

[0015] In a preferred example, provided is a polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises a sequence that is at least 95% identical to a sequence selected from the group consisting of SEQ ID NOs: 9, 14, and 26. In a even more preferred example, provided is a polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 9, 14, and 26.

[0016] In one aspect, provided is a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises:

ACX₁GGGGTCCGGCX₂TX₃TTCATTTGGCX₄CCGGTGAGAX₅X₆AX₇ACCCTTX₈X₉X₁₀X₁₁CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:1),
wherein

X₁ is A or G;

X₂ is C or no nucleotide;

X₃ is T or no nucleotide;

X₄ is A or G;

X₅ is any nucleotide;

X₆ is any nucleotide;

X₇ is any nucleotide;

X₈ is any nucleotide;

X₉ is C, G, or T;

X₁₀ is any nucleotide;

X₁₁ is A or T;

X₁₂ is C or T; and

X₁₃ is C or T.

[0017] In examples, the aptamer encoding sequence has one or more of the following properties: X₁ is not A; X₂ is not C; X₃ is not T; X₄ is not G; X₅ is not G; X₆ is not C; X₇ is not C; X₈ is not T; X₉ is not G; X₁₀ is not A; X₁₁ is not A; X₁₂ is not T; and X₁₃ is not C.

[0018] In examples, all of the following are not simultaneously present in the aptamer encoding sequence: X₁ is A; X₂ is C; X₃ is T; X₄ is G; X₅ is G; X₆ is C; X₇ is C; X₈ is T; X₉ is G; X₁₀ is A; X₁₁ is A; X₁₂ is T; and X₁₃ is C.

[0019] In one aspect, provided is a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises:

ACX₁GGGGTCCGGCX₂TX₃TTCATTTGGCGCCGGTGAGAX₅X₆AX₇ACCCTTX₈
X₉X₁₀X₁₁CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:2),

wherein

X₁ is A or G;

X₂ is C or no nucleotide;

X₃ is T or no nucleotide;

X₅ is G or T;

X₆ is C or T;

X₇ is C or T;

X₈ is any nucleotide;

X₉ is G or T;

X₁₀ is A, G, or T;

X₁₁ is A or T;

X₁₂ is C or T; and

X₁₃ is C or T.

[0020] In examples, the aptamer encoding sequence has one or more of the following properties: X₁ is not A; X₂ is not C; X₃ is not T; X₄ is not G; X₅ is not G; X₆ is not C; X₇ is not C; X₈ is not T; X₉ is not G; X₁₀ is not A; X₁₁ is not A; X₁₂ is not T; and X₁₃ is not C. In examples, all of the following are not simultaneously present in the aptamer encoding sequence: X₁ is A; X₂ is C; X₃ is T; X₅ is G; X₆ is C; X₇ is C; X₈ is T; X₉ is G; X₁₀ is A; X₁₁ is A; X₁₂ is T; and X₁₃ is C.

[0021] In one aspect, provided is a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises:

ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTTGAA
CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:3),

wherein

X₁₂ is C or T; and

X₁₃ is C or T.

[0022] In examples, the aptamer encoding sequence has one or more of the following properties: X₁₂ is not T; and X₁₃ is not C. In examples, all of the following are not simultaneously present in the aptamer encoding sequence: X₁₂ is T and X₁₃ is C.

[0023] In one aspect, provided is a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises:

ACX₁GGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTX₈X₉X₁₀X₁₁CCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:4),

wherein

X₁ is A or G;

X₈ is any nucleotide;

X₉ is G or T;

X₁₀ is A, G, or T; and

X₁₁ is A or T.

[0024] In examples, the aptamer encoding sequence has one or more of the following properties: X₁ is not A; X₈ is not T; X₉ is not G; X₁₀ is not A; X₁₁ is not A. In examples, all of the following are not simultaneously present in the aptamer encoding sequence: X₁ is A; X₈ is T; X₉ is G; X₁₀ is A; X₁₁ is A; X₁₂ is T; and X₁₃ is C.

[0025] In one aspect, provided is a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises:

ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAX₅X₆AX₇ACCCTTTGA
ACCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:5),

wherein

X₅ is G or T;

X₆ is C or T; and

X₇ is C or T.

[0026] In examples, the aptamer encoding sequence has one or more of the following properties: X₅ is not G; X₆ is not C; and X₇ is not C. In examples, all of the following are not simultaneously present in the aptamer encoding sequence: X₅ is G; X₆ is C; and X₇ is C.

[0027] In one aspect, provided is a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises:

ACAGGGGTCCGGCCTTTTCATTTGGCX₄CCGGTGAGAX₅X₆AX₇ACCCTTX₈X₉X₁₀ACCTGTTTACGGATAATGCCGCTGCAGGGAGT (SEQ ID NO:6),

wherein

X₄ is A or G;

X₅ is any nucleotide;

X₆ is any nucleotide;

X₇ is any nucleotide;

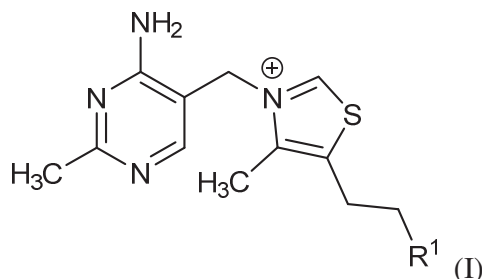
X₈ is any nucleotide;

X₉ is C or G; and

X₁₀ is any nucleotide.

[0028] In one aspect, provided is a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises a sequence that is at least 95% identical to a sequence selected from the group consisting of SEQ ID NOs:7-36 (see Table 1). In some examples, the aptamer encoding sequence comprises a sequence selected from the group consisting of SEQ ID NOs:8, 9, 14-18, 21, 25, 26, and 30. In one example, the aptamer encoding sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 9, 14, and 26.

