Analysis of lysine recognition and specificity of the Bacillus subtilis L box riboswitch

Sharnise N. Wilson-Mitchell1,2, Frank J. Grundy1,3 and Tina M. Henkin1,2,3,*

1Center for RNA Biology, 2The Ohio State Biochemistry Program and 3Department of Microbiology, the Ohio State University, Columbus, OH 43210, USA

Received December 2, 2011; Revised February 15, 2012; Accepted February 16, 2012

ABSTRACT

The ever-changing environment of a bacterial cell requires sophisticated mechanisms to adjust gene expression in response to changes in nutrient availability. L box riboswitch RNAs regulate gene expression in response to cellular lysine (lys) concentrations in the absence of additional regulatory factors. In Bacillus subtilis, binding of lysine (lys) to the L box RNA causes premature transcription termination in the leader region upstream of the lysC coding sequence. To date, little is known about the specific RNA–lys interactions required for transcription termination. In this study, we characterize features of the B. subtilis lysC leader RNA responsible for lys specificity, and structural elements of the lys molecule required for recognition. The wild-type lysC leader RNA can recognize and discriminate between lys and lys analogs. We identified leader RNA variants with mutations in the lys-binding pocket that exhibit changes in the specificity of ligand recognition. These data demonstrate that lysC leader RNA specificity is the result of recognition of ligand features through a series of distinct interactions between lys and nucleotides that comprise the lys-binding pocket, and provide insight into the molecular mechanisms employed by L box riboswitch RNAs to bind and recognize lys.

INTRODUCTION

Recent work has uncovered a set of bacterial regulatory RNAs, designated riboswitches, in which conserved cis–acting RNA elements located in the ‘leader region’ between the promoter and the regulated coding sequence respond to a variety of physiological signals without additional proteins or co-factors (1). Typical riboswitches are composed of two domains: the aptamer or ligand–sensing domain and the expression platform (2). The aptamer binds the cognate ligand and initiates a series of structural rearrangements that affect the fate of the downstream gene through structural modulation of the expression platform. Regulation of gene expression by these conserved leader RNAs can occur by transcription attenuation, modulation of translation initiation and effects on RNA stability (2,3).

A large number of genes that control essential processes are under riboswitch regulation in Bacillus subtilis (1). One of the earliest identified riboswitches, designated the L box, regulates transcription of the lysC gene in response to lysine (lys) (4–6). L box leader RNAs have been found upstream of a variety of genes involved in lys biosynthesis, transport and utilization (6–8). Lys biosynthesis in most bacteria and some plants occurs through the diaminopimelate pathway. The first step of lys biosynthesis is the phosphorylation of aspartate by aspartokinase. There are three differentially regulated aspartokinase isozyms in Escherichia coli and B. subtilis. However, only aspartokinase II, the product of the lysC gene, responds to lys (9). Formation of aspartate-4-phosphate is critical, as it is also a precursor for methionine and threonine biosynthesis (9). Other L box-regulated genes synthesize metabolites in the lys pathway that are required for other essential physiological processes, such as synthesis of cell wall peptidoglycan (10) and formation of bacterial endospores (11). Additional phylogenetic analyses suggest that L box leader RNAs are also involved in regulation of genes responsible for lys transport and the regulation of cellular pH (6,8).

When lys concentrations are high, lys binds to the B. subtilis lysC leader RNA and promotes the formation of an intrinsic terminator that prevents transcription of the lysC coding region (4,5). Low lys concentrations result in the formation of an alternate antiterminator structure that allows transcription to continue (Figure 1). In contrast, the E. coli lysC leader RNA is predicted to regulate expression at the translational level through the formation of a Shine–Dalgarno sequestering helix when lys concentrations are high (4,6,9). The lys analog aminoethylcysteine (AEC) is 10-fold less effective...

*To whom correspondence should be addressed. Tel: +1 614 688 3831; Fax: +1 614 292 8120; Email: henkin.3@osu.edu

© The Author(s) 2012. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
than lys in promotion of *B. subtilis* **lysC** termination in vitro (4). AEC differs from lys by a single substitution of carbon with sulfur (Figure 3); the observed differential sensitivity demonstrates specific lys recognition by the **lysC** leader RNA (4). Lys analog-binding studies using the aptamer region of the **lysC** leader RNA further suggested that leader RNA binding is specific for lys (12).

Previous studies of riboswitch specificity identified several unique features adopted by riboswitch RNAs to recognize their cognate ligands. For example, the purine responsive riboswitches use canonical Watson–Crick base pairing for recognition (13). A single mutation of a conserved base in the guanine responsive riboswitch results in a specificity switch that allows recognition of adenine (14). Natural variants of the guanine and adenine riboswitches were identified that selectively bind 2′-deoxyguanosine (dG) and discriminate against guanine (15). The discovery of these variants demonstrates that riboswitch RNAs can use similar RNA architectures to recognize different molecules.

