APTAMERS THAT BIND THIAMINE ANALOGS AND DERIVATIVES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Application No. 62/994,135, filed March 24, 2020, which is encorparted herein by reference in its entirety.

FIELD

[0002] The present disclosure relates to oligonucleotide aptamers that bind to certain small molecules and methods of generating aptamers that bind to the small molecules. Also contemplated are riboswitches and polynucleotide cassettes for regulating the expression of a target gene, wherein the polynucleotide cassettes comprise the aptamers disclosed herein. Further provided are small molecules that are modulators of target gene expression where the target gene contains a riboswitch comprising an aptamer described herein.

BACKGROUND

[0003] Aptamers are oligonucleotides that bind to a target ligand with high affinity and specificity. These nucleic acid sequences have proven to be of high therapeutic and diagnostic value with recent FDA approval of the first aptamer drug and additional ones in the clinical pipelines. Their high degree of specificity and versatility have established RNA aptamers as one of the pivotal tools of the emerging RNA nanotechnology field in the fight against human diseases including cancer, viral infections and other diseases.

[0004] In addition, aptamers may be utilized as part of a riboswitch that has certain effects in the presence or absence of an aptamer ligand. For example, riboswitches may be used to regulate gene expression in response to the presence or absence of the aptamer ligand.

[0005] However, aptamers derived from prokaryotic sources or generated using *in vitro* selection methods often fail to demonstrate the functionality required for the expression of therapeutic targets genes in eukaryotic systems. As such, novel aptamer sequences able to regulate gene expression in response to the presence or absence of a small molecule ligand are needed.

SUMMARY

[0006] Provided herein are aptamer sequences that bind to small molecules, such as thiamine pyrophosphate (TPP) and analogs or derivatives thereof. Also contemplated are

riboswitches and polynucleotide cassettes for regulating the expression of a target gene, wherein the polynucleotide cassettes comprise the aptamers disclosed herein. Further provided are methods of using said aptamers, riboswitches, and/or polynucleotide cassettes for the regulation of target genes, including therapeutic genes. Also provided herein are small molecules that are modulators of target gene expression where the target gene contains a riboswitch comprising an aptamer described herein.

[0007] In one aspect, 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:

$ACX_1GGGGTCCGGCX_2TX_3TTCATTTGGCX_4CCGGTGAGAX_5X_6AX_7ACCCTTX_8X_9X_{10}X_{11}CCTGTTX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT$ (SEQ ID NO:1),

wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is T or no nucleotide;

X_4 is A or G;

X_5 is any nucleotide;

X_6 is any nucleotide;

X_7 is any nucleotide;

X_8 is any nucleotide;

X_9 is C, G, or T;

X_{10} is any nucleotide;

X_{11} is A or T;

X_{12} is C or T; and

X_{13} is C or T.

[0008] In one aspect, 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

$ACX_1GGGGTCCGGCX_2TX_3TTCATTTGGCGCCGGTGAGAX_5X_6AX_7ACCCTTX_8X_9X_{10}X_{11}CCTGTTX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT$ (SEQ ID NO:2),

wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is T or no nucleotide;

X_5 is G or T;

X_6 is C or T;

X_7 is C or T;

X_8 is any nucleotide;

X_9 is G or T;

X_{10} is A, G, or T;

X_{11} is A or T;

X_{12} is C or T; and

X_{13} is C or T.

[0009] In one aspect, 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

ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTTGAA CCTGTTX$_{12}$ACGGATAATGCCGCX$_{13}$GCAGGGAGT (SEQ ID NO:3),

wherein

X_{12} is C or T; and

X_{13} is C or T.

[0010] In one aspect, 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

ACX$_1$GGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTX$_8$X$_9$X$_{10}$X$_{11}$CCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:4),

wherein

X_1 is A or G;

X_8 is any nucleotide;

X_9 is G or T;

X_{10} is A, G, or T; and

X_{11} is A or T.

[0011] In one aspect, 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

ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAX$_5$X$_6$AX$_7$ACCCTTTGA ACCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:5),

wherein

X_5 is G or T;

X_6 is C or T; and

X_7 is C or T.

[0012] In one aspect, 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

ACAGGGGTCCGGCCTTTTCATTTGGCX$_4$CCGGTGAGAX$_5$X$_6$AX$_7$ACCCTTX$_8$X$_9$X$_{10}$ACCTGTTCACGGATAATGCCGCTGCAGGGAGT (SEQ ID NO:6),

wherein

X_4 is A or G;

X_5 is any nucleotide;

X_6 is any nucleotide;

X_7 is any nucleotide;

X_8 is any nucleotide;

X_9 is C or G; and

X_{10} is any nucleotide.

[0013] In one aspect, provided is a polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprisises 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 one aspect, provided is a polynucleotide cassette for regulating the expression of a target gene, wherein the polynucleotide cassette comprisises a sequence encoding an aptamer that binds to a small molecule, wherein the aptamer encoding sequence comprises a sequence sequence selected from the group consisting of SEQ ID NOs:7-36 (see Table 1).

[0014] In one aspect, 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:8, 9, 14-18, 21, 25, 26, and 30. In one aspect, 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:8, 9, 14-18, 21, 25, 26, and 30.

[0015] In a preferred embodiment, 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 embodiment, 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:

$ACX_1GGGGTCCGGCX_2TX_3TTCATTTGGCX_4CCGGTGAGAX_5X_6AX_7ACCCTTX_8X_9X_{10}X_{11}CCTGTTX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT$ (SEQ ID NO:1),

wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is T or no nucleotide;

X_4 is A or G;

X_5 is any nucleotide;

X_6 is any nucleotide;

X_7 is any nucleotide;

X_8 is any nucleotide;

X_9 is C, G, or T;

X_{10} is any nucleotide;

X_{11} is A or T;

X_{12} is C or T; and

X_{13} is C or T.

[0017] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_1 is not A; X_2 is not C; X_3 is not T; X_4 is not G; X_5 is not G; X_6 is not C; X_7 is not C; X_8 is not T; X_9 is not G; X_{10} is not A; X_{11} is not A; X_{12} is not T; and X_{13} is not C.

[0018] In embodiments, all of the following are not simultaneously present in the aptamer encoding sequence: X_1 is A; X_2 is C; X_3 is T; X_4 is G; X_5 is G; X_6 is C; X_7 is C; X_8 is T; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is T; and X_{13} 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:

$ACX_1GGGGTCCGGCX_2TX_3TTCATTTGGCGCCGGTGAGAX_5X_6AX_7ACCCTTX_8$
$X_9X_{10}X_{11}CCTGTTX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT$ (SEQ ID NO:2),
wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is T or no nucleotide;

X_5 is G or T;

X_6 is C or T;

X_7 is C or T;

X_8 is any nucleotide;

X_9 is G or T;

X_{10} is A, G, or T;

X_{11} is A or T;

X_{12} is C or T; and

X_{13} is C or T.

[0020] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_1 is not A; X_2 is not C; X_3 is not T; X_4 is not G; X_5 is not G; X_6 is not C; X_7 is not C; X_8 is not T; X_9 is not G; X_{10} is not A; X_{11} is not A; X_{12} is not T; and X_{13} is not C. In embodiments, all of the following are not simultaneously present in the aptamer encoding sequence: X_1 is A; X_2 is C; X_3 is T; X_5 is G; X_6 is C; X_7 is C; X_8 is T; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is T; and X_{13} 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
$CCTGTTX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT$ (SEQ ID NO:3),
wherein

X_{12} is C or T; and

X_{13} is C or T.

[0022] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_{12} is not T; and X_{13} is not C. In embodiments, all of the following are not simultaneously present in the aptamer encoding sequence: X_{12} is T and X_{13} 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:

ACX$_1$GGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTX$_8$X$_9$X$_{10}$X$_{11}$CCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:4),

wherein

X_1 is A or G;

X_8 is any nucleotide;

X_9 is G or T;

X_{10} is A, G, or T; and

X_{11} is A or T.

[0024] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_1 is not A; X_8 is not T; X_9 is not G; X_{10} is not A; X_{11} is not A. In embodiments, all of the following are not simultaneously present in the aptamer encoding sequence: X_1 is A; X_8 is T; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is T; and X_{13} 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:

ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAX$_5$X$_6$AX$_7$ACCCTTTGAACCTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:5),

wherein

X_5 is G or T;

X_6 is C or T; and

X_7 is C or T.

[0026] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_5 is not G; X_6 is not C; and X_7 is not C. In embodiments, all of the following are not simultaneously present in the aptamer encoding sequence: X_5 is G; X_6 is C; and X_7 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:

ACAGGGGTCCGGCCTTTTCATTTGGCX$_4$CCGGTGAGAX$_5$X$_6$AX$_7$ACCCTTX$_8$X$_9$X$_{10}$ACCTGTTCACGGATAATGCCGCTGCAGGGAGT (SEQ ID NO:6),

wherein

X_4 is A or G;

X_5 is any nucleotide;

X_6 is any nucleotide;

X_7 is any nucleotide;

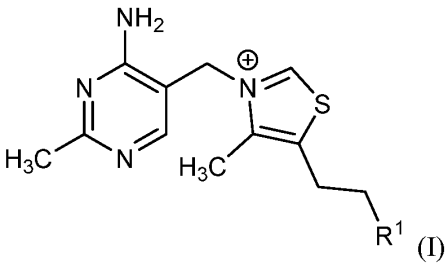
X_8 is any nucleotide;

X_9 is C or G; and

X_{10} 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 embodiments, 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 embodiment, the aptamer encoding sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 9, 14, and 26.

[0029] In some embodiments, 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 embodiments, the small molecule has the structure according to Formula I:

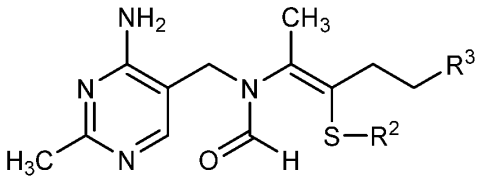


(I)

wherein:

R^1 is selected from the group consisting of OH, amino, F, Cl, Br, phosphate, pyrophosphate, -O-C(=O)-C_1-C_6 alkyl, -O-C(=O)-C_2-C_6 alkenyl, -O-C(=O)-phenyl, -O-C(=O)-heterocycle, -O-C(=O)-O-C_1-C_6 alkyl, -O-C(=O)-O-C_2-C_6 alkenyl, -O-C(=O)-O-phenyl, and -O-C(=O)-O-heterocycle.

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



(II)

wherein:

R^2 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, -$(CH_2)_n$-R^6,

-C(=O)-R^4, -C(=O)-O-R^4, -CHR^5-O-C(=O)-R^4, -S-C_1-C_6 alkyl, -S-C_2-C_6 alkenyl, -S-heterocycle, and -S-CH_2-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)-C_1-C_6 alkyl, -O-C(=O)-C_1-C_6 alkenyl, -O-C(=O)-phenyl, -O-C(=O)-heterocycle, -O-C(=O)-O-C_1-C_6 alkyl, -O-C(=O)-O-C_1-C_6 alkenyl, -O-C(=O)-O-phenyl, -O-C(=O)-O-heterocycle;

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

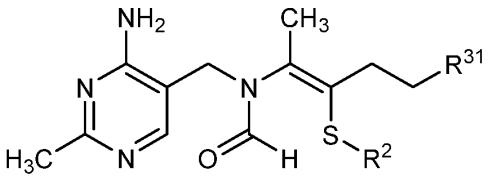
R^5 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, aryl, heteroaryl and hetercyclyl;

R^6 is hydroxyl, amino, amido, C_1-C_6 alkoxy, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, aryl, heteroaryl and hetercyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

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



(III)

wherein:

R^2 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, -$(CH_2)_n$-R^6, -C(=O)-R^4, -C(=O)-O-R^4, -CHR^5-O-C(=O)-R^4, -S-C_1-C_6 alkyl, -S-C_2-C_6 alkenyl, -S-heterocycle, and -S-CH_2-heterocycle;

R^{31} is selected from the group consisting of OH and phosphate;

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

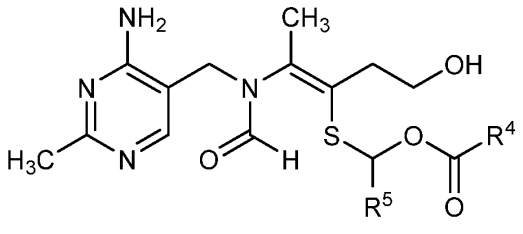
R^5 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, aryl, heteroaryl and hetercyclyl;

R^6 is hydroxyl, amino, amido, C_1-C_6 alkoxy, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, aryl, heteroaryl and hetercyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

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



(IV)

wherein:

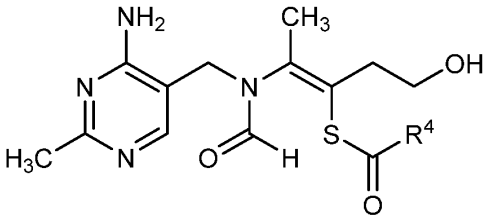
R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl ,-(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

R^5 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, aryl, heteroaryl and hetercyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

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



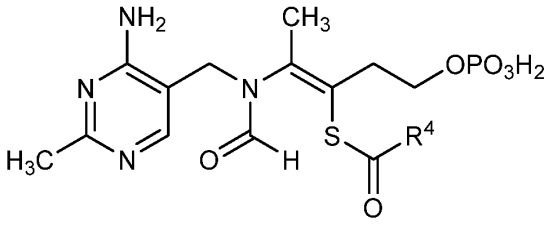
(V)

wherein:

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

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



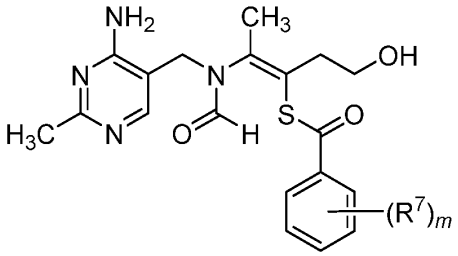
(VI)

wherein:

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

[0035] In some embodiments, the small molecule has the structure according to Formula VII:



(VII)

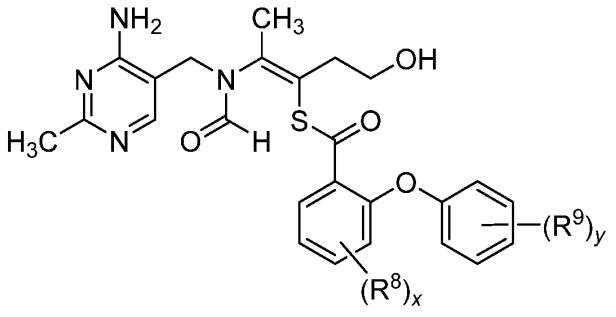
wherein

each R^7 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} bicyclyl, 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 hetercyclyl;

additionally or alternatively, two adjacent R^7 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} bicyclyl, 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 hetercyclyl; and

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

[0036] In some embodiments, the small molecule has the structure according to Formula VIII:



(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} bicyclyl, 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 hetercyclyl;

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} bicyclyl, 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 hetercyclyl;

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} bicyclyl, 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 hetercyclyl;

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} bicyclyl, 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 hetercyclyl;

x is 0, 1, 2 or 3; and

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

[0037] In some embodiments, 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 embodiment, 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 embodiments, the aptamer binds to, or otherwise responds to the presence of, benfotiamine. In embodiments, the aptamer binds to, or otherwise responds to the presence of, fursultiamine. In embodiments, the aptamer binds to, or otherwise responds to the presence of, prosultiamine.

[0038] In some embodiments, 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 embodiments 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 embodiment, 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 embodiment, 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 embodiment, 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 embodiment, the polynucleotide cassette is located in the protein coding sequence of the target gene. In one embodiment, 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 embodiment, the vector is a viral vector. In some embodiments, 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 derivate, 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 embodiments, the parent aptamer sequence is a TTP aptamer (including a putative TPP aptamer or a known TPP aptamer). In embodiments, the parent aptamer sequence is selected from Rfam TPP riboswitch family RF00059. In embodiments, 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 embodiments, 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 embodiments, the compound of interest is a thiamine analog or derivative. In embodiments, the thiamine analog is fursultiamine, benfotiamine or prosultiamine. In one embodiment, 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 embodiments, the compound of interest is a compound according to Formulas I-VIII, In embodiments, the compound of interest is a compound according to Formulas I-III.

[0047] In aspects, the disclosure provides compounds of Formulas I-VIII. In embodiments, 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) ± 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.

[0050] **Figure 3** shows the predicted structure of parent aptamer 14G4 with the native P1 stem (SEQ ID NO:38) (**Figure 3A**) and the aptamer 14G4 encoding sequence (SEQ ID NO:7) and libraries derived therefrom (**Figure 3B**). The predicted structure of the 14G4 aptamer, including positions of the stems and unpaired regions was determined as described herein. The structure shading around each base denotes its base-pairing probabilities with darker shading indicating higher probability. For unpaired regions the darker shading denotes the probability of being unpaired. The nucleotides that were randomly mutated to obtain aptamer libraries A1 (SEQ ID NO:39), A2 (SEQ ID NO:40), or A3 (SEQ ID NO:41), respectively are indicated with "N" in the aptamer coding sequences in **Figure 3B** and correspond to nucleotides that are unpaired or adjacent to unpaired regions.

[0051] **Figure 4** illustrates the improvement of gene regulation activity through mutagenesis of the aptamers contained within riboswitches. Riboswitch constructs comprising randomly mutated aptamers (derived from aptamer libraries A1, A2, or A3) were screened for improved gene regulation activity in HEK cells. Transfected HEK cells were treated with 50 μM fursultiamine or left untreated. Examples of riboswitches comprising re-engineered aptamer sequences that were isolated in the screen and that demonstrated improved gene regulation activity as compared to riboswitch 14G4 are shown in **Figure 4A**. **Figure 4B** illustrates that riboswitches comprising re-engineered aptamer sequences are useful for the dose-dependent induction of a target gene in different types of cells. AML12, C2C12 and ARPE-19 cells were transfected with re-engineered riboswitch constructs using TransIT-X2 reagent. The transfected cells were treated with fursultiamine at the indicated doses. **Figure 4C** illustrates that riboswitches comprising re-engineered aptamer sequences regulate luciferase expression in response to different TPP analogs. AML12 cells were transfected with riboswitches comprising re-engineered aptamer sequences and were treated with fursultiamine, benfotiamine, or prosultiamine at the indicated doses. **Figures 4A-4C**: The 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 to a thiamine analog.

[0052] **Figure 5** shows the nucleotide sequence of the 3H4 aptamer encoding sequence (SEQ ID NO:9), which was obtained through mutagenesis of 14G4. Nucleotides that were randomly mutated to obtain aptamer library A4 (SEQ ID NO:42) are indicated with "N" in the aptamer encoding sequence.

[0053] **Figure 6A** illustrates that mutagenesis of select aptamer nucleotides involved in aptamer/ligand binding can improve the gene regulation activity of aptamer-based

riboswitches in mammalian cells. HEK 293, ARPE-19 and AML12 cells were transfected with riboswitch constructs comprising re-engineered aptamer sequences isolated from aptamer library A4. The transfected cells were treated with fursultiamine at the doses indicated or left untreated. **Figure 6B** illustrates the ability of riboswitches comprising re-engineered aptamer sequences to induce expression of luciferase in response to different thiamine derivatives that share chemical structural features with thiamine. AML12 cells were transfected with the indicated riboswitch constructs and treated with fursultiamine, benfotiamine, or prosultiamine at the doses indicated or left untreated. The fold induction was calculated as the quotient of the luciferase activity obtained from cells treated with thiamine analog compounds divided by the luciferase activity obtained from cells without thiamine analog treatment.

[0054] **Figure 7** illustrates the ability of riboswitches comprising re-engineered aptamer sequences to induce expression of a variety of target genes in response to a ligand. **Figures 7A and 7B** illustrate that a riboswitch comprising aptamer 6B4 regulates enhanced green fluorescent protein (EGFP) expression in response to fursultiamine treatment. Stably transfected HEK 293 cells containing the EGFP-6B4 construct were treated with fursultiamine at 50 μM (**Figure 7A**) or at the indicated doses (**Figure 7B**) for 24 hours or left untreated. **Figures 7B and 7C** illustrate that a riboswitch comprising 3H4 or 15D10, respectively, regulate mouse erythropoietin (mEpo) expression in response to fursultiamine treatment. AML12 cells were transfected with Epo-3H4 or Epo-15D10 constructs. The transfected cells were treated with fursultiamine at the indicated doses or left untreated. The expression of mEpo was detected and quantified using mEpo ELISA and was expressed as mean ± S.D. (**Figure 7C**). The fold induction was calculated as the quotient of the amount of mEpo produced from cells treated with fursultiamine divided by the amount of mEpo produced from cells without fursultiamine treatment (**Figure 7D**).

[0055] **Figure 8** illustrates the inducible expression of luciferase by riboswitches comprising re-engineered aptamer sequences *in vivo*. Adeno-associated AAV2/8 viral particles were produced containing an intron-alternative exon-intron cassette with (1) a non-regulatable riboswitch without aptamer ("Control 1"), (2) a riboswitch cassette comprising aptamer 3H4 ("Switch 3H4" or "3H4"), or (3) a riboswitch cassette comprising aptamer 6B4 ("Switch 6B4" or "6B4"). Balb/c mice (n=5 each group) were injected with single tail vein injection of 1.0×10^{11} (**Figures 8A, 8B, and 8D**) or 2.5×10^{11} (**Figures 8C and 8E**) viral particles per mouse. Twenty-eight days after AAV delivery, mice were injected with 50 mg/kg prosultiamine intraperitoneally. Luciferase activity was measured the day prior to drug

dosing, as well as 6 h, 24 h, 48 h, and 72 h after drug dosing. After the first administration of prosultiamine, the mice were subjected to three additional rounds of dosing and imaging cycles as follows: Day 36 (after AAV administration): 100 mg/kg; day 43: 200 mg/kg; and day 51: 400 mg/kg. **Figure 8A** shows luciferase expression at the time points and for the doses of prosultiamine indicated for mice transfected with AAV vectors comprising the luciferase gene with the re-engineered riboswitches 3H4 or 6B4. **Figures 8B** and **8C** show the luciferase fluorescence intensity observed for the different groups the day prior to drug dosing, as well as 6 h, 24 h, 48 h and 72 after dosing. **Figures 8D** and **8E** show the relative increase in luciferase expression upon prosultiamine treatment after drug dosing for the different drug doses and groups. The fold induction of luciferase expression was calculated as the quotient of photon/s obtained from mice treated with prosultiamine divided by the value obtained from mice one day before prosultiamine treatment.

[0056] **Figure 9** shows induction of luciferase target gene expression in response to fursultiamine or benfotiamine compared to additional thiamine analogs of Formulas IV, V, and VI. The luciferase target gene contained a sequence encoding a riboswitch comprising the 15D10 aptamer. Group IV compounds were tested for their ability to induce luciferase expression in HEK 293 (**Figure 9A**) and AML 12 (**Figure 9B**) cells. Group V and Group VI compounds were tested for their ability to induce luciferase expression in HEK 293 (**Figure 9C**) and AML 12 (**Figure 9D**) cells. The most potent compounds were further tested for their ability to induce luciferase expression in ARPE-19 (**Figure 9E**) and Hep G2 (**Figure 9F**) cells.

DETAILED DESCRIPTION

[0057] Provided herein are aptamer sequences that bind to, or otherwise respond to the presence of, small molecules, such as thiamine or TPP and analogs or derivatives of thiamine or TPP. In some embodiments, the aptamer sequences provided herein are useful for the regulation of the expression of a target gene in response to an analog or derivative of thiamine or TPP. Also contemplated are riboswitches comprising the aptamer sequences disclosed herein, as well as polynucleotide cassettes for regulating the expression of a target gene, wherein the polynucleotide cassettes comprise sequences encoding the riboswitches disclosed herein. Also provided herein are methods of using the aptamers, riboswitches, and/or polynucleotide cassettes for the regulation of target genes, including therapeutic genes, and for the treatment of subjects in need thereof.

[0058] **Aptamers**

[0059] Aptamers are single-stranded nucleic acid molecules that non-covalently bind to specific ligands with high affinity and specificity by folding into three-dimensional structures. Aptamer ligands include ions, small molecules, proteins, viruses, and cells. Aptamer ligands can be, for example, an organic compound, amino acid, steroid, carbohydrate, or nucleotide. Non-limiting examples of small molecule aptamer ligands include antibiotics, therapeutics, dyes, cofactors, metabolites, molecular markers, neurotransmitters, pollutants, toxins, food adulterants, carcinogens, drugs of abuse. As such, aptamers are useful for the detection of small molecules. Application of small-molecule detection by aptamers include environmental monitoring, food safety, medicine (including diagnostics), microbiology, analytical chemistry, forensic science, agriculture, and basic biology research.

[0060] The term "aptamer" as used herein refers to an RNA polynucleotide (or DNA sequence encoding the RNA polynucleotide) that specifically binds to a class of ligands. The term "ligand" refers to a molecule that is specifically bound by an aptamer. Aptamers have binding regions that are capable of forming complexes with an intended target molecule (i.e., the ligand). An aptamer will typically be between about 15 and about 200 nucleotides in length. More commonly, an aptamer will be between about 30 and about 100 nucleotides in length, for example, 70 to 90 nucleotides in length. Aptamers typically comprise multiple paired (P) regions in which the aptamer forms a stem and unpaired regions where the aptamer forms a joining (J) region or a loop (L) region. The paired regions can be numbered sequentially starting at the 5′ end (P1) and numbering each stem sequentially (P2, P3, etc.). The loops (L1, L2, etc.) are numbered based on the adjacent paired region and the joining regions are numbered according to the paired regions that they link.

[0061] In one embodiment, the aptamer encoding sequence comprises $ACX_1GGGGTCCGGCX_2TX_3TTCATTTGGCX_4CCGGTGAGAX_5X_6AX_7ACCCTTX_8X_9X_{10}X_{11}CCTGTTX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT$ (SEQ ID NO:1),

wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is T or no nucleotide;

X_4 is A or G;

X_5 is any nucleotide;

X_6 is any nucleotide;

X_7 is any nucleotide;

X_8 is any nucleotide;

X_9 is C, G, or T;

X_{10} is any nucleotide;

X_{11} is A or T;

X_{12} is C or T; and

X_{13} is C or T.

[0062] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_1 is not A; X_2 is not C; X_3 is not T; X_4 is not G; X_5 is not G; X_6 is not C; X_7 is not C; X_8 is not T; X_9 is not G; X_{10} is not A; X_{11} is not A; X_{12} is not T; and X_{13} is not C. In embodiments, all of the following are not simultaneously present in the aptamer encoding sequence: X_1 is A; X_2 is C; X_3 is T; X_5 is G; X_6 is C; X_7 is C; X_8 is T; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is T; and X_{13} is C.

[0063] In one embodiment, the aptamer encoding sequence comprises $ACX_1GGGGTCCGGCX_2TX_3TTCATTTGGCGCCGGTGAGAX_5X_6AX_7ACCCTTX_8X_9X_{10}X_{11}CCTGTTX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT$ (SEQ ID NO:2),

wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is T or no nucleotide;

X_5 is G or T;

X_6 is C or T;

X_7 is C or T;

X_8 is any nucleotide;

X_9 is G or T;

X_{10} is A, G, or T;

X_{11} is A or T;

X_{12} is C or T; and

X_{13} is C or T.

[0064] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_1 is not A; X_2 is not C; X_3 is not T; X_4 is not G; X_5 is not G; X_6 is not C; X_7 is not C; X_8 is not T; X_9 is not G; X_{10} is not A; X_{11} is not A; X_{12} is not T; and X_{13} is not C. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_1 is A; X_2 is C; X_3 is T; X_5 is G; X_6 is C; X_7 is C; X_8 is T; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is T; and X_{13} is C.

[0065] In one embodiment, the aptamer encoding sequence comprises: ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTTTGAACCTGT TX_{12}ACGGATAATGCCGCX_{13}GCAGGGAGT (SEQ ID NO:3),

wherein

X_{12} is C or T; and

X_{13} is C or T.

[0066] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_{12} is not T; and X_{13} is not C. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X12 is T and X13 is C.

[0067] In one embodiment, the aptamer encoding sequence comprises: ACX_1GGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACACCCTT$X_8X_9X_{10}X_{11}$C CTGTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:4),

wherein

X_1 is A or G;

X_8 is any nucleotide;

X_9 is G or T;

X_{10} is A, G, or T; and

X_{11} is A or T.

[0068] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_1 is not A; X_8 is not T; X_9 is not G; X_{10} is not A; X_{11} is not A. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_1 is A; X_8 is T; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is T; and X_{13} is C.

[0069] In one embodiment, the aptamer encoding sequence comprises: ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAX_5X_6AX_7ACCCTTTGAACCT GTTTACGGATAATGCCGCCGCAGGGAGT (SEQ ID NO:5),

wherein

X_5 is G or T;

X_6 is C or T; and

X_7 is C or T.

[0070] In embodiments, the aptamer encoding sequence has one or more of the following properties: X_5 is not G; X_6 is not C; and X_7 is not C. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_5 is G; X_6 is C; and X_7 is C.

[0071] In one embodiment, the aptamer encoding sequence comprises: ACAGGGGTCCGGCCTTTTCATTTGGCX$_4$CCGGTGAGAX$_5$X$_6$AX$_7$ACCCTTX$_8$X$_9$X$_{10}$AC CTGTTCACGGATAATGCCGCTGCAGGGAGT (SEQ ID NO:6),

wherein

X$_4$ is A or G;

X$_5$ is any nucleotide;

X$_6$ is any nucleotide;

X$_7$ is any nucleotide;

X$_8$ is any nucleotide;

X$_9$ is C or G; and

X$_{10}$ is any nucleotide.