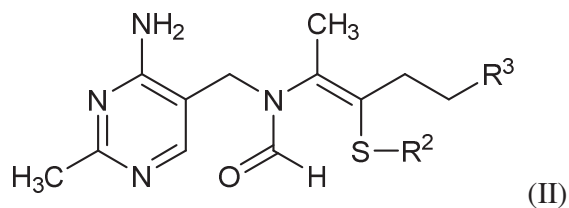
[0029] In some examples, the aptamer binds to, or otherwise responds to the presence of, a small molecule disclosed herein including small molecules having the structure according to Formula I-VIII. In examples, the small molecule has the structure according to Formula I:



wherein:

R¹ is selected from the group consisting of OH, amino, F, Cl, Br, phosphate, pyrophosphate, -O-C(=O)-C₁-C₆ alkyl, -O-C(=O)-C₂-C₆ alkenyl, -O-C(=O)-phenyl, -O-C(=O)-heterocycle, -O-C(=O)-O-C₁-C₆ alkyl, -O-C(=O)-O-C₂-C₆ alkenyl, -O-C(=O)-O-phenyl, and -O-C(=O)-O-heterocycle.

[0030] In examples, the small molecule has the structure according to Formula II:



wherein:

R² is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, -(CH₂)_n-R⁶,

-C(=O)-R⁴, -C(=O)-O-R⁴, -CHR⁵-O-C(=O)-R⁴, -S-C₁-C₆ alkyl, -S-C₂-C₆ alkenyl,
-S-heterocycle, and -S-CH₂-heterocycle;

or R² is -S-[Formula II] such that the compound forms a dimer of two molecules of Formula II connected through a disulfide (-S-S-) linkage;

R³ is selected from the group consisting of OH, amino, F, Cl, Br, phosphate, pyrophosphate,
-O-C(=O)-C₁-C₆ alkyl, -O-C(=O)-C₁-C₆ alkenyl, -O-C(=O)-phenyl,
-O-C(=O)-heterocycle, -O-C(=O)-O-C₁-C₆ alkyl, -O-C(=O)-O-C₁-C₆ alkenyl,
-O-C(=O)-O-phenyl, -O-C(=O)-O-heterocycle;

R⁴ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl,
C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl,
heteroaryl and heterocyclyl;

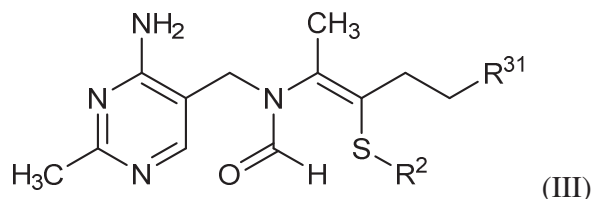
R⁵ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl,
aryl, heteroaryl and heterocyclyl;

R⁶ is hydroxyl, amino, amido, C₁-C₆ alkoxy, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, aryl,
heteroaryl and heterocyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocyclyl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl.

[0031] In examples, the small molecule has the structure according to Formula III:



wherein:

R² is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, -(CH₂)_{*n*}-R⁶,
-C(=O)-R⁴, -C(=O)-O-R⁴, -CHR⁵-O-C(=O)-R⁴, -S-C₁-C₆ alkyl, -S-C₂-C₆ alkenyl,
-S-heterocycle, and -S-CH₂-heterocycle;

R³¹ is selected from the group consisting of OH and phosphate;

R⁴ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;

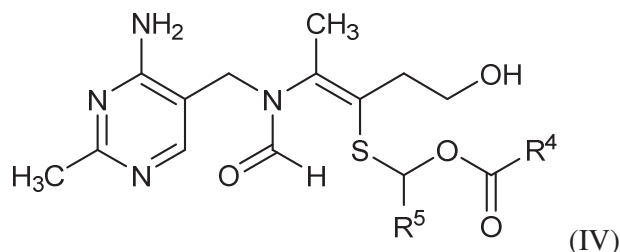
R⁵ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, aryl, heteroaryl and heterocyclyl;

R⁶ is hydroxyl, amino, amido, C₁-C₆ alkoxy, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, aryl, heteroaryl and heterocyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocyclyl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl.

[0032] In examples, the small molecule has the structure according to Formula IV:



wherein:

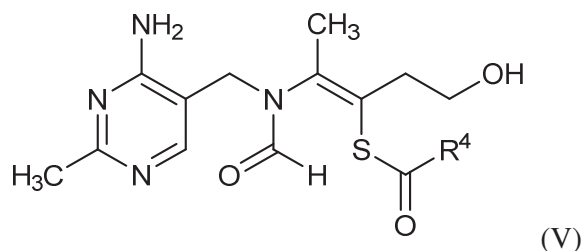
R⁴ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;

R⁵ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, aryl, heteroaryl and heterocyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocyclyl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆

perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl.

[0033] In examples, the small molecule has the structure according to Formula V:

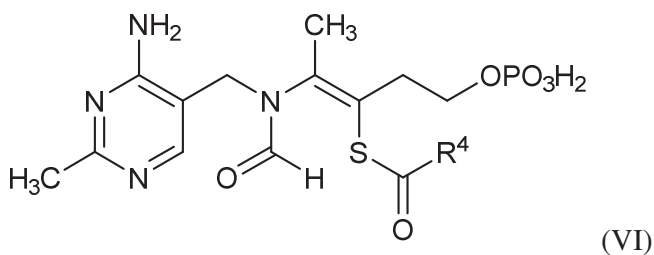


wherein:

R⁴ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;

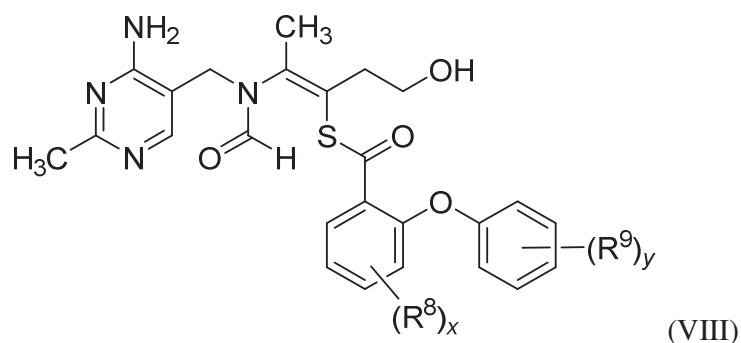
and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocyclyl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl.

