a ligand (Figure 2A). As the rounds progressed, the concentration of GlcN6P was decreased from 200 to 2 μM to favor isolation of ribozyme variants that react fast even when low concentrations of ligand are provided. The observed rate constant (k_{obs}) values for each population were established at the concentrations used for the corresponding in vitro selection reactions (Figure 2). A total of 10 of these additional rounds of in vitro selection were carried out in the presence of 2 μM GlcN6P. However, the k_{obs} values for the resulting populations never substantially exceeded the k_{obs} value exhibited by the wild-type ribozyme (0.015 min^{-1}) under the same assay conditions (Figure 2B). These results suggest that there is not an abundance of sequence variants of glmS ribozymes that can bind GlcN6P more tightly and retain robust RNA cleavage activity.

Similarly, lineage II was extended by an additional round of in vitro selection in the presence of 200 μM each of the seven GlcN6P analogs (Figure 3A population G3). We observed that the RNA population at G4 of lineage II exhibited activity only when presented with GlcN6S (data not shown). This population was split into two to further propagate lineage II and to create lineage III (see below).

An additional nine rounds of selection were conducted with lineage II (G4–G12) using only 200 μM GlcN6S as the ligand. As was observed with lineage I, RNA populations from each round exhibited k_{obs} values in the presence of 200 μM GlcN6S that were no greater than that observed for the wild-type ribozyme (k_{obs} ~0.02 min^{-1}) under the same conditions.

Figure 2. In vitro selection of glmS ribozyme variants that respond to GlcN6P. (A) Histogram representing the fraction of precursor RNAs cleaved in the absence (none) and presence of compounds indicated for RNAs isolated from lineage I. G0 through G15 represent the original random sequence RNA population through generation 15, respectively. Time values reflect the length of incubation for the ribozyme cleavage assay for the data sets identified. (B) Plot depicting the k_{obs} values for successive RNA populations in the presence of decreasing concentrations of GlcN6P or no ligand. Dashed line represents the k_{obs} value for the wild-type glmS ribozyme from B.cereus in the presence of 2 μM GlcN6P.
in the parent glmS RNA that bind ligand and promote catalysis will reside in the most highly conserved portion of the ribozyme. Furthermore, if the ribozymes isolated in this study do not exhibit altered kinetic or molecular recognition characteristics towards GlcN6P or its analogs, then we expect that variants comprising the in vitro selection populations should not carry mutations in the conserved core.

We investigated this hypothesis by cloning and sequencing individual ribozyme variants from G16 of lineage I and G12 of lineage II (Figure 4B; see also Supplementary Figure S1). Of 20 clones examined from lineage I, only one retains the parental B. cereus glmS ribozyme sequence. The remaining 19 individuals have an average of 4.4 mutations per molecule. One variant from this lineage (clone A10) was assayed and was found to exhibit essentially wild-type activity with GlcN6P (data not shown), which is consistent with the activity of the G16 population from lineage I.

None of the lineage I variants carries a mutation at any of the nucleotides conserved at ≈97%, and none of the mutations occurs at residues that are predicted (38) to be critical for binding the phosphate of the ligand. Furthermore, many of the mutations cluster in regions that form stems P3, P3.1, P4 and P4.1. This accessory domain allows the ribozyme to attain maximal activity in low Mg2+ concentrations (27), but is not essential for ligand-induced ribozyme activity (24). Given that the in vitro selection reactions were conducted at a relatively high concentration of Mg2+ (10 mM), the isolation of numerous variants that tolerate mutations in this accessory domain is expected based on its proposed function. Overall, these findings are consistent with our expectation that ribozyme variants with wild-type-like activity should not carry mutations in the core of the ribozyme.

Similarly, sequencing of 17 clones from G12 of lineage II (Figure 4C and Supplementary Figure S2) produced a single example of the parental B. cereus glmS ribozyme. Only one individual (clone B13) carries a mutation at a site within the catalytic core that is most highly conserved. However, this RNA fails to exhibit measurable self-cleavage activity (data not shown). The remaining 15 individuals each contain an average of 3.4 mutations per molecule and, as with the variants from lineage I, the locations of these mutations are compatible with the structural and functional models for this ribozyme class.

It is notable that nearly half the clones examined in both lineages have a mutation from G to A at nt 61 (Figure 4). Consistent changes in nucleotide sequence at specific sites usually indicate that the mutation provides some selective advantage, and therefore could be reflective of a phenotypic change from the parent glmS ribozyme. Nucleotide sequences recorded in GenBank (39) reveal that B. cereus and Bacillus anthracis bacterial strains carry two extra nucleotides located in the junction between P2 and P2.2 compared to glmS ribozymes from other bacteria (24,27). These 2 nt, corresponding to positions 61 and 62 of the B. cereus parent sequence (Figure 1A), are A and C, respectively, in nearly all strains of B. cereus and B. anthracis in GenBank. A single exception is the strain used in our studies (B. cereus ATCC 10987), which carries G and C at these sites. Therefore, the frequent mutation of G to A at position 61 results in a nucleotide sequence that matches most natural glmS sequences from B. cereus and B. anthracis. Therefore, this
commonly occurring mutation is not indicative of a fundamental change in ribozyme activity, but is a mutation that yields a sequence more commonly observed in natural glmS ribozyme isolates.