[0072] In one embodiment, the aptamer encoding sequence comprises a sequence selected from the group consisting of SEQ ID NOs:7-36 (see Table 1). In one embodiment, the aptamer encoding sequence comprises a sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to a sequence selected from the group consisting of SEQ ID NOs:7-36. "Percent sequence identity" with respect to a reference polypeptide or nucleic acid sequence is defined as the percentage of amino acid residues or nucleotides in a candidate sequence that are identical with the amino acid residues or nucleotides in the reference polypeptide or nucleic acid sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for purposes of determining percent amino acid or nucleic acid sequence identity can be achieved in ways known to the ordinarily-skilled artisan, for example, using publicly available computer software programs including BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software.

Table 1. Aptamer coding sequences

SEQ ID NO:	Aptamer	Aptamer encoding sequence
7	14G4 (parent)	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTTGAACCTGTTTACGGATAATGCCGCCGCAGGGAGT
8	1D10	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATTATACC CTTTGAACCTGTTTACGGATAATGCCGCCGCAGGGAGT
9	3H4	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTTGAACCTGTTCACGGATAATGCCGCTGCAGGGAGT
10	3F10	ACGGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTCGGACCTGTTTACGGATAATGCCGCCGCAGGGAGT

11	3H9	ACAGGGGTCCGGCTTTTCATTTGGCGCCGGTGAGAGCACACC CTTATGACCTGTTTACGGATAATGCCGCCGCAGGGAGT
12	4G2	ACGGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTTGGACCTGTTTACGGATAATGCCGCCGCAGGGAGT
13	6D2	ACGGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTTGTTCCTGTTTACGGATAATGCCGCCGCAGGGAGT
14	6B4	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTGTGACCTGTTTACGGATAATGCCGCCGCAGGGAGT
15	4H2	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTCGGACCTGTTTACGGATAATGCCGCCGCAGGGAGT
16	6C4	ACGGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTCGAACCTGTTTACGGATAATGCCGCCGCAGGGAGT
17	6G12	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTTGGACCTGTTTACGGATAATGCCGCCGCAGGGAGT
18	8F1	ACGGGGGTCCGGCCTTTCATTTGGCGCCGGTGAGAGCACACC CTTCGGACCTGTTTACGGATAATGCCGCCGCAGGGAGT
19	10A7	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATGATAC CCTTTGGACCTGTTCACGGATAATGCCGCTGCAGGGAGT
20	12D5	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGCACAC CCTTTGGACCTGTTCACGGATAATGCCGCTGCAGGGAGT
21	12G7	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATTATACC CTTCGGACCTGTTCACGGATAATGCCGCTGCAGGGAGT
22	12H3	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATCACAC CCTTACTACCTGTTCACGGATAATGCCGCTGCAGGGAGT
23	13H7	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAACAGAC CCTTTGCACCTGTTCACGGATAATGCCGCTGCAGGGAGT
24	13B6	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAACACAC CCTTTGTACCTGTTCACGGATAATGCCGCTGCAGGGAGT
25	15A5	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATTATACC CTTACTACCTGTTCACGGATAATGCCGCTGCAGGGAGT
26	15D10	ACAGGGGTCCGGCCTTTTCATTTGGCACCGGTGAGAACATAC CCTTCGGACCTGTTCACGGATAATGCCGCTGCAGGGAGT
27	15F9	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATTACACC CTTAGCACCTGTTCACGGATAATGCCGCTGCAGGGAGT
28	16E5	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATCAAAC CCTTGGCACCTGTTCACGGATAATGCCGCTGCAGGGAGT
29	16G8	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGTACAC CCTTCGCACCTGTTCACGGATAATGCCGCTGCAGGGAGT
30	16G6	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATCACAC CCTTGGTACCTGTTCACGGATAATGCCGCTGCAGGGAGT
31	17E2	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATTACACC CTTTGGACCTGTTCACGGATAATGCCGCTGCAGGGAGT
32	17G1	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGATCACAC CCTTGGAACCTGTTCACGGATAATGCCGCTGCAGGGAGT
33	17D3	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGGATAC CCTTCGGACCTGTTCACGGATAATGCCGCTGCAGGGAGT
34	17F5	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGTATAC CCTTAGTACCTGTTCACGGATAATGCCGCTGCAGGGAGT

35	17G3	ACAGGGGTCCGGCCTTTTCATTTGGCACCGGTGAGACAATAC CCTTGGTACCTGTTCACGGATAATGCCGCTGCAGGGAGT
36	18G9	ACAGGGGTCCGGCCTTTTCATTTGGCGCCGGTGAGAGAATAC CCTTGGTACCTGTTCACGGATAATGCCGCTGCAGGGAGT

[0073] The ordinarily-skilled artisan would understand that the aptamers described herein may be ribonucleic acid (RNA) molecules. In embodiments, the aptamers described herein are part of a longer RNA polynucleotide, including, for example, hnRNA, mRNA, siRNA, or miRNA.

[0074] In one aspect, provided is an aptamer comprising the sequence: $ACX_1GGGGUCCGGCX_2UX_3UUCAUUUGGCX_4CCGGUGAGAX_5X_6AX_7ACCCUUX_8X_9X_{10}X_{11}CCUGUUX_{12}ACGGAUAAUGCCGCX_{13}GCAGGGAGU$ (SEQ ID NO:43),

wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is U or no nucleotide;

X_4 is A or G;

X_5 is any nucleotide;

X_6 is any nucleotide;

X_7 is any nucleotide;

X_8 is any nucleotide;

X_9 is C, G, or U;

X_{10} is any nucleotide;

X_{11} is A or U;

X_{12} is C or U; and

X_{13} is C or U.

[0075] In embodiments, the aptamer sequence has one or more of the following properties: X_1 is not A; X_2 is not C; X_3 is not U; X_4 is not G; X_5 is not G; X_6 is not C; X_7 is not C; X_8 is not U; X_9 is not G; X_{10} is not A; X_{11} is not A; X_{12} is not U; and X_{13} is not C.

[0076] In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_1 is A; X_2 is C; X_3 is U; X_4 is G; X_5 is G; X_6 is C; X_7 is C; X_8 is U; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is U; and X_{13} is C.

[0077] In one aspect, provided is an aptamer comprising the sequence: $ACX_1GGGGUCCGGCX_2UX_3UUCAUUUGGCGCCGGUGAGAX_5X_6AX_7ACCCUUX_8X_9X_{10}X_{11}CCUGUUX_{12}ACGGAUAAUGCCGCX_{13}GCAGGGAGU$ (SEQ ID NO:44),

wherein

X_1 is A or G;

X_2 is C or no nucleotide;

X_3 is U or no nucleotide;

X_5 is G or U;

X_6 is C or U;

X_7 is C or U;

X_8 is any nucleotide;

X_9 is G or U;

X_{10} is A, G, or U;

X_{11} is A or U;

X_{12} is C or U; and

X_{13} is C or U.

[0078] In embodiments, the aptamer sequence has one or more of the following properties: X_1 is not A; X_2 is not C; X_3 is not U; X_4 is not G; X_5 is not G; X_6 is not C; X_7 is not C; X_8 is not U; X_9 is not G; X_{10} is not A; X_{11} is not A; X_{12} is not U; and X_{13} is not C. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_1 is A; X_2 is C; X_3 is U; X_5 is G; X_6 is C; X_7 is C; X_8 is U; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is U; and X_{13} is C.

[0079] In one aspect, provided is an aptamer comprising the sequence: ACAGGGGUCCGGCCUUUUCAUUUGGCGCCGGUGAGAGCACACCCUUUGAACCU GUUX$_{12}$ACGGAUAAUGCCGCX$_{13}$GCAGGGAGU (SEQ ID NO:45),

wherein

X_{12} is C or U; and

X_{13} is C or U.

[0080] In embodiments, the aptamer sequence has one or more of the following properties: X_{12} is not U; and X_{13} is not C. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_{12} is U and X_{13} is C.

[0081] In one aspect, provided is an aptamer comprising the sequence: ACX$_1$GGGGUCCGGCCUUUUCAUUUGGCGCCGGUGAGAGCACACCCUUX$_8$X$_9$X$_{10}$ X$_{11}$CCUGUUUACGGAUAAUGCCGCCGCAGGGAGU (SEQ ID NO:46),

wherein

X_1 is A or G;

X_8 is any nucleotide;

X_9 is G or U;

X_{10} is A, G, or U; and

X_{11} is A or U.

[0082] In embodiments, the aptamer sequence has one or more of the following properties: X_1 is not A; X_8 is not U; X_9 is not G; X_{10} is not A; and X_{11} is not A. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_1 is A; X_8 is U; X_9 is G; X_{10} is A; X_{11} is A; X_{12} is U; and X_{13} is C.

[0083] In one aspect, provided is an aptamer comprising the sequence: ACAGGGGUCCGGCCUUUUCAUUUGGCGCCGGUGAGAX$_5$X$_6$AX$_7$ACCCUUUGAAC CUGUUUACGGAUAAUGCCGCCGCAGGGAGU (SEQ ID NO:47),

wherein

X_5 is G or U;

X_6 is C or U; and

X_7 is C or U.

[0084] In embodiments, the aptamer sequence has one or more of the following properties: X_5 is not G; X_6 is not C; and X_7 is not C. In embodiments, all of the following are not simultaneously present in the aptamer sequence: X_5 is G; X_6 is C; and X_7 is C.

[0085] In one aspect, provided is an aptamer comprising the sequence: ACAGGGGUCCGGCCUUUUCAUUUGGCX$_4$CCGGUGAGAX$_5$X$_6$AX$_7$ACCCUUX$_8$X$_9$X$_{10}$ ACCUGUUCACGGAUAAUGCCGCUGCAGGGAGU (SEQ ID NO:48),

wherein

X_4 is A or G;

X_5 is any nucleotide;

X_6 is any nucleotide;

X_7 is any nucleotide;

X_8 is any nucleotide;

X_9 is C or G; and

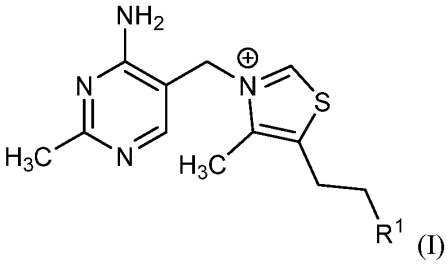
X_{10} is any nucleotide.

[0086] In one aspect, provided is an aptamer that binds to a small molecule, wherein the aptamer comprises a sequence that is at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, identical to the sequence selected from the group consisting of SEQ ID NOs:49-78. In some embodiments, the aptamer sequence comprises a sequence selected from the group consisting of SEQ ID NOs:49-78. In embodiments, the aptamer sequence comprises a sequence selected from the group consisting of SEQ ID NOs:51, 56, and 68.

[0087] Aptamer Ligands

[0088] In one embodiment, the aptamer ligand is thiamine (vitamin B1) or is a thiamine analog and/or is a derivative of thiamine. As used herein, the term "thiamine analog" refers to a molecule that has similar physical, chemical, biochemical, or pharmacological properties compared to thiamine, and includes, for example, amprolium or cycotiamine. A thiamine analog may be a thiamine derivative. The term "thiamine derivative", as used herein, refers to a compound derived from thiamine, or its thiazole ring-opened form, by modification or substitution. Thiamine derivatives may include, for example, acefurtiamine, acetiamine, allithiamine, beclotiamine, benfotiamine, bentiamine, bisbentiamine, cetotiamine, fursultiamine, monophosphothiamine, octotiamine, prosultiamine, sulbutiamine, thiamine, thiamine pyrophosphate, or vintiamol.

[0089] In some embodiments, the aptamer binds to, or otherwise responds to the presence or addition of, a small molecule disclosed herein including small molecules having the structure according to Formula I-VIII. In embodiments, the small molecule has the structure according to Formula I:

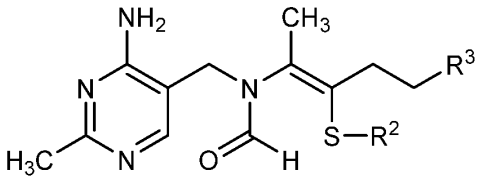


(I)

wherein:

R^1 is selected from the group consisting of OH, amino, F, Cl, Br, phosphate, pyrophosphate, -O-C(=O)-C_1-C_6 alkyl, -O-C(=O)-C_2-C_6 alkenyl, -O-C(=O)-phenyl, -O-C(=O)-heterocycle, -O-C(=O)-O-C_1-C_6 alkyl, -O-C(=O)-O-C_2-C_6 alkenyl, -O-C(=O)-O-phenyl, and -O-C(=O)-O-heterocycle.

[0090] In embodiments, the small molecule has the structure according to Formula II:



(II)

wherein:

R^2 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, -$(CH_2)_n$-R^6,

-C(=O)-R^4, -C(=O)-O-R^4, -CHR^5-O-C(=O)-R^4, -S-C_1-C_6 alkyl, -S-C_2-C_6 alkenyl, -S-heterocycle, and -S-CH_2-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)-C_1-C_6 alkyl, -O-C(=O)-C_1-C_6 alkenyl, -O-C(=O)-phenyl, -O-C(=O)-heterocycle, -O-C(=O)-O-C_1-C_6 alkyl, -O-C(=O)-O-C_1-C_6 alkenyl, -O-C(=O)-O-phenyl, -O-C(=O)-O-heterocycle;

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

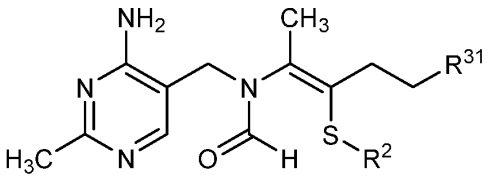
R^5 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, aryl, heteroaryl and hetercyclyl;

R^6 is hydroxyl, amino, amido, C_1-C_6 alkoxy, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, aryl, heteroaryl and hetercyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

[0091] In embodiments, the small molecule has structure according to Formula III:



(III)

wherein:

R^2 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, -$(CH_2)_n$-R^6, -C(=O)-R^4, -C(=O)-O-R^4, -CHR^5-O-C(=O)-R^4, -S-C_1-C_6 alkyl, -S-C_2-C_6 alkenyl, -S-heterocycle, and -S-CH_2-heterocycle;

R^{31} is selected from the group consisting of OH and phosphate;

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

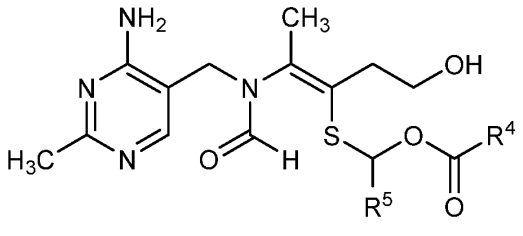
R^5 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, aryl, heteroaryl and hetercyclyl;

R^6 is hydroxyl, amino, amido, C_1-C_6 alkoxy, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, aryl, heteroaryl and hetercyclyl; and

n is 1 to 8;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

[0092] In embodiments, the small molecule has the structure according to Formula IV:



(IV)

wherein:

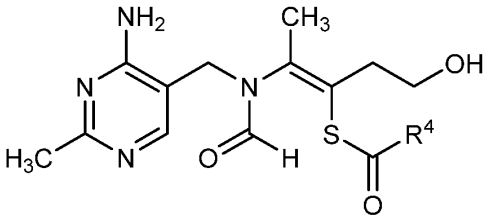
R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

R^5 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, aryl, heteroaryl and hetercyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

[0093] In embodiments, the small molecule has the structure according to Formula V:



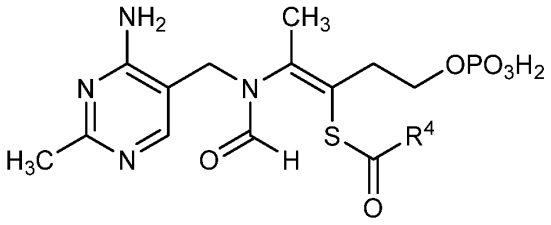
(V)

wherein:

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

[0094] In embodiments, the small molecule has the structure according to Formula VI:



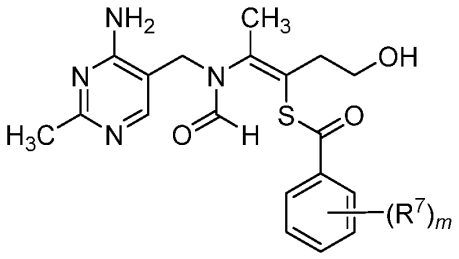
(VI)

wherein:

R^4 is selected from the group consisting of H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, C_9-C_{14} tricyclyl, -(C_1-C_6 alkyl)-aryl, -(C_2-C_6 alkenyl)-aryl, aryl, heteroaryl and hetercyclyl;

and wherein each of the alkyl, alkenyl, cycloalkyl, bicyclyl, tricyclyl, aryl, heteroaryl and hetercyclyl 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, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_3-C_7 cycloalkyl, C_6-C_{10} bicyclyl, 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 hetercyclyl.

[0095] In embodiments, the small molecule has the structure according to Formula VII:



(VII)

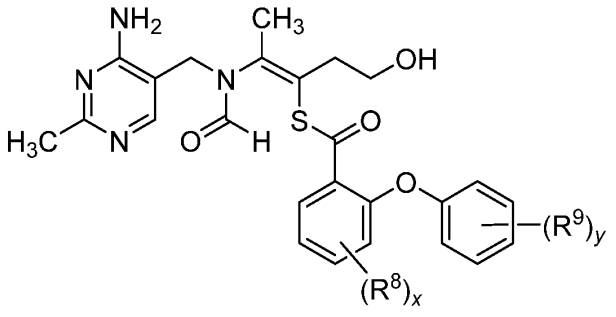
wherein

each R^7 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} bicyclyl, 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 hetercyclyl;

additionally or alternatively, two adjacent R^7 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} bicyclyl, 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 hetercyclyl; and

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

[0096] In some embodiments, the small molecule has the structure according to Formula VIII:



(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} bicyclyl, 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 hetercyclyl;

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} bicyclyl, 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 hetercyclyl;

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} bicyclyl, 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 hetercyclyl;

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} bicyclyl, 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 hetercyclyl;

x is 0, 1, 2 or 3; and

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

[0097] The term "alkyl" refers to the radical of saturated aliphatic groups, including straight-chain alkyl groups and branched-chain alkyl groups. In preferred embodiments, a straight chain or branched chain alkyl has 6 or fewer carbon atoms in its backbone (e.g., C_1-C_6 for straight chain, C_3-C_6 for branched chain). Alkyl groups include methyl, ethyl, propyl, isopropyl, n-butyl, iso-butyl, tert-butyl, pentyl, isopentyl, hexyl, and the like.

[0098] The term "cycloalkyl" refers to saturated, carbocyclic groups having from 3 to 6 carbons in the ring. Cycloalkyl groups include cyclopropyl, cyclobutyl, cyclopentyl and cyclohexyl.

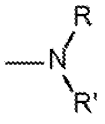
[0099] The term “bicyclyl” refers to saturated carbocyclic groups having two joined ring systems, which may be fused or bridged. Bicyclic groups include bicycle[2.1.1]hexane, bicycle[2.2.1]heptane, decalin, and the like. The term “tricyclyl” refers to saturated carbocyclic groups having three joined ring systems, which may be fused and/or bridged. Tricyclic groups include adamantane and the like.

[0100] The term “alkenyl” refers to unsaturated aliphatic groups, including straight-chain alkenyl groups and branched-chain alkenyl groups, having at least one carbon-carbon double bond. In preferred embodiments, the alkenyl group has two to six carbon atoms (e.g., C_2-C_6 alkenyl).

[0101] As used herein, the term "halogen" or "halo" designates -F, -Cl, -Br or -I, and preferably -F, -Cl or -Br.

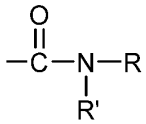
[0102] The terms "alkoxyl" or "alkoxy" as used herein refers to an alkyl group, as defined above, that is attached through an oxygen atom. Representative alkoxyl groups include methoxy, ethoxy, propyloxy, tert-butoxy and the like.

[0103] The terms "amine" and "amino" refer to both unsubstituted and substituted amines, e.g., a moiety that can be represented by the general formula:



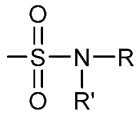
wherein R and R' are each independently selected from H and C_1-C_3 alkyl.

[0104] The terms "amido" refer to both unsubstituted and substituted amide substituents, e.g., a moiety that can be represented by the general formula:



wherein R and R' are each independently selected from H and C_1-C_3 alkyl.

[0105] The terms "sulfonamide" or "sulfonamido" refer to both unsubstituted and substituted sulfonamide substituents, e.g., a moiety that can be represented by the general formula:



wherein R and R' are each independently selected from H and C_1-C_3 alkyl.

[0106] The term "aryl" as used herein includes 5- and 6-membered single-ring aromatic groups that may include from zero to four heteroatoms, for example, benzene, pyrene, pyrrole, furan, thiophene, imidazole, oxazole, thiazole, triazole, pyrazole, pyridine, pyrazine, pyridazine and pyrimidine. Those aryl groups having heteroatoms in the ring structure may also be referred to as "aryl heterocycles" or "heteroaryl" groups. The term "aryl" also includes polycyclic ring systems having two or more cyclic rings in which two or more carbons are common to two adjoining rings (the rings are "fused rings") wherein at least one of the rings is aromatic. Accordingly, aryl includes 8- to 10-membered fused bicyclic aromatic groups that may include from zero to five heteroatoms, in which one or both rings are aromatic, for example napthylene, quinolone, isoquinoline, benzo[b]thiophene, tetrahydronapthelene, and the like. Each aryl group may be unsubstituted or may be substituted with 1 to 5 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} bicyclyl, 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 hetercyclyl.

[0107] The term "heterocycle" of "heterocyclyl" refer to non-aromatic heterocycles having from 1 to 3 ring heteroatoms. Preferred heterocycles are 5- and 6-membered heterocyclic groups having from 1 to 3 heteroatoms selected from the group consisting of O, N and S.

[0108] The term "heteroatom" as used herein means an atom of any element other than carbon or hydrogen. Preferred heteroatoms are nitrogen, oxygen, and sulfur.

[0109] As used herein, the definition of each expression, e.g. alkyl, R^1, R^2, etc., when it occurs more than once in any structure, is intended to be independent of its definition elsewhere in the same structure.

[0110] It will be understood that "substitution" or "substituted with" includes the implicit proviso that such substitution is in accordance with permitted valence of the substituted atom and the substituent, and that the substitution results in a stable compound, e.g., which does not spontaneously undergo transformation such as by rearrangement, cyclization, elimination, etc.

[0111] The aptamer ligands disclosed herein may exist in particular geometric or stereoisomeric forms well as mixtures thereof. Such geometric or stereoisomeric forms include, but not limited to, cis- and trans-isomers, R- and S-enantiomers, diastereomers, (D)-isomers, (L)-isomers, the racemic mixtures thereof, and other mixtures thereof.. Additional asymmetric carbon atoms may be present in a substituent such as an alkyl group.

[0112] The compounds according to Formulas I to VIII may contain an acidic or basic functional group, and accordingly may be present in a salt form. Preferably, the salt form is a pharmaceutically acceptable salt. The term "pharmaceutically-acceptable salts" in this respect, refers to the relatively non-toxic, inorganic and organic acid and base addition salts of the compounds disclosed herein.

[0113] The compounds according to Formulas I to VIII may contain one or more basic functional group, such as amino or alkylamino, and are, thus, capable of forming pharmaceutically-acceptable salts with pharmaceutically-acceptable acids. These salts can be prepared *in situ* in the administration vehicle or the dosage form manufacturing process, or by separately reacting a purified compound disclosed herein in its free base form with a suitable organic or inorganic acid, and isolating the salt thus formed during subsequent purification. Representative salts include the hydrobromide, hydrochloride, sulfate, bisulfate, phosphate, nitrate, acetate, valerate, oleate, palmitate, stearate, laurate, benzoate, lactate, phosphate, tosylate, citrate, maleate, fumarate, succinate, tartrate, napthylate, mesylate, glucoheptonate, lactobionate, and laurylsulphonate salts and the like (*see, e.g.*, Berge et al. (1977) "Pharmaceutical Salts", J. Pharm. Sci. 66:1-19).

[0114] The pharmaceutically acceptable salts of the subject compounds include the conventional nontoxic salts or quaternary ammonium salts of the compounds, e.g., from non-toxic organic or inorganic acids. For example, such conventional nontoxic salts include those

derived from inorganic acids such as hydrochloride, hydrobromic, sulfuric, sulfamic, phosphoric, nitric, and the like; and the salts prepared from organic acids such as acetic, propionic, succinic, glycolic, stearic, lactic, malic, tartaric, citric, ascorbic, palmitic, maleic, hydroxymaleic, phenylacetic, glutamic, benzoic, salicyclic, sulfanilic, 2-acetoxybenzoic, fumaric, toluenesulfonic, methanesulfonic, ethane disulfonic, oxalic, isothionic, and the like.

[0115] In other cases, the compounds according to Formulas I to VIII may contain one or more acidic functional groups and, thus, are capable of forming pharmaceutically-acceptable salts with pharmaceutically-acceptable bases. These salts can likewise be prepared *in situ* in the administration vehicle or the dosage form manufacturing process, or by separately reacting the purified compound in its free acid form with a suitable base, such as the hydroxide, carbonate or bicarbonate of a pharmaceutically-acceptable metal cation, with ammonia, or with a pharmaceutically-acceptable organic primary, secondary or tertiary amine. Representative alkali or alkaline earth salts include the lithium, sodium, potassium, calcium, magnesium, and aluminum salts and the like. Representative organic amines useful for the formation of base addition salts include ethylamine, diethylamine, ethylenediamine, ethanolamine, diethanolamine, piperazine and the like (*see, e.g.*, Berge et al., supra).

[0116] In embodiments, the aptamers provided herein bind to, or otherwise respond to the presence of, one or more thiamine or TPP analogs provided herein, and/or bind to, or otherwise respond to, a metabolite of the thiamine or TPP analog or derivative provided herein, including for example TPP and/or thiamine.

[0117] The specificity of the binding of an aptamer to its ligand can be defined in terms of the comparative dissociation constants (K_d) of the aptamer for its ligand as compared to the dissociation constant of the aptamer for unrelated molecules. Thus, the ligand is a molecule that binds to the aptamer with greater affinity than to unrelated material. Typically, the K_d for the aptamer with respect to its ligand will be at least about 10-fold less than the K_d for the aptamer with unrelated molecules. In other embodiments, the K_d will be at least about 20-fold less, at least about 50-fold less, at least about 100-fold less, and at least about 200-fold less, at least about 500-fold less, at least about 1000-fold less, or at least about 10,000-fold less than the K_d for the aptamer with unrelated molecules.

[0118] In one embodiment, the aptamer binds to TPP with an affinity that is least 5-fold, at least 10-fold, at least about 20-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1000-fold, or at least 10,000-fold lower than the affinity of said aptamer to a compound of Formula I-VI1 or to acefurtiamine, acetiamine, allithiamine, beclotiamine, benfotiamine, bentiamine, bisbentiamine, cetotiamine, cycotiamine, fursultiamine, monophosphothiamine,

octotiamine, prosultiamine, sulbutiamine, or vintiamol. In one embodiment, the aptamer binds to thiamine with an affinity that is least 5-fold, at least 10-fold, at least about 20-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1000-fold, or at least 10,000 fold lower than the affinity of said aptamer a compound of Formula I-VI1 or to acefurtiamine, acetiamine, allithiamine, beclotiamine, benfotiamine, bentiamine, bisbentiamine, cetotiamine, cycotiamine, fursultiamine, monophosphothiamine, octotiamine, prosultiamine, sulbutiamine, or vintiamol.

[0119] Aptamers for the regulation of gene expression

[0120] In some embodiments, the aptamers contemplated by the disclosure are used for the regulation of gene expression. Regulation of the expression of a target gene (e.g., a therapeutic transgene) is advantageous in a variety of situations. In the context of the therapeutic expression of genes, for example, techniques that enable regulated expression of transgenes in response to the presence of a small molecule can enhance safety and efficacy by allowing for the regulation of the level of target gene expression and its timing. In a research setting, the regulation of gene expression allows a systematic investigation of different experimental conditions.

[0121] In embodiments, the sequence encoding the aptamer is part of a gene regulation cassette that provides the ability to regulate the expression level of a target gene in response to the presence or absence of a small molecule described herein. In embodiments, the gene regulation cassette further comprises a target gene. As used herein, “target gene” refers to a transgene that is expressed in response to the presence or absence of the small molecule ligands disclosed herein due to the small molecule binding to the aptamers disclosed herein. In embodiments, the target gene comprises the coding sequence for a protein (e.g., a therapeutic protein), a miRNA, or a siRNA. The target gene is heterologous to the aptamer used for the regulation of target gene expression, is heterologous to the poynucleotide cassette used for the regulation of target gene and/or is heterologous to a portion of the polynucleotide cassette used for the regulation of target gene.

[0122] When used to regulate the expression of a target gene in response to the presence/absence of a ligand, the aptamers described herein can be part of a polynucleotide cassette that encodes the aptamer as part of a riboswitch. The terms “gene regulation cassette”, “regulatory cassette”, or “polynucleotide cassette” are used interchangeably herein.