[0034] In examples, the small molecule has the structure according to Formula VI:



wherein:

R⁴ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;



wherein:

each R^8 is independently selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_3 - C_7 cycloalkyl, C_6 - C_{10} bicycyl, C_1 - C_6 haloalkyl, C_1 - C_6 perhaloalkyl, -O-(C_1 - C_6 alkyl), O-(C_3 - C_7 cycloalkyl), -O-(C_1 - C_6 haloalkyl), -O-(C_1 - C_6 perhaloalkyl), aryl, -O-aryl, -(C_1 - C_6 alkyl)-aryl, -O-(C_1 - C_6 alkyl)-aryl, -S-(C_1 - C_6 alkyl), -S-(C_3 - C_7 cycloalkyl), -S-(C_1 - C_6 haloalkyl), -S-(C_1 - C_6 perhaloalkyl), -S-aryl, -S-(C_1 - C_6 alkyl)-aryl, heteroaryl and heterocyclyl;

additionally or alternatively, two adjacent R^8 groups may be taken together to form a fused 5- or 6-membered aromatic or non-aromatic ring, which contains 0 to 2 ring heteroatoms, and which is unsubstituted or is substituted by up to four substituents selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_3 - C_7 cycloalkyl, C_6 - C_{10} bicycyl, C_1 - C_6 haloalkyl, C_1 - C_6 perhaloalkyl, -O-(C_1 - C_6 alkyl), O-(C_3 - C_7 cycloalkyl), -O-(C_1 - C_6 haloalkyl), -O-(C_1 - C_6 perhaloalkyl), aryl, -O-aryl, -(C_1 - C_6 alkyl)-aryl, -O-(C_1 - C_6 alkyl)-aryl, -S-(C_1 - C_6 alkyl), -S-(C_3 - C_7 cycloalkyl), -S-(C_1 - C_6 haloalkyl), -S-(C_1 - C_6 perhaloalkyl), -S-aryl, -S-(C_1 - C_6 alkyl)-aryl, heteroaryl and heterocyclyl;

each R^9 is independently selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_3 - C_7 cycloalkyl, C_6 - C_{10} bicycyl, C_1 - C_6 haloalkyl, C_1 - C_6 perhaloalkyl, -O-(C_1 - C_6 alkyl), O-(C_3 - C_7 cycloalkyl), -O-(C_1 - C_6 haloalkyl), -O-(C_1 - C_6 perhaloalkyl), aryl, -O-aryl, -(C_1 - C_6 alkyl)-aryl, -O-(C_1 - C_6 alkyl)-aryl, -S-(C_1 - C_6 alkyl), -S-(C_3 - C_7 cycloalkyl), -S-(C_1 - C_6 haloalkyl), -S-(C_1 - C_6 perhaloalkyl), -S-aryl, -S-(C_1 - C_6 alkyl)-aryl, heteroaryl and heterocyclyl;

additionally or alternatively, two adjacent R^9 groups may be taken together to form a fused 5- or 6-membered aromatic or non-aromatic ring, which contains 0 to 2 ring heteroatoms, and which is unsubstituted or is substituted by up to four substituents selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_3 - C_7 cycloalkyl, C_6 - C_{10} bicycyl, C_1 - C_6 haloalkyl, C_1 - C_6 perhaloalkyl, -O-(C_1 - C_6 alkyl), O-(C_3 - C_7 cycloalkyl), -O-(C_1 - C_6 haloalkyl), -O-(C_1 - C_6 perhaloalkyl), aryl, -O-aryl, -(C_1 - C_6 alkyl)-aryl,

-O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl;

x is 0, 1, 2 or 3; and

y is 0, 1, 2, 3, or 4.

[0037] In some examples, the aptamer binds to, or otherwise responds to the presence of, a small molecule selected from the group consisting of acefurtiamine, acetiamine, allithiamine, amprolium, beclotiamine, benfotiamine, bentiamine, bisbentiamine, cetotiamine, cycotiamine, fursultiamine, monophosphothiamine, octotiamine, oxythiamine, prosultiamine, sulbutiamine, thiamine, thiamine pyrophosphate, and vintiamol. In a preferred example, the aptamer binds to, or otherwise responds to the presence of, a small molecule selected from the group consisting of benfotiamine, fursultiamine, and prosultiamine. In some examples, the aptamer binds to, or otherwise responds to the presence of, benfotiamine. In examples, the aptamer binds to, or otherwise responds to the presence of, fursultiamine. In examples, the aptamer binds to, or otherwise responds to the presence of, prosultiamine.

[0038] In some examples, the aptamer has reduced binding and/or shows a reduced response to thiamine pyrophosphate (TPP) compared to equimolar amounts of fursultiamine, benfotiamine or prosultiamine. In some examples where the aptamer is in the context of a riboswitch encoded as part of a polynucleotide cassette for regulating the expression of a target gene, the aptamer has reduced response to TPP compared to equimolar amounts of fursultiamine, benfotiamine or prosultiamine.

[0039] Also provided is a riboswitch for the regulation of target gene expression in response to a small molecule, wherein the riboswitch comprises an aptamer disclosed herein. In one example, the riboswitch encoding sequence comprises the sequence of SEQ ID NO:37.