**Implications of in vitro selection results on the structure and action of glmS ribozymes**

In previous studies, in vitro selection strategies have been used to find mutations within aptamers that are able to compensate for changes in ligand structure. For example, variants of an RNA aptamer that selectively binds citrulline were isolated that preferentially bind arginine (40). The global fold of both aptamers are strikingly similar and changes at only 3 nt are required to alter ligand specificity (41). In another example, only a single nucleotide difference is required to change the ligand-binding preference of a theophylline aptamer to favor binding of 3-methylxanthine (42). Natural aptamers also can acquire altered ligand-binding characteristics with only modest changes in nucleotide sequence. Specifically, a single C-to-U mutation in the core of guanine riboswitches can switch ligand preference to adenine (19,43). These engineered and natural aptamers provide examples that prove RNAs can routinely adapt to small changes in ligand composition.

Our observation that the *B. cereus* glmS ribozyme has difficulty adapting to new ligands has implications for the structure and action of this ribozyme class. It has been suggested that glmS ribozymes might use GlcN6P as a coenzyme to promote RNA cleavage (24,26), perhaps functioning as a general acid or general base catalyst (26,31). If true, it might be more difficult for glmS RNAs to acquire mutations that alter molecular recognition characteristics while preserving the ability of the RNA to properly position or activate the cofactor for catalytic function.

Furthermore, GlcN6P analogs that lack an intact ring structure, that lack the 2-amino group, or that have an altered pKₐ of the 2-amino group exhibit little or no ability to induce glmS ribozyme activity (31). Compounds that either have an open ring structure (GlcNol6P) or that lack the 2-amino group (Glc; Glc6P) were present in our in vitro selection reactions, but no variants were found that could use these compounds. If the 2-amino group is directly involved in the chemical step of ribozyme cleavage, then GlcN6P analogs that lack an amine are not likely to substitute as a cofactor even for variants of this ribozyme class. However, the failure
of in vitro selection to enrich for mutations that improve even modestly the ribozyme’s activity in the presence of GlicNol6P. GlicN, GlicN6S or even GlicN6P suggests that the simultaneous demands that molecular recognition and catalytic activity place on the ribozyme cannot be altered with only a few mutations.

Moreover, it is notable that none of the three glmS ribozymes derived from B. cereus (26,27,37), B. subtilis (24,25,27,31,32) or B. anthracis (30) exhibits \( k_{\text{obs}} \) values greater then \( \sim 3 \text{ min}^{-1} \) when saturated with GlicN6P, although the rate constants for these ribozymes might be limited by folding and not by the chemical step of the reaction (30). Regardless, our in vitro selection results suggest that improvements in the overall speed of RNA cleavage might also be difficult to achieve within the context of this ribozyme class. It has been noted that glmS ribozymes have kinetic characteristics that are consistent with a ribozyme that uses in-line positioning of attacking and leaving groups, and uses activation of the 2'-oxygen nucleophile (27) to accelerate RNA cleavage. If glmS ribozymes use these two catalytic strategies to the exclusion of all other mechanisms for RNA cleavage, then the maximum \( k_{\text{obs}} \) values observed for the constructs tested might be close to the maximum theoretically possible for such enzymes (44,45).

CONCLUSIONS

Our exploration into the possibility of altering the ligand-binding specificity of glmS ribozymes contributes to several aspects of riboswitch research and application. Variant ribozymes that retain function with the natural ligand GlicN6P carry patterns of mutations that are easily reconciled with the nucleotide sequence phylogeny of natural glmS ribozymes. This indicates that the nucleotide sequences and structures that are most conserved in the natural examples are important for molecular recognition and catalysis, as opposed to the possibility that they are present exclusively for some other biological function.

Similarly, the fact that no variant ribozymes could be isolated that exhibit improved ligand affinity or altered ligand specificity suggests that glmS ribozymes might not be very receptive to changes at the ligand-binding site. If this ribozyme class uses the ligand as a coenzyme versus as an allosteric effector, then the simultaneous demands that molecular recognition and accurate coenzyme positioning place on an RNA molecule should make it more difficult to encounter variants that productively alter both functions.

Finally, engineered fusions between aptamers and ribozymes can be made to function as precision biosensors for various ligands (46–48). Constructs adapted from glmS ribozymes also have been used as fluorescent reporter systems for conducting high-throughput screens for compounds that trigger riboswitch function (49,50). Similar assays with other GlicN6P analogs, including glucose or glucose-6-phosphate could find utility as reporters for these compounds both in vitro and in vivo. However, our data suggest that glmS ribozymes are not be predisposed to serve as the best starting points for molecular engineers who seek to create such sensors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Adam Roth for design of the construct for in vitro selection, and Jinsoo Lim for synthesis of GlicNol6P, \( \alpha \)-MeOglcN6P and \( \beta \)-MeOglcN6P. The authors also thank Gail A. M. Emilsson and Adam Roth for critically reading the manuscript. Jeffrey Barrick for the updated glmS ribozyme consensus sequence, and other members of the Breaker laboratory for helpful discussion. This work was supported by the Howard Hughes Medical Institute (HHMI), and by grants from the NSF (EIA 0323510) and the NIH (R33 DK07027). Funding to pay the Open Access publication charges for this article was provided by HHMI.

Conflict of interest statement. None declared.

REFERENCES