[0123] In embodiments, the presence of a small molecule that binds to an aptamer disclosed herein leads to an increase in expression of a target gene as compared to the expression of the target gene in absence of the small molecule. In such an embodiment, the

aptamer constitutes an "on" switch. In embodiments, the expression of the target gene is increased by at least 3-fold, by at least 5-fold, by at least 10-fold, by at least 15-fold, by at least 20-fold, by at least 25-fold, by at least 30-fold, by at least 40-fold, by at least 50-fold, by at least 100-fold, by at least 1000-fold, or by at least 10,000-fold in presence of the small molecule that binds to an aptamer disclosed herein as compared to in absence of the small molecule. In embodiments, the expression of the target gene is increased by between 2-fold and 10-fold, between 5-fold and 10-fold, between 5-fold and 15-fold, between 5-fold and 20-fold, between 5-fold and 25-fold, between 5-fold and 30-fold, between 10-fold and 20-fold, between 10-fold and 30-fold, between 10-fold and 40-fold, between 10-fold and 50-fold, between 10-fold and 100-fold, between 10-fold and 500-fold, between 10-fold and 1,000-fold, between 50-fold and 100-fold, between 50-fold and 500-fold, between 50-fold and 100-fold, between 50-fold and 1,000-fold, between 100-fold and 1,000-fold, or between 100-fold and 10,000-fold in presence of the small molecule that binds to an aptamer disclosed herein as compared to in absence of the small molecule.

[0124] In embodiments, the presence of a small molecule that binds to an aptamer disclosed herein leads to a decrease in expression of a target gene as compared to the expression of the target gene in the absence of the small molecule. In such embodiments, the aptamer constitutes an "off" switch. In embodiments, the expression of the target gene is decreased by at least 3-fold, by at least 5-fold, by at least 10-fold, by at least 15-fold, by at least 20-fold, by at least 25-fold, by at least 30-fold, by at least 40-fold, by at least 50-fold, by at least 100-fold, by at least 1000-fold, or by at least 10,000-fold in presence of the small molecule that binds to an aptamer disclosed herein as compared to in absence of the small molecule. In one embodiment, the expression of the target gene is decreased by between 2-fold and 10-fold, between 5-fold and 10-fold, between 5-fold and 15-fold, between 5-fold and 20-fold, between 5-fold and 25-fold, between 5-fold and 30-fold, between 10-fold and 20-fold, between 10-fold and 30-fold, between 10-fold and 40-fold, between 10-fold and 50-fold, between 10-fold and 100-fold, between 10-fold and 500-fold, between 10-fold and 1,000-fold, between 50-fold and 100-fold, between 50-fold and 500-fold, between 50-fold and 100-fold, between 50-fold and 1,000-fold, between 100-fold and 1,000-fold, or between 100-fold and 10,000-fold in presence of the small molecule that binds to an aptamer disclosed herein as compared to in absence of the small molecule.

[0125] In embodiments, the aptamer is part of a riboswitch. Riboswitches are regulatory segments of an RNA polynucleotide that regulate the stability of the RNA polynucleotide and/or regulate the production of a protein from the RNA polynucleotide in response to the

presence or absence of aptamer-specific ligand molecules. In embodiments, the riboswitch comprises a sensor region (e.g., the aptamer region) and an effector region that together are responsible for sensing the presence of a ligand (e.g., a small molecule) and causing an effect that leads to increased or decreased expression of the target gene. The riboswitches described herein are recombinant, utilizing polynucleotides from two or more sources. In embodiments, the sensor and effector regions are joined by a polynucleotide linker. In embodiments, the polynucleotide linker forms a RNA stem or paired region (i.e., a region of the RNA polynucleotide that is double-stranded). In embodiments, the paired region linking the aptamer to the effector region comprises all, or some of an aptamer stem (e.g., for example all, or some of the aptamer P1 stem.).

[0126] Riboswitches comprising aptamer sequences may be used, for example, to control the formation of rho-independent transcription termination hairpins leading to premature transcription termination. Riboswitches comprising aptamer sequences may also induce structural changes in the RNA, leading to sequestration for the ribosome binding site and inhibition of translation. Alternative riboswitch structures comprising the aptamer sequences disclosed herein can further affect the splicing of mRNA in response to the presence of the small molecule ligand.

[0127] **Alternative splicing riboswitch**

[0128] In one embodiment, the aptamers described herein are encoded as part of a gene regulation cassette for the regulation of a target gene by aptamer/ligand mediated alternative splicing of the resulting RNA (e.g., pre-mRNA). In this context, the gene regulation cassette comprises a riboswitch comprising a sensor region (e.g., the aptamers described herein) and an effector region that together are responsible for sensing the presence of a small molecule ligand and altering splicing to an alternative exon. Splicing refers to the process by which an intronic sequence is removed from the nascent pre-messenger RNA (pre-mRNA) and the exons are joined together to form the mRNA. Splice sites are junctions between exons and introns, and are defined by different consensus sequences at the 5′ and 3′ ends of the intron (i.e., the splice donor and splice acceptor sites, respectively). Splicing is carried out by a large multi-component structure called the spliceosome, which is a collection of small nuclear ribonucleoproteins (snRNPs) and a diverse array of auxiliary proteins. By recognizing various cis regulatory sequences, the spliceosome defines exon/intron boundaries, removes intronic sequences, and splices together the exons into a final message (e.g., the mRNA). In the case of alternative splicing, certain exons can be included or excluded to vary the final coding message thereby changing the resulting expressed protein.

[0129] In one embodiment, the regulation of target gene expression is achieved by using any of the DNA constructs disclosed in WO2016/126747, which is hereby incorporated by reference in its entirety.In embodiments of the present disclosure, the riboswitches and polynucleotide cassettes disclosed in WO2016/126747 comprise an aptamer encoding sequence described herein in place of the aptmer sequence disclosed in WO2016/126747.

[0130] In one embodiment, the polynucleotide cassette comprises (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. In embodiments, the effector region comprises the intronic 5′ splice site ("5′ ss") sequence of the intron that is immediately 3′ of the alternative exon, as well as the sequence complimentary to the 5′ ss sequence of the 3′ intron. When the aptamer binds its ligand, the effector region forms a stem and thus prevents splicing to the splice donor site at the 3′ end of the alternative exon. Under certain conditions (for example, when the aptamer is not bound to its ligand), the effector region is in a context that provides access to the splice donor site at the 3′ end of the alternative exon, leading to inclusion of the alternative exon in the target gene mRNA. In some embodiments, the polynucleotide cassette is placed in the target gene to regulate expression of the target gene in response to a ligand. In one embodiment, 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.

[0131] In one embodiment, the gene regulation cassette comprises the sequence of SEQ ID NO:37, wherein -X- represents an aptamer encoding sequence disclosed herein. Lower case letters indicate paired stem sequence linking the aptamer to the remainder of the riboswitch. In one embodiment, the alternative exon (underlined in SEQ ID NO:37, below) is replaced with another alternative exon sequence.

[0132] SEQ ID NO:37 -

GTGAGTCTATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCA
TGTCATAGGAAGGGGAGAAGTAACAGGGTACACATATTGACCAAATCAGGGTAA
TTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACTTTTTTGTTTAT
CTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATGT
ATCATGCCGAGTAACGCTGTTTCTCTAACTTGTAGGAATGAATTCAGATATTTCC
AGAGAATGAAAAAAAAATCTTCAGTAGAAGgtaatgt-**X**-
acattacGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTTAAGGCAATAGCA
ATATTTCTGCATATAAATATTTCTGCATATAAATTGTAACTGATGTAAGAGGTTTC

ATATTGCTAATAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTATGGTTGG GATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGCTAATCATGTTCAT ACCTCTTATCTTCCTCCCACAG.

[0133] The alternative exon is flanked by 5′ and 3′ intronic sequences. The 5′ and 3′ intronic sequences that can be used in the gene regulation cassettes disclosed herein can be any sequence that can be spliced out of the target gene creating either the target gene mRNA or the target gene comprising the alternative exon in the mRNA, depending upon the presence or absence of a ligand that binds the aptamer. The 5′ and 3′ intronic sequences each have the sequences necessary for splicing to occur, i.e., splice donor, splice acceptor and branch point sequences. In one embodiment, the 5′ and 3′ intronic sequences of the gene regulation cassette are derived from one or more naturally occurring introns or portions thereof. In one embodiment, the 5′ and 3′ intronic sequences are derived from a truncated human beta-globin intron 2 (IVS2Δ), from intron 2 of the human 03-globin gene, from the SV40 mRNA intron (used in pCMV-LacZ vector from Clontech Laboratories, Inc.), from intron 6 of human triose phosphate isomerase (TPI) gene (Nott Ajit, et al. RNA. 2003, 9:6070617), from an intron from human factor IX (Sumiko Kurachi, et al. J. Bio. Chem. 1995, 270(10), 5276), from the target gene's own endogenous intron, or from any genomic fragment or synthetic introns (Yi Lai, et al. Hum Gene Ther. 2006:17(10): 1036) that contain elements that are sufficient for regulated splicing (Thomas A. Cooper, Methods 2005 (37):331).

[0134] In one embodiment, the alternative exon and riboswitch are engineered to be in an endogenous intron of a target gene. That is, the intron (or a substantially similar intronic sequence) naturally occurs at that position of the target gene. In this case, the intronic sequence immediately upstream of the alternative exon is referred to as the 5′ intron or 5′ intronic sequence, and the intronic sequence immediately downstream of the alternative exon is referred to as the 3′ intron or 3′ intronic sequence. In this case, the endogenous intron is modified to contain a splice acceptor sequence and splice donor sequence flanking the 5′ and 3′ ends of the alternative exon. In one embodiment, the 5′ and/or 3′ introns are exogenous to the target gene.

[0135] The splice donor and splice acceptor sites in the alternative splicing gene regulation cassette can be modified to be strengthened or weakened. That is, the splice sites can be modified to be closer to the consensus for a splice donor or acceptor by standard cloning methods, site directed mutagenesis, and the like. Splice sites that are more similar to the splice consensus tend to promote splicing and are thus strengthened. Splice sites that are

less similar to the splice consensus tend to hinder splicing and are thus weakened. The consensus for the splice donor of the most common class of introns (U2) is A/C A G||G T A/G A G T (where || denotes the exon/intron boundary). The consensus for the splice acceptor is C A G||G (where || denotes the exon/intron boundary). The frequency of particular nucleotides at the splice donor and acceptor sites are described in the art (see, e.g., Zhang, M. Q., Hum Mol Genet. 1988. 7(5):919-932). The strength of 5′ and 3′ splice sites can be adjusted to modulate splicing of the alternative exon.

[0136] Additional modifications to 5′ and 3′ introns present in the alternative splicing gene regulation cassette that can be made to modulate splicing include modifying, deleting, and/or adding intronic splicing enhancer elements, intronic splicing suppressor elements and or splice sites, and/or modifying the branch site sequence.

[0137] In one embodiment, the 5′ intron has been modified to contain a stop codon that will be in frame with the target gene. The 5′ and 3′ intronic sequences can also be modified to remove cryptic slice sites, which can be identified with publicly available software (see, e.g., Kapustin, Y. et al. Nucl. Acids Res. 2011. 1-8).

[0138] The lengths of the 5′ and 3′ intronic sequences can be adjusted in order to, for example, meet the size requirements for viral expression constructs. In one embodiment, the 5′ and/or 3′ intronic sequences are about 50 to about 300 nucleotides in length. In one embodiment, the 5′ and/or 3′ intronic sequences are about 125 to about 240 nucleotides in length.

[0139] The stem portion of the effector region should be of a sufficient length (and GC content) to substantially prevent alternative splicing of the alternative exon upon ligand binding the aptamer, while also allowing access to the splice site when the ligand is not present in sufficient quantities. In embodiments, the stem portion of the effector region comprises a stem sequence in addition to the 5′ splice site sequence of the 3′ intron and its complementary sequence of the 5′ splice site sequence. In embodiments, this additional stem sequence comprises a sequence from the aptamer stem. The length and sequence of the stem portion can be modified using known techniques in order to identify stems that allow acceptable background expression of the target gene when no ligand is present and acceptable expression levels of the target gene when the ligand is present. In one embodiment, the effector region stem of the riboswitch is about 7 to about 20 base pairs in length. In one embodiment, the effector region stem is 8 to 11 base pairs in length. In addition to the length

of the stem, the GC base pair content of the stem can be altered to modify the stability of the stem.

[0140] In one embodiment, the alternative exon that is part of the alternative splicing gene regulation cassettes disclosed herein is a polynucleotide sequence capable of being transcribed to a pre-mRNA and alternatively spliced into the mRNA of the target gene. In one embodiment, the alternative exon contains at least one sequence that inhibits translation such that when the alternative exon is included in the target gene mRNA, expression of the target gene from that mRNA is prevented or reduced. In a preferred embodiment, the alternative exon contains a stop codon (TGA, TAA, TAG) that is in frame with the target gene when the alternative exon is included in the target gene mRNA by splicing. In embodiments, the alternative exon comprises, in addition to a stop codon, or as an alternative to a stop codon, another sequence that reduces or substantially prevents translation when the alternative exon is incorporated by splicing into the target gene mRNA including, e.g., a microRNA binding site, which leads to degradation of the mRNA. In one embodiment, the alternative exon comprises a miRNA binding sequence that results in degradation of the mRNA. In one embodiment, the alternative exon encodes a polypeptide sequence which reduces the stability of the protein containing this polypeptide sequence. In one embodiment, the alternative exon encodes a polypeptide sequence which directs the protein containing this polypeptide sequence for degradation.

[0141] The basal or background level of splicing of the alternative exon can be optimized by altering exon splice enhancer (ESE) sequences and exon splice suppressor (ESS) sequences and/or by introducing ESE or ESS sequences into the alternative exon. Such changes to the sequence of the alternative exon can be accomplished using methods known in the art, including, but not limited to site directed mutagenesis. Alternatively, oligonucleotides of a desired sequence (e.g., comprising all or part of the alternative exon) can be obtained from commercial sources and cloned into the gene regulation cassette. Identification of ESS and ESE sequences can be accomplished by methods known in the art, including, for example using ESEfinder 3.0 (Cartegni, L. et al. ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acid Research, 2003, 31(13): 3568-3571) and/or other available resources.

[0142] In one embodiment, the alternative exon is a naturally-occurring exon. In another embodiment, the alternative exon is derived from all or part of a known exon. In this context, "derived" refers to the alternative exon containing sequence that is substantially homologous to a naturally occurring exon, or a portion thereof, but may contain various mutations, such a

mutations generated by altering exon splice enhancer (ESE) sequences and exon splice suppressor (ESS) sequences and/or by introducing ESE or ESS sequences into the alternative exon. “Homology” and “homologous” as used herein refer to the percent of identity between two polynucleotide sequences or between two polypeptide sequences. The correspondence between one sequence to another can be determined by techniques known in the art. For example, homology can be determined by a direct comparison of two polypeptide molecules by aligning their sequences and using readily available computer programs. Alternatively, homology can be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. Two polynucleotide or two polypeptide sequences are “substantially homologous” to each other when, after optimally aligned with appropriate insertions or deletions, at least about 80%, at least about 85%, at least about 90%, and at least about 95% of the nucleotides or amino acids, respectively, match over a defined length of the molecules, as determined using the methods above.

[0143] In one embodiment, the alternative exon is exogenous to the target gene, although it may be derived from a sequence originating from the organism where the target gene will be expressed. As used herein, “exogenous” means derived from a genotypically distinct entity from that of the rest of the entity to which it is compared or into which it is introduced or incorporated. For example, a polynucleotide introduced by genetic engineering techniques into a different cell type is a heterologous polynucleotide (and, when expressed, can encode a heterologous polypeptide). In one embodiment, the alternatively-spliced exon is derived from exon 2 of the human dihydrofolate reductase gene (DHFR), mutant human Wilms tumor 1 exon 5, mouse calcium/calmodulin-dependent protein kinase II delta exon 16, or SIRT1 exon 6. In embodiments, the alternatively-spliced exon is, or comprises, the modified DHFR exon 2 in SEQ ID NO:79 (GAATGAATTCAGATATTTCCAGAGAATGAAAAAAAAATCTTCAGTAGAAG). In embodiments, the alternatively-spliced exon is, or comprises, the modified DHFR exon 2 in SEQ ID NO:98 GAATGAATTCAGATATTTCCAGAGAATGAAAAAAAATCTTCAGTAGAAG.

[0144] Aptamer-mediated cleavage by self-cleaving ribozymes

[0145] In one embodiment, the aptamer-mediated expression of the target gene is regulated by an aptamer-mediated modulation of small endonucleolytic ribozymes. A ribozyme is an RNA enzyme that catalyzes a chemical reaction. In the nucleic acids and

methods disclosed herein, a ribozyme may be any small endonucleolytic ribozyme that will self-cleave in the target cell type including, but not limited to a hammerhead, hairpin, the hepatitis delta virus, the Varkud satellite, twister, twister sister, pistol or hatchet ribozyme. Accordingly, in one embodiment, provided is a riboswitch, and a gene expression cassette comprising the riboswitch that contains a ribozyme linked to an aptamer disclosed herein. WO2017/136608, which is incorporated in its entirety by reference herein, describes such riboswitches that activate ribozyme self-cleavage in the presence of aptamer ligand ("off" switch) or riboswitches that inhibit ribozyme self-cleavage in the presence of aptamer ("on" switch).

[0146] In an "off" switch scenario, aptamer/ligand binding increases the ribonuclease function of the ribozyme, leading to cleavage of the target gene RNA that contains the polynucleotide cassette, thereby reducing target gene expression. Examples of such an off switch include a polynucleotide cassette for the regulation of the expression of a target gene comprising a riboswitch that comprises a twister ribozyme linked by a stem to an aptamer, wherein the stem linking the twister ribozyme to the aptamer attaches to the ribozyme at the location of the P3 stem of the twister ribozyme and wherein the target gene is linked to the P1 stem of the twister ribozyme (see, e.g. Figs. 1a, 1b, or 3a of WO2017/136608 and the associated text, incorporated herein by reference).

[0147] In an "on" switch scenario, aptamer/ligand binding inhibits the ribonuclease function of the ribozyme, decreasing cleavage of the target gene RNA that contains the polynucleotide cassette, thereby increasing target gene expression in the presence of ligand. Examples of an on switch include a riboswitch that comprises a twister ribozyme linked to an aptamer, wherein the aptamer is linked to the 3′ or 5' end of the twister ribozyme P1 stem, wherein when the aptamer is linked to the 3' end of the twister ribozyme P1 stem, a portion of the 3' arm of the twister ribozyme P1 stem is alternatively the 5' arm of the aptamer P1 stem, and wherein when the aptamer is linked to the 5' end of the twister ribozyme P1 stem, a portion of the 5' arm of the twister ribozyme P1 stem is alternatively the 3' arm of the aptamer P1 stem (see, e.g., Figs. 6a-6b of WO2017/136608 and the associated text, incorporated herein by reference).

[0148] **Aptamer modulation of polyadenylation**

[0149] In embodiments, the expression of a target gene is regulated by aptamer-modulated polyadenylation. The 3' end of almost all eukaryotic mRNAs comprises a poly(A) tail—a homopolymer of 20 to 250 adenosine residues. Because addition of the poly(A) tail to mRNA

protects it from degradation, expression of a gene can be influenced by modulating the polyadenylation the corresponding mRNA.

[0150] In one embodiment, the expression of the target gene is regulated through aptamer-modulated accessibility of polyadenylation signals as described in and WO2018/156658, which is incorporated in its entirety by reference herein. In such embodiments, the riboswitch comprises an effector stem-loop and an aptamer described herein, wherein the effector stem-loop comprises a polyadenylation signal, and wherein the aptamer and effector stem-loop are linked by an alternatively shared stem arm comprising a sequence that is complementary to the unshared arm of the aptamer stem and to the unshared arm of the effector stem loop (see, e.g., Figs 1a, 1b, 2a, and 5a of WO2018/156658 and the associated text, incorporated herein by reference). In one embodiment, the effector stem-loop is positioned 3′ of the aptamer such that the alternatively shared stem arm comprises all or a portion of the 3′ aptamer stem arm and all or a portion of the 5′ arm of the effector stem. In one embodiment, the effector stem-loop is positioned 5′ of the aptamer such that the alternatively shared stem arm comprises all or a portion of the 5′ aptamer stem arm and all or a portion of the 3′ arm of the effector stem. In one embodiment, the polyadenylation signal is AATAAA or ATTAAA. In one embodiment, the polyadenylation signal is a downstream element (DSE). In one embodiment, the polyadenylation signal is an upstream sequence element (USE). In one embodiment, the polynucleotide cassette comprises two riboswitches, wherein the effector stem loop of the first riboswitch comprises all or part of the polyadenylation signal AATAAA or ATTAAA and the effector stem loop of the second riboswitch comprises all or part of the downstream element (DSE). In one embodiment, the two riboswitches each comprise aptamers that bind the same ligand. In one embodiment, the two riboswitches comprise different aptamers that bind different ligands.

[0151] In some embodiments, the riboswitch comprises a sensing region (e.g., an aptamer described herein) and an effector region comprising a binding site for the small nuclear ribonucleoprotein (snRNP) U1, which is part of the spliceosome. WO2017/136591 describes riboswitches wherein the effector region comprises a U1 snRNP binding site, and is incorporated herein by reference in its entirety. When the aptamer binds its ligand, the effector region forms a stem and sequesters the U1 snRNP binding site from binding a U1 snRNP. Under certain conditions (for example, when the aptamer is not bound to its ligand), the effector region is in a context that provides access to the U1 snRNP binding site, allowing U1 snRNP to bind the mRNA and inhibit polyadenylation leading to degradation of the message. The U1 snRNP binding site can be any polynucleotide sequence that is capable of

binding the U1 snRNP, thereby recruiting the U1 snRNP to the 3′ UTR of a target gene and suppressing polyadenylation of the target gene message. In one embodiment, the U1 snRNP binding site is the consensus site CAGGTAAGTA (SEQ ID NO:80) (CAGGUAAGUA, SEQ ID NO:81, when in the mRNA). In some embodiments, the U1 snRNP binding site is a variation of this consensus sequence, including for example sequences that are shorter or have one or more nucleotides changed from the consensus sequence. In one embodiment, the U1 snRNP binding site contains the sequence CAGGTAAG. In some embodiments, the binding site is encoded by the sequence selected from CAGGTAAGTA (SEQ ID NO:80), CAGGTAAGT, and CAGGTAAG. The U1 snRNP binding site can be any 5′ splice site from a gene, e.g., the 5′ splice site from human DHFR exon 2.

[0152] Aptamer-mediated modulation of ribonuclease cleavage

[0153] In one embodiment, the expression of the target gene is regulated through aptamer-modulated ribonuclease cleavage. Ribonucleases (RNases) recognize and cleave specific ribonuclease substrate sequences. Provided herein are recombinant DNA constructs that, when incorporated into the DNA of a target gene, provide the ability to regulate expression of the target gene by aptamer/ligand mediated ribonuclease cleavage of the resulting RNA. In some embodiments, the aptamer encoding sequence described herein is part of a construct that contains or encodes a ribonuclease substrate sequence and a riboswitch comprising an effector region and the aptamer such that when the aptamer binds a ligand, target gene expression occurs (as described in WO2018/161053, which is incorporated in its entirety by reference herein). In embodiments, an RNase P substrate sequence is linked to a riboswitch wherein the riboswitch comprises an effector region and an aptamer described herein, wherein the effector region comprises a sequence complimentary to a portion of the RNase P substrate sequence. Binding of a suitable ligand to the aptamer induces structural changes in the aptamer and effector region, altering the accessibility of the ribonuclease substrate sequence for cleavage by the ribonuclease.

[0154] In one embodiment, the aptamer sequence is located 5′ to the RNase P substrate sequence and the effector region comprises all or part of the leader sequence and all or part of the 5′ acceptor stem sequence of the RNase P substrate sequence. See, e.g., Figs. 1a, 1b, and 3b of WO2018/161053 and the associated text, incorporated herein by reference. In further embodiments, the acceptor stem of the RNase P substrate and the riboswitch effector region are separated by 0, 1, 2, 3, or 4 nucleotides. In other embodiments, the effector region stem includes, in addition to leader sequence (and its complement), one or more nucleotides of the

acceptor stem of the RNase P substrate, and sequence complementary to the one or more nucleotides of the acceptor stem.

[0155] In one embodiment, the aptamer sequence of the polynucleotide cassette is located 3′ to the RNase P substrate sequence and the effector region comprises sequence complimentary to the all or part of the 3′ acceptor stem of the RNase P substrate sequence. See, e.g., Fig. 3a of WO2018/161053 and the associated text, incorporated herein by reference. In further embodiments, the effector region sequence complimentary to the 3′ acceptor stem of the RNase P substrate is 1 to 7 nucleotides. In other words, the effector region stem includes 1 to 7 nucleotides of the acceptor stem and includes sequence that is complementary to this 1 to 7 nucleotides of the acceptor stem. In embodiments, the riboswitch is located 3′ of the RNase P substrate so the effector region stem and the acceptor stem of the RNase P substrate do not overlap. In embodiments, the effector region and the acceptor stem of the RNase P substrate are immediately adjacent (i.e., not overlapping). In other embodiments, the effector region and the acceptor stem of the RNase P substrate are separated by 1, 2, 3, 4, 5 or more nucleotides.

[0156] The aptamers and gene regulation cassettes disclosed herein can be used to regulate the expression of any target gene that can be expressed in a target cell, tissue or organism. The term “target gene” refers to a polynucleotide that is introduced into a cell and is capable of being transcribed into RNA and translated and/or expressed under appropriate conditions. Alternatively, the target gene is endogenous to the target cell and the gene regulation cassette is positioned into the target gene (for example into an existing untranslated region or intron of the endogenous target gene). An example of a target gene is a polynucleotide encoding a therapeutic polypeptide. In one embodiment, the target gene is exogenous to the cell in which the recombinant DNA construct is to be transcribed. In another embodiment, the target gene is endogenous to the cell in which the recombinant DNA construct is to be transcribed. The target gene may be a gene encoding a protein, or a sequence encoding a non-protein coding RNA. The target gene may be, for example, a gene encoding a structural protein, an enzyme, a cell signaling protein, a mitochondrial protein, a zinc finger protein, a hormone, a transport protein, a growth factor, a cytokine, an intracellular protein, an extracellular protein, a transmembrane protein, a cytoplasmic protein, a nuclear protein, a receptor molecule, an RNA binding protein, a DNA binding protein, a transcription factor, translational machinery, a channel protein, a motor protein, a cell adhesion molecule, a mitochondrial protein, a metabolic enzyme, a kinase, a phosphatase,

exchange factors, a chaperone protein, and modulators of any of these. In embodiments, the target gene encodes erythropoietin (Epo), human growth hormone (hGH), transcription activator-like effector nucleases (TALEN), human insulin, CRISPR associated protein 9 (cas9), or an immunoglobulin (or portion thereof), including, e.g., a therapeutic antibody.

[0157] In embodiments, the aptamers and gene regulation cassettes disclosed herein are used to regulate the expression of a target gene in eukaryotic cells for example, mammalian cells and more particularly human cells. In embodiments, the aptamers and gene regulation cassettes disclosed herein are used to regulate the expression of a target gene in the eye (including cornea and retina), central nervous system (including the brain), liver, kidney, pancreas, heart, airway, muscle, skin, lung, cartilage, testes, arteries, thymus, bone marrow, or in tumors.

[0158] In one aspect, provided are recombinant vectors and their use for the introduction of a polynucleotide comprising a target gene and a gene regulation cassette, wherein the gene regulation cassette comprises an aptamer disclosed herein. In some embodiments, the recombinant DNA constructs include additional DNA elements including DNA segments that provide for the replication of the DNA in a host cell and expression of the target gene in target cells at appropriate levels. The ordinarily skilled artisan appreciates that expression control sequences (promoters, enhancers, and the like) are selected based on their ability to promote expression of the target gene in the target cell. "Vector" means a recombinant plasmid, yeast artificial chromosome (YAC), mini chromosome, DNA mini-circle or virus (including virus derived sequences) that comprises a polynucleotide to be delivered into a host cell, either in vitro or in vivo. In one embodiment, the recombinant vector is a viral vector or a combination of multiple viral vectors.

[0159] Viral vectors for the expression of a target gene in a target cell, tissue, or organism are known in the art and include adenoviral (AV) vectors, adeno-associated virus (AAV) vectors, retroviral and lentiviral vectors, and Herpes simplex type 1 (HSV1) vectors.

[0160] Adenoviral vectors include, for example, those based on human adenovirus type 2 and human adenovirus type 5 that have been made replication defective through deletions in the E1 and E3 regions. The transcriptional cassette can be inserted into the E1 region, yielding a recombinant E1/E3-deleted AV vector. Adenoviral vectors also include helper-dependent high-capacity adenoviral vectors (also known as high-capacity, "gutless" or "gutted" vectors), which do not contain viral coding sequences. These vectors, contain the cis-acting elements needed for viral DNA replication and packaging, mainly the inverted terminal repeat sequences (ITR) and the packaging signal (CY). These helper-dependent AV

vector genomes have the potential to carry from a few hundred base pairs up to approximately 36 kb of foreign DNA.