[0040] Also provided is a polynucleotide cassette for the regulation of the expression of a target gene in response to a small molecule, the polynucleotide cassette comprising:

- (a) a riboswitch; and
- (b) an alternatively-spliced exon, flanked by a 5' intron and a 3' intron, wherein the riboswitch comprises (i) an effector region comprising a stem that includes the 5' splice site sequence of the 3' intron, and (ii) an aptamer disclosed herein; and wherein the alternatively-spliced exon comprises a stop codon that is in-frame with the target gene when the alternatively-spliced exon is spliced into the target gene mRNA.

[0041] In one example, the polynucleotide cassette comprises a riboswitch encoding sequence comprising the sequence of SEQ ID NO:37 and further comprising an aptamer encoding sequence, wherein the aptamer sequence is selected from an aptamer sequence disclosed herein.

[0042] In one example, provided is a nucleic acid molecule comprising an aptamer, riboswitch, and/or polynucleotide cassette disclosed herein. Also provided is a nucleic acid molecule comprising a target gene containing a riboswitch or a polynucleotide cassette disclosed herein. In one example, the polynucleotide cassette is located in the protein coding sequence of the target gene. In one example, the polynucleotide cassette is located in an untranslated region of the target gene or in an intron of the target gene.

[0043] Also provided is a vector comprising any of the nucleic acid molecules disclosed herein. In one example, the vector is a viral vector. In some examples, the viral vector is selected from the group consisting of an adenoviral vector, an adeno-associated virus vector, and a lentiviral vector.

[0044] In one aspect, provided is a method for identifying an aptamer that modulates target gene expression in response to a compound of interest (e.g., a thiamine analog or derivative, such as a compound according to Formula I-VIII), the method comprising the steps of:

- (i) selecting a parent aptamer sequence;
- (ii) generating a riboswitch library comprising a sequences encoding all or part of the aptamer selected in step (i) wherein the aptamer encoding sequences comprise one or more randomly mutated nucleotides in one or more unpaired regions in the aptamer, wherein the mutated aptamer sequences are in the context of a riboswitch that controls the expression of a reporter gene;
- (iii) screening the library from (ii) for aptamers having increased regulation (e.g., higher fold induction or repression) of the target gene expression in response to exposure to the compound of interest, compared to the parent aptamer sequence;
- (iv) optionally repeating steps (ii) and (iii) (on an aptamer identified in step (iii) instead of an aptamer selected in step (i)).

[0045] In examples, the parent aptamer sequence is a TTP aptamer (including a putative TPP aptamer or a known TPP aptamer). In examples, the parent aptamer sequence is selected from Rfam TPP riboswitch family RF00059. In examples, the one or more unpaired regions of the aptamer sequence are junction (J) regions. In addition or alternatively, the unpaired region may be a loop (L) region. In further examples, the

sequence encoding the aptamer having one or more nucleotides in one or more unpaired regions that are randomly mutated, also has one or more nucleotides in a paired (P) region mutated, for example in one or more paired nucleotides adjacent to an unpaired region.

[0046] In examples, the compound of interest is a thiamine analog or derivative. In examples, the thiamine analog is fursultiamine, benfotiamine or prosultiamine. In one example, the aptamer has reduced binding and/or shows a reduced response to thiamine pyrophosphate (TPP) compared to equimolar amounts of fursultiamine, benfotiamine or prosultiamine. In examples, the compound of interest is a compound according to Formulas I-VIII, In examples, the compound of interest is a compound according to Formulas I-III.

[0047] In aspects, the disclosure provides compounds of Formulas I-VIII. In examples, the disclosure provides compounds of Formulas I-VIII in a pharmaceutical composition.

BRIEF DESCRIPTION OF THE FIGURES

[0048] **Figure 1** illustrates that a synthetic riboswitch TPPz or TPPm comprising the thiC and thiM TPP aptamer, respectively, induce luciferase expression in response to treatment with thiamine pyrophosphate (TPP) or a thiamine analog. HEK 293 cells were transfected with constructs containing the luciferase gene containing the TPPz or TPPm riboswitch. Transfected cells were treated with TPP (**Figure 1A**), fursultiamine (**Figure 1B**), prosultiamine (**Figure 1C**), bisbentiamine, beclotiamine or sulbutiamine (**Figure 1D**), at the doses indicated or left untreated. The luciferase activity was expressed as mean arbitrary light units (ALU) $\pm$ S.D., and fold induction was calculated as the quotient of the luciferase activity recorded for cells exposed to TPP or a thiamine analog divided by the luciferase activity recorded for cells that were not exposed to TPP or a thiamine analog.

[0049] **Figure 2** illustrates that a putative aptamer sequence (14G4) subtracted from a family of TPP aptamers (Rfam family RF00059), regulates luciferase expression in response to fursultiamine. Cells were transfected with a luciferase-riboswitch construct comprising putative TPP aptamer, 14G4 (accession number AACY023654033.1/903-800 from Rfam RF00059). The transfected cells were treated with fursultiamine at the indicated doses in HEK 293 cells (**Figure 2A**), or in other types of human and mouse cell lines (**Figure 2B**). The fold induction was calculated as the quotient of the luciferase activity recorded for cells exposed to fursultiamine divided by the luciferase activity recorded for cells that were not exposed to fursultiamine.

We claim:

1. A polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
ACX₁GGGGTCCGGCX₂TX₃TTCATTTGGCX₄CCGGTGAGAX₅X₆AX₇ACCCTTX₈X₉X₁₀X₁₁CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:1),
and wherein:
X₁ is A or G;
X₂ is C or no nucleotide;
X₃ is T or no nucleotide;
X₄ is A or G;
X₅ is any nucleotide;
X₆ is any nucleotide;
X₇ is any nucleotide;
X₈ is any nucleotide;
X₉ is C, G, or T;
X₁₀ is any nucleotide;
X₁₁ is A or T;
X₁₂ is C or T; and
X₁₃ is C or T.
2. A polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
ACX₁GGGGTCCGGCX₂TX₃TTCATTTGGCGCCGGTGAGAX₅X₆AX₇ACCCTTX₈X₉X₁₀X₁₁CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:2)
and wherein:
X₁ is A or G;
X₂ is C or no nucleotide;
X₃ is T or no nucleotide;
X₅ is G or T;
X₆ is C or T;

X₇ is C or T;
 X₈ is any nucleotide;
 X₉ is G or T;
 X₁₀ is A, G, or T;
 X₁₁ is A or T;
 X₁₂ is C or T; and
 X₁₃ is C or T.