Several riboswitches use similar mechanisms to recognize nucleotide-like molecules. In each example, the RNA uses common features such as the face of the nucleotide to assist in recognition. Recognition of S-adenosylmethionine...
(SAM) by the S box (SAM-I) riboswitch occurs through a base triple interaction that includes the adenine ring of the SAM molecule (16). Similarly, the aminopyrimidine ring of the thiamine pyrophosphate (TPP) molecule is recognized by the THI-box RNA through molecular stacking with conserved residues (17). The nucleotide face of the cyclic di-GMP (c-di-GMP) molecule is also recognized using asymmetric RNA interactions (18). Although the RNAs recognize similar features of different ligands, each riboswitch utilizes a different mechanism for recognition.

Another common mechanism for ligand recognition incorporates divalent metal ions for both binding and recognition of negatively charged groups. RNAs that have been shown to use this method include the S box (17), THI-box (18), glycine riboswitch (19) and the glmS ribozyme (20). Analysis of glycine recognition by the RNA and is buried within a lattice formed from conserved core residues (22,23). Recent analysis of the lysC leader RNA has highlighted the importance of peripheral elements and long-range interactions in the organization of the lys-binding pocket; the significance of peripheral elements has also been demonstrated for the S-box riboswitch (24,25). Residues in direct contact with lys or in the binding pocket are the most highly conserved. Additionally, a K’ ion was buried between lysys and the RNA (Figure 2) (23). This arrangement is similar to the Mg’ ions found in the glycine-binding pocket of the glycine riboswitch (20). Currently, the role of K’ in the binding and function of L box leader RNAs has not been established, but it was suggested that it might assist in lys recognition (23).

With the variety of genes under L box control, it is essential that the RNA selectively binds lys and efficiently regulates expression under appropriate conditions. The current structural model of the lys-binding site does not clearly demonstrate how the RNA can specifically recognize lys and discriminate against substitutions in the ligand. Previous studies of lys analog recognition focused on a subset of available compounds (12). In this study, we test a broad range of lys analogs and identify mutations that affect ligand recognition; these mutations reveal molecular interactions responsible for recognition and specificity. We also investigate the role of K’ in the binding site and the potential effects on leader RNA function.

MATERIALS AND METHODS

Construction of DNA templates

Templates for in vitro transcription by B. subtilis RNA polymerase (RNAP) were generated by fusion of the B. subtilis glyQS promoter sequence to the lysC leader RNA sequence. PCR was carried out using DNA oligonucleotide primers that included the glyQS promoter and hybridized within the lysC leader region. A transcriptional fusion plasmid, pFG328, with a wild-type lysC insertion (4), was the DNA source for PCR amplification using Taq DNA polymerase as per the manufacturer’s instructions (Invitrogen, Carlsbad, CA, USA). The resulting PCR fragment contained the lysC leader template downstream of the glyQS promoter such that transcription would start at +17 relative to the lysC transcription start-site and included lysC sequence extending to 14 nt downstream of the leader region termination site. The transcription start site was designed to allow transcription initiation with the dinucleotide ApC and a halt at position G42 (relative to the native lysC transcript) during transcription in the absence of cytidine 5’-triphosphate (CTP). The 3’-end of the construct included an additional 106 bp of random sequence to allow resolution between terminated (252 nt) and readthrough (372 nt) transcriptions. PCR products were purified using a QiAquick PCR clean up kit (Qiagen, Chatsworth, CA, USA) and sequenced by Genewiz (South Plainfield, NJ, USA).

DNA templates used for T7 RNAP transcription were generated using complementary pairs of overlapping DNA oligonucleotides as previously described (25,26).
Briefly, the 5' pair contained the phage T7 RNAP promoter sequence fused to the +18 position of the lysC leader region. The remaining complementary pairs contained the wild-type lysC sequence, each with a 5-nt 3' overhang complementary to the 5' region of the adjacent pair. The terminal pair contained the 3' leader RNA sequence that ended at +234, 3 nt prior to the 5' position of the transcription terminator. This position was selected so that helix 1 formation could be monitored without the competing antiterminator structure. Each internal oligonucleotide pair was phosphorylated using T4 polynucleotide kinase (New England Biolabs, Beverly, MA, USA). The pairs were then mixed, incubated at 95°C and slow cooled to room temperature for annealing. The paired oligonucleotides were ligated using T4 DNA ligase per the manufacturer’s instructions (New England Biolabs, Beverly, MA, USA). The resulting DNA template was amplified using the flanking 5' and 3'DNA oligonucleotides as primers for PCR using Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA). The final template DNA was purified and sequenced as described above.