[0161] Recombinant adeno-associated virus "rAAV" vectors include any vector derived from any adeno-associated virus serotype, including, without limitation, AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-7 and AAV-8, AAV-9, AAV-10, and the like. rAAV vectors can have one or more of the AAV wild-type genes deleted in whole or in part, preferably the Rep and/or Cap genes, but retain functional flanking ITR sequences. Functional ITR sequences are retained for the rescue, replication, packaging and potential chromosomal integration of the AAV genome. The ITRs need not be the wild-type nucleotide sequences, and may be altered (e.g., by the insertion, deletion or substitution of nucleotides) so long as the sequences provide for functional rescue, replication and packaging.

[0162] Alternatively, other systems such as lentiviral vectors can be used. Lentiviral-based systems can transduce nondividing as well as dividing cells making them useful for applications targeting, for examples, the nondividing cells of the CNS. Lentiviral vectors are derived from the human immunodeficiency virus and, like that virus, integrate into the host genome providing the potential for very long-term gene expression.

[0163] Polynucleotides, including plasmids, YACs, minichromosomes and minicircles, carrying the target gene containing the gene regulation cassette can also be introduced into a cell or organism by nonviral vector systems using, for example, cationic lipids, polymers, or both as carriers. Conjugated poly-L-lysine (PLL) polymer and polyethylenimine (PEI) polymer systems can also be used to deliver the vector to cells. Other methods for delivering the vector to cells includes hydrodynamic injection and electroporation and use of ultrasound, both for cell culture and for organisms. For a review of viral and non-viral delivery systems for gene delivery see Nayerossadat, N. et al. (Adv Biomed Res. 2012; 1:27) incorporated herein by reference.

[0164] In one aspect, this disclosure provides a method of modulating the expression of a target gene (e.g., a therapeutic gene) comprising (a) inserting the polynucleotide cassette comprising an aptamer disclosed herein into the target gene, (b) introducing the target gene comprising the polynucleotide cassette into a cell, and (c) exposing the cell to a small molecule ligand that specifically binds the aptamer in an amount effective to induce expression of the target gene. In aspects, expression of the target gene in target cells confers a desired property to a cell into which it was introduced, or otherwise leads to a desired therapeutic outcome.

[0165] In one embodiment, a gene regulation cassette comprising an aptamer disclosed herein is inserted into the protein coding sequence of the target gene (rather than in the 5′ or 3′ untranslated regions). In one embodiment, a single gene regulation cassette comprising an aptamer disclosed herein is inserted into the target gene. In other embodiments 2, 3, 4, or more gene regulation cassettes are inserted in the target gene, wherein one or more gene regulation cassettes comprise an aptamer disclosed herein. In one embodiment, two gene regulation cassettes are inserted into the target gene, wherein one or both gene regulation cassettes comprise an aptamer disclosed herein. When multiple gene regulation cassettes are inserted into a target gene, they each can contain the same aptamer such that a single ligand can be used to modulate target gene expression. In other embodiments, multiple gene regulation cassettes are inserted into a target gene, each can contain a different aptamer so that exposure to multiple different small molecule ligands modulates target gene expression.

[0166] Methods of Treatment and Pharmaceutical Compositions

[0167] In one aspect, provided is a method of regulating the level of a therapeutic protein delivered by gene therapy. The therapeutic gene sequence containing a regulatory cassette comprising an aptamer disclosed herein is delivered to the target cells in the body, e.g., by a vector. The cell specificity of the target gene expression may be controlled by a promoter and/or other elements within the vector and/or by the capsid of the viral vector. Delivery of the vector construct containing the target gene, and the transfection of the target tissues resulting in stable transfection of the regulated target gene, is the first step in producing the therapeutic protein. However, due to an aptamer within the target gene sequence, the target gene is not expressed at significant levels, i.e., it is in the "off state" in the absence of the specific ligand that binds to the aptamer contained within in the regulatory cassette riboswitch. Only when the aptamer specific ligand is administered is the target gene expression activated.

[0168] The delivery of the vector construct containing the target gene and the delivery of the activating ligand generally are separated in time. The delivery of the activating ligand will control when the target gene is expressed, as well as the level of protein expression. The ligand may be delivered by a number of routes including, but not limited to, intravitreal, intraocular, inhalation, subcutaneous, intramuscular, intradermal, intralesion, topical, intraperitoneal, intravenous (IV), intra-arterial, perivascular, intracerebral, intracerebroventricular, oral, sublingual, sublabial, buccal, nasal, intrathoracic, intracardiac, intrathecal, epidural, intraosseous, or intraarticular.

[0169] The timing of delivery of the ligand will depend on the requirement for activation of the target gene. For example, if the therapeutic protein encoded by the target gene is required constantly, an oral small molecule ligand may be delivered daily, or multiple times a day, to ensure continual activation of the target gene, and thus continual expression of the therapeutic protein. If the target gene has a long acting effect, the inducing ligand may be dosed less frequently, for example, once a week, every other week, once a month.

[0170] This aptamers described herein in the context of a gene regulation cassette comprising a riboswitch allow the expression of a therapeutic transgene to be controlled temporally, in a manner determined by the temporal dosing of the ligand specific to the aptamer. The expression of the therapeutic transgene only on ligand administration, increases the safety of a gene therapy treatment by allowing the target gene to be off in the absence of the ligand.

[0171] Different aptamers can be used in multiple riboswitches to allow different ligands to up-regulate or down-regulate the expression of a target gene. In certain embodiments, each therapeutic gene containing a regulatory cassette will have a specific aptamer within the cassette that will be activated by a specific small molecule. This means that each therapeutic gene can be activated only by the ligand specific to the aptamer housed within it. In these embodiments, each ligand will only activate one therapeutic gene. This allows for the possibility that several different "target genes" may be delivered to one individual and each will be activated on delivery of the specific ligand for the aptamer contained within the regulatory cassette housed in each target gene.

[0172] The aptamers disclosed herein in the context of a riboswitch allow any therapeutic protein whose gene can be delivered to the body (such as erythropoietin (EPO) or a therapeutic antibody) to be produced by the body when the activating ligand is delivered. This method of therapeutic protein delivery may replace the manufacture of such therapeutic proteins outside of the body which are then injected or infused, e.g., antibodies used in cancer or to block inflammatory or autoimmune disease. The body containing the regulated target gene becomes the biologics manufacturing factory, which is switched on when the gene-specific ligand is administered.

[0173] In one embodiment, the target protein may be a nuclease that can target and edit a particular DNA sequence. Such nucleases include Cas9, zinc finger containing nucleases, or TALENs. In the case of these nucleases, the nuclease protein may be required for only a short period of time that is sufficient to edit the target endogenous genes. However, if an unregulated nuclease gene is delivered to the body, this protein may be present for the rest of

the life of the cell. In the case of nucleases, there is an increasing risk of off-target editing the longer the nuclease is present. Regulation of expression of such proteins has a significant safety advantage. In this case, vector containing the nuclease target gene containing a regulatory cassette could be delivered to the appropriate cells in the body. The target gene is in the "off" state in the absence of the cassette-specific ligand, so no nuclease is produced. Only when the activating ligand is administered, is the nuclease produced. When sufficient time has elapsed allowing sufficient editing to occur, the ligand will be withdrawn and not administered again. Thus the nuclease gene is thereafter in the "off" state and no further nuclease is produced and editing stops. This approach may be used to correct genetic conditions, including a number of inherited retinopathies such as LCA10 caused by mutations in CEP290 and Stargardts disease caused by mutations in ABCA4.

[0174] Administration of a regulated target gene encoding a therapeutic protein which is activated only on specific ligand administration may be used to regulate therapeutic genes to treat many different types of diseases, e.g., cancer with therapeutic antibodies, immune disorders with immune modulatory proteins or antibodies, metabolic diseases, rare diseases such as PNH with anti-C5 antibodies or antibody fragments as the regulated gene, or ocular angiogenesis with therapeutic antibodies, and dry AMD with immune modulatory proteins.

[0175] A wide variety of specific target genes, allowing for the treatment of a wide variety of specific diseases and conditions, are suitable for use as a target gene whose expression can be regulated using an aptamer/ligand described herein. For example, insulin or an insulin analog (preferably human insulin or an analog of human insulin) may be used as the target gene to treat type I diabetes, type II diabetes, or metabolic syndrome; human growth hormone may be used as the target gene to treat children with growth disorders or growth hormone-deficient adults; erythropoietin (preferably human erythropoietin) may be used as the target gene to treat anemia due to chronic kidney disease, anemia due to myelodysplasia, or anemia due to cancer chemotherapy. Additional target genes compatibles with the aptamers and gene expression cassettes disclosed herein include, but are not limited to, cyclic nucleotide-gated cation channel alpha-3 (CNGA3) and cyclic nucleotide-gated cation channel beta-3 (CNGB3) for the treatment of achromatopsia, retinoid isomerohydrolase (RPE65) for the treatment of retinitis pigmentosa or Leber's congential amaurosis, X-linked retinitis pigmentosa GTPase regulator (RPGR) for the treatment of X-linked retinitis pigmentosa, glutamic acid decarboxylase (GAD) including for the treatment of Parkinson's disease, regulator of nonsense transcripts 1 (UPF1) for the treatment amyotrophic lateral sclerosis, and aquaporin for the treatment of radiation-induced xerostomia and Sjogren's syndrome. Additional target

genes include ArchT (archaerhodopsin from *Halorubrum* strain TP009), Jaws (cruxhalorhodopsin derived from *Haloarcula* (Halobacterium) *salinarum* (strain Shark)), iC1C2 (a variant of a C1C2 chimaera between channel rhodopsins ChR1 and ChR2 from *Chlamydomonas reinhardtii*), or Rgs9-anchor protein (R9AP), a critical component of GTPase complex that mediates the deactivation of phototransduction cascade.

[0176] The expression constructs comprising an aptamer disclosed herein may be especially suitable for treating diseases caused by single gene defects such as cystic fibrosis, hemophilia, muscular dystrophy, thalassemia, or sickle cell anemia. Thus, human β-, γ-, δ-, or ζ-globin may be used as the target gene to treat β-thalassemia or sickle cell anemia; human Factor VIII or Factor IX may be used as the target gene to treat hemophilia A or hemophilia B.

[0177] The small molecules described herein are generally combined with one or more pharmaceutically acceptable carriers to form pharmaceutical compositions suitable for administration to a patient. Pharmaceutically acceptable carriers include solvents, binders, diluents, disintegrants, lubricants, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, generally used in the pharmaceutical arts. Pharmaceutical compositions may be in the form of tablets, pills, capsules, troches, and the like, and are formulated to be compatible with their intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, intranasal, subcutaneous, oral, inhalation, transdermal (topical), transmucosal,

[0178] The pharmaceutical compositions comprising thiamine analog or derivative are administered to a patient in a dosing schedule such that an amount of thiamine analog or derivative sufficient to desirably regulate the target gene is delivered to the patient. When the dosage form is a tablet, pill, or the like, preferably the pharmaceutical composition comprises from 0.1 mg to 10 g of thiamine analog or derivative; from 0.5 mg to 5 g of thiamine analog or derivative; from 1 mg to 1 g of thiamine analog or derivative; from 2 mg to 750 mg of thiamine analog or derivative; from 5 mg to 500 mg of thiamine analog or derivative; from 10 mg to 250 mg of thiamine analog or derivative; or from 150 mg to 300 mg of thiamine analog or derivative.

[0179] The pharmaceutical compositions may be dosed once per day or multiple times per day (e.g., 2, 3, 4, 5, or more times per day). Alternatively, pharmaceutical compositions may be dosed less often than once per day, e.g., once every 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 days, or once a month or once every few months. In some embodiments, the

pharmaceutical compositions may be administered to a patient only a small number of times, e.g., once, twice, three times, etc.

[0180] Provided herein is a method of treating a patient in need of increased expression of a therapeutic protein encoded by a target gene, the method comprising administering to the patient a pharmaceutical composition comprising a ligand, which an aptamer disclosed herein binds to or otherwise responds to, wherein the patient previously had been administered a recombinant DNA comprising the target gene, and where the target gene contains a gene regulation cassette disclosed herein that provides the ability to regulate expression of the target gene by the ligand of the aptamer. Provided herein is a pharmaceutical composition comprising a ligand, which an aptamer disclosed herein binds to or otherwise responds to, for use in a method of treating a patient in need of increased expression of a therapeutic protein encoded by a target gene, wherein the patient previously had been administered a recombinant DNA comprising the target gene, and where the target gene contains a gene regulation cassette disclosed herein that provides the ability to regulate expression of the target gene by the ligand of the aptamer.

[0181] Aptamers for detection and/or diagnostic uses

[0182] A wide range of detection and diagnostic agents can be linked to aptamers through chimerical or physical conjugation. Further, aptamers can be incorporated in biosensors, microfluidic devices and other detection platforms. In some embodiments, the aptamer is conjugated to a polyalkylene glycol moiety, including, but not limited to, polyethylene glycol (PEG), polypropylene glycol (PPG), polyoxyethylated glycerol (POG) and other polyoxyethylated polyols, polyvinyl alcohol (PVA) and other polyalkylene oxides, polyoxyethylated sorbitol, or polyoxyethylated glucose.

[0183] In some embodiments, the aptamer is conjugated to a detectable moiety, including, but not limited to, fluorescent moieties or labels, imaging agents, radioisotopic moieties, radiopaque moieties, and the like, e.g. detectable labels such as biotin, fluorophores, chromophores, spin resonance probes, nanoparticles (including, but not limited to gold, magnetic, and superparamagnetic nanoparticles), quantum dots, radiolabels. Exemplary fluorophores include fluorescent dyes (e.g. fluorescein, rhodamine, and the like) and other luminescent molecules (e.g. luminal). A fluorophore may be environmentally-sensitive such that its fluorescence changes if it is located close to one or more residues in the modified protein that undergo structural changes upon binding a substrate (e.g. dansyl probes). Exemplary radiolabels include small molecules containing atoms with one or more low

sensitivity nuclei (^{13}C, ^{15}N, ^{2}H, ^{125}I, ^{123}I, ^{99}Tc, ^{43}K, ^{52}Fe, ^{67}Ga, ^{68}Ga, ^{111}In and the like). Other useful moieties are known in the art.

[0184] In some embodiments, the aptamer is conjugated to a therapeutic moiety, including, but not limited to, an anti-inflammatory agent, anti-cancer agent, anti-neurodegenerative agent, anti-infective agent, or generally a therapeutic agent.

[0185] Methods for Identifying an Aptamer That Binds to a Compound

[0186] Disclosed herein are methods for identifying an aptamer that binds to a compound of interest (such a small molecule including a thiamine analog, TPP analog, or derivatives thereof), or otherwise modulates target gene expression when part of a riboswitch in response to the addition of, or exposure to, the compound of interest. In one embodiment, the method comprises the steps of:

(i) selecting a parent aptamer sequence;

(ii) generating an aptamer library comprising sequence encoding the aptamer selected in (i), wherein one or more nucleotides in the aptamer encoding sequence are randomly mutated at one or more positions that correspond to 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 the thiamine analog compared to the parent aptamer sequence;

(iv) optionally repeating steps (ii) and (iii) on an aptamer identified in step (iii) rather than an aptamer selected in step (i).

[0187] The parent aptamer sequence may be a TPP aptamer, including known TPP aptamer sequence or may be a putative TPP aptamer identified by searching for homologous sequences in available databases. The parent aptamer sequence may be an aptamer sequence disclosed herein.

[0188] The step of selecting a parent aptamer sequence can involve, for example, (i) identifying a putative TPP aptamer; (ii) inserting the aptamer into a riboswitch that modulates the expression of a target gene (for example a reporter gene); and (iii) exposing the riboswitch/target gene construct to a thiamine or TPP analog or derivative (e.g., the compounds described herein).

[0189] Putative TPP aptamers can be identified from an appropriate sequence database such as the Rfam database, which is a collection of RNA families, each represented by

multiple sequence alignments, consensus secondary structures and covariance models (CMs). In embodiments, the putative TPP aptamer is identified from the Rfam TPP riboswitch family RF00059. In embodiments, the putative TPP aptamer has a sequence that is at least 85%, at least 90%, at least 95%, at least 96%, at least 97% at least 98% or at least 99% identical to SEQ ID NO:94 or SEQ ID NO:95 (thiC or thiM aptamer with stems, respectively).

SEQ ID NO:94:

GUAAUGUGUCGGAGUGCCUUAGGGAUUAUUCCCCUAAAGCUGAGACCGCAUU GCGGGAUCCGUUGAACCUGAUCAGGCUAAUACCUGCGAAGGGAACACAUUAC

SEQ ID NO:95:

GUAAUGUCUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACC UGAUCUGGAUAAUGCCAGCGUAGGGAAGACAUUAC

[0190] The putative TPP aptamer can be inserted into a riboswitch using techniques known to the ordinarily skilled artisan. The responsiveness of the aptamer to the presence of TPP and one or more thiamine or TPP analogs or derivatives (e.g., the compounds described herein) can be tested in cell culture and/or in a cell-free system. In particular, the cell culture system is a eukaryotic cell culture including, e.g., a mammalian, a plant, or an insect cell culture.

[0191] In order to identify aptamers that respond to a thiamine or TPP analog or derivative (e.g., the compounds described herein), one or more nucleotide positions of the sequence encoding the aptamer (i.e., the parent aptamer) are randomized. The nucleotide positions for randomization can be selected based on the structure of the parent aptamer sequence. The predicted secondary structure can be obtained using available programs such as RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and/or by comparison to the crystal structure of a related aptamer (e.g., the *E. coli* thiM riboswitch in Edwards, TE & Ferré-D'Amaré, AR, Structure. 2006 Sep;14(9):1459-68). For example, unpaired regions of the aptamer, including loop (L) regions (e.g., L3 and/or L5) and joining (J) regions (e.g., J3-2 (joining paired regions P3 and P2), J2-4, and/or J4-5), can be identified, and one or more nucleotides in one or more unpaired regions can be randomized to generate a library of aptamers. In embodiments, one or more nucleotides adjacent to one or more unpaired regions are randomized. Additionally, one or more nucleotides in a paired (P) region can be randomized. Further, one or more nucleotides in an unpaired or paired region can be added or deleted. The mutagenized aptamer sequences can be provided as a library of aptamer sequences in the context of a riboswitch. In embodiments, the aptamer library is provided in the context of a riboswitch as part of a gene expression cassette disclosed herein.

[0192] The aptamer encoding sequences containing one or more mutations can be tested for responsiveness to the presence of TPP or one or more thiamine or TPP analogs or derivatives as described above. In embodiments, the aptamer containing one or more mutations is responsive to the analog or derivative, but has reduced responsiveness to thiamine and/or TPP than the parent aptamer from which it is derived (in the context of the same riboswitch/gene expression cassette).

[0193] Aptamers that are responsive to the desired compound, can be further mutagenized by randomizing nucleotides. The nucleotides at selected positions, for example unpaired regions, can be randomized and a library created as described above.

[0194] In some embodiments, the compound of interest is thiamine analog, such as acefurtiamine, acetiamine, allithiamine, amprolium, beclotiamine, benfotiamine, bentiamine, bisbentiamine, cetotiamine, cycotiamine, fursultiamine, monophosphothiamine, octotiamine, oxythiamine, prosultiamine, sulbutiamine, or vintiamol. In embodiments, the thiamine or TPP analog is a compound of Formula I-VIII, including, but not limited to the compounds M10-M99. In embodiments, the thiamine or TPP analog is one of M10, M16, M18, M19, M21, M26, M27, M28, M29, M30, M31, M32, M33, and M34.

[0195] Reporter proteins encoded by the reporter genes used in the methods disclosed herein are proteins that can be assayed by detecting characteristics of the reporter protein, such as enzymatic activity or spectrophotometric characteristics, or indirectly, such as with antibody-based assays. Examples of reporter gene products that are readily detectable include, but are not limited to, puromycin resistance marker (pac), 3-galactosidase, luciferase, orotidine 5'-phosphate decarboxylase (URA3), arginine permease CAN1, galactokinase (GAL1), beta-galactosidase (LacZ), or chloramphenicol acetyl transferase (CAT). Other examples of detectable signals include cell surface markers, including, but not limited to CD4. Reporter genes suitable for the use in the methods for identifying aptamers disclosed herein also include fluorescent proteins (e.g., green fluorescent protein (GFP) and its derivatives), or proteins fused to a fluorescent tag. Examples of fluorescent tags and proteins include, but are not limited to, (3-F)Tyr-EGFP, A44-KR, aacuGFP1, aacuGFP2, aceGFP, aceGFP-G222E-Y220L, aceGFP-h, AcGFP1, AdRed, AdRed-C148S, aeurGFP, afraGFP, alajGFP1, alajGFP2, alajGFP3, amCyan1, amFP486, amFP495, amFP506, amFP515, amilFP484, amilFP490, amilFP497, amilFP504, amilFP512, amilFP513, amilFP593, amilFP597, anm1GFP1, anm1GFP2, anm2CP, anobCFP1, anobCFP2, anobGFP, apulFP483, AQ14, AQ143, Aquamarine, asCP562, asFP499, AsRed2, asulCP, atenFP, avGFP, avGFP454, avGFP480, avGFP509, avGFP510, avGFP514, avGFP523, AzamiGreen, Azurite,

BDFP1.6, bfloGFPa1, bfloGFPc1, BFP, BFP.A5, BFP5, bsDronpa (On), ccalGFP1, ccalGFP3, ccalOFP1, ccalRFP1, ccalYFP1, cEGFP, cerFP505, Cerulean, CFP, cFP484, cfSGFP2, cgfmKate2, CGFP, cgfTagRFP, cgigGFP, cgreGFP, CheGFP1, CheGFP2, CheGFP4, Citrine, Citrine2, Clomeleon, Clover, cp-mKate, cpCitrine, cpT-Sapphire174-173, CyOFP1, CyPet, CyRFP1 (CyRFP1), d-RFP618, D10, d1EosFP (Green), d1EosFP (Red), d2EosFP (Green), d2EosFP (Red), deGFP1, deGFP2, deGFP3, deGFP4, dendFP (Green), dendFP (Red), Dendra (Green), Dendra (Red), Dendra2 (Green), Dendra2 (Red), Dendra2-M159A (Green), Dendra2-M159A (Orange), Dendra2-T69A (Green), Dendra2-T69A (Orange), dfGFP, dimer1, dimer2, dis2RFP, dis3GFP, dKeima, dKeima570, dLanYFP, DrCBD, Dreiklang (On), Dronpa (On), Dronpa-2 (On), Dronpa-3 (On), dsFP483, DspR1, DsRed, DsRed-Express, DsRed-Express2, DsRed-Max, DsRed.M1, DsRed.T3, DsRed.T4, DsRed2, DstC1, dTFP0.1, dTFP0.2, dTG, dTomato, dVFP, E2-Crimson, E2-Orange, E2-Red/Green, EaGFP, EBFP, EBFP1.2, EBFP1.5, EBFP2, ECFP, ECFPH148D, ECGFP, eechGFP1, eechGFP2, eechGFP3, eechRFP, efasCFP, efasGFP, eforCP, EGFP, eGFP203C, eGFP205C, Emerald, Enhanced Cyan-Emitting GFP, EosFP (Green), EosFP (Red), eqFP578, eqFP611, eqFP611V124T, eqFP650, eqFP670, EYFP, EYFP-Q69K, fabdGFP, ffDronpa (On), FoldingReporterGFP, FP586, FPrfl2.3, FR-1, FusionRed, FusionRed-M, G1, G2, G3, Gamillus (On), Gamillus0.1, Gamillus0.2, Gamillus0.3, Gamillus0.4, GCaMP2, gfasGFP, GFP(S65T), GFP-151pyTyrCu, GFP-Tyr151pyz, GFPmut2, GFPmut3, GFPxm16, GFPxm161, GFPxm162, GFPxm163, GFPxm18, GFPxm181uv, GFPxm18uv, GFPxm19, GFPxm191uv, GFPxm19uv, H9, HcRed, HcRed-Tandem, HcRed7, hcriGFP, hmGFP, HriCFP, HriGFP, iFP1.4, iFP2.0, iLov, iq-EBFP2, iq-mApple, iq-mCerulean3, iq-mEmerald, iq-mKate2, iq-mVenus, iRFP670, iRFP682, iRFP702, iRFP713, iRFP720, IrisFP (Green), IrisFP (Orange), IrisFP-M159A (Green), Jred, Kaede (Green), Kaede (Red), Katushka, Katushka-9-5, Katushka2S, KCY, KCY-G4219, KCY-G4219-38L, KCY-R1, KCY-R1-158A, KCY-R1-38H, KCY-R1-38L, KFP1 (On), KikGR1 (Green), KikGR1 (Red), KillerOrange, KillerRed, KO, Kohinoor (On), laesGFP, laGFP, LanFP1, LanFP2, lanRFP-ΔS83l, LanYFP, laRFP, LSS-mKate1, LSS-mKate2, LSSmOrange, M355NA, mAmetrine, mApple, Maroon0.1, mAzamiGreen, mBanana, mBeRFP, mBlueberry1, mBlueberry2, mc1, mc2, mc3, mc4, mc5, mc6, McaG1, McaG1ea, McaG2, mCardinal, mCarmine, mcavFP, mcavGFP, mcavRFP, mcCFP, mCerulean, mCerulean.B, mCerulean.B2, mCerulean.B24, mCerulean2, mCerulean2.D3, mCerulean2.N, mCerulean2.N(T65S), mCerulean3, mCherry, mCherry2, mCitrine, mClavGR2 (Green), mClavGR2 (Red), mClover3, mCyRFP1, mECFP, meffCFP, meffGFP, meffRFP, mEGFP, meleCFP, meleRFP, mEmerald, mEos2 (Green),

mEos2 (Red), mEos2-A69T (Green), mEos2-A69T (Orange), mEos3.1 (Green), mEos3.1 (Red), mEos3.2 (Green), mEos3.2 (Red), mEos4a (Green), mEos4a (Red), mEos4b (Green), mEos4b (Red), mEosFP (Green), mEosFP (Red), mEosFP-F173S (Green), mEosFP-F173S (Red), mEosFP-M159A (Green), mEYFP, MfaG1, mGarnet, mGarnet2, mGeos-C (On), mGeos-E (On), mGeos-F (On), mGeos-L (On), mGeos-M (On), mGeos-S (On), mGinger1, mGinger2, mGrape1, mGrape2, mGrape3, mHoneydew, MiCy, mIFP, miniSOG, miniSOGQ103V, miniSOG2, miRFP, miRFP670, miRFP670nano, miRFP670v1, miRFP703, miRFP709, miRFP720, mIrisFP (Green), mIrisFP (Red), mK-GO (Early), mK-GO (Late), mKalama1, mKate, mKateM41GS158C, mKateS158A, mKateS158C, mKate2, mKeima, mKelly1, mKelly2, mKG, mKikGR (Green), mKikGR (Red), mKillerOrange, mKO, mKO2, mKOκ, mLumin, mMaple (Green), mMaple (Red), mMaple2 (Green), mMaple2 (Red), mMaple3 (Green), mMaple3 (Red), mMaroon1, mmGFP, mMiCy, mmilCFP, mNectarine, mNeonGreen, mNeptune, mNeptune2, mNeptune2.5, mNeptune681, mNeptune684, Montiporasp.#20-9115, mOrange, mOrange2, moxBFP, moxCerulean3, moxDendra2 (Green), moxDendra2 (Red), moxGFP, moxMaple3 (Green), moxMaple3 (Red), moxNeonGreen, moxVenus, mPapaya, mPapaya0.7, mPlum, mPlum-E16P, mRaspberry, mRed7, mRed7Q1, mRed7Q1S1, mRed7Q1S1BM, mRFP1, mRFP1-Q66C, mRFP1-Q66S, mRFP1-Q66T, mRFP1.1, mRFP1.2, mRojoA, mRojoB, mRouge, mRtms5, mRuby, mRuby2, mRuby3, mScarlet, mScarlet-H, mScarlet-I, mStable, mStrawberry, mT-Sapphire, mTagBFP2, mTangerine, mTFP0.3, mTFP0.7 (On), mTFP1, mTFP1-Y67W, mTurquoise, mTurquoise2, muGFP, mUkG, mVenus, mVenus-Q69M, mVFP, mVFP1, mWasabi, Neptune, NijiFP (Green), NijiFP (Orange), NowGFP, obeCFP, obeGFP, obeYFP, OFP, OFPxm, oxBFP, oxCerulean, oxGFP, oxVenus, P11, P4, P4-1, P4-3E, P9, PA-GFP (On), Padron (On), Padron(star) (On), Padron0.9 (On), PAmCherry1 (On), PAmCherry2 (On), PAmCherry3 (On), PAmKate (On), PATagRFP (On), PATagRFP1297 (On), PATagRFP1314 (On), pcDronpa (Green), pcDronpa (Red), pcDronpa2 (Green), pcDronpa2 (Red), PdaC1, pdae1GFP, phiYFP, phiYFPv, pHluorin,ecliptic, pHluorin,ecliptic (acidic), pHluorin,ratiometric (acidic), pHluorin,ratiometric (alkaline), pHluorin2 (acidic), pHluorin2 (alkaline), pHuji, PlamGFP, pmeaGFP1, pmeaGFP2, pmimGFP1, pmimGFP2, Pp2FbFP, Pp2FbFPL30M, ppluGFP1, ppluGFP2, pporGFP, pporRFP, PS-CFP (Cyan), PS-CFP (Green), PS-CFP2 (Cyan), PS-CFP2 (Green), psamCFP, PSmOrange (Far-red), PSmOrange (Orange), PSmOrange2 (Far-red), PSmOrange2 (Orange), ptilGFP, R3-2+PCB, RCaMP, RDSmCherry0.1, RDSmCherry0.2, RDSmCherry0.5, RDSmCherry1, rfloGFP, rfloRFP, RFP611, RFP618, RFP630, RFP637, RFP639, roGFP1, roGFP1-R1, roGFP1-R8, roGFP2,

rrenGFP, RRvT, rsCherry (On), rsCherryRev (On), rsCherryRev1.4 (On), rsEGFP (On), rsEGFP2 (On), rsFastLime (On), rsFolder (Green), rsFolder2 (Green), rsFusionRed1 (On), rsFusionRed2 (On), rsFusionRed3 (On), rsTagRFP (ON), Sandercyanin, Sapphire, sarcGFP, SBFP1, SBFP2, SCFP1, SCFP2, SCFP3A, SCFP3B, scubGFP1, scubGFP2, scubRFP, secBFP2, SEYFP, sg11, sg12, sg25, sg42, sg50, SGFP1, SGFP2, SGFP2(206A), SGFP2(E222Q), SGFP2(T65G), SHardonnay, shBFP, shBFP-N158S/L173I, ShG24, Sirius, SiriusGFP, Skylan-NS (On), Skylan-S (On), smURFP, SNIFP, SOPP, SOPP2, SOPP3, SPOON (on), stylGFP, SuperfolderGFP, SuperfoldermTurquoise2, SuperfoldermTurquoise2ox, SuperNovaGreen, SuperNovaRed, SYFP2, T-Sapphire, TagBFP, TagCFP, TagGFP, TagGFP2, TagRFP, TagRFP-T, TagRFP657, TagRFP675, TagYFP, td-RFP611, td-RFP639, tdimer2(12), tdKatushka2, TDsmURFP, tdTomato, tKeima, Topaz, TurboGFP, TurboGFP-V197L, TurboRFP, Turquoise-GL, Ultramarine, UnaG, usGFP, Venus, VFP, vsfGFP-0, vsfGFP-9, W1C, W2, W7, WasCFP, Wi-Phy, YPet, zFP538, zoan2RFP, ZsGreen, ZsYellow1, αGFP, 10B, 22G, 5B, 6C, A1a, aacuCP, acanFP, ahyaCP, amilCP, amilCP580, amilCP586, amilCP604, apulCP584, BFPsol, Blue102, CFP4, cgigCP, CheGFP3, Clover1.5, cpasCP, Cy11.5, dClavGR1.6, dClover2, dClover2A206K, dhorGFP, dhorRFP, dPapaya0.1, Dronpa-C62S, DsRed-Timer, echFP, echiFP, EYFP-F46L, fcFP, fcomFP, Fpaagar, Fpag_frag, Fpcondchrom, FPmann, FPmcavgr7.7, Gamillus0.5, gdjiCP, gfasCP, GFPhal, gtenCP, hcriCP, hfriFP, KikG, LEA, mcFP497, mcFP503, mcFP506, mCherry1.5, mClavGR1, mClavGR1.1, mClavGR1.8, mClover1.5, mcRFP, meffCP, mEos2-NA, meruFP, mKate2.5, mOFP.T.12, mOFP.T.8, montFP, moxEos3.2, mPA-GFP, mPapaya0.3, mPapaya0.6, mRFP1.3, mRFP1.4, mRFP1.5, mTFP0.4, mTFP0.5, mTFP0.6, mTFP0.8, mTFP0.9, mTFP1-Y67H, mTurquoise-146G, mTurquoise-146S, mTurquoise-DR, mTurquoise-GL, mTurquoise-GV, mTurquoise-RA, mTurquoise2-G, NpR3784g, PDM1-4, psupFP, Q80R, rfloGFP2, RpBphP1, RpBphP2, RpBphP6, rrGFP, RSGFP1, RSGFP2, RSGFP3, RSGFP4, RSGFP6, RSGFP7, Rtms5, scleFP1, scleFP2, spisCP, stylCP, sympFP, TeAPCα, tPapaya0.01, Trp-lessGFP, vsGFP, Xpa, yEGFP, YFP3, zGFP, and zRFP.