3. A polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
 ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTTGAA
 CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:3) and
 wherein:
 X₁₂ is C or T; and
 X₁₃ is C or T.

4. A polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
 ACX₁GGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTX₈X₉
 X₁₀X₁₁CCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:4) and
 wherein
 X₁ is A or G;
 X₈ is any nucleotide;
 X₉ is G or T;
 X₁₀ is A, G, or T; and
 X₁₁ is A or T.

5. A polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
 ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAX₅X₆AX₇ACCCTTTGA
 ACCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:5) and wherein:

X₅ is G or T;

X₆ is C or T; and

X₇ is C or T.

6. A polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises

ACAGGGGTCCGGCCTTTTCATTTGGCX₄CCGGTGAGAX₅X₆AX₇ACCCTTX₈

X₉X₁₀ACCTGTTACGGATAATGCCGCTGCAGGGAGT (SEQ ID NO:6) and

wherein:

X₄ is A or G;

X₅ is any nucleotide;

X₆ is any nucleotide;

X₇ is any nucleotide;

X₈ is any nucleotide;

X₉ is C or G; and

X₁₀ is any nucleotide.

7. A polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises a sequence that is at least 95% identical to a sequence selected from the group consisting of SEQ ID NOs:7-36.

8. The polynucleotide cassette of claim 7, wherein the aptamer encoding sequence is selected from the group consisting of SEQ ID NOs:7-36.

9. The polynucleotide cassette of claim 8, wherein the aptamer encoding sequence is selected from the group consisting of SEQ ID NOs:8, 9, 14-18, 21, 25, 26, and 30.

10. The polynucleotide cassette of claim 9, wherein the aptamer encoding sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 9, 14, and 26.

11. A nucleic acid sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
ACX₁GGGGTCCGGCX₂TX₃TTCATTTGGCX₄CCGGTGAGAX₅X₆AX₇ACCCTTX₈X₉X₁₀X₁₁CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:1)
and wherein:
X₁ is A or G;
X₂ is C or no nucleotide;
X₃ is T or no nucleotide;
X₄ is A or G;
X₅ is any nucleotide;
X₆ is any nucleotide;
X₇ is any nucleotide;
X₈ is any nucleotide;
X₉ is C, G, or T;
X₁₀ is any nucleotide;
X₁₁ is A or T;
X₁₂ is C or T; and
X₁₃ is C or T.

12. The nucleic acid sequence of claim 11, wherein the aptamer encoding sequence has one or more of the following properties: X₁ is not A; X₂ is not C; X₃ is not T; X₄ is not G; X₅ is not G; X₆ is not C; X₇ is not C; X₈ is not T; X₉ is not G; X₁₀ is not A; X₁₁ is not A; X₁₂ is not T; and X₁₃ is not C.

13. The nucleic acid sequence of claim 11, wherein all of the following are not simultaneously present in the aptamer encoding sequence: X₁ is A; X₂ is C; X₃ is T; X₄ is G; X₅ is G; X₆ is C; X₇ is C; X₈ is T; X₉ is G; X₁₀ is A; X₁₁ is A; X₁₂ is T; and X₁₃ is C.

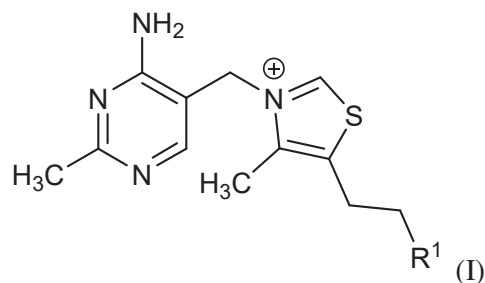
14. A nucleic acid sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
ACX₁GGGGTCCGGCX₂TX₃TTCATTTGGCGCCGGTGAGAX₅X₆AX₇ACCCTTX₈X₉X₁₀X₁₁CCTGTTX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:2)
and wherein:

X₁ is A or G;
 X₂ is C or no nucleotide;
 X₃ is T or no nucleotide;
 X₅ is G or T;
 X₆ is C or T;
 X₇ is C or T;
 X₈ is any nucleotide;
 X₉ is G or T;
 X₁₀ is A, G, or T;
 X₁₁ is A or T;
 X₁₂ is C or T; and
 X₁₃ is C or T.

15. The nucleic acid sequence of claim 14, wherein the aptamer encoding sequence has one or more of the following properties: X₁ is not A; X₂ is not C; X₃ is not T; X₄ is not G; X₅ is not G; X₆ is not C; X₇ is not C; X₈ is not T; X₉ is not G; X₁₀ is not A; X₁₁ is not A; X₁₂ is not T; and X₁₃ is not C.
16. The nucleic acid sequence of claim 14, wherein all of the following are not simultaneously present in the aptamer encoding sequence: X₁ is A; X₂ is C; X₃ is T; X₅ is G; X₆ is C; X₇ is C; X₈ is T; X₉ is G; X₁₀ is A; X₁₁ is A; X₁₂ is T; and X₁₃ is C.
17. A nucleic acid sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
 ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTTGAACCTGT
 TX₁₂ACGGATAATGCCGCX₁₃GCAGGGAGT (SEQ ID NO:3) and wherein:
 X₁₂ is C or T; and
 X₁₃ is C or T.
18. The nucleic acid sequence of claim 17, wherein the aptamer encoding sequence has one or more of the following properties: X₁₂ is not T; and X₁₃ is not C.
19. The nucleic acid sequence of claim 17, wherein all of the following are not simultaneously present in the aptamer encoding sequence: X₁₂ is T and X₁₃ is C.