Site-directed mutagenesis
The transcriptional fusion vector pFG328 containing the wild-type lysC DNA template (4) was used as a template for oligonucleotide-directed mutagenesis. DNA oligonucleotides containing the desired mutations were designed as primers for PCR amplification of pFG328-lysC DNA using Pfu DNA polymerase (Stratagene, La Jolla, CA, USA). The resulting products were subjected to digestion with DpnI (New England Biolabs, Beverly, MA, USA) to remove the starting wild-type template and introduced into XL-2 blue ultracompetent cells by transformation as per the manufacturer’s instructions (Stratagene, La Jolla, CA, USA). The plasmid DNA was isolated using a Promega Wizard prep kit and sequenced to confirm the mutations (Genewiz, South
Plainfield, NJ, USA). The final constructs contained the glyQS promoter sequence upstream of the lysC leader RNA with the desired mutations, and were used as the DNA template for in vitro transcription.

The DNA templates for transcription by the T7 phage RNAP contained the wild-type lysC leader RNA sequence downstream of the T7 RNAP promoter sequence. Mutation of the lysC construct was performed by the substitution of complementary oligonucleotide pairs containing the desired mutations. The wild-type and mutant pairs were phosphorylated, ligated and amplified by PCR using Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA). The resulting DNA templates with the appropriate mutations were used in T7 RNAP transcription.

**In vitro transcription termination assays**

Single round transcription of the wild-type lysC leader RNA template was performed as previously described by Grundy et al. (27). A reaction mixture of 20 mM Tris–HCl, pH 8, 20 mM NaCl, 10 mM MgCl2, 100 μM EDTA, 150 μM ApC (Sigma), 2.5 μM GTP and ATP, 0.75 μM UTP, [α-32P]-UTP (GE Healthcare; 0.25 μM, 800 Ci/mmol [30 TBq/mmoll]), DNA template (10 nM) and His-tagged purified B. subtilis RNAP (6 nM) (28) was incubated for 15 min at 37°C. CTP was excluded from the reaction mixture to generate a transcriptional halt at +42 (relative to the native lysC transcription start site). After incubation, Heparin was added to block reinitiation and synchronize transcript synthesis. KCl (27 mM) was added where indicated. Lys or lys analogs were added, and elongation was resumed by the addition of 10 μM rNTPs and 40 mM MgCl2 followed by incubation for 15 min at 37°C. The transcription reactions were stopped by phenol/chloroform extraction, and the products were resolved by denaturing PAGE and visualized by PhosphorImager analysis. Percent termination represents terminated transcripts relative to the total amount of RNA transcribed; this value was plotted as a function of ligand concentration. Non-linear regression analysis was used to determine the concentration required for half-maximal termination (Trm1/2) (Table 1) was used to identify features of the lys

<table>
<thead>
<tr>
<th>Lysine modification</th>
<th>Ligand</th>
<th>Wild-type</th>
<th>Term.</th>
<th>Ratio</th>
<th>Trm1/2 (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Lys</td>
<td>96</td>
<td>3.2</td>
<td>0.17 ± 0.02</td>
<td>&gt;13</td>
</tr>
<tr>
<td>d-lys</td>
<td>46</td>
<td>1.5</td>
<td>&gt;13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diaminopentane</td>
<td>43</td>
<td>1.4</td>
<td>&gt;13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AEC</td>
<td>84</td>
<td>2.7</td>
<td>0.80 ± 0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydroxy lys</td>
<td>35</td>
<td>1.1</td>
<td>&gt;13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-ornithine</td>
<td>32</td>
<td>1.0</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ε-amino group</td>
<td>Lys methyl</td>
<td>73</td>
<td>2.4</td>
<td>1.3 ± 0.40</td>
<td>&gt;13</td>
</tr>
<tr>
<td>t-arg</td>
<td>35</td>
<td>1.1</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAP</td>
<td>30</td>
<td>0.97</td>
<td>&gt;13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carboxyl group</td>
<td>lysNH</td>
<td>90</td>
<td>2.9</td>
<td>0.30 ± 0.02</td>
<td>&gt;13</td>
</tr>
<tr>
<td>lysOH</td>
<td>87</td>
<td>2.8</td>
<td>1.9 ± 0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lysEE</td>
<td>76</td>
<td>2.5</td>
<td>1.0 ± 0.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Termination (term.) is the percent of RNAs terminated in the presence of 10 mM ligand relative to the total amount of RNA transcribed.