[0196] Methods for screening an aptamer library disclosed herein may include measuring the activity of the reporter gene under the control of the aptamer and/or comparing the activity of the reporter gene in presence of the thiamine or TPP analog used for the screen as compared to the activity of the reporter gene in absence of the thiamine or TPP analog used for the screen.

[0197] Articles of manufacture and kits

[0198] Also provided are kits or articles of manufacture for use in the methods described herein. In aspects, the kits comprise the compositions described herein (e.g., compositions for delivery of a vector comprising the target gene containing the gene regulation cassette) in suitable packaging. Suitable packaging for compositions (such as ocular compositions for injection) described herein are known in the art, and include, for example, vials (such as sealed vials), vessels, ampules, bottles, jars, flexible packaging (e.g., sealed Mylar or plastic bags), and the like. These articles of manufacture may further be sterilized and/or sealed.

[0199] Also provided are kits comprising the compositions described herein. These kits may further comprise instruction(s) on methods of using the composition, such as uses described herein. The kits described herein may further include other materials desirable from a commercial and user standpoint, including buffers, diluents, filters, needles, syringes, and package inserts with instructions for performing the administration of the composition or performing any methods described herein. For example, in some embodiments, the kit comprises an rAAV for the expression of a target gene comprising a gene regulation cassette containing an aptamer sequence described herein, a pharmaceutically acceptable carrier suitable for injection, and one or more of: a buffer, a diluent, a filter, a needle, a syringe, and a package insert with instructions for performing the injections. In some embodiments, the kit is suitable for intraocular injection, intramuscular injection, intravenous injection and the like.

[0200] It is to be understood and expected that variations of the compositions of matter and methods herein disclosed can be made by one skilled in the art and it is intended that such modifications are to be included within the scope of the present disclosure. The following Examples further illustrate the invention, but should not be construed to limit the scope of the invention in any way.

[0201] All references cited herein are hereby incorporated by reference in their entirety. All nucleotide sequences provided herein are in a 5′ to 3′ orientation unless stated otherwise. A Sequence Listing is filed herewith, the contents of which are encorporated herein by reference in its entirety.

EXAMPLES

[0202] Example 1. Synthetic riboswitch comprising thiamine pyrophosphate (TPP)-responsive aptamers can regulate gene expression in response to TPP and thiamine analogs

[0203] *Experimental procedures*

[0204] Riboswitch constructs: Aptamers were synthesized by Integrated DNA Technologies, Inc. and Golden Gate cloning strategy (New England Biolabs, NEB) was used to clone the synthesized aptamer sequences into intron-exon-intron cassette to replace the guanine aptamer in the G17 riboswitch cassette (see SEQ ID NO: 15 recited in WO 2016/126747, which is incorporated herein in its entirety) with a TPP aptamer from the *Alishewanella tabrizica thiC* gene (Microbiol Res. 2017 Jan; 195:71-80) or a TPP aptamer from the *Escherichia coli* thiM gene (Structure. 2006 Sep; 14(9):1459-68), generating riboswitch TPPz and TPPm, respectively.

[0205] SEQ ID NO: 82 was obtained by inserting the TPPz riboswitch into the luciferase reporter gene. Capital letters indicate the luciferase encoding sequence. Lower case letters indicate the intron/alternative exon/intron and riboswitch sequence. The thiC aptamer encoding sequence (SEQ ID NO: 96) is underlined. In one embodiment, provided is a riboswitch comprising SEQ ID NO: 82, wherein the aptamer encoding sequence (SEQ ID NO: 96) in SEQ ID NO:82 is replaced with another aptamer sequence disclosed herein.

[0206] SEQ ID NO: 82

[0207] ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCG
CTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCC
CTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTT
ACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGG
GCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTT
ATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACA
TTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGT
GTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCC
AATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAG
TCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTT
TGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCT
GGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGA
GATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGC
GATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGAT
ATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTT
TCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTA
TTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACA
CGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCC

AAGAGGTTCCATCTGCCAGGTATCAGGgtgagtctatgggacccttgatgttttctttccccttctttttctatggtt
aagttcatgtcataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttct
tcttttaatatacttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccgagtaacgctg
tttctctaacttgtaggaatgaattcagatatttccagagaatgaaaaaaaaatcttcagtagaaggtaatgtgtcggagtgccttagggat
tattcccctaaagctgagaccgcattgcgggatccgttgaacctgatcaggctaatacctgcgaagggaacacattacgcaccattcta
aagaataacagtgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggttt
catattgctaatagcagctacaatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagctaggccct
tttgctaatcatgttcatacctcttatcttcctcccacagCAAGGATATGGGCTCACTGAGACTACATCAGC
TATTCTGATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTT
CCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTA
ATCAAAGAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAA
CAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGG
AGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAG
TCTCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCT
TGCTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGA
CGCCGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACG
GAAAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTG
CGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTC
GACGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGAT
CGCCGTGTAA

[0208] SEQ ID NO: 83 was obtained by inserting the TPPm riboswitch into the luciferase reporter gene. Capital letters indicate the luciferase encoding sequence. Lower case letters indicate the intron/alternative exon/intron and riboswitch sequence. The thiM aptamer encoding sequence (SEQ ID NO:97) is underlined. In one embodiment, provided is a riboswitch comprising SEQ ID NO: 83, wherein the aptamer encoding sequence (SEQ ID NO: 97) in SEQ ID NO:83 is replaced with another aptamer sequence disclosed herein.

[0209] SEQ ID NO:83

[0210] ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCG
CTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCC
CTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTT
ACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGG
GCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTT
ATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACA
TTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGT

GTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCC
AATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAG
TCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTT
TGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCT
GGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGA
GATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGC
GATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGAT
ATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTT
TCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTA
TTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACA
CGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCC
AAGAGGTTCCATCTGCCAGGTATCAGGgtgagtctatgggacccttgatgttttctttccccttcttttctatggtt
aagttcatgtcataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttct
tcttttaatatacttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccgagtaacgctg
tttctctaacttgtaggaatgaattcagatatttccagagaatgaaaaaaaaatcttcagtagaaggtaatgtctcggggtgcccttctgcgt
gaaggctgagaaatacccgtatcacctgatctggataatgccagcgtagggaagacattacgcaccattctaaagaataacagtgata
atttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcag
ctacaatccagctaccattctgcttttatttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcatgttcat
acctcttatcttcctcccacagCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATT
ACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTG
AAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAG
GCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGA
AGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAGC
TTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATT
AAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAAC
ACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGA
ACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGA
GATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGG
AGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAG
AAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGT
AA

[0211] Transfection: 3.5 x10^4 human embryonic kidney (HEK) 293 cells were plated in a 96-well flat bottom plate the day before transfection. Plasmid DNA (500 ng) was added to a tube or a 96-well U-bottom plate. Separately, TransIT-293 reagent (Mirus; 1.4 µL) was

added to 50 µL Optimem I media (Life Technologies) and allowed to sit for 5 minutes at room temperature (RT). Then, 50 µL of this diluted transfection reagent was added to the DNA, mixed, and incubated at RT for 20 min. Finally, 7 µL of this solution was added to a well of cells in the 96-well plate. Four hours after transfection, medium containing transfection solution was replaced by medium containing either TPP, fursultiamine, prosultiamine, bisbentiamine, beclotiamine hydrochloride, or sulbutiamine as aptamer inducers.

[0212] Firefly luciferase assay of cultured cells: Twenty-four hours after media change, plates were removed from the incubator, and equilibrated to RT for several minutes on a lab bench, then aspirated. Glo-lysis buffer (Promega, 100 µL, RT) was added, and the plates allowed to remain at RT for at least 5 minutes. Then, the well contents were mixed by 50 µL trituration, and 20 µL of each sample was mixed with 20 µL of bright-glo reagent (Promega) that had been diluted to 10% in glo-lysis buffer. 96 wells were spaced on an opaque white 384-well plate. Following a 5 min incubation at RT, luminescence was measured using a Tecan machine with 500 ms read time. The luciferase activity was expressed as mean arbitrary light units (ALU) ± S.D., and fold induction was calculated as the quotient of the luciferase activity obtained from cells with TPP or analog treatment divided by the luciferase activity obtained from cells without TPP or analog treatment.

[0213] *Results*

[0214] Gene expression cassettes comprising TPP-responsive riboswitches were generated by inserting TPP aptamers from either the *A. tabrizica thiC* riboswitch (for riboswitch TPPz) or *E. coli thiM* riboswitch (for riboswitch TPPm), respectively, into a synthetic riboswitch gene expression cassette. Here, the aptamer sequence was inserted into an intron downstream of an alternative exon containing an in-frame stop codon as described in WO2016/126747, incorporated herein by reference in its entirety. Ligand binding to the aptamer controls the accessibility of the 5′ splice site of the 3′ intron, therefore allowing for regulation of the expression of a target gene through modulating alternative splicing.

[0215] As shown in **Figure 1A**, both synthetic riboswitches TPPz and TPPm regulate luciferase expression in response to TPP. TPPz induces target gene expression at even lower concentrations than TPPm, indicating that the *A. tabrizica* thiC aptamer has a higher TPP binding affinity in mammalian cells as compared to the *E. coli thiM* aptamer.

[0216] To determine whether the synthetic riboswitches TPPz and TPPm also respond to thiamine analogues, a group of thiamine analogs was surveyed. As shown in **Figures 1B and 1C**, both TPPz and TPPm induce luciferase expression in response to fursultiamine and

prosultiamine treatment, while treatment with bisbentiamine, beclotiamine hydrochloride, and sulbutiamine resulted in weak or no effects on luciferase expression (see **Figure 1D**). In all cases, riboswitch TPPz induced target gene expression at a lower concentration than TPPm.

[0217] These data indicate that synthetic riboswitches comprising heterologous aptamer sequences can effectively induce target gene expression in response to a variety of thiamine-related molecules in a dose dependent manner in mammalian cells.

[0218] Example 2. A TPP aptamer homologous sequence regulates gene expression in mammalian cells in response to thiamine analogs.

[0219] *Experimental Procedures:*

[0220] Riboswitch construction: TPP aptamer homologous sequences were obtained from Rfam 12.0 (https://rfam.xfam.org/) and synthesized (Twister Biotech). To make riboswitch constructs containing TPP aptamer homologous sequences, the synthesized oligos were used as PCR templates and replaced the guanine aptamer in G17 riboswitch construct (WO 2016/126747) using Golden Gate cloning (NEB).

[0221] The transfection and firefly luciferase assay were performed as described in Example 1.

[0222] *Results:*

[0223] For the regulation of genes for therapeutic purposes, e.g. in a human subject, the use of synthetic riboswitches in combination with synthetic compounds that naturally do not occur in the subject to be treated (such as fursultiamine) is particularly useful. This is because the absence of the regulatory compound in the patient allows for a stringently controlled expression of therapeutic genes.

[0224] To identify aptamers with increased gene regulation activity in response to thiamine analogs as compared to TPPz and TPPm, a putative TPP aptamer with homologous sequence was obtained from Rfam 12.0 (RNA family database RF00059, http://rfam.xfam.org/family/RF00059). This putative TPP aptamer (accession number AACY023654033.1, with the sequence starting at position 895 and ending at position 815; referred to herein as 14G4) was inserted into the alternative splicing-based gene regulation cassette as described in Example 1 to generate aptamer riboswitch 14G4.

[0225] SEQ ID NO: 84 was obtained by inserting the 14G4 riboswitch into the luciferase reporter gene. Capital letters indicate the luciferase encoding sequence. Lower case indicates the intron/alternative exon/intron and riboswitch sequence. The 14G4 aptamer encoding

sequence (SEQ ID NO:7) is underlined. In one embodiment, provided is a riboswitch comprising SEQ ID NO: 84, wherein the aptamer encoding sequence (SEQ ID NO: 7) in SEQ ID NO:84 is replaced with another aptamer sequence disclosed herein.

[0226] SEQ ID NO:84

[0227] ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCG
CTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCC
CTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTT
ACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGG
GCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTT
ATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACA
TTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGT
GTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCC
AATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAG
TCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTT
TGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCT
GGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGA
GATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGC
GATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGAT
ATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTT
TCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTA
TTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACA
CGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCC
AAGAGGTTCCATCTGCCAGGTATCAGGgtgagtctatgggacccttgatgttttctttccccttcttttctatggtt
aagttcatgtcataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttct
tcttttaatatacttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccgagtaacgctg
tttctctaacttgtaggaatgaattcagatatttccagagaatgaaaaaaaaatcttcagtagaaggtaatgtacaggggtccggccttttc
atttggcgccggtgagagcacacccttgaacctgtttacggataatgccgccgcagggagtacattacgcaccattctaaagaataac
agtgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaa
tagcagctacaatccagctaccattctgcttttatttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcat
gttcatacctcttatcttcctcccacagCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTG
ATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTT
TTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAA
GAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCC
GGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACAT

AGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTG
ATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCC
AACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGG
TGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAA
AGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGG
AGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGC
AAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCG
TGTAA

[0228] This riboswitch was tested in HEK 293 cells for its ability to regulate luciferase gene expression. As shown in **Figure 2**, luciferase expression increased 100-fold upon treatment with fursultiamine in a dose dependent manner. Further, the 14G4 riboswitch regulated luciferase gene expression in response to treatment with fursultiamine in multiple types of cells.

[0229] This experiment illustrates the successful generation of mammalian riboswitches comprising aptamers that are capable of significantly inducing target gene expression in mammalian cells in response to synthetic small molecules.

[0230] Example 3. Generation of riboswitches comprising re-engineered aptamer sequences that have increased sensitivity to thiamine analogs.

[0231] Riboswitch 14G4 was chosen for further improvement due to its selectivity for fursultiamine as compared to TPP. After comparing the predicted secondary structure (RNAfold, http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) of the 14G4 riboswitch with the crystallography structure of the *E. coli* thiM riboswitch, three regions in the 14G4 sequence were identified that do not appear to be involved in helical formation but may participate in tertiary structure upon ligand binding. These three regions were chosen for sequence randomization to generate riboswitches with re-engineered aptamer sequences with improved activity.

[0232] Three aptamer libraries A1, A2 and A3, were generated by randomizing nucleotides at 6 positions in, or adjacent to, regions J4-5, J2-4 and J3-2, respectively (see **Figure 3**). Single bacterial colonies were picked and plasmids containing riboswitch constructs were screened in HEK 293 cells for improved gene regulation activity in response to 50 μM fursultiamine as compared to riboswitch 14G4. As shown in **Figure 4A**, several of the aptamer constructs that were isolated, including 1D10, 3H4, 4H2, 6B4, 6C4, and 6G12, show an increase in fursultiamine-dependent induction as compared to the parent construct

14G4. These riboswitches comprising re-engineered aptamer sequences also lead to significant enhancements of target gene expression in other cells types, including alpha mouse liver 12 (AML12), C2C12 (murine myoblast cell line), and adult retinal pigment epithelial cell line-19 (ARPE-19) (see **Figure 4B**). Further, the selected riboswitches comprising re-engineered aptamer sequences also increased target gene expression in response to prosultiamine and benfotiamine (in AML 12 cells) treatment (see **Figure 4C**).

[0233] This data demonstrates that the isolated riboswitches comprising re-engineered aptamer sequences are useful for inducing target gene expression in a variety of cell types in response to a variety of synthetic small molecules.

[0234] To further enhance gene regulation activity, a second round of mutagenesis was performed using library A4 (see **Figure 5**). For this library, the aptamer sequence of the 3H4 variant riboswitch construct was randomized at six nucleotide positions, namely at three bases in the J3-2 region and three bases in the J2-4 region of the predicted secondary structure of 3H4 aptamer. Single bacterial colonies were picked and plasmids containing riboswitch constructs were screened in HEK 293 cells for improved gene regulation activities as compared to riboswitch 3H4, using 50 μM fursultiamine.

[0235] Several riboswitches comprising different aptamer variants that were isolated from library A4 showed further improved sensitivity to fursultiamine as compared to both 14G4 and 3H4, demonstrating that mutagenesis of select nucleotides involved in aptamer/ligand binding can improve the gene regulation activity of aptamer-based riboswitches in mammalian cells (see **Figure 6A**).

[0236] Next, the ability of the riboswitches comprising re-engineered aptamer sequences isolated from aptamer library A4 to respond to other thiamine analogs that share chemical structural features with thiamine was determined. As shown in **Figure 6B**, 3H4-derived riboswitches comprising aptamers 15A5 and 15D10 robustly regulate luciferase expression in response to synthetic molecules fursultiamine, prosultiamine, and benfotiamine, but respond very poorly to naturally occurring TPP. Further, the riboswitches comprising re-engineered aptamer sequences 15A5 and 15D10 exhibit significantly improved dynamic ranges in response to fursultiamine, prosultiamine, and benfotiamine as compared to the parent riboswitch 3H4.

[0237] This example illustrates that, through multiple rounds of mutagenesis, improved riboswitches comprising re-engineered aptamer sequences can be generated that enhance target gene expression in response to treatment with several synthetic thiamine analogs.

[0238] Example 4. Synthetic thiamine analog riboswitches can regulate expression of various target genes in response to fursultiamine

[0239] As discussed in Example 3, isolated riboswitches comprising re-engineered aptamer sequences efficiently induce expression of the reporter protein luciferase in response to various thiamine analogs. To test the ability of the isolated aptamers to regulate expression of other target genes, several of the riboswitches comprising re-engineered aptamer sequences were inserted into the cDNA sequence of murine erythropoietin (mEpo) and the cDNA sequence of enhanced green fluorescent protein (EGFP).

[0240] *Experimental Procedures:*

[0241] Riboswitch constructs: Alternative splicing riboswitches containing aptamers 3H4 or 15D10, respectively, were inserted at position 308 into the mouse erythropoietin cDNA sequence in construct Con8-Epo (SEQ ID NO:85), resulting in constructs Epo-3H4 (SEQ ID NO:86) and Epo-15D10 (SEQ ID NO:87). Expression of the erythropoietin gene was driven by a cytomegalovirus (CMV) promoter. The riboswitch cassette containing aptamer 6B4 was inserted at position 276 into the cDNA sequence encoding enhanced green fluorescent protein (EGFP) in vector pEGFP-C1 to generate the EGFP-6B4 construct (SEQ ID NO:88). The intron-exon-intron cassette without aptamer sequence was inserted into Con8-Epo to create construct Epo-Con1 (SEQ ID NO:89), serving as a control for constitutive target gene expression.

[0242] SEQ ID NO: 86 was obtained by inserting the 3H4 riboswitch into the erythropoietin gene. Capital letters indicate the erythropoietin encoding sequence (see SEQ ID NO:85). Lower case letters indicate the intron/alternative exon/intron and riboswitch sequence. The 3H4 aptamer encoding sequence (SEQ ID NO:9) is underlined.

[0243] SEQ ID NO:86

[0244] ATGGGGGTGCCCGAACGTCCCACCCTGCTGCTTTTACTCTCCTTGCTAC
TGATTCCTCTGGGCCTCCCAGTCCTCTGTGCTCCCCCACGCCTCATCTGCGACAGT
CGAGTTCTGGAGAGGTACATCTTAGAGGCCAAGGAGGCAGAAAATGTCACGATG
GGTTGTGCAGAAGGTCCCAGACTGAGTGAAAATATTACAGTCCCAGATACCAAA
GTCAACTTCTATGCTTGGAAAAGAATGGAGGTGGAAGAACAGGCCATAGAAGTT
TGGCAAGGCCTGTCCCTGCTCTCAGAAGCCATCCTGCAGGgtgagtctatgggacccttgatgttt
tctttccccttcttttctatggttaagttcatgtcataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatt
tgtaattttaaaaaatgctttcttcttttaatatactttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaat
gtatcatgccgagtaacgctgtttctctaacttgtaggaatgaattcagatatttccagagaatgaaaaaaaatcttcagtagaaggtaatg
tacaggggtccggccttttcatttggcgccggtgagagcacacccttgaacctgttcacggataatgccgctgcagggagtacattac

gcaccattctaaagaataacagtgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgat
gtaagaggtttcatattgctaatagcagctacaatccagctaccattctgcttttatttatggttgggataaggctggattattctgagtcca
agctaggccctttgctaatcatgttcatacctcttatcttcctcccacagCCCAGGCCCTGCTAGCCAATTCCTCC
CAGCCACCAGAGACCCTTCAGCTTCATATAGACAAAGCCATCAGTGGTCTACGTA
GCCTCACTTCACTGCTTCGGGTACTGGGAGCTCAGAAGGAATTGATGTCGCCTCC
AGATACCACCCCACCTGCTCCACTCCGAACACTCACAGTGGATACTTTCTGCAAG
CTCTTCCGGGTCTACGCCAACTTCCTCCGGGGGGAAACTGAAGCTGTACACGGGAG
AGGTCTGCAGGAGAGGGGACAGGTGA

[0245] SEQ ID NO: 87 was obtained by inserting the 15D10 riboswitch into the erythropoietin gene. Capital letters indicate the erythropoietin encoding sequence (see SEQ ID NO:85). Lower case letters indicate the intron/alternative exon/intron and riboswitch sequence. The 15D10 aptamer encoding sequence (SEQ ID NO:26) is underlined.

[0246] SEQ ID NO:87

ATGGGGGTGCCCGAACGTCCCACCCTGCTGCTTTTACTCTCCTTGCTACTGATTCC
TCTGGGCCTCCCAGTCCTCTGTGCTCCCCCACGCCTCATCTGCGACAGTCGAGTTC
TGGAGAGGTACATCTTAGAGGCCAAGGAGGCAGAAAATGTCACGATGGGTTGTG
CAGAAGGTCCCAGACTGAGTGAAAATATTACAGTCCCAGATACCAAAGTCAACT
TCTATGCTTGGAAAAGAATGGAGGTGGAAGAACAGGCCATAGAAGTTTGGCAAG
GCCTGTCCCTGCTCTCAGAAGCCATCCTGCAGGgtgagtctatgggacccttgatgttttctttccccttctt
ttctatggttaagttcatgtcataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaa
aatgctttcttcttttaatatactttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccga
gtaacgctgtttctctaacttgtaggaatgaattcagatatttccagagaatgaaaaaaaatcttcagtagaaggtaatgtacaggggtcc
ggccttttcatttggcaccggtgagaacataccttcggacctgttcacggataatgccgctgcagggagtacattacgcaccattctaa
agaataacagtgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttc
atattgctaatagcagctacaatccagctaccattctgcttttatttatggttgggataaggctggattattctgagtccaagctaggcccttt
tgctaatcatgttcatacctcttatcttcctcccacagCCCAGGCCCTGCTAGCCAATTCCTCCCAGCCAC
CAGAGACCCTTCAGCTTCATATAGACAAAGCCATCAGTGGTCTACGTAGCCTCAC
TTCACTGCTTCGGGTACTGGGAGCTCAGAAGGAATTGATGTCGCCTCCAGATACC
ACCCCACCTGCTCCACTCCGAACACTCACAGTGGATACTTTCTGCAAGCTCTTCC
GGGTCTACGCCAACTTCCTCCGGGGGGAAACTGAAGCTGTACACGGGAGAGGTCT
GCAGGAGAGGGGACAGGTGA

[0247] SEQ ID NO: 88 was obtained by inserting the 6B4 riboswitch into the EGFP gene. Capital letters indicate the EGFP encoding sequence. Lower case indicate the

intron/alternative exon/intron and riboswitch sequence. The 6B4 aptamer encoding sequence (SEQ ID NO:14) is underlined.

[0248] SEQ ID NO:88

[0249] ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTG
GTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGC
GAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGC
AAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT
GCCCGAAGGgtgagtctatgggacccttgatgttttctttccccttctttttctatggttaagttcatgtcataggaaggggagaagta
acagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttctttttaatatacttttttgtttatcttatttctaa
tactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccgagtaacgctgtttctctaacttgtaggaatgaattcagata
tttccagagaatgaaaaaaaatcttcagtagaaggtaatgtacaggggtccggccttttcatttggcgccggtgagagcacacccttgtg
acctgtttacggataatgccgccgcagggagtacattacgcaccattctaaagaataacagtgataatttctgggttaaggcaatagcaa
tatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagctacaatccagctaccattctgctt
ttattttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcatgttcatacctcttatcttcctcccacagCT
ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCG
CCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCA
TCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACA
ACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGA
ACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACT
ACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
TGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT
GTACAAGTAA

[0250] SEQ ID NO: 89 was obtained by inserting the intron-exon-intron cassette without aptamer sequence into the erythropoietin gene. Capital letters indicate the erythropoietin encoding sequence (see SEQ ID NO:85). Lower case letters indicate the intron/alternative exon/intron sequence.