20. A nucleic acid sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
ACX₁GGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTX₈X₉
X₁₀X₁₁CCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:4) and
wherein:
X₁ is A or G;
X₈ is any nucleotide;
X₉ is G or T;
X₁₀ is A, G, or T; and
X₁₁ is A or T.
21. The nucleic acid sequence of claim 20, wherein the aptamer encoding sequence has one or more of the following properties: X₁ is not A; X₈ is not T; X₉ is not G; X₁₀ is not A; X₁₁ is not A.
22. The nucleic acid sequence of claim 20, wherein all of the following are not simultaneously present in the aptamer encoding sequence: X₁ is A; X₈ is T; X₉ is G; X₁₀ is A; X₁₁ is A; X₁₂ is T; and X₁₃ is C.
23. A nucleic acid sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAX₅X₆AX₇ACCCTTTGA
ACCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:5) and wherein:
X₅ is G or T;
X₆ is C or T; and
X₇ is C or T.
24. The nucleic acid sequence of claim 23, wherein the aptamer encoding sequence has one or more of the following properties: X₅ is not G; X₆ is not C; and X₇ is not C.
25. The nucleic acid sequence of claim 23, wherein all of the following are not simultaneously present in the aptamer encoding sequence: X₅ is G; X₆ is C; and X₇ is C.

26. A nucleic acid sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises
ACAGGGGTCCGGCCTTTTCATTTGGCX₄CCGGTGAGAX₅X₆AX₇ACCCTTX₈
X₉X₁₀ACCTGTTACGGATAATGCCGCTGCAGGGAGT (SEQ ID NO:6) and
wherein:
X₄ is A or G;
X₅ is any nucleotide;
X₆ is any nucleotide;
X₇ is any nucleotide;
X₈ is any nucleotide;
X₉ is C or G; and
X₁₀ is any nucleotide.
27. A nucleic acid sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises a sequence that is at least 95% identical to a sequence selected from the group consisting of SEQ ID NOs:7-36.
28. The nucleic acid sequence of claim 27, wherein the aptamer encoding sequence is selected from the group consisting of SEQ ID NOs:7-36.
29. The nucleic acid sequence of claim 28, wherein the aptamer encoding sequence is selected from the group consisting of SEQ ID NOs:8, 9, 14-18, 21, 25, 26, and 30.
30. The nucleic acid sequence of claim 29, wherein the aptamer encoding sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 9, 14, and 26.
31. The polynucleotide cassette of any one of claims 1-10 or the nucleic acid sequence of any one of claims 11-30, wherein the aptamer binds to, or otherwise responds to the presence of, a small molecule having the structure according to any one of:
- a) Formula I:

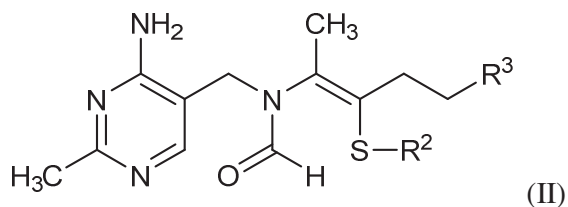


wherein:

R^1 is selected from the group consisting of OH, amino, F, Cl, Br, phosphate, pyrophosphate,

-O-C(=O)-C₁-C₆ alkyl, -O-C(=O)-C₂-C₆ alkenyl, -O-C(=O)-phenyl, -O-C(=O)-heterocycle, -O-C(=O)-O-C₁-C₆ alkyl, -O-C(=O)-O-C₂-C₆ alkenyl, -O-C(=O)-O-phenyl, and -O-C(=O)-O-heterocycle;

b) Formula II:



wherein:

R^2 is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, -(CH₂)_n-R⁶, -C(=O)-R⁴, -C(=O)-O-R⁴, -CHR⁵-O-C(=O)-R⁴, -S-C₁-C₆ alkyl, -S-C₂-C₆ alkenyl, -S-heterocycle, and -S-CH₂-heterocycle;

or R^2 is -S-[Formula II] such that the compound forms a dimer of two molecules of Formula II connected through a disulfide (-S-S-) linkage;

R^3 is selected from the group consisting of OH, amino, F, Cl, Br, phosphate, pyrophosphate,

-O-C(=O)-C₁-C₆ alkyl, -O-C(=O)-C₁-C₆ alkenyl, -O-C(=O)-phenyl, -O-C(=O)-heterocycle, -O-C(=O)-O-C₁-C₆ alkyl, -O-C(=O)-O-C₁-C₆ alkenyl, -O-C(=O)-O-phenyl, -O-C(=O)-O-heterocycle;

R^4 is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;

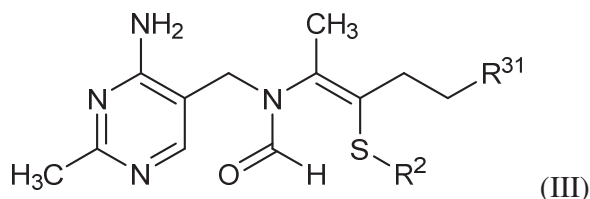
R⁵ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, aryl, heteroaryl and heterocyclyl;

R⁶ is hydroxyl, amino, amido, C₁-C₆ alkoxy, C₃-C₇ cycloalkyl, C₆-C₁₀ bicyclyl, aryl, heteroaryl and heterocyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and heterocyclyl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicyclyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl;

c) Formula III:



wherein:

R² is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, -(CH₂)_{*n*}-R⁶, -C(=O)-R⁴, -C(=O)-O-R⁴, -CHR⁵-O-C(=O)-R⁴, -S-C₁-C₆ alkyl, -S-C₂-C₆ alkenyl, -S-heterocycle, and -S-CH₂-heterocycle;

R³¹ is selected from the group consisting of OH and phosphate;

R⁴ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicyclyl, C₉-C₁₄ tricyclyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;