**RESULTS**

Wild-type lysC transcription termination in response to lys analogs

In vitro transcription of the wild-type lysC leader by B. subtilis RNAP in the presence of lys or lys analogs (Table 1) was used to identify features of the lys molecule required for ligand-dependent transcription termination. Lys and lys analogs with the ability to promote transcription of the lysC leader RNA were further tested in vitro using a broad range of ligand concentrations (0–12.8 mM) (Table 1). Transcription of the lysC leader region in the presence of lys (10 mM) resulted in 96% termination and the concentration required for Trm1/2 was 0.17 mM. In vitro transcription in the presence of 11 different lys analogs (added at 10 mM final concentration) showed that alterations to the main chain functional groups, as seen in d-lys and diaminopentane, resulted in a reduction of ligand-dependent termination (Table 1). The Trm1/2 for both d-lys and diaminopentane was >13 mM. These results suggest that changes to functional groups that participate in RNA–ligand interactions are
The increase in Trm1/2 in the presence of lys methyl substitution in the side chain of the molecule suggests reduced ligand recognition. The Trm1/2 in the presence of AEC (0.8 mM) increased 5-fold as compared to lys. The increase in Trm1/2 in the presence of lys demonstrates sensitivity to the presence of the sulfur molecule in AEC. Transcription of the leader RNA in the presence of ligands with more complex functional groups, as seen in L-arginine (L-arg) and L,L-diaminopimelate (DAP), resulted in a loss of recognition, possibly because the changes result in a molecule too large to fit in the lys-binding pocket.

Alteration to the global structure of the molecule, such as the shape of the molecule (hydroxy lys), and the length of the side chain (L-ornithine) caused reduced termination efficiency (Table 1). Modification of the α-amino group was tolerated only in the context of the addition of a methyl group, as in lysine methyl (lys methyl). Transcription in the presence of lys methyl resulted in a Trm1/2 of 1.3 mM, an 8-fold increase as compared to lys. The increase in Trm1/2 in the presence of lys methyl suggests that the methyl addition to the α-amino group is deleterious to ligand recognition. The Trm1/2 in the presence of lys methyl was 1.9 mM relative to lys. Similar to the observed increase in Trm1/2 of lysOH as compared to lysEE, modifications to the carboxyl and α-amino groups of lys (23).

The carboxy-modified ligands were more effective than analogs with α-amino or side chain modifications and exhibited lower Trm1/2 concentrations (Table 1). Transcription in the presence of lysinamide (lysNH), which contains a nitrogen substitution of the hydroxyl carbon, resulted in a Trm1/2 of 0.3 mM, a 2-fold increase as compared to lys. In the presence of lysNH, the leader RNA exhibited a smaller increase in Trm1/2 relative to lys methyl and AEC; this may be indicative of a larger or more stable network of RNA–ligand interactions near the carboxyl group. Substitution of the amide (lysNH) with an ethyl ester (lysEE) caused a 3-fold increase in Trm1/2 (1.0 mM) as compared to lysNH, and a 6-fold increase relative to lys. The ethyl ester extension of the oxygen may disrupt other essential interactions required for lys-dependent termination. The observed decrease in Trm1/2 in the presence of lysNH as compared to lysEE may reflect a direct interaction between the leader RNA and the substituted amide. Transcription in the presence of lysine hydroxamate (lysOH) resulted in a 10-fold increase in Trm1/2 (1.9 mM) relative to lys. Similar to the other carboxy-modified analogs, lysOH contains a hydroxy amide substitution of the hydroxyl carbon (Figure 3). The increase in Trm1/2 of lysOH as compared to lysNH suggests that the addition of a hydroxyl group causes a reduction in ligand-dependent termination. The observed increase in Trm1/2 of lysOH as compared to lysEE emphasizes the importance of functional group recognition as compared to minor alterations in ligand length. This indicates that both the identity of the molecular substitution and additional interactions within the binding pocket affect ligand recognition.

Effects of substitutions at G111 and A112

In an effort to identify sequence elements within the lysC leader that are important for ligand recognition, we used site-directed mutagenesis to probe RNA requirements for lys-dependent termination. The nucleotides chosen for mutation are highly conserved (Figure 1) and predicted to interact with lys (Figure 2). Universally conserved G111 and A112 are positioned nearest to the α-amino group in the binding pocket (Figure 2). Substitution of G111 with C resulted in reduced termination (12%) in the absence of ligand as compared to wild-type (Table 1; 31%) and caused a reduction in sensitivity to lys and AEC (Table 2). Each G111 variant exhibited a 5-fold increase in termination in response to lys and AEC. However, the G111A substitution resulted in a 2-fold increase in termination in response to DAP (Table 2), which is inactive in the context of the wild-type lysC sequence (Table 1). Similarly, the G111U substitution resulted in a 5-fold increase in termination in response to lys methyl, as compared to the no ligand control (Table 2). The reduction in lys-dependent termination suggests that G111 plays a pivotal role in lys recognition and/or binding. However, the termination response in the presence of DAP (G111A) and lys methyl (G111U) suggests that G111 is not required for formation. The sensitivity of G111A and G111U to α-amino-modified analogs is in good agreement with the published crystal structure, as G111 is positioned nearest the α-amino group of the lys molecule (23).