[0251] SEQ ID NO:89

[0252] ATGGGGGTGCCCGAACGTCCCACCCTGCTGCTTTTACTCTCCTTGCTAC
TGATTCCTCTGGGCCTCCCAGTCCTCTGTGCTCCCCCACGCCTCATCTGCGACAGT
CGAGTTCTGGAGAGGTACATCTTAGAGGCCAAGGAGGCAGAAAATGTCACGATG
GGTTGTGCAGAAGGTCCCAGACTGAGTGAAAATATTACAGTCCCAGATACCAAA
GTCAACTTCTATGCTTGGAAAAGAATGGAGGTGGAAGAACAGGCCATAGAAGTT

TGGCAAGGCCTGTCCCTGCTCTCAGAAGCCATCCTGCAGGgtgagtctatgggacccttgatgttt
tctttccccttcttttctatggttaagttcatgtcataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatt
tgtaattttaaaaaatgctttcttcttttaatatactttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaat
gtatcatgcctctttgcaccattctaaagaataacagtgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatat
aaattgtaactgatgtaagaggtttcatattgctaatagcagctacaatccagctaccattctgcttttattttatggttgggataaggctgga
ttattctgagtccaagctaggccctttgctaatcatgttcatacctcttatcttcctcccacagCCCAGGCCCTGCTAGCCA
ATTCCTCCCAGCCACCAGAGACCCTTCAGCTTCATATAGACAAAGCCATCAGTGG
TCTACGTAGCCTCACTTCACTGCTTCGGGTACTGGGAGCTCAGAAGGAATTGATG
TCGCCTCCAGATACCACCCCACCTGCTCCACTCCGAACACTCACAGTGGATACTT
TCTGCAAGCTCTTCCGGGTCTACGCCAACTTCCTCCGGGGGAAACTGAAGCTGTA
CACGGGAGAGGTCTGCAGGAGAGGGGACAGGTGA.

[0253] Enzyme-linked immunosorbent assay (ELISA) for mouse erythropoietin: AML12 cells were transfected as described in Example 1 with TransIT-X2 transfection reagent (Mirus Bio). Four hours after transfection, AML12 cells were treated with or without fursultiamine at the indicated doses. The supernatants from the transfected cells were collected 24 hours after fursultiamine treatment and were subjected to ELISA for the detection of mEpo in the supernatant following the manufacturer's instruction (R&D).

[0254] Generation of a cell line expressing EGFP-6B4: A stable cell line containing the EGFP-6B4 construct was generated by electroporating HEK 293 cells with 100 ng plasmid DNA using a Gene Pulser Xcell (Bio-Rad) and applying the default parameters for HEK 293 cells. 48 hours after electroporation, the cell culture was treated with 800 μg/ml of the antibiotic G418 for two weeks to select for cells that stably express the EGFP-6B4 cassette, which carries a G418 resistance gene. Cells were trypsinized. Intensity of EGFP fluorescence in the cell suspension was determined by flow cytometry using a Guava EasyCyte 8HT machine. The resulting data was analyzed using GuavaSoft 3.3. The fold increase in induction was calculated as the quotient of mean fluorescent intensity (MFI) obtained from cells after treated with fursultiamine divided by the MFI obtained from cells without fursultiamine treatment.

[0255] Flow cytometry analysis: 1.5×10^5 HEK 293 cells stably transduced with EGFP-6B4 construct were plated in 24-well plate one day before fursultiamine treatment. Cells were treated with fursultiamine for 24 hours. The intensity of EGFP fluorescence was determined by flow cytometry as described above.

[0256] *Results:*

[0257] Cell expressing construct EGFP-6B4 exhibited low expression of the reporter protein EGFP in the absence of fursultiamine treatment, but showed a 14 fold increase in EGFP expression in the presence of fursultiamine (see **Figure 7A**), illustrating the ability of the riboswitch comprising re-engineered aptamer 6B4 to induce expression of reporter proteins other than luciferase in response to the synthetic ligand fursultiamine. Further, the induction of EGFP expression in response to fursultiamine treatment was dose-dependent (see **Figure 7B**).

[0258] Next, the ability of riboswitches comprising aptamers 3H4 and 15D10 to regulate gene expression of mEpo was examined. In the absence of fursultiamine, cells containing mEpo-3H4 or the mEpo-15D10 expressed very low levels of mEpo. However, upon treatment with fursultiamine, expression of mEpo was enhanced in a dose-dependent manner in cells containing mEpo-3H4 or the mEpo-15D10 constructs. As expected, the control construct mEpo-Con1 expressed mEpo constitutively, irrespective of the presence or absence of fursultiamine (see **Figure 7C**). In response to treatment with 12.5 μM fursultiamine, expression of mEpo was induced by about 40-fold (Epo-15D10) or 9-fold (Epo-3H4) as compared to expression in absence of fursultiamine (see **Figure 7D)**, the level of mEpo from mEpo-15D10 at 12.5 μM of inducer is approximately 61.3% of that from the constitutive and non-regulatable mEpo-Con1 construct.

[0259] These results demonstrate that the ability of riboswitches comprising re-engineered aptamer sequences to induce gene expression in response to small molecules is not restricted to specific target gene sequences, indicating a general applicability of these aptamer riboswitches in regulating target gene expression.

[0260] Example 5: Synthetic riboswitches regulate gene expression *in vivo* in mice

[0261] To assess the ability of re-engineered aptamers to induce gene expression *in vivo*, mice were transfected with an adeno-associated viral vector (AAV) carrying a re-engineered riboswitch, which was inserted into the gene for the reporter protein luciferase. Prosultiamine was used as the aptamer ligand to induce luciferase expression.

[0262] *Experimental Procedures:*

[0263] AAV2.8 viral particle production: The AAV2.8 particles used for the transfection of mice comprised a viral genome derived from AAV2 and a capsid derived from AAV8. The luciferase gene containing an intron-exon-intron cassette with (1) a non-regulatable riboswitch without aptamer ("luci-Con1", SEQ ID NO:90), (2) a riboswitch cassette comprising aptamer 3H4 ("luci-3H4", SEQ ID NO:91), or (3) a riboswitch cassette

comprising aptamer 6B4 ("luci-6B4", SEQ ID NO:92), respectively, was cloned into an AAV2 plasmid vector. Expression of the luciferase gene was controlled by a CASI promoter, which includes CMV and ubiquitin C enhancer elements and the chicken β-actin promoter. The viral vector was packaged into AAV8 capsid and produced following manufacture's protocol (Vigene Biosciences).

[0264] SEQ ID NO: 90 was obtained by inserting an intron-exon-intron cassette without aptamer sequence into the luciferase reporter gene. Capital letters indicate the luciferase encoding sequence. Lower case letters indicate the intron/alternative exon/intron sequence.

[0265] SEQ ID NO:90

ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAA
GATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTT
CCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTTACGCTG
AGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGA
ATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCC
GGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTAT
AATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCG
TTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCA
TCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGAT
GTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGC
CAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATC
TACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATTC
TCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTTT
AAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGA
TATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAG
GAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCC
TTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAA
TTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAG
GTTCCATCTGCCAGGTATCAGGgtgagtctatgggacccttgatgttttctttccccttctttttctatggttaagttcatgt
cataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaatttttaaaaaatgctttcttcttttaatat
actttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgcctctttgcaccattctaaagaat
aacagtgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattg
ctaatagcagctacaatccagctaccattctgcttttatttttatggttgggataaggctggattattctgagtccaagctaggccctttttgcta
atcatgttcatacctcttatcttcctcccacagCAAGGATATGGGCTCACTGAGACTACATCAGCTATT
CTGATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCA

TTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATC
AAAGAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACA
ATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAG
ACATAGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTC
TCTGATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTG
CTCCAACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACG
CCGGTGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGA
AAAAGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCG
CGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGA
CGCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCG
CCGTGTAA

[0266] SEQ ID NO: 91 was obtained by inserting the 3H4 riboswitch into the luciferase reporter gene. Capital letters indicate the luciferase encoding sequence. Lower case letters indicate the intron/alternative exon/intron and riboswitch sequence. The 3H4 aptamer encoding sequence (SEQ ID NO:9) is underlined.

[0267] SEQ ID NO:91

ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAA
GATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTT
CCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTTACGCTG
AGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGA
ATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCC
GGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTAT
AATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCG
TTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCA
TCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGAT
GTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGC
CAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATC
TACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATTC
TCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTTT
AAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGA
TATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAG
GAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCC
TTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAA
TTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAG

GTTCCATCTGCCAGGTATCAGGgtgagtctatgggacccttgatgttttctttccccttcttttctatggttaagttcatgt
cataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttcttttaatat
actttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccgagtaacgctgtttctctaac
ttgtaggaatgaattcagatatttccagagaatgaaaaaaaatcttcagtagaaggtaatgtacaggggtccggccttttcatttggcgcc
ggtgagagcacacccttgaacctgttcacggataatgccgctgcagggagtacattacgcaccattctaaagaataacagtgataattt
ctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagcta
caatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcatgttcatacc
tcttatcttcctcccacagCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTAC
ACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAA
GCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGC
GAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGAAG
CGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAGCTTA
CTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATTAAG
TACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAACACC
CCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACT
TCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGAT
CGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGT
TGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAA
AATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAA

[0268] SEQ ID NO: 92 was obtained by inserting the 6B4 riboswitch into the luciferase reporter gene. Capital letters indicate the luciferase encoding sequence. Lower case letters indicate the intron/alternative exon/intron and riboswitch sequence. The 6B4 aptamer encoding sequence (SEQ ID NO:14) is underlined.

[0269] SEQ ID NO:92

ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCGCTGGAA
GATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTT
CCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTTACGCTG
AGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGA
ATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTTATGCC
GGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACATTTAT
AATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGTGTTCG
TTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCCAATCA
TCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAGTCGAT
GTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTTTGTGC

CAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCTGGATC
TACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGAGATTC
TCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTTT
AAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGA
TATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTTTCTGAG
GAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTATTCTCC
TTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAA
TTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCCAAGAG
GTTCCATCTGCCAGGTATCAGGgtgagtctatgggacccttgatgttttctttccccttcttttctatggttaagttcatgt
cataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttcttttaatat
actttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccgagtaacgctgtttctctaac
ttgtaggaatgaattcagatatttccagagaatgaaaaaaaatcttcagtagaaggtaatgtacaggggtccggccttttcatttggcgcc
ggtgagagcacacccttgtgacctgtttacggataatgccgccgcagggagtacattacgcaccattctaaagaataacagtgataattt
ctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaatagcagcta
caatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcatgttcatacc
tcttatcttcctcccacagCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTAC
ACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTTTTGAA
GCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAAGAGGC
GAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCCGGAAG
CGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACATAGCTTA
CTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTGATTAAG
TACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCCAACACC
CCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGGTGAACT
TCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAAAGAGAT
CGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGGAGGAGT
TGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGCAAGAAA
AATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCGTGTAA

[0270] Animal study: Male Balb/c mice received a single tail vein injection of 1.0×10^{11} or 2.5×10^{11} genome copies of the receptive AAV2.8 viral particle. Twenty-eight days after AAV vector delivery, mice were treated intraperitoneally (I.P.) with 50 mg/kg prosultiamine. Luciferase activity was measured the day prior to drug dosing, as well as 6 h, 24 h, 48 h and 72 h after drug dosing. After the first administration of prosultiamine, the mice were subjected to three additional rounds of dosing and imaging cycles as follows: Day 36 (after AAV administration):100 mg/kg; day 43: 200 mg/kg; and day 51: 400 mg/kg.

[0271] Noninvasive live animal bioluminescence imaging: Before imaging, mice were anesthetized with 2% isoflurane, and injected with 150 mg/kg body weight of luciferin. At the indicated time point post drug dosing, images were taken within 2 to 5 minutes after luciferin injection using a Bruker Xtreme system. Luciferase activity was expressed as mean photon/s ± S.D. (n=5). The fold induction of luciferase gene expression was calculated as the quotient of photon/s obtained from mice treated with prosultiamine divided by the value obtained from mice the day before prosultiamine treatment.

[0272] *Results:*

[0273] To test the riboswitch in regulating gene expression in animals, AAV vectors harboring luciferase gene with or without riboswitch were delivered into mice intravenously. Mice were treated with prosultiamine intraperitoneally (I.P.) 4 weeks after AAV injection. Six hours after a single dose of prosultiamine (50 mg/kg) treatment, luciferase activity was significantly increased in mice injected AAV vectors containing a luciferase gene comprising riboswitches 3H4 or 6B4, but not in the group of mice injected with the same dose of non-regulatable control vector Con1 (see **Figures 8A-8E**). Further, the increase in luciferase activity upon prosultiamine treatment was dose-dependent. In the group of mice that were treated with a single dose of 50 mg/kg or with two doses of 50 mg/kg and 100 mg/kg prosultiamine, respectively, the induced luciferase activity attenuated to the level before inducer treatment 24 hours after single dose treatment. However, in mice that were additionally treated with higher doses of the inducer prosultiamine (i.e., 200 mg/kg and 400 mg/kg) the induced luciferase activity remained for 48 h or longer.

[0274] These results demonstrate that riboswitches comprising re-engineered aptamer sequences selectively induce target gene expression upon treatment with a thiamine analog in a dose-dependent manner *in vivo*.

[0275] **Examples 6 to 99: Thiamine analog riboswitch regulates gene expression in mammalian cells in response to novel thiamine analogs**

[0276] *Experimental Procedures:*

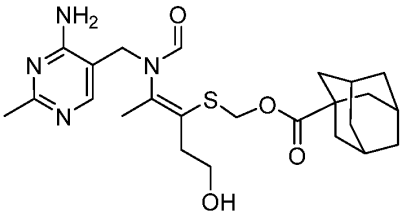
[0277] Generation of thiamine analogs

[0278] All solvents and reagents were obtained commercially and used as received. 1H NMR spectra were recorded on a Bruker instrument (300MHz or 400MHz) in the cited deuterated solvents. Chemical shifts are given in ppm, and coupling constants are in hertz. All final compounds were purified by flash chromatography using 220-400 mesh silica gel or reverse-phase HPLC with CH3CN/water as the solvents. Thin-layer chromatography was

done on silica gel 60 F-254 (0.25-nm thickness) plates. Visualization was accomplished with UV light and/or 10% phosphomolybdic acid in ethanol. Nominal (low resolution) mass spectra were acquired on either a Waters LCT or an Applied Biosystems API 3000 mass spectrometer. High resolution mass spectra (HRMS) were acquired on either a Waters LCT or an Agilent TOF mass spectrometer. All other LC-MS experiments were done on an Agilent 1100 HPLC coupled with an Agilent single quadrupole mass spectrometer. Compound purity was determined by a LC-MS with 230 nM and 254 nM wavelengths. All final compounds reported here have purity ≥ 95%.

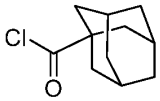
[0279] **Example 6 (M19)**

[0280] (((Z)-2-(N-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl (3r,5r,7r)-adamantane-1-carboxylate



[0281] Step 1

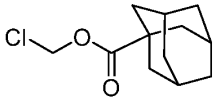
[0282] (3r,5r,7r)-Adamantane-1-carbonyl chloride



[0283] A mixture of (3r,5r,7r)-adamantane-1-carboxylic acid (10.0 g, 55.4 mmol) and $SOCl_2$ (32.8 g, 275 mmol, 20.0 mL) was stirred at 25°C for 1 h. TLC (dichloromethane: methanol = 10: 1, bromocresol green, R_f = 0.62) showed that the starting material was consumed completely. The mixture was concentrated to give the title compound (11.0 g, 99.7%) as colorless crystal which was used directly for the next step reaction without further purification.

[0284] Step 2

[0285] Chloromethyl (3r,5r,7r)-adamantane-1-carboxylate

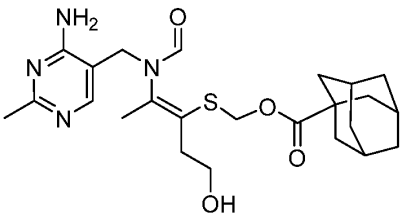


[0286] To a mixture of (3r,5r,7r)-adamantane-1-carboxylic acid (15 g, 83.2 mmol, 1 equiv), tetrabutylammonium bromide (TBAB) (2.68 g, 8.32 mmol, 0.1 equiv) and $NaHCO_3$ (21.0 g, 250.0 mmol, 3 equiv) in DCM (150 mL) and H_2O (150 mL) was added

chloro(chlorosulfonyloxy)methane (16.5 g, 99.9 mmol, 1.2 equiv) drop-wise at 20°C. The mixture was stirred at 20°C for 15 h (gas evolution). Upon standing, the mixture was separated into two layers and the aqueous layer was extracted with methylene dichloride (100 mL × 2). The combined organic phases were dried with anhydrous Na_2SO_4, filtered, and concentrated to give a crude colorless oil. The crude product was diluted with petroleum ether (500 mL) and filtered through a pad of silica gel. The filtrate was concentrated to give the title compound (15.5 g, 81.4%) as a colorless oil. ^{1}H NMR (400 MHz, DMSO-d_6) δ 5.72 (s, 2H), 2.05-1.92 (m, 6H), 1.74-1.70 (m, 2H).

[0287] Step 3

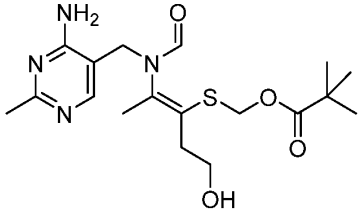
[0288] (((*Z*)-2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl (3r,5r,7r)-adamantane-1-carboxylate



[0289] To a mixture of vitamin B1 (10 g, 33.2 mmol, 1 equiv) and KI (551.0 mg, 3.32 mmol, 0.1 equiv) in H_2O (150 mL) and THF (150 mL) was added NaOH (2.66 g, 66.5 mmol, 2 equiv) in portions and the mixture was stirred at 20°C for 0.5 h. Chloromethyl (3r,5r,7r)-adamantane-1-carboxylate (15.2 g, 66.5 mmol, 2 equiv) was added drop-wise to the mixture and the resulting mixture was stirred at 20°C for another 12 h. LCMS showed completion of the reaction. The reaction mixture separated into two layers. The aqueous layer was extracted with ethyl acetate (200 mL × 3). The combined organic phases were dried with anhydrous Na_2SO_4, filtered, and concentrated to give a crude product (TLC: ethyl acetate/ethanol = 10/1, product R_f = 0.2). The crude product was purified by silica gel chromatography (methylene dichloride/methanol = 10/1) to give a product as a yellow solid which was further purified by pre-HPLC (column: Waters Xbridge BEH C18 250 × 50mm, 10 μm; mobile phase: water-acetonitrile with 0.05% ammonia hydroxide v/v, 20 min) to give the title compound (580.24 mg, 4%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.93 (s, 1H), 7.82 (s, 1H), 6.96 (br, 2H), 4.95 (s, 2H), 4.68 (t, J = 5.2 Hz, 1H), 4.32 (s, 2H), 3.52-3.48 (m, 2H), 2.58 (m, 2H), 2.28 (s, 3H), 1.94 (s, 6H), 1.80-1.78 (m, 6H), 1.68-1.62 (m, 6H). MS (ES^+) m/e 475.2 $(M+H)^+$.

[0290] **Example 7 (M10)**

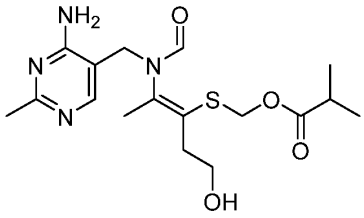
[0291] (*Z*)-((2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl pivalate



[0292] To a solution of vitamin B1 (1.00 g, 3.78 mmol, 1.00 equiv) in EtOH (10.0 mL) was added NaOEt (257mg, 3.78 mmol, 1.00 equiv), chloromethyl pivalate (569 mg, 3.78 mmol, 547 uL, 1.00 equiv) and NaOH (45.4 mg, 1.13 mmol, 0.30 equiv). The mixture was stirred at 20°C for 2 h. The reaction mixture was concentrated under reduced pressure to give a residue. The residue was purified by prep-HPLC to give the title compound (61.0 mg, 3.4%) as a light yellow solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.93 (s, 1H), 7.83 (s, 1H), 6.71 (br d, J = 0.9 Hz, 2H), 4.96 (s, 2H), 4.69 (t, J = 5.5 Hz, 1H), 4.34 (br s, 2H), 3.60 - 3.42 (m, 2H), 2.59 (t, J = 6.9 Hz, 2H), 2.28 (s, 3H), 1.96 (s, 3H), 1.12 (s, 9H). MS (ES$^+$) m/e 397.3 $(M+H)^+$.

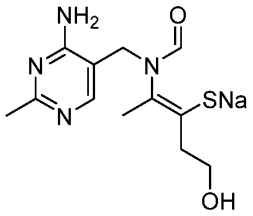
[0293] **Example 8 (M16)**

[0294] (*Z*)-((2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl isobutyrate



[0295] Step 1

[0296] Sodium (*Z*)-2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-ene-3-thiolate

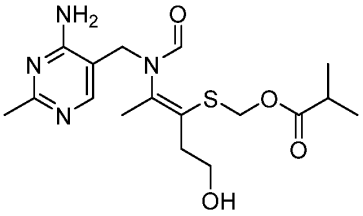


[0297] To a solution of vitamin B1 (40.0 g, 132 mmol) in EtOH (80.0 mL) was added NaOEt (40.0 g, 123mmol) in EtOH at -10 °C. The mixture was stirred at 10 °C for 30 min.

The solid formed was collected by filtration and dried to give the title compound (36.0 g, 88.9%) as a yellow solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 9.51 (s, 1H), 8.05 (s, 1H), 7.18 (s, 2 H), 5.41 (s, 3 H), 3.65-3.62 (m, 2H), 3.01-2.95 (m, 2H), 2.48 (s, 3H), 2.34 (s, 3H).

[0298] Step 2

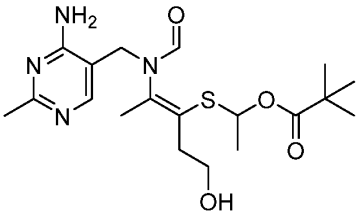
[0299] (Z)-((2-(N-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl isobutyrate



[0300] To a mixture of sodium (Z)-2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-ene-3-thiolate (3.00 g, 9.86 mmol) in EtOH (30.0 mL) was added NaOH (120 mg, 3.00 mmol) at 25°C followed by chloromethyl isobutyrate (1.35 g, 9.88 mmol). The mixture was stirred at 25°C for 1 h and was concentrated to give a residue. The residue was purified by prep-HPLC (basic condition) and lyophilized to give 150 mg of crude brown solid which was further purified by prep-HPLC (buffered with NH4HCO3) followed by lyophilization to provide the title compound (50.1 mg, 29%) as a yellow solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.92 (s, 1H), 7.83 (s, 1H), 6.70 (br, 2H), 4.96 (s, 2H), 4.69 (t, J = 5.2 Hz, 1H), 4.34 (s, 2H), 3.41-3.52 (m, 2H), 2.58 (t, J = 6.8 Hz, 2H), 2.50-2.40 (m, 1H), 2.28 (s, 3H), 1.95 (s, 3H), 1.08 (d, J = 7.2 Hz, 6H). MS (ES$^+$) m/e 383.3 $(M+H)^+$.

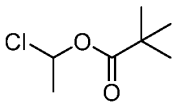
[0301] **Example 9 (M18)**

[0302] (*Z*)-1-((2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)ethyl pivalate



[0303] Step 1

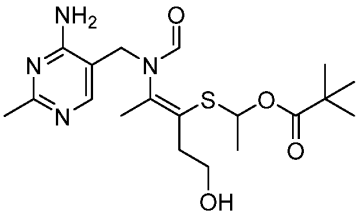
[0304] 1-Chloroethyl pivalate



[0305] To a mixture of pivaloyl chloride (9.80 g, 81.2 mmol, 10.0 mL) and 2,4,6-trimethyl-1,3,5-trioxane (4.95 g, 37.4 mmol, 5.00 mL) was added $ZnCl_2$ (1 M, 2.50 mL) and the mixture was stirred at 90 °C for 1 h. The reaction mixture was cooled to room temperature and diluted with EtOAc (100 mL). The resulting organic solution was washed with ice-cooled $NaHCO_3$ (20.0 mL × 3) solution and brine, dried over Na_2SO_4, filtered, and concentrated to give a residue. The residue was distilled at 90 °C (gage pressure: -0.09 MPa) to give the title compound (1.60 g) as a colorless oil. ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.52-6.56 (m, 1H), 1.80 (d, J = 5.6 Hz, 3H), 1.22 (s, 9H).

[0306] Step 2

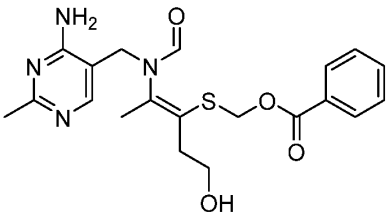
[0307] (*Z*)-1-((2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)ethyl pivalate



[0308] To a mixture of sodium (*Z*)-2-(*N*-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-ene-3-thiolate (3.00 g, 9.86 mmol) in EtOH (30.0 mL) was added NaOEt (3.20 g, 9.88 mmol) in EtOH at 25 °C followed by 1-chloroethyl pivalate (1.60 g, 9.72 mmol) and NaOH (120 mg, 3.00 mmol). The mixture was stirred at 25 °C for 12 h and was concentrated to give a residue. The residue was purified by prep-HPLC followed by lyophilization to give the title compound (60.0 mg, 2% yield) as a yellow solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.92 (s, 1H), 7.85 (s, 1H), 6.48 (br, 2H), 5.74-5.79 (m, 1H), 4.46 (d, J = 15.2 Hz, 1H), 4.37 (s, 1H), 4.28 (d, J = 15.2 Hz, 1H), 3.53-3.55 (m, 2H), 2.52-2.57 (m, 2H), 2.30 (s, 3H), 1.95 (s, 3H), 1.29 (d, J = 6.4 Hz, 3H), 1.33 (s, 9H). MS (ES^+) m/e 411.3 $(M+H)^+$.

[0309] **Example 10 (M21)**

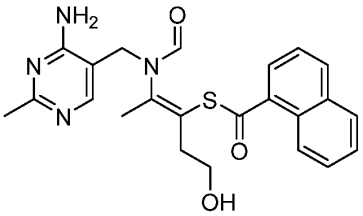
[0310] (*Z*)-((2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl benzoate



[0311] To a mixture of vitamin B1 (5 g, 16.6 mmol, 1 equiv) and KI (138.0 mg, 831.0 μmol, 0.05 equiv) in H_2O (25 mL) was added NaOH (1.33 g, 33.2 mmol, 2 equiv) in portions. The mixture was stirred at 20°C for 0.5 h and then a solution of chloromethyl benzoate (2.84 g, 16.6 mmol, 1 equiv) in THF (25 mL) was added drop-wise. The reaction mixture was then stirred at 20°C for another 10 h. LCMS showed a new product formed. The mixture was quenched by adding 5 mL of methanol and pH adjusted to ~7 with saturated sodium bicarbonate aqueous solution. The aqueous phase was then extracted with ethyl acetate (200 mL × 2). The combined organic phases were dried by anhydrous Na_2SO_4, filtered, and concentrated to give a crude product which was triturated with ethyl acetate (50 mL) to provide the title compound (700 mg, 10%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.96-7.92 (m, 3H), 7.81 (s, 1H), 7.73-7.67 (m, 1H), 7.57-7.53 (m, 2H), 6.76 (brs, 2H), 5.28 (s, 2H), 4.74-4.70 (m, 1H), 4.32 (s, 2H), 3.54-3.48 (m, 2H), 2.69-2.66 (m, 2H), 2.26 (s, 3H), 1.96 (s, 3H). MS (ES^+) m/e 417.2 $(M+H)^+$.

[0312] **Example 11 (M34)**

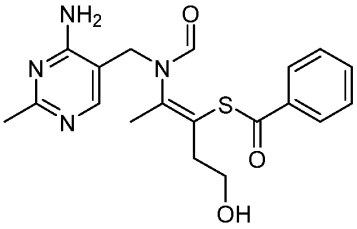
[0313] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) naphthalene-1-carbothioate



[0314] To a mixture of vitamin B1 (9 g, 29.9 mmol, 1 equiv) and KI (497.0 mg, 2.99 mmol, 0.1 equiv) in H_2O (150 mL) and THF (150 mL) was added NaOH (2.39 g, 59.8 mmol, 2 equiv). The mixture was stirred at 20 °C for 0.5 h. 1-Naphthoyl chloride (11.1 g, 58.0 mmol, 8.71 mL, 1.94 equiv) was added drop-wise. The reaction mixture was stirred at 20°C for another 12 h. LCMS showed completion of the reaction. The reaction mixture was extracted with ethyl acetate (100 mL × 3). The aqueous phase was adjusted to pH 7~8 with saturated sodium bicarbonate solution and extracted with ethyl acetate (100 mL × 3). The combined organic phases were dried with anhydrous Na_2SO_4, filtered and concentrated to give a crude product which was triturated with ethyl acetate (10 mL) to provide the title compound (420 mg, 962 μmol, 3% yield) as an off-white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 8.21-8.20 (m, 2H), 8.05-8.02 (m, 1H), 7.99 (s, 1H), 7.92-7.91 (m, 1H), 7.88 (s, 1H), 7.68-7.59 (m, 3H), 6.74 (brs, 2H), 4.70-4.68 (m, 1H), 4.44 (s, 2H), 3.53 (m, 2H), 2.70-2.68 (m, 2H), 2.19 (s, 3H), 2.13 (s, 3H). MS (ES^+) m/e 437.1 $(M+H)^+$.