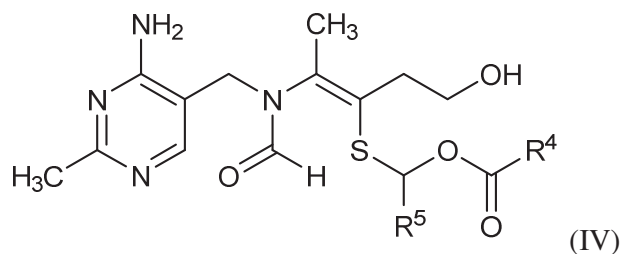
R⁵ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, aryl, heteroaryl and heterocyclyl;

R⁶ is hydroxyl, amino, amido, C₁-C₆ alkoxy, C₃-C₇ cycloalkyl, C₆-C₁₀ bicyclyl, aryl, heteroaryl and heterocyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocycl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocycl;

d) Formula IV:



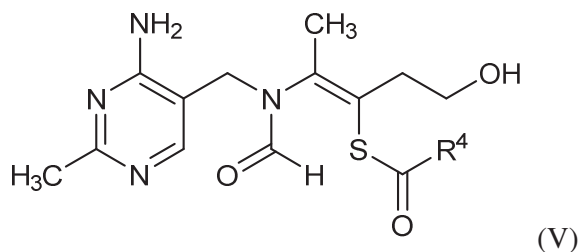
wherein:

R⁴ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocycl;

R⁵ is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, aryl, heteroaryl and heterocycl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocycl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocycl;

e) Formula V:

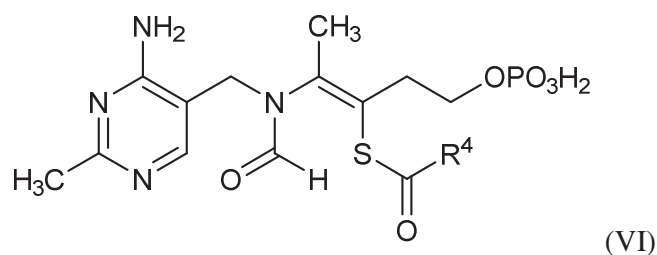


wherein:

R^4 is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocyclyl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl;

f) Formula VI:



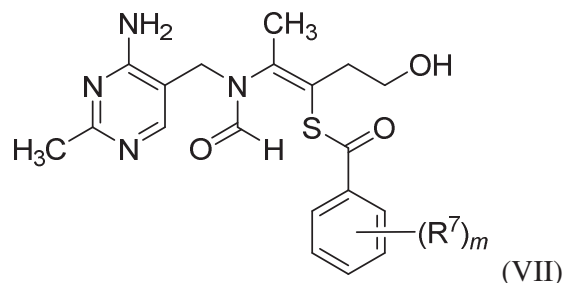
wherein:

R^4 is selected from the group consisting of H, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₉-C₁₄ tricycyl, -(C₁-C₆ alkyl)-aryl, -(C₂-C₆ alkenyl)-aryl, aryl, heteroaryl and heterocyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicycyl, tricycyl, aryl, heteroaryl and heterocyclyl groups may be unsubstituted or substituted by 1 to 3 substituents selected from the group consisting of halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆

alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl;

g) Formula VII



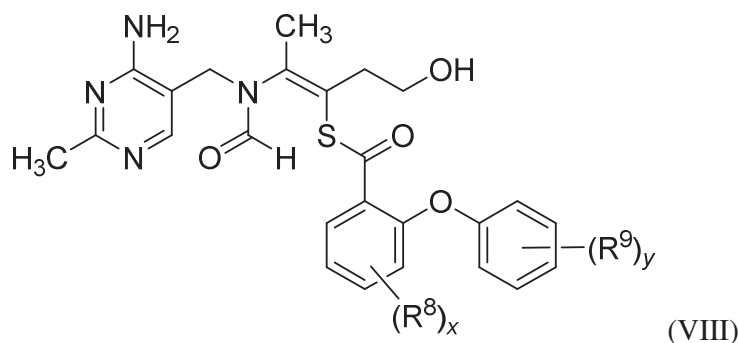
wherein:

each R⁷ is independently selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl;

additionally or alternatively, two adjacent R⁷ groups may be taken together to form a fused 5- or 6-membered aromatic or non-aromatic ring, which contains 0 to 2 ring heteroatoms, and which is unsubstituted or is substituted by up to four substituents selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl; and

m is 0, 1, 2, 3 or 4; and

h) Formula VIII



wherein:

each R^8 is independently selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_3 - C_7 cycloalkyl, C_6 - C_{10} bicycyl, C_1 - C_6 haloalkyl, C_1 - C_6 perhaloalkyl, -O-(C_1 - C_6 alkyl), O-(C_3 - C_7 cycloalkyl), -O-(C_1 - C_6 haloalkyl), -O-(C_1 - C_6 perhaloalkyl), aryl, -O-aryl, -(C_1 - C_6 alkyl)-aryl, -O-(C_1 - C_6 alkyl)-aryl, -S-(C_1 - C_6 alkyl), -S-(C_3 - C_7 cycloalkyl), -S-(C_1 - C_6 haloalkyl), -S-(C_1 - C_6 perhaloalkyl), -S-aryl, -S-(C_1 - C_6 alkyl)-aryl, heteroaryl and heterocyclyl;

additionally or alternatively, two adjacent R^8 groups may be taken together to form a fused 5- or 6-membered aromatic or non-aromatic ring, which contains 0 to 2 ring heteroatoms, and which is unsubstituted or is substituted by up to four substituents selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_3 - C_7 cycloalkyl, C_6 - C_{10} bicycyl, C_1 - C_6 haloalkyl, C_1 - C_6 perhaloalkyl, -O-(C_1 - C_6 alkyl), O-(C_3 - C_7 cycloalkyl), -O-(C_1 - C_6 haloalkyl), -O-(C_1 - C_6 perhaloalkyl), aryl, -O-aryl, -(C_1 - C_6 alkyl)-aryl, -O-(C_1 - C_6 alkyl)-aryl, -S-(C_1 - C_6 alkyl), -S-(C_3 - C_7 cycloalkyl), -S-(C_1 - C_6 haloalkyl), -S-(C_1 - C_6 perhaloalkyl), -S-aryl, -S-(C_1 - C_6 alkyl)-aryl, heteroaryl and heterocyclyl;