Unlike G111, in which no substitution allowed normal function, A112 was amenable to substitution with a guanine residue, exhibiting a ligand response similar to that of wild-type, although termination in the presence of α-amino and carboxyl modified analogs was reduced (Tables 1 and 2). When A112 was substituted with either pyrimidine base (Table 2), there was a decrease in termination in both the presence and absence of ligand (~7% for A112C and ~19% for A112U). Both A112 and G111 were intolerant to pyrimidine substitutions, which suggests that purines are required for ligand recognition and lys-dependent termination. In contrast to A112G, the purine–purine substitution of G111 with A was deleterious to the function of the RNA. This suggests that there are specific interactions between G111 and the lys molecule, which directly contribute to lys-dependent termination.

Mutations near the lys carboxyl group result in altered ligand specificity

Lys-binding pocket variants were generated to assess the interaction between the leader RNA and the carboxyl...
region of the lys molecule. Crystal structure analysis revealed base pairing between G40 and C110 of the leader RNA, which is in agreement with previous studies that identified mutations at this position that result in deregulation of the lysC leader and cause AEC resistance (29). The conserved G40–C110 and G144–U170 base pairs were substituted with other canonical base pairs and tested for termination in the presence of lys or lys analogs. The G40U–C110A substitution inhibited lys-dependent termination (Table 3). Lys or lys analogs with substitutions in the side chain (AEC) or e-amino group were not recognized by the G40A–C110U variant (Table 3). However, termination of the G40A–C110U variant was observed in the presence of carboxyl-modified analogs. Transcription of the G40A–C110U variant in the presence of lysNH, which replaces a carboxyl oxygen with an amide, resulted in a reduction in termination (Table 3) as compared to wild-type (Table 1). Termination of the G40A–C110U variant increased 4-fold in the presence of lysOH (75%) as compared to the no ligand control (Table 4). The Trm1/2 of the G40A–C110U variant in the presence of lysOH was 2.1 mM as compared to >13 mM for lys (Table 4). The Trm1/2 exhibited by the G40A–C110U variant in the presence of lysOH was similar to wild-type; however, the Trm1/2 was increased, as compared to wild-type, in the presence of lys (Table 4). The response to lysOH and lack of response to lys suggests that the G40A–C110U mutations cause a change in ligand recognition specificity. Termination of the G40A–C110U variant in the presence of lysNH, as compared to lysOH, indicates that the hydroxyl substitution at the carbonyl more effectively promotes termination than an amide at the same position. It is plausible that the change in specificity is due to the formation of new molecular interactions between G40A and the amide or hydroxyl group of the lys analog.

The G144–U170 leader variants also exhibited specificity for carboxyl-modified analogs. Substitution of the wild-type G–U pair with U–G caused a loss of lys recognition and a reduction in termination in the presence of lys analogs (Table 5). The U170G mutation, which replaces the G144–U170 pair with a G–G pair, resulted in complete loss of ligand-dependent termination in vitro, suggesting
that base pairing is required at this position (Table 5). Variants that maintained base pairing between positions 144 and 170 showed an increase in termination in the presence of lys methyl, lysNH and lysEE. With the exception of the A–U and C–G constructs, all other variants resulted in termination efficiencies lower than that of wild-type in the presence of lys methyl, lysNH and lysEE (Tables 1 and 5). The C–G variant (G144–U170G) showed a 14% increase in termination in the absence of ligand (Table 5) as compared to wild-type (C–G). In contrast to the C–G variant, the G–C variant (G144–U170C) resulted in a reduction of ligand-dependent termination as compared to wild-type in the absence of ligand (G–C) (Tables 1 and 5). Termination of the G–C variant in the presence of lys and lys analogs was reduced as compared to wild-type. These results indicate that the U170C mutation does not abolish function of the RNA but perturbs ligand recognition. Transcription of the G144A–U170 variant in the presence of lysEE resulted in increased termination (94%) (Table 5; A–U) as compared to wild-type termination in the presence of lysEE (76%) (Table 1). The Trm1/2 of the G144A variant for lysEE was 0.85 mM, similar to that of wild-type (Table 4). However, Trm1/2 of the G144A variant in the presence of lys was >13 mM, suggesting a loss in lys recognition (Table 4). The G144A substitution, a G–U to A–U base pair exchange, caused an alternate ligand specificity where the variant leader RNA responds to lysEE and discriminates against lys. Substitution of the G144A–U170 pair with U–A resulted in a reduction of termination efficiency in the presence of the carboxyl-modified analogs; however, ligand specificity was similar to that of the A–U variant. Recognition of lys methyl by the G144–U170 variants is surprising as these positions were not predicted to reside near the \(\epsilon\)-amino group. One explanation is that the G144–U170 base pair is not primarily responsible for \(\epsilon\)-amino group recognition. Termination in the presence of lys analogs as opposed to lys and AEC suggests that substitution of the G40–C110 and G144–U170 base pairs resulted in a loss of lys recognition. However, G144–U170 variants retained specificity for \(\epsilon\)-amino and carboxyl-modified analogs while G40A–C110U exhibited specificity for analogs with carboxyl modifications. Together, these results indicate that specific mutations of the RNA sequence can result in alternate ligand specificity whereby the leader RNA responds to lys analogs while discriminating against lys.