[0315] Example 12 (M26)

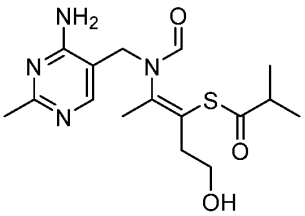
[0316] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) benzothioate



[0317] To a solution of vitamin B1 (1.00 g, 3.32 mmol, 1.00 equiv) and NaOH (133.0 mg, 3.32 mmol, 1.00 equiv) in H_2O (50.0 mL) and THF (5.00 mL) was added benzoyl chloride (935.0 mg, 6.65 mmol, 772.0 µL, 2.00 equiv) at 0 °C. The reaction mixture was stirred at 0 °C for 0.5 h. LCMS showed that the starting material was consumed and a product with the desired mass was detected. The reaction mixture was quenched with MeOH (5.00 mL) and adjusted to pH = 7, and was extracted with 10:1 DCM:MeOH (25.0 mL × 2). The combined organic layers were dried over Na_2SO_4, filtered and concentrated under a reduced pressure to give a crude product, which upon trituration with EtOAc/EtOH/DCM (2:2:1) provided the title compound (251 mg, 18.6%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.89 (s, 1H), 7.82 (s, 1H), 7.73 (d, J = 7.2 Hz, 2H), 7.68 (t, J = 7.6 Hz, 1H), 7.53 (t, J = 7.6 Hz, 2H), 6.65 (brs, 2H), 4.64 (t, J = 4.2 Hz, 1H), 4.39 (s, 2H), 3.46 (m, 2H), 2.57 (m, 2H), 2.16 (s, 3H), 2.14 (s, 3H). MS (ES$^+$) m/e 387.0 $(M+H)^+$.

[0318] Example 13 (M27)

[0319] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-methylpropanethioate

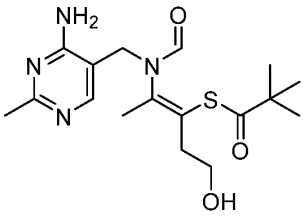


[0320] To a solution of vitamin B1 (1.00 g, 3.32 mmol, 1.00 equiv) and NaOH (266 mg, 6.65 mmol, 2.00 equiv) in H_2O (50.0 mL) and THF (5.00 mL) was added isobutyryl chloride (708 mg, 6.65 mmol, 695.0 µL, 2.00 equiv) at 0 °C. The reaction mixture was stirred at 0 °C for 0.5 h, quenched with MeOH (5.00 mL), and adjusted to pH = 7. The mixture was extracted with 10:1 DCM:MeOH (25.0 mL × 2). The combined organic layers were dried over Na_2SO_4, filtered and concentrated under reduced pressure to give a crude product. The

crude product was triturated with EtOAc/EtOH/DCM (2:2:1) to give the title compound (224.0 mg, 19.0%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.79 (s, 1H), 7.77 (s, 1H), 6.68 (brs, 2H), 4.61 (t, J = 4.2 Hz, 1H), 4.35 (s, 2H), 3.40 (q, J = 6.8 Hz, 2H), 2.60-2.50 (m, 1H), 2.44 (t, J = 7.2 Hz, 1H), 2.26 (s, 3H), 2.07 (s, 3H), 1.02 (s, 3H), 1.00 (s, 3H). MS (ES$^+$) m/e 353.3 (M+H)$^+$.

[0321] Example 14 (M28)

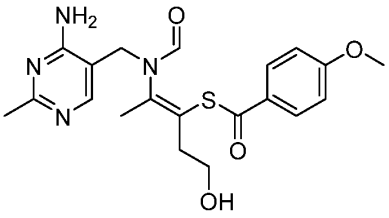
[0322] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2,2-dimethylpropanethioate



[0323] To a mixture of sodium (*Z*)-2-(*N*-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-ene-3-thiolate (0.50 g, 1.64 mmol, 1.00 equiv) and NaOH (65.7 mg, 1.64 mmol, 1.00 equiv) in H_2O (5.00 mL) was added pivaloyl chloride (198 mg, 1.64 mmol, 202 µL, 1 equiv) at 0 °C and the resulting mixture was stirred at 0° C for 1 h. The mixture was concentrated and purified by column chromatography (SiO_2, dichloromethane: methanol = 100:1 to 10:1, TLC: dichloromethane: methanol= 10:1, R_f = 0.4) to give the title compound (110 mg, 18.2%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.79 (s, 1 H) 7.75 (s, 1 H) 6.70 (br s, 2 H) 4.60 (t, J = 5.6 Hz, 1 H) 4.35 (br s, 2 H) 3.39 (q, J = 6.8 Hz, 2 H) 2.42 (br t, J = 6.8 Hz, 2 H) 2.26 (s, 3 H) 2.08 (s, 3 H) 1.07 (s, 9 H). MS (ES$^+$) m/e 367.2 (M+H)$^+$.

[0324] Example 15 (M29)

[0325] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-methoxybenzothioate

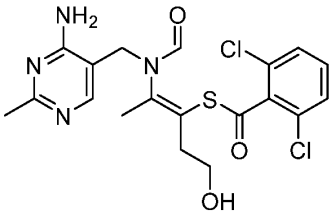


[0326] To a mixture of sodium (*Z*)-2-(*N*-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-ene-3-thiolate (0.50 g, 1.64 mmol, 1.00 equiv) and NaOH (65.7 mg, 1.64 mmol, 1.00 equiv) in H_2O (4 mL) was added compound 4-

methoxybenzoyl chloride (280 mg, 1.64 mmol, 226 uL, 1.00 equiv) at 0 °C and resulting mixture was stirred at 0 °C for 1 h. LCMS showed that the starting material was consumed and a product with the desired mass was detected. The mixture was then concentrated, and the residue was purified by silica gel column chromatography (100:0 to 20:1 dichloromethane:methanol) to give the title compound (60.0 mg, 9%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.87 (s, 1 H), 7.81 (s, 1 H), 7.72 (br d, J = 8.8 Hz, 2 H), 7.05 (br d, J = 8.8 Hz, 2 H), 6.64 (m, 2 H), 4.64 (br s, 1 H), 4.37 (br s, 2 H), 3.84 (s, 3 H), 3.46 (br d, J = 4.0 Hz, 2 H), 2.58 (m, 2 H), 2.19 (s, 3 H), 2.12 (s, 3 H). MS (ES$^+$) m/e 417.1 (M+H)$^+$.

[0327] Example 16 (M30)

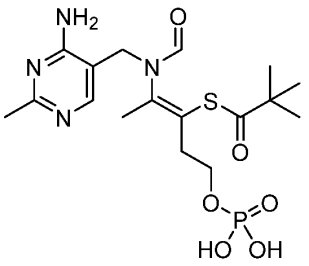
[0328] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2,6-dichlorobenzothioate



[0329] To a mixture of sodium (*Z*)-2-(*N*-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-ene-3-thiolate (0.50 g, 1.64 mmol, 1.00 equiv) and NaOH (65.7 mg, 1.64 mmol, 1.00 equiv) in H_2O (4.00 mL) was added 2,6-dichlorobenzoyl chloride (344 mg, 1.64 mmol, 235 µL, 1.00 equiv) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and was then concentrated to provide a residue. The residue was purified by prep-HPLC to provide the title compound (150 mg, 20.1%) as a white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 7.96 (s, 1 H), 7.84 (s, 1 H), 7.60 - 7.54 (m, 3 H), 6.68 (m, 2 H), 4.68 - 4.75 (m, 1 H), 4.41 (br s, 2 H), 3.46 - 3.52 (m, 2 H), 2.62 - 2.69 (m, 2 H), 2.26 (s, 3 H), 2.14 (s, 3 H). MS (ES$^+$) m/e 455.1 (M+H)$^+$.

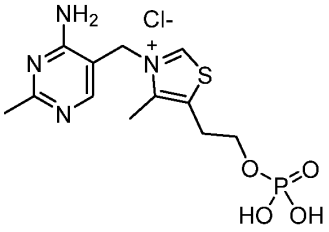
[0330] Example 17 (M31)

[0331] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-(phosphonooxy)pent-2-en-3-yl) 2,2-dimethylpropanethioate



[0332] Step 1

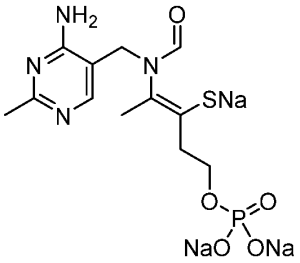
[0333] 3-((4-Amino-2-methylpyrimidin-5-yl)methyl)-4-methyl-5-(2-(phosphonooxy)ethyl)thiazol-3-ium chloride



[0334] To a solution of polyphosphoric acid (150 g, 49.8 mmol, 1.00 equiv) at 130 °C was added vitamin B1 (15.0 g, 49.8 mmol, 1.00 equiv) in portions. The mixture was stirred at 100-130 °C for 2 h. Water (250 mL) was added, and the mixture was stirred at 100 °C for additional 2 h, cooled to 25 °C, and extracted by trioctylamine/MTBE (1:1, 250 mL × 2). The water layer was separated, diluted with 400 mL of EtOH, and stirred at 25 °C for 2 h. The white solid formed was collected by filtration and washed by EtOH (200 mL × 2) and dried to provide the title compound (17.0 g, 98.7%).

[0335] Step 2

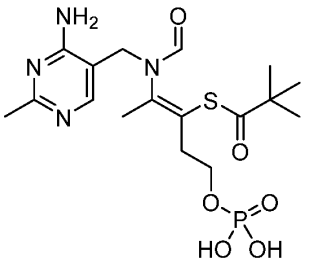
[0336] Sodium (*Z*)-4-(*N*-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-3-sulfidopent-3-en-1-yl phosphate



[0337] To a mixture of 3-((4-amino-2-methylpyrimidin-5-yl)methyl)-4-methyl-5-(2-(phosphonooxy)ethyl)thiazol-3-ium (2.00 g, 5.79 mmol, 1.00 equiv) in H_2O (20 mL) was added NaOH (5.64 g, 42.3 mmol, 7.30 equiv) slowly at 0-5 °C. The mixture was stirred at 5-10 °C for 0.5 h and concentrated to provide the title compound (2.40 g, 97%) as a yellow solid which was used directly for the next step reaction without further purification.

[0338] Step 3

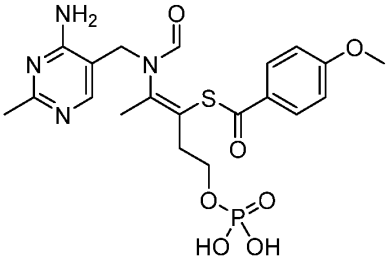
[0339] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-(phosphonooxy)pent-2-en-3-yl) 2,2-dimethylpropanethioate



[0340] To a mixture of sodium (*Z*)-4-(*N*-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-3-sulfidopent-3-en-1-yl phosphate (0.800 g, 1.87 mmol, 1.00 equiv) in NaOH (1 M, 3.74 mL, 2.00 equiv) was added pivaloyl chloride (840 mg, 6.97 mmol, 857 µL, 3.73 equiv). The resulting mixture was stirred at 25 °C for 2 h. TLC (ethyl acetate: dichloromethane = 2: 1) showed the starting material was consumed and a new spot was formed. The reaction mixture was concentrated and purified by reversed-phase HPLC to give the title compound (132 mg, 15.7%) as a white solid. ^{1}H NMR (400 MHz, MeOD) δ 7.99 (s, 1 H), 7.79 (s, 1 H), 3.93 (br d, *J* = 5.6 Hz, 2 H), 2.56 (s, 3 H), 2.25 (s, 3 H), 1.14 (s, 9 H). MS (ES$^+$) m/e 447.1 (M+H)$^+$.

[0341] **Example 18 (M32)**

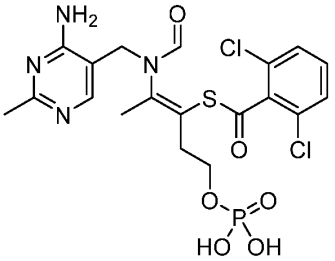
[0342] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-(phosphonooxy)pent-2-en-3-yl) 4-methoxybenzothioate



[0343] To a mixture of sodium (Z)-4-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-3-sulfidopent-3-en-1-yl phosphate (0.80 g, 1.87 mmol, 1.00 equiv) in NaOH (1 M, 3.74 mL, 2.00 equiv) was added 4-methoxybenzoyl chloride (1.19 g, 6.97 mmol, 958 µL, 3.73 equiv). The resulting mixture was stirred at 25 °C for 2 h. TLC (ethyl acetate: dichloromethane= 2: 1) showed the starting material was consumed, and a new spot formed. The mixture was concentrated and purified by reversed-phase HPLC to give the title compound (174 mg, 18.1%) as a white solid. ^{1}H NMR (400 MHz, MeOD) δ 8.04 (s, 1 H), 7.99 (s, 1 H), 7.75 (d, *J* = 8.8 Hz, 2 H), 7.00 (d, *J* = 8.8 Hz, 2 H), 4.74 - 4.35 (m, 2 H), 4.03 (q, *J* = 5.6 Hz, 2 H), 3.88 (s, 3 H), 2.81 (m, 2 H), 2.40 (s, 3 H), 2.30 (s, 3 H). MS (ES$^+$) m/e 497.1 (M+H)$^+$.

[0344] **Example 19 (M33)**

[0345] (*Z*)-S-(2-(*N*-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-(phosphonooxy)pent-2-en-3-yl) 2,6-dichlorobenzothioate

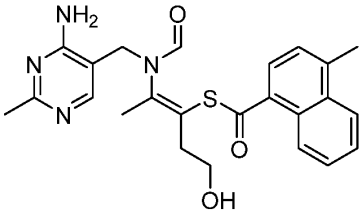


[0346] To a mixture of sodium (*Z*)-4-(*N*-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-3-sulfidopent-3-en-1-yl phosphate (0.800 g, 1.87 mmol, 1.00 equiv) in NaOH (1 M, 3.74 mL, 2.00 equiv) was added 2,6-dichlorobenzoyl chloride (391 mg, 1.87 mmol, 268 µL, 1.00 eq). The mixture was stirred at 25°C for 2 h. TLC (ethyl acetate: dichloromethane = 2: 1) showed that the starting material was consumed, and a new spot formed. The mixture was concentrated and purified by reversed-phase HPLC to provide the title compound (72.0 mg, 7.20%) as a white solid. ^{1}H NMR (400 MHz, MeOD) δ 8.06 (s, 1 H), 8.02 (s, 1 H), 7.44 (s, 3 H), 4.07 - 4.04 (m, 2 H), 2.49 (s, 3 H), 2.32 (s, 3 H). MS (ES$^+$) m/e 535.0 (M+H)$^+$.

[0347] **Examples 20 to 59**: The following compounds were synthesized following the procedure for the preparation of M34 (Example 11) with appropriate starting material.

[0348] **Example 20** (M37)

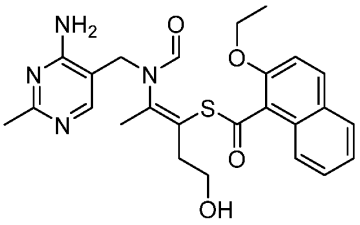
[0349] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-methylnaphthalene-1-carbothioate



[0350] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.30 - 8.25 (m, 1H), 8.13 (dd, *J* = 3.2, 6.6 Hz, 1H), 7.98 (s, 1H), 7.88 - 7.80 (m, 2H), 7.69 - 7.64 (m, 2H), 7.47 (d, *J* = 7.6 Hz, 1H), 6.71 (br s, 2H), 4.69 (br t, *J* = 5.6 Hz, 1H), 4.43 (br s, 2H), 3.55 - 3.47 (m, 2H), 2.72 (s, 3H), 2.66 (br t, *J* = 6.8 Hz, 2H), 2.18 (s, 3H), 2.13 (s, 3H). MS (ES$^+$) m/e 451 (M+H)$^+$.

[0351] **Example 21** (M38)

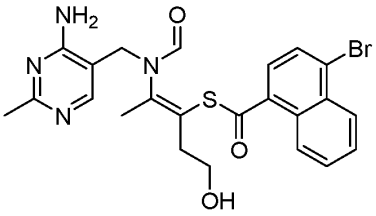
[0352] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-ethoxynaphthalene-1-carbothioate



[0353] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.06 - 7.92 (m, 4H), 7.57 - 7.42 (m, 4H), 6.67 (br s, 2H), 4.73 (br t, *J* = 4.8 Hz, 1H), 4.42 (br s, 2H), 4.23 (q, *J* = 6.8 Hz, 2H), 3.62 - 3.51 (m, 2H), 2.73 (br s, 2H), 2.23 (s, 3H), 2.13 (s, 3H), 1.32 (br t, *J* = 6.8 Hz, 3H).MS (ES$^+$) m/e 481 (M+H)$^+$.

[0354] **Example 22** (M39)

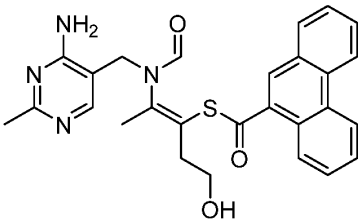
[0355] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-bromonaphthalene-1-carbothioate



[0356] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.27 - 8.23 (m, 2H), 8.02 - 7.98 (m, 2H), 7.86 - 7.78 (m, 4H), 6.72 (br d, *J* = 1.2 Hz, 2H), 4.71 (t, *J* = 5.6 Hz, 1H), 4.44 (br s, 2H), 3.53 (q, *J* = 6.4 Hz, 2H), 2.67 (br t, *J* = 6.8 Hz, 2H), 2.19 (s, 3H), 2.08 (s, 3H). MS (ES$^+$) m/e 515 (M+H)$^+$.

[0357] **Example 23** (M40)

[0358] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) phenanthrene-9-carbothioate

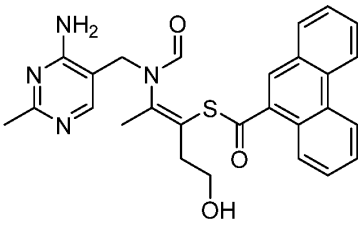


[0359] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.92 (d, *J* = 8.0 Hz, 1H), 8.88 (d, *J* = 8.4 Hz, 1H), 8.26-8.19 (m, 3H), 8.03 (s, 1H), 7.90 (s, 1H), 7.80-7.75 (m, 1H), 7.80-7.77 (m, 2H), 7.74-7.72 (m, 1H), 6.80 (s, 2H), 4.74 (t, *J* = 4.8 Hz, 1H), 4.47 (s, 2H), 4.11 (d, *J* = 3.6 Hz,

2H), 3.58 (q, *J* = 6.0 Hz, 2H), 3.17 (d, *J* = 2.4 Hz, 2H), 2.72 (t, *J* = 6.4 Hz, 2H), 2.21 (s, 3H), 2.09 (s, 3H). MS (ES^+) m/e 487 $(M+H)^+$.

[0360] **Example 24** (M43)

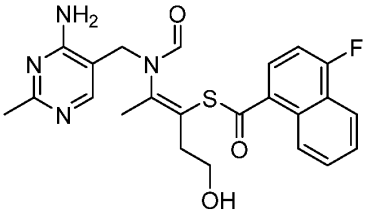
[0361] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) phenanthrene-9-carbothioate



[0362] ^1H NMR (400 MHz, DMSO-d6) δ 7.88 - 7.83 (m, 2H), 7.32 - 7.20 (m, 3H), 6.73 (br d, *J* = 1.2 Hz, 2H), 4.66 (t, *J* = 5.6 Hz, 1H), 4.40 (br s, 2H), 3.47 (q, *J* = 6.4 Hz, 2H), 2.75 (br s, 2H), 2.67 (br s, 2H), 2.58 (br t, *J* = 6.8 Hz, 2H), 2.21 (s, 3H), 2.13 (s, 3H), 1.70 (m, 4H). MS (ES^+) m/e 441 $(M+H)^+$.

[0363] **Example 25** (M44)

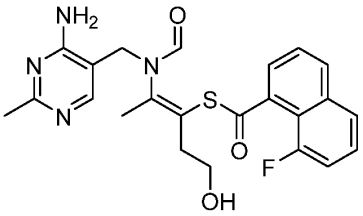
[0364] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-fluoronaphthalene-1-carbothioate



[0365] ^1H NMR (400 MHz, DMSO-d6) δ 8.33 (br d, *J* = 8.0 Hz, 1H), 8.16 (br d, *J* = 7.6 Hz, 1H), 8.08 - 7.94 (m, 2H), 7.87 (s, 1H), 7.82 - 7.70 (m, 2H), 7.45 (dd, *J* = 8.0, 10.1 Hz, 1H), 6.71 (br s, 2H), 4.69 (br t, *J* = 5.2 Hz, 1H), 4.43 (br s, 2H), 3.66 - 3.47 (m, 2H), 2.66 (br t, *J* = 6.8 Hz, 2H), 2.25 - 2.15 (m, 3H), 2.13 - 1.96 (m, 3H). MS (ES^+) m/e 455 $(M+H)^+$.

[0366] **Example 26** (M45)

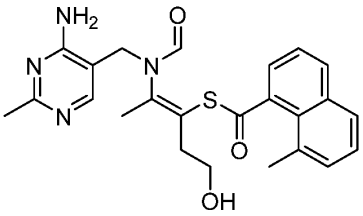
[0367] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 8-fluoronaphthalene-1-carbothioate



[0368] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.20 (br d, J = 8.4 Hz, 1H), 7.98 (s, 1H), 7.91 - 7.83 (m, 2H), 7.73 - 7.58 (m, 2H), 7.56 - 7.29 (m, 2H), 6.81 (br s, 2H), 4.73 (br s, 1H), 4.46 (br s, 2H), 3.55 (br s, 2H), 2.69 (br t, J = 6.4 Hz, 2H), 2.20 (s, 3H), 2.18 (s, 3H). MS (ES^+) m/e 455 $(M+H)^+$.

[0369] **Example 27** (M46)

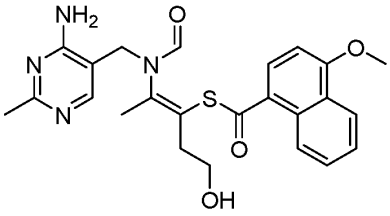
[0370] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 8-methylnaphthalene-1-carbothioate



[0371] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.14 (dd, J_1 = 1.6, J_2 = 8.0 Hz, 1H), 7.98 (s, 1H), 7.91 - 7.86 (m, 2H), 7.59 - 7.47 (m, 4H), 6.83 (br s, 2H), 4.75 (br t, J = 5.2 Hz, 1H), 4.46 (br s, 2H), 3.54 (q, J = 6.4 Hz, 2H), 2.70 - 2.64 (m, 2H), 2.48 (s, 3H), 2.19 (d, J = 5.6 Hz, 6H). MS (ES^+) m/e 451 $(M+H)^+$.

[0372] **Example 28** (M47)

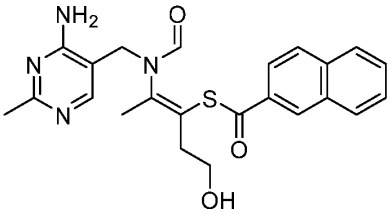
[0373] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-methoxynaphthalene-1-carbothioate



[0374] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.42 (d, J = 8.0 Hz, 1H), 8.26 (d, J = 0.4 Hz, 1H), 8.05 (d, J = 8.4 Hz, 1H), 7.96 (s, 1H), 7.87 (s, 1H), 7.68 - 7.65 (m, 1H), 7.65 - 7.61 (m, 1H), 7.08 (d, J = 8.4 Hz, 1H), 6.71 (br s, 2H), 4.69 (br t, J = 5.6 Hz, 1H), 4.43 (br s, 2H), 4.08 (s, 3H), 3.55 - 3.50 (m, 2H), 2.66 - 2.63 (m, 2H), 2.18 (s, 3H), 2.14 (s, 3H). MS (ES^+) m/e 467 $(M+H)^+$.

[0375] **Example 29** (M49)

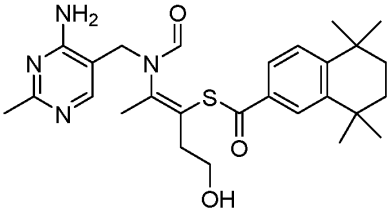
[0376] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) naphthalene-2-carbothioate



[0377] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.41 (s, 1H), 8.18 (br d, *J* = 8.0 Hz, 1H), 8.06 - 8.02 (m, 2H), 7.95 (s, 1H), 7.85 (s, 1H), 7.76 - 7.64 (m, 3H), 6.64 (br s, 2H), 4.67 (br t, *J* = 5.6 Hz, 1H), 4.41 (br s, 2H), 3.52 (q, *J* = 6.4 Hz, 2H), 2.62 (br t, *J* = 6.8 Hz, 2H), 2.18 (s, 3H), 2.14 (s, 3H). MS (ES$^+$) m/e 437 (M+H)$^+$.

[0378] **Example 30** (M50)

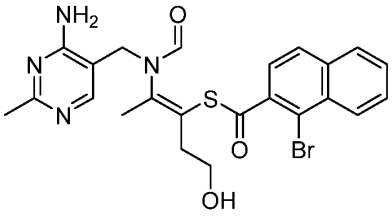
[0379] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 5,5,8,8-tetramethyl-5,6,7,8-tetrahydronaphthalene-2-carbothioate



[0380] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.84 (d, *J* = 14.4 Hz, 2H), 7.64 (s, 1H), 7.51-7.46 (m, 2H), 6.63 (s, 2H), 4.63 (t, *J* = 4.2 Hz, 1H), 4.38 (s, 2H), 3.45 (d, *J* = 5.6 Hz, 2H), 2.56 (t, *J* = 6.4 Hz, 2H), 2.20 (s, 2H), 2.13 (s, 2H), 1.65 (s, 3H), 1.25 (s, 12H). MS (ES$^+$) m/e 497 (M+H)$^+$.

[0381] **Example 31** (M51)

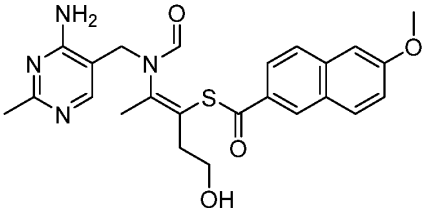
[0382] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 1-bromonaphthalene-2-carbothioate



[0383] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.28 (d, *J* = 8.8 Hz, 2H), 8.12-8.09 (m, 2H), 7.98 (s, 1H), 7.88 (s, 1H), 7.80-7.76 (m, 1H), 7.74-7.72 (m, 1H), 7.46 (d, *J* = 8.4 Hz, 2H), 6.75 (s, 2H), 4.72 (t, *J* = 5.6 Hz, 1H), 4.45 (s, 2H), 3.55 (q, *J* = 6.8 Hz, 2H), 2.68 (t, *J* = 6.4 Hz, 2H), 2.20 (s, 3H), 2.17 (s, 3H). MS (ES$^+$) m/e 516 (M+H)$^+$.

[0384] **Example 32** (M52)

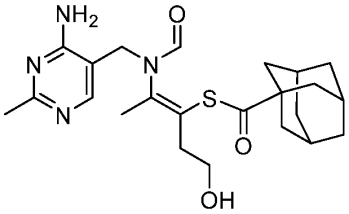
[0385] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 6-methoxynaphthalene-2-carbothioate



[0386] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.33 (d, J = 1.2 Hz, 1H), 8.08 (d, J = 8.8 Hz, 1H), 7.94 - 7.83 (m, 3H), 7.70 (dd, J_1 = 1.6, J_2 = 8.4 Hz, 1H), 7.43 (d, J = 2.0 Hz, 1H), 7.28 (dd, J_1 = 2.4, J_2 = 8.8 Hz, 1H), 6.76 - 6.55 (m, 2H), 4.65 (t, J = 5.6 Hz, 1H), 4.40 (br s, 2H), 3.92 (s, 3H), 3.53 - 3.46 (m, 2H), 2.60 (br t, J = 6.8 Hz, 2H), 2.16 (s, 3H), 2.14 (s, 3H). MS (ES$^+$) m/e 467 (M+H)$^+$.

[0387] **Example 33** (M53)

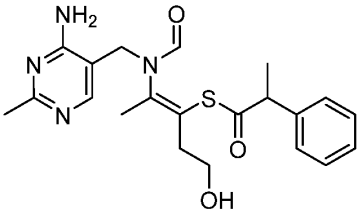
[0388] S-((Z)-2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) (3r,5r,7r)-adamantane-1-carbothioate



[0389] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.79 (s, 1H), 7.74 (s, 1H), 6.75 (br s, 2H), 4.60 (t, J = 5.6 Hz, 1H), 4.35 (br s, 2H), 3.44 - 3.38 (m, 2H), 2.41 (t, J = 6.8 Hz, 2H), 2.28 (s, 3H), 2.08 (s, 3H), 1.98 (br s, 3H), 1.84 - 1.75 (m, 1H), 1.71 - 1.62 (m, 11H). MS (ES$^+$) m/e 445 (M+H)$^+$.

[0390] **Example 34** (M54)

[0391] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-phenylpropanethioate

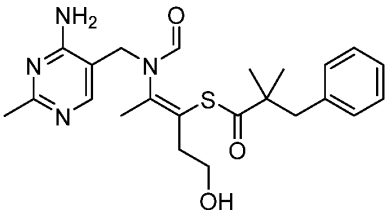


[0392] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.77 (s, 1H), 7.73 (s, 1H), 7.34 - 7.24 (m, 5H), 6.64 (s, 2H), 4.55 (t, J = 5.6 Hz, 1H), 4.40-4.24 (m, 2H), 3.82 (m, 1H), 3.29 - 3.28 (m, 2H),

2.48 - 2.39 (m, 1H), 2.32 - 2.28 (m, 1H), 2.28 (s, 3H), 1.34-1.30 (d, *J* = 7.6 Hz, 3H). MS (ES^+) m/e 415 $(M+H)^+$.