each R^9 is independently selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_3 - C_7 cycloalkyl, C_6 - C_{10} bicycyl, C_1 - C_6 haloalkyl, C_1 - C_6 perhaloalkyl, -O-(C_1 - C_6 alkyl), O-(C_3 - C_7 cycloalkyl), -O-(C_1 - C_6 haloalkyl), -O-(C_1 - C_6 perhaloalkyl), aryl, -O-aryl, -(C_1 - C_6 alkyl)-aryl, -O-(C_1 - C_6 alkyl)-aryl, -S-(C_1 - C_6 alkyl), -S-(C_3 - C_7 cycloalkyl), -S-(C_1 - C_6 haloalkyl), -S-(C_1 - C_6 perhaloalkyl), -S-aryl, -S-(C_1 - C_6 alkyl)-aryl, heteroaryl and heterocyclyl;

additionally or alternatively, two adjacent R^9 groups may be taken together to form a fused 5- or 6-membered aromatic or non-aromatic ring, which contains 0 to 2

ring heteroatoms, and which is unsubstituted or is substituted by up to four substituents selected from halogen, hydroxyl, amino, cyano, amido, sulfonamide, nitro, -SH, C₁-C₆ alkyl, C₂-C₆ alkenyl, C₃-C₇ cycloalkyl, C₆-C₁₀ bicycyl, C₁-C₆ haloalkyl, C₁-C₆ perhaloalkyl, -O-(C₁-C₆ alkyl), O-(C₃-C₇ cycloalkyl), -O-(C₁-C₆ haloalkyl), -O-(C₁-C₆ perhaloalkyl), aryl, -O-aryl, -(C₁-C₆ alkyl)-aryl, -O-(C₁-C₆ alkyl)-aryl, -S-(C₁-C₆ alkyl), -S-(C₃-C₇ cycloalkyl), -S-(C₁-C₆ haloalkyl), -S-(C₁-C₆ perhaloalkyl), -S-aryl, -S-(C₁-C₆ alkyl)-aryl, heteroaryl and heterocyclyl;

x is 0, 1, 2 or 3; and

y is 0, 1, 2, 3, or 4.

32. The polynucleotide cassette of any one of claims 1-10 or the nucleic acid sequence of any one of claims 11-30, wherein the aptamer binds to, or otherwise responds to the presence of, a small molecule selected from the group consisting of acefurfuramine, acetiamine, allithiamine, amprolium, beclotiamine, benfotiamine, bentiamine, bisbentiamine, cetotiamine, cycotiamine, fursultiamine, monophosphothiamine, octotiamine, oxythiamine, prosultiamine, sulbutiamine, thiamine, thiamine pyrophosphate, and vintiamol.
33. The polynucleotide cassette or the nucleic acid sequence of claim 32, wherein the aptamer binds to, or otherwise responds to the presence of, a small molecule selected from the group consisting of benfotiamine, fursultiamine, and prosultiamine.
34. The polynucleotide cassette or the nucleic acid sequence of claim 33, wherein the aptamer binds to, or otherwise responds to the presence of, benfotiamine.
35. The polynucleotide cassette or the nucleic acid sequence of claim 33, wherein the aptamer binds to, or otherwise responds to the presence of, fursultiamine.
36. The polynucleotide cassette or the nucleic acid sequence of claim 33, wherein the aptamer binds to, or otherwise responds to the presence of, prosultiamine.

37. The polynucleotide cassette of any one of claims 1-10 or the nucleic acid sequence of any one of claims 11-30, wherein the aptamer has reduced binding and/or shows a reduced response to thiamine pyrophosphate (TPP) compared to equimolar amounts of fursultiamine, benfotiamine or prosultiamine.
38. The polynucleotide cassette of or the nucleic acid sequence of claim 37, wherein the aptamer is in the context of a riboswitch encoded as part of a polynucleotide cassette for regulating the expression of a target gene.
39. A riboswitch for the regulation of target gene expression in response to a small molecule, wherein the riboswitch comprises the aptamer of any one of claims 1-38.
40. The riboswitch of claim 39, the riboswitch comprises the sequence of SEQ ID NO:37.
41. A polynucleotide cassette for the regulation of the expression of a target gene in response to a small molecule, the polynucleotide cassette comprising:
(a) a riboswitch; and
(b) an alternatively-spliced exon, flanked by a 5' intron and a 3' intron,
wherein the riboswitch comprises (i) an effector region comprising a stem that includes the 5' splice site sequence of the 3' intron, and (ii) the aptamer of any one of claims 1-38;
and
wherein the alternatively-spliced exon comprises a stop codon that is in-frame with the target gene when the alternatively-spliced exon is spliced into the target gene mRNA.
42. The polynucleotide cassette of claim 41, wherein the riboswitch comprises the sequence of SEQ ID NO:37.
43. The polynucleotide cassette of claim 41, wherein the polynucleotide cassette is located in the protein coding sequence of the target gene.
44. The polynucleotide cassette of claim 41, wherein the polynucleotide cassette is located in an untranslated region of the target gene or in an intron of the target gene.

- 45. A vector comprising the polynucleotide cassette of any one of claims 1-10, 31-38 or 41-44, the nucleic acid of any one of claims 11-38, or the riboswitch of any one of claims 39 or 40.
- 46. The vector of claim 45, wherein the vector is a viral vector.
- 47. The vector of claim 46, wherein the viral vector is selected from the group consisting of an adenoviral vector, an adeno-associated virus vector, and a lentiviral vector.