Table 4. Termination efficiency of lysC leader variants in the presence of lys or lys analogs

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Variant</th>
<th>Trm1/2 (mM)</th>
<th>Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>lys</td>
<td>A112Gb</td>
<td>0.23 ± 0.07</td>
<td>&gt;1.0</td>
</tr>
<tr>
<td>lys</td>
<td>G40A–C110Ub</td>
<td>&gt;1.0</td>
<td>&gt;1.0</td>
</tr>
<tr>
<td>lys</td>
<td>G144A–U170b</td>
<td>&gt;1.0</td>
<td>&gt;1.0</td>
</tr>
<tr>
<td>AEC</td>
<td>G144A–U170b</td>
<td>&gt;1.0</td>
<td>&gt;1.0</td>
</tr>
<tr>
<td>lysOH</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>lysEE</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>ND</td>
<td>lysOH</td>
<td>2.1 ± 0.20</td>
<td>ND</td>
</tr>
<tr>
<td>ND</td>
<td>lysEE</td>
<td>0.85 ± 0.10</td>
<td>ND</td>
</tr>
</tbody>
</table>

*Trm1/2 is the concentration of ligand required to promote 50% termination.
ND, not determined.

Table 5. In vitro transcription termination of G144–U170 lysC leader variants in the presence of 10 mM lys or lys analogs

<table>
<thead>
<tr>
<th>Lysine modification</th>
<th>Ligand</th>
<th>U-G</th>
<th>G-G</th>
<th>C-G</th>
<th>G-C</th>
<th>A-U</th>
<th>U-A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Term.</td>
<td>Ratio</td>
<td>Term.</td>
<td>Ratio</td>
<td>Term.</td>
<td>Ratio</td>
<td>Term.</td>
</tr>
<tr>
<td>None</td>
<td>33</td>
<td>22</td>
<td>33</td>
<td>24</td>
<td>45</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>Lys</td>
<td>33</td>
<td>22</td>
<td>33</td>
<td>24</td>
<td>45</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>d-lys</td>
<td>41</td>
<td>28</td>
<td>41</td>
<td>28</td>
<td>48</td>
<td>29</td>
<td>26</td>
</tr>
<tr>
<td>Diaminopentane</td>
<td>43</td>
<td>28</td>
<td>43</td>
<td>28</td>
<td>69</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>AEC</td>
<td>30</td>
<td>21</td>
<td>41</td>
<td>21</td>
<td>41</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>Hydroxy lys</td>
<td>38</td>
<td>23</td>
<td>40</td>
<td>21</td>
<td>40</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>1-ornithine</td>
<td>38</td>
<td>21</td>
<td>40</td>
<td>21</td>
<td>40</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>L-ornithine</td>
<td>38</td>
<td>21</td>
<td>40</td>
<td>21</td>
<td>40</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>L-arg</td>
<td>38</td>
<td>21</td>
<td>40</td>
<td>21</td>
<td>40</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>DAP</td>
<td>32</td>
<td>21</td>
<td>43</td>
<td>21</td>
<td>43</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>lysNH</td>
<td>49</td>
<td>21</td>
<td>57</td>
<td>23</td>
<td>57</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>lysOH</td>
<td>47</td>
<td>20</td>
<td>57</td>
<td>23</td>
<td>57</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>lysEE</td>
<td>72</td>
<td>24</td>
<td>94</td>
<td>42</td>
<td>94</td>
<td>42</td>
<td>94</td>
</tr>
</tbody>
</table>

*Termination (term.) is the percent of RNAs terminated relative to the total transcription.
Ratio is the ratio of percent terminated transcripts in the presence of 10 mM ligand to the percent terminated transcripts in the absence of ligand.
K+ decreases the concentration of lys required for transcription termination

The B. subtilis lysC leader RNA does not require K+ for lys-dependent termination in vitro (4,5). However, recent analyses suggest that K+ is important for lys recognition and affinity (7,30). We, therefore, tested the effect of addition of K+ to our purified in vitro transcription system. In the presence of K+, the wild-type leader RNA showed a 5-fold decrease in Trm1/2 for lys (Fig. 4A). Other monovalent (NaCl) and divalent salts (CaCl2) were also assayed to test whether the observed effect was specific to K+ and no observable change in Trm1/2 was detected (our unpublished results). The 5-fold decrease in Trm1/2 for lys in the presence of K+ suggests that K+ plays a key role in lys-dependent termination by increasing the affinity of the RNA for lys.

Variant leader RNAs and lys analogs were further tested to determine if the K+ is involved in ligand recognition. The A112G variant, which does not show altered ligand recognition but exhibited a 2-fold decrease in lys-dependent termination relative to the wild-type (Table 2), exhibited no observable change in lys sensitivity in response to K+ (Figure 4A). The lack of response to the presence of K+ may be the result of changes in the structure of the binding pocket imposed by the A112G mutation. Based on the position of K+ in the structure, it is plausible that the G40–C110 or G144–U170 base pairs of the B. subtilis lysC binding pocket interact directly with K+ (Figure 2). The G40A–C110U variant, which exhibited increased termination only in the presence of lysOH (Table 3), exhibited no change in Trm1/2 for lysOH in the presence of K+ (Figure 4B). Similarly, the G144A variant, which responds to lysEE, showed no reduction in Trm1/2 for lysEE in the presence of K+ (Figure 4B). Transcription of the G144A variant in the presence of lys demonstrated that K+ was unable to enhance lys recognition (our unpublished results). These results indicate that K+ does not play a role in the specificity of ligand recognition but rather affects affinity of the wild-type leader RNA for lys.

lysC leader RNA variants undergo ligand-dependent structural changes

Structural rearrangement of the leader RNA when the regulatory ligand is present is the hallmark of riboswitch regulation. This structural change is crucial as it directly controls gene expression. The current model of the lysC leader RNA structural transition indicates that formation of the transcription terminator requires the stabilization of the anti-antiterminator (AAT; Figure 1) through the formation of helix 1 (4). To monitor the structural change of the variant leader RNAs in response to lys analogs, we used RNase H probing of helix 1. Addition of a DNA oligonucleotide complementary to the 3′ side of helix 1 will result in cleavage only when helix 1 is unpaired; stabilization of the helix prevents binding of the oligonucleotide, which results in protection of the RNA from cleavage by RNase H, an endonuclease that is specific for RNA–DNA hybrids. In the absence of ligand, 31% of the wild-type RNAs were protected from RNase H cleavage as compared to 89% protection in the presence of lys (Figure 5, lanes 1 and 2). These results indicate that the presence of lys promotes the formation of helix 1. When wild-type RNAs were transcribed in the presence of high concentrations of AEC, 84% protection was observed (Figure 5, lane 3). Similarly, wild-type RNAs exhibited 82% protection in the presence of lysOH and 60% protection in the presence of lysEE (Figure 5, lanes 4 and 5). The reduced protection of wild-type helix 1 in the presence of lys analogs, as compared to lys, is consistent with lower wild-type termination in vitro in the presence of lys analogs. For the A112G variant, 48% of the transcribed RNAs were protected from cleavage in the no ligand control as compared to lys, is consistent with lower wild-type termination in vitro in the presence of lys analogs. For the A112G variant, 48% of the transcribed RNAs were protected from cleavage in the no ligand control as compared to 65% in the presence of lys (Figure 5, lanes 6 and 7). A reduction in protection of the A112G variant, as compared to wild-type, was observed in the presence of AEC, lysOH, and lysEE (Figure 5, lanes 8–10). This demonstrates that protection...
of helix 1 in the A112G construct was more effective in the presence of lys than protection in the presence of lys analogs. RNase H probing of the G40A–C110U variant (Figure 5, lanes 11–15) resulted in 38% protection in the absence of ligand and 51% protection in the presence of lysOH. Protection of the G40A–C110U variant in presence of lys (40%) and AEC (33%) decreased 2-fold as compared to protection of wild-type in presence of lys or AEC (Figure 5). The lack of helix 1 formation by the G40A–C110U variant is in agreement with in vitro termination results that suggest that the G40A–C110U variant is specific for lysOH and discriminates against lys and AEC. Similarly, protection of the G144A variant increased 2-fold in the presence of lysEE as compared to the no ligand control (Figure 5, lane 20). Protection of the G144A and wild-type constructs in the presence of lysEE was 48% and 60%, respectively. For each lysC leader variant, stabilization of helix 1 was specific for the lys analogs to which they responded in the in vitro transcription assay and was reduced in the presence of lys. These results confirm that the ligand-dependent termination effects described above correlate with stabilization of helix 1.

### DISCUSSION

Previous studies of riboswitch specificity have identified RNA elements involved in ligand recognition (25,30,31). However, these features are variable and often specific to the RNA and its cognate ligand. Structural analysis of ligand-bound riboswitches has revealed mechanistic themes employed for ligand recognition. Ligand features such as negatively charged groups and nucleotide-like structures are targets for recognition. Shape complementarity, base pairing, nucleotide stacking and metal ion coordination are known mechanisms of ligand recognition by riboswitches. The molecular structure of lys makes a number of these mechanisms impossible.

Recent analysis of the aptamer region of the *T. maritima* L box leader RNA has identified the nucleotide composition of the lys-binding pocket. Although the nucleotides that directly contact lys were identified, the molecular interactions and their effect on function are poorly understood. In this study, we use the lys-responsive *B. subtilis* lysC leader RNA to analyze recognition determinants and their effects on ligand-dependent structural changes and transcription termination.

Figure 5. RNase H cleavage of wild-type and variant lysC leader RNAs. Radiolabeled RNAs were transcribed in the presence of 7.5 mM lys, AEC, lysOH and lysEE and incubated with a DNA oligonucleotide complementary to the 3' side of helix 1 to allow hybridization followed by RNase H digestion. Helix 1 protection of wild-type in the presence of lys and lys analogs is shown in lanes 1–5; lanes 6–10, A112G variant; lanes 11–15, G40A–C110U variant; and lanes 16–20, G144A variant. FL, full-length transcripts protected from RNase H cleavage. C, transcripts cleaved by RNase H. Percent protection is amount of full-length RNAs relative to total amount of RNA transcribed.
triple interactions that may be essential for formation of the lys-binding pocket (23). The data presented here provide evidence that G111 is required for formation of the lys-binding pocket as well as lys recognition. It is possible that G111 must be precisely positioned to interact with other nucleotides within the RNA and the bound lys molecule. Additionally, our data suggest that A112 is not essential for ligand recognition but is primarily involved in lys affinity. This effect on affinity may be achieved through extensive contacts with A112 that stabilize the structure of the lys-binding pocket. Variants of the lys-binding pocket near the carboxyl region of the lys-molecule (G40A–C110U and G144A–U170) exhibited preferential specificity for certain lys analogs and discriminated against lys. The alternate specificity of leader RNA variants suggests that each base pair is responsible for the recognition of molecular features of the lys carboxyl group. Additionally, mutations in the carboxyl region of the binding pocket did not inhibit recognition of the ε-amino region, as G144–U170 variants retained the ability to recognize lys methyl. The leader RNA is able to use alternate molecular interactions with lys analogs to promote termination. However, the increased concentration requirement for lys analogs suggests that the new RNA–analogue interactions do not perfectly mimic the interactions with lys. This analysis confirms that the sequence of lysC leader RNA is optimized for lys recognition as all of the variants in this study, with the exception of A112G, exhibited a loss of lys-dependent termination. In total, these data indicate that recognition of lys by the L box riboswitch is a modular process in which individual nucleotides recognize independent features of the lys molecule.

The presence of a K⁺ ion in the L box crystal structure (23) raised questions about the role of this ligand in binding. Analysis of the effects of the addition of K⁺ to the in vitro transcription termination assay revealed that the potassium ion may increase the affinity of the leader RNA for lys. The increased affinity was observed only in the presence of the wild-type RNA and lys. This suggests that both the sequence and structure of the binding pocket are necessary for stimulation of tighter lys binding by K⁺. Additional studies are required to determine the direct effects of K⁺ on lys binding.

The discovery of new riboswitches and characterization of RNA–ligand complexes continue to diversify our understanding of the mechanisms used by RNAs to recognize and bind small molecules. Small molecule recognition by riboswitches is essential to regulation, as gene expression must occur only in response to the appropriate stimulus. Recognition of the appropriate ligand by riboswitch RNAs can occur in a variety of ways with some common themes distributed amongst the different classes. The lysC leader RNA not only utilizes some common mechanisms, such as shape complementarity and metal ion assistance for lys binding and recognition, but also makes nucleotide-specific interactions for ligand identification. These interactions differ from the mechanism used by the glycine riboswitch to recognize similar amino acid features (19). The alternate specificities exhibited by lysC variants demonstrate that the leader RNA uses a unique arrangement of contacts to recognize the ligand and regulate gene expression. This characterization has provided a more detailed view of how the lysC leader RNA specifically recognizes the lys molecule and contributes to our current understanding of small molecule recognition by RNAs. However, the specific features responsible for lys binding affinity are still unclear and require further analysis.

ACKNOWLEDGEMENTS

We would like the thank members of the Henkin lab for helpful discussions and Dr Mike Ibba for providing the lys analogs.

FUNDING

National Institutes of Health (R01GM063615 to T.M.H. and F31GM076894 to S.M.). Funding for open access charge: NIH.

Conflict of interest statement. None declared.

REFERENCES