[0393] **Example 35** (M55)

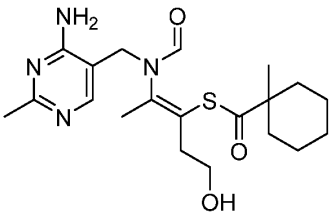
[0394] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2,2-dimethyl-3-phenylpropanethioate



[0395] ^1H NMR (400 MHz, DMSO-d6) δ 7.78 (s, 1H), 7.68 (s, 1H), 7.27 - 7.20 (m, 3H), 7.09 (m, 2H), 6.71 (s, 1H), 4.61 (t, *J* = 5.6 Hz, 1H), 4.34 (s, 1H), 3.42 -3.36 (m, 2H), 2.73 (s, 2H), 2.43 (t, *J* = 6.8 Hz, 2H), 2.25 (s, 2H), 2.07 (s, 3H), 1.02 (s, 6H).. MS (ES^+) m/e 443 $(M+H)^+$.

[0396] **Example 36** (M57)

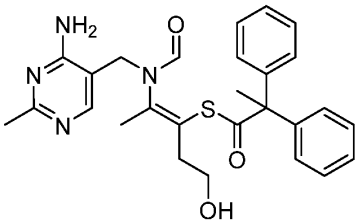
[0397] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 1-methylcyclohexane-1-carbothioate



[0398] ^1H NMR (400 MHz, DMSO-d6) δ 7.79 (s, 1H), 7.75 (s, 1H), 6.70 (s, 2H), 4.61 (t, *J* = 5.6 Hz, 1H), 4.34 (s, 2H), 3.45 - 3.35 (m, 2H), 2.43 (br t, *J* = 6.8 Hz, 2H), 2.27 (s, 3H), 2.08 (s, 3H), 1.78 (dd, *J* = 6.8, 9.8 Hz, 2H), 1.49 - 1.24 (m, 8H), 1.05 (s, 3H). MS (ES^+) m/e 407 $(M+H)^+$.

[0399] **Example 37** (M58)

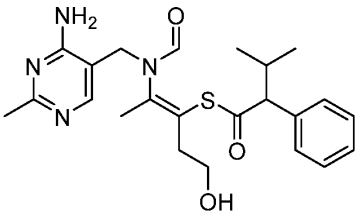
[0400] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2,2-diphenylpropanethioate



[0401] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.77 (d, *J* = 3.2 Hz, 2H), 7.35 - 7.25 (m, 6H), 7.16 - 7.08 (m, 4H), 6.73 (br s, 2H), 4.58 (t, *J* = 5.6 Hz, 1H), 4.31 (br s, 2H), 3.32 - 3.28 (m, 2H), 2.41 (br t, *J* = 6.8 Hz, 2H), 2.26 (s, 3H), 2.05 (s, 3H), 1.84 (s, 3H). MS (ES$^+$) m/e 491 (M+H)$^+$.

[0402] **Example 38** (M59)

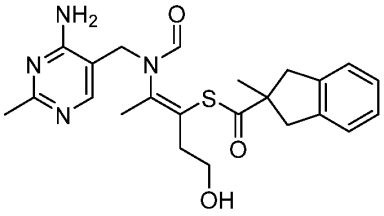
[0403] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 3-methyl-2-phenylbutanethioate



[0404] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.77 (s, 1H), 7.73 (s, 1H), 7.35 - 7.30 (m, 2H), 7.29 - 7.21 (m, 3H), 6.64 (s, 2H), 4.55 (t, *J* = 5.6 Hz, 1H), 4.25 (s, 2H), 3.45 - 3.41 (m, 1H), 3.29 - 3.17 (m, 2H), 2.48 - 2.39 (m, 1H), 2.27-2.20 (m, 5H), 1.99 (s, 3H), 0.92 (d, *J* = 6.4 Hz, 3H), 0.61 (d, *J* = 6.8 Hz, 3H). MS (ES$^+$) m/e 443 (M+H)$^+$.

[0405] **Example 39** (M62)

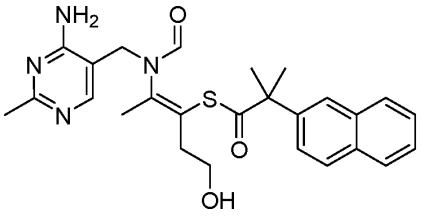
[0406] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-methyl-2,3-dihydro-1H-indene-2-carbothioate



[0407] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.79 (s, 2H), 7.21 - 7.13 (m, 4H), 6.72 (s, 2H), 4.62-4.59 (t, *J* = 5.6 Hz, 1H), 4.35 (s, 1H), 3.43 - 3.38 (m, 2H), 3.20 (d, *J* = 16.0 Hz, 2H), 2.73 (d, *J* = 16.0 Hz, 2H), 2.44 (t, *J* = 6.8 Hz, 2H), 2.22 (s, 3H), 2.09 (s, 3H), 1.21 (s, 3H). MS (ES$^+$) m/e 441 (M+H)$^+$.

[0408] **Example 40** (M63)

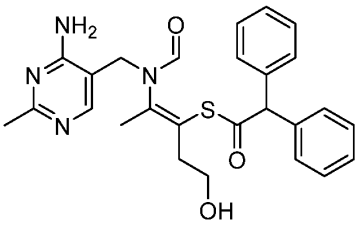
[0409] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-methyl-2-(naphthalen-2-yl)propanethioate



[0410] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.90-7.89 (m, 1H), 7.88-7.85 (m, 3H), 7.77 (s, 1H), 7.73 (s, 1H), 7.53-7.51 (m, 2H), 7.35-7.33 (m, 1H), 6.71 (s, 2H), 4.50 (t, *J* = 5.6 Hz, 1H), 4.31 (s, 2H), 3.26 (q, *J* = 5.6 Hz, 2H), 2.38 (t, *J* = 6.8 Hz, 2H), 2.22 (s, 3H), 2.04 (s, 3H), 1.53 (s, 6H). MS (ES$^+$) m/e 479 (M+H)$^+$.

[0411] **Example 41** (M64)

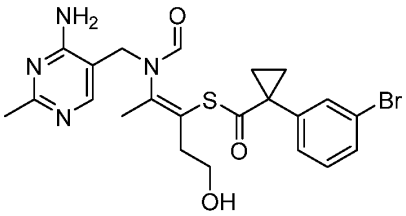
[0412] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2,2-diphenylethanethioate



[0413] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.77 (d, *J* = 3.2 Hz, 2H), 7.39 - 7.23 (m, 12H), 6.78 - 6.57 (m, 2H), 5.29 (s, 1H), 4.63 - 4.51 (m, 1H), 4.31 (br s, 2H), 2.21 (s, 3H), 2.02 (s, 3H). MS (ES$^+$) m/e 477 (M+H)$^+$.

[0414] **Example 42** (M65)

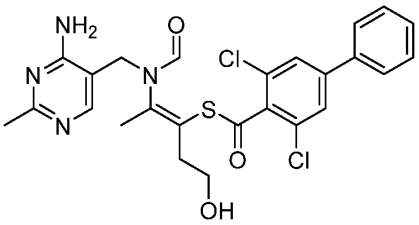
[0415] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 1-(3-bromophenyl)cyclopropane-1-carbothioate



[0416] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.74 (s, 1H), 7.71 (s, 1H), 7.55-7.52(m, 2H), 7.39-7.37 (m, 2H), 6.71 (s, 2H), 4.55-4.52 (m, 1H), 4.31 (s, 2H), 3.37-3.35 (m, 1H), 2.41-2.34 (m, 2H), 2.37 (s, 3H), 2.04 (s, 3H), 1.40-1.37 (m, 2H), 1.22-1.19 (s, 2H). MS (ES$^+$) m/e 506 (M+H)$^+$.

[0417] **Example 43** (M66)

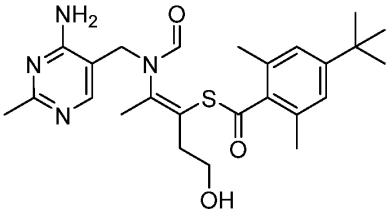
[0418] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 3,5-dichloro-[1,1'-biphenyl]-4-carbothioate



[0419] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.98 (s, 1H), 7.89 (s, 3H), 7.78 (dd, J_1 = 1.6, J_2 = 8.0 Hz, 2H), 7.58 - 7.45 (m, 3H), 6.92 (m, 2H), 4.74 (t, J = 5.6 Hz, 1H), 4.43 (br s, 2H), 3.55 - 3.48 (m, 2H), 2.67 (br d, J = 1.6 Hz, 2H), 2.22 (s, 3H), 2.15 (s, 3H). MS (ES$^+$) m/e 532 (M+H)$^+$.

[0420] **Example 44** (M67)

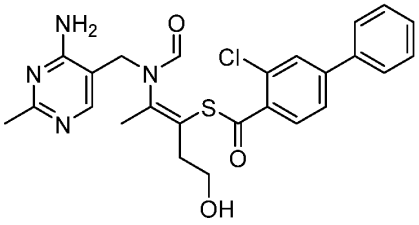
[0421] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-(tert-butyl)-2,6-dimethylbenzothioate



[0422] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.95 (s, 1H), 7.83 (s, 1H), 7.13-7.08(m, 2H), 6.75 (s, 2H),4.75-4.72 (m, 1H), 4.40 (s, 2H), 3.49-3.44 (m, 2H), 2.68 (m, 1H), 2.26 (s, 3H), 2.20 (s, 6H), 2.15 (s, 3H), 1.24 (s, 9H). MS (ES$^+$) m/e 471 (M+H)$^+$.

[0423] **Example 45** (M68)

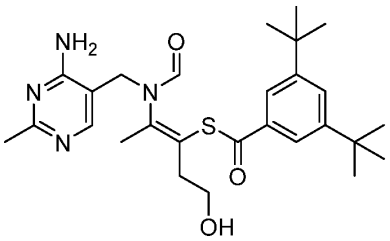
[0424] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 3-chloro-[1,1'-biphenyl]-4-carbothioate



[0425] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.92 (s, 1H), 7.90 - 7.84 (m, 2H), 7.80 - 7.73 (m, 3H), 7.65 (d, J = 8.0 Hz, 1H), 7.57 - 7.45 (m, 3H), 6.96 - 6.56 (m, 2H), 4.70 (t, J = 5.6 Hz, 1H), 4.43 (br s, 2H), 3.53 - 3.46 (m, 2H), 2.61 (br t, J = 6.8 Hz, 2H), 2.21 (s, 3H), 2.16 (s, 3H). MS (ES$^+$) m/e 498 (M+H)$^+$.

[0426] **Example 46** (M70)

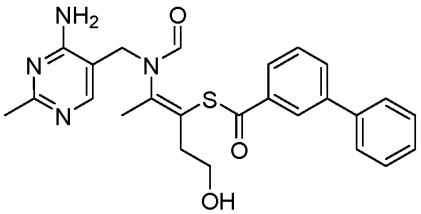
[0427] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 3,5-di-tert-butylbenzothioate



[0428] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.88 (s, 1H), 7.83 (s, 1H), 7.74 (t, *J* = 2.0 Hz, 1H), 7.55 (d, *J* = 2.0 Hz, 1H), 6.60 (s, 2H), 4.65 (t, *J* = 5.6 Hz, 1H), 4.39 (s, 2H), 3.57 - 3.42 (m, 2H), 2.57 (t, *J* = 6.8 Hz, 2H), 2.20 (s, 3H), 2.15 (s, 3H), 1.32 (s, 18H). MS (ES$^+$) m/e 499 (M+H)$^+$.

[0429] **Example 47** (M72)

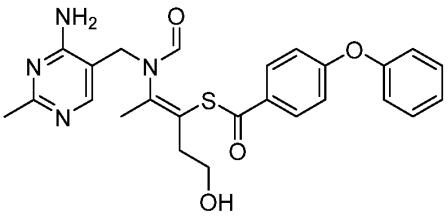
[0430] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) [1,1'-biphenyl]-3-carbothioate



[0431] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.98 (m, 1H), 7.93 (s, 1H), 7.91-7.90 (m, 1H), 7.84 (s, 1H), 7.73-7.71 (m, 3H), 7.63 (m, 1H), 7.55-7.51 (m, 2H), 7.44 (m, 1H), 6.64 (s, 2H), 4.64-4.65 (m, 2H), 4.40 (s, 2H), 3.52-3.48 (m, 2H), 2.61-2.51 (m, 2H), 2.21 (s, 3H), 2.13 (s, 3H). MS (ES$^+$) m/e 463 (M+H)$^+$.

[0432] **Example 48** (M73)

[0433] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-phenoxybenzothioate

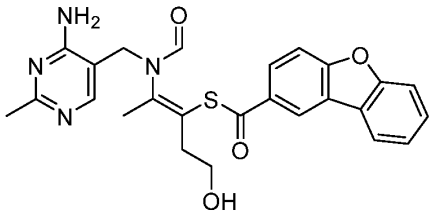


[0434] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.87 (s, 1H), 7.82 (s, 1H), 7.77 (s, 1H), 7.75 (s, 1H), 7.50-7.46 (m, 2H), 7.27 (m, 1H), 7.15 (d, *J* = 8 Hz, 2H), 7.04 (d, *J* = 8 Hz, 2H), 6.64 (s,

2H), 4.64-4.61 (m, 2H), 4.38 (s, 2H), 3.48-3.43 (m, 2H), 2.58-2.52 (m, 2H), 2.18 (s, 3H), 2.13 (s, 3H). MS (ES^+) m/e 463 $(M+H)^+$.

[0435] **Example 49** (M76)

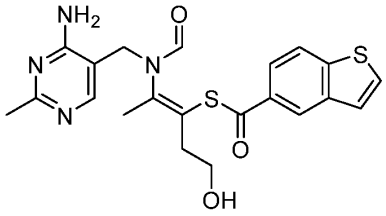
[0436] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) dibenzo[b,d]furan-2-carbothioate



[0437] ^1H NMR (400 MHz, DMSO-d6) δ 8.57 (s, 1H), 8.35-8.34 (d, J = 4Hz, 1H), 7.96 - 7.90 (m, 5H), 7.79-7.77 (m, 1H), 7.61-7.48 (m, 1H), 6.65 (m, 2H), 4.70 (m, 1H), 4.41 (s, 1H), 3.54 (m, 2H), 2.63 (m, 2H), 2.17 (s, 3H), 2.16 (s, 6H). MS (ES^+) m/e 477 $(M+H)^+$.

[0438] **Example 50** (M79)

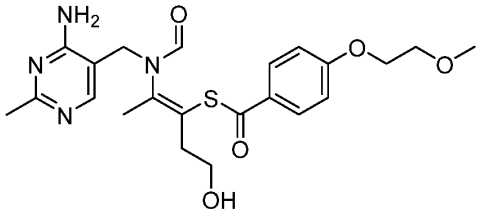
[0439] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) benzo[b]thiophene-5-carbothioate



[0440] ^1H NMR (400 MHz, DMSO-d6) δ 8.31 (s,1 H), 8.15 (d, J = 8.4 Hz, 1H), 7.94-7.93 (m, 2H), 7.84 (s, 1H), 7.67-7.65 (m, 2H), 6.63 (s, 2H), 4.67-4.64 (m, 2H), 4.40 (s, 2H), 3.51-3.47 (m, 2H), 2.67-2.58 (m, 2H), 2.16 (s, 3H), 2.14 (s, 3H).. MS (ES^+) m/e 443 $(M+H)^+$.

[0441] **Example 51** (M82)

[0442] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-(2-methoxyethoxy)benzothioate

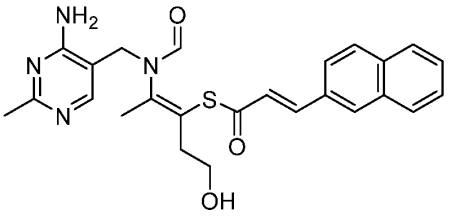


[0443] ^1H NMR (400 MHz, DMSO-d6) δ 7.88 (s, 1H), 7.83 (s, 1H), 7.73-7.69 (m, 2H), 7.08-7.04 (m, 2H), 6.64 (s, 2H), 4.66-4.63 (m, 2H), 4.38 (s, 2H), 4.21-4.18 (m, 2H), 3.69-

3.68 (m, 2H), 3.47 (m, 2H), 3.34 (s, 3H), 2.56 (m, 2H), 2.25 (s, 3H), 2.20 (s, 3H). MS (ES^+) m/e 461 $(M+H)^+$.

[0444] **Example 52** (M88)

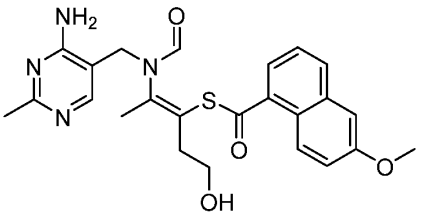
[0445] S-((Z)-2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) (E)-3-(naphthalen-2-yl)prop-2-enethioate



[0446] ^1H NMR (400 MHz, DMSO-d6) δ 8.26 (s, 1H), 8.00-7.94 (m, 4H), 7.90 (s, 1H), 7.84 (s, 1H), 7.59-7.56 (m, 3H), 6.95 (d, J = 15.6 Hz, 1H), 6.68 (s, 2H), 4.69 (s, 1H), 4.39 (s, 2H), 3.48 (t, J = 7.2 Hz, 2H), 2.57 (t, J = 6.4 Hz, 2H), 2.22 (s, 3H), 2.12 (s, 3H). MS (ES^+) m/e 463 $(M+H)^+$.

[0447] **Example 53** (M108)

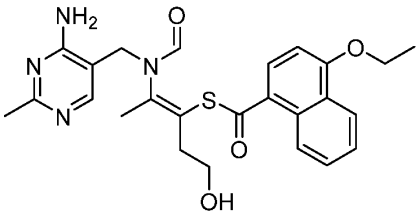
[0448] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 6-methoxynaphthalene-1-carbothioate



[0449] ^1H NMR (400 MHz, DMSO-d6) δ 8.13 (d, J = 9.6 Hz, 1H), 8.09 (d, J = 8.4 Hz, 1H), 7.97 (s, 1H), 7.87 (s, 1H), 7.74 (d, J = 6.8 Hz, 1H), 7.55 (t, J = 6.8 Hz, 1H), 7.45 (d, J = 2.4 Hz, 1H), 7.29 (dd, J_1 = 2.8 Hz, J_2 = 9.2 Hz, 1H), 6.73 (s, 2H), 4.71 (s, 1H), 4.43 (s, 2H), 3.89 (s, 3H), 3.52 (s, 2H), 2.66 (t, J = 6.8 Hz, 2H), 2.18 (s, 3H), 2.13 (s, 3H). MS (ES^+) m/e 467 $(M+H)^+$.

[0450] **Example 54** (M109)

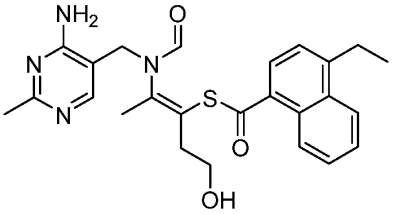
[0451] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-ethoxynaphthalene-1-carbothioate



[0452] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.42 (d, *J* = 8.4 Hz, 1H), 8.26 (d, *J* = 8.0 Hz, 1H), 8.03 (d, *J* = 8.4 Hz, 1H), 7.96 (s, 1H), 7.87 (s, 1H), 7.67 (t, *J* = 1.2 Hz, 1H), 7.65-7.60 (m, 1H), 7.05 (d, *J* = 8.4 Hz, 1H), 6.70 (s, 2H), 4.68 (t, *J* = 9.2 Hz, 1H), 4.42 (s, 1H), 4.32 (q, *J* = 7.2 Hz, 2H), 3.51 (q, *J* = 6.4 Hz, 2H), 2.64 (t, *J* = 6.8 Hz, 2H), 2.17 (s, 3H), 2.13 (s, 3H), 1.50 (t, *J* = 6.8 Hz, 3H). MS (ES$^+$) m/e 481 (M+H)$^+$.

[0453] **Example 55** (M110)

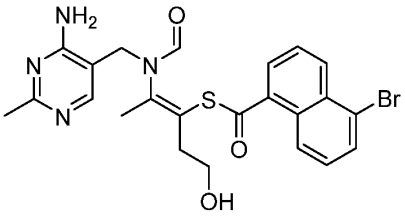
[0454] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-ethylnaphthalene-1-carbothioate



[0455] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.28-8.26 (m, 1H), 8.19-8.18 (m,1H), 7.97 (s, 1H), 7.87-7.85 (m, 1H), 7.67-7.64 (m, 1H), 7.48 (d, *J* = 7.6 Hz, 1H), 6.74 (s, 2H), 4.70 (t, *J* = 5.6 Hz, 1H), 4.43 (s, 1H), 3.52 (q, *J* = 6.0 Hz, 2H), 3.14 (q, *J* = 7.2 Hz, 2H), 2.66 (t, *J* = 6.8 Hz, 2H), 2.18 (s, 3H), 2.14 (s, 3H), 1.31 (t, *J* = 7.6 Hz, 3H). MS (ES$^+$) m/e 465 (M+H)$^+$.

[0456] **Example 56** (M111)

[0457] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 5-bromonaphthalene-1-carbothioate

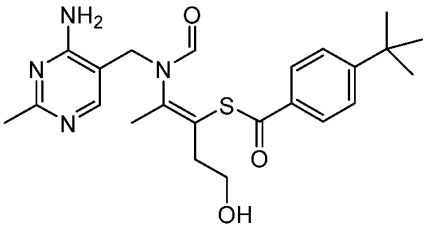


[0458] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.44 (d, *J* = 8.8 Hz, 1H), 8.20 (d, *J* = 8.8 Hz, 1H), 8.06 - 7.95 (m, 3H), 7.88 (s, 1H), 7.79 (dd, J_1 = 7.2, J_2 = 8.4 Hz, 1H), 7.58 (dd, J_1 = 7.6, J_2 = 8.4 Hz, 1H), 6.74 (br s, 2H), 4.70 (t, *J* = 5.6 Hz, 1H), 4.45 (br s, 2H), 3.56 - 3.49 (m,

2H), 2.68 (br t, J = 6.8 Hz, 2H), 2.21 - 2.16 (m, 3H), 2.11 (s, 3H). MS (ES^+) m/e 516 $(M+H)^+$.

[0459] **Example 57** (M114)

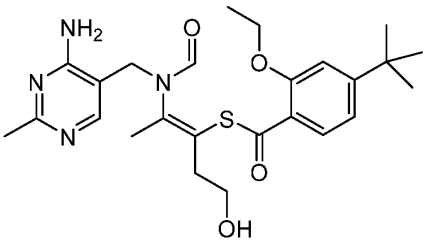
[0460] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-(tert-butyl)benzothioate



[0461] ^1H NMR (400 MHz, DMSO-d6) δ 7.90 - 7.82 (m, 2H), 7.68 (d, J = 8.4 Hz, 2H), 7.59 - 7.52 (m, 2H), 6.84 - 6.54 (m, 2H), 4.67 (br s, 1H), 4.38 (br s, 2H), 4.03 (q, J = 7.2 Hz, 1H), 3.45 (br s, 2H), 2.56 (br t, J = 6.8 Hz, 2H), 2.27 - 2.07 (m, 6H), 1.30 (s, 9H). MS (ES^+) m/e 443 $(M+H)^+$.

[0462] **Example 58** (M115)

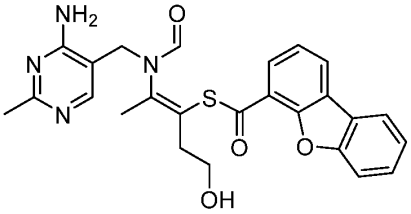
[0463] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-(tert-butyl)-2-ethoxybenzothioate



[0464] ^1H NMR (400 MHz, DMSO-d6) δ 7.83-7.82 (m, 2H), 7.52 (d, J = 8.0 Hz, 1H), 7.06-7.03 (m, 2H), 6.68 (s, 2H), 4.62 (t, J = 5.6 Hz, 1H), 4.36 (s, 1H), 4.17 (q, J = 7.2 Hz, 2H), 3.44 (q, J = 5.6 Hz, 2H), 2.54-2.52 (m, 2H), 2.20 (s, 3H), 2.10 (s, 3H), 1.32 (t, J = 6.8 Hz, 3H), 1.28 (s, 9H).. MS (ES^+) m/e 487 $(M+H)^+$.

[0465] **Example 59** (M148)

[0466] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) dibenzo[b,d]furan-4-carbothioate

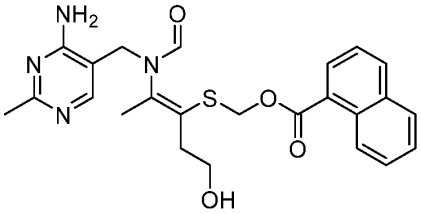


[0467] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.47 (d, *J* = 6.80 Hz, 1H), 8.22 (d, *J* = 7.60 Hz, 1H), 7.95 (s, 1H), 7.90 - 7.86 (m, 3H), 7.62 (t, *J* = 8.00 Hz, 1H), 7.58 - 7.47 (m, 2H), 6.64 (s, 2H), 4.67 (t, *J* = 5.60 Hz, 1H), 4.42 (s, 2H), 3.54 - 3.48 (m, 2H), 2.65 (t, *J* = 6.80 Hz, 2H), 2.18 (s, 3H), 2.13 (s, 3H). MS (ES$^+$) m/e 477 (M+H)$^+$.

[0468] **Examples 60 to 68**: The following compounds were synthesized following the procedure for the preparation of M19 (example 6) with appropriate starting material.

[0469] **Example 60** (M97)

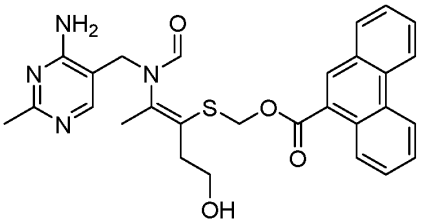
[0470] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl 1-naphthoate



[0471] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.75 (d, 1H), 8.29 - 8.22 (m, 1H), 8.14 - 8.12 (m, 1H), 8.12 - 8.05 (m, 1H), 7.94 (s, 1H), 7.81 (s, 1H), 7.69 - 7.63 (m, 3H), 6.74 – 6.71 (m, 2H), 5.37 (s, 2H), 4.73 (t, *J* = 5.6 Hz, 1H), 4.34 (br s, 2H), 3.57 - 3.52 (m, 2H), 2.73 - 2.69 (m, 2H), 2.25 (s, 3H), 1.95 (s, 3H). MS (ES$^+$) m/e 467 (M+H)$^+$.

[0472] **Example 61** (M98)

[0473] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl phenanthrene-9-carboxylate

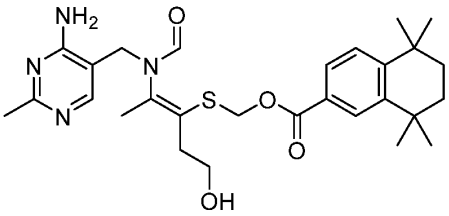


[0474] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.97 - 8.87 (m, 2H), 8.77 - 8.70 (m, 1H), 8.52 (s, 1H), 8.19 (d, *J* = 8.0 Hz, 1H), 7.98 (s, 1H), 7.90 - 7.81 (m, 2H), 7.80 - 7.73 (m, 3H), 6.74

(br s, 2H), 5.41 (s, 2H), 4.78 (t, J = 5.6 Hz, 1H), 4.35 (br s, 2H), 3.63 - 3.50 (m, 2H), 2.74 (br t, J = 7.2 Hz, 2H), 2.24 (s, 3H), 1.98 (s, 3H). MS (ES^+) m/e 517 $(M+H)^+$.

[0475] **Example 62** (M99)

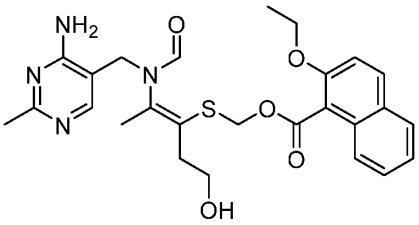
[0476] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl 5,5,8,8-tetramethyl-5,6,7,8-tetrahydronaphthalene-2-carboxylate



[0477] ^1H NMR (400 MHz, DMSO-d6) δ 7.93 (s, 1H), 7.89 - 7.85 (m, 1H), 7.77 (s, 1H), 7.66 (dd, J_1 = 2.0, 1H), 7.51 - 7.46 (m, 1H), 6.73 (m, 2H), 5.24 (s, 2H), 4.69 (t, J = 5.6 Hz, 1H), 4.44 - 4.18 (m, 2H), 3.57 - 3.44 (m, 2H), 2.67 (br t, J = 6.8 Hz, 2H), 2.29 - 2.24 (m, 3H), 1.94 (s, 3H), 1.66 (s, 4H), 1.25 (s, 12H). MS (ES^+) m/e 527 $(M+H)^+$.

[0478] **Example 63** (M100)

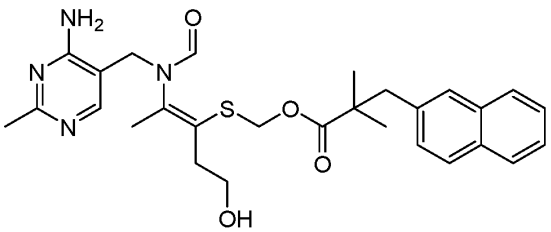
[0479] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl 2-ethoxy-1-naphthoate



[0480] ^1H NMR (400 MHz, DMSO-d6) δ 8.07 (d, J = 8.8 Hz, 1H), 7.95 - 7.90 (m, 2H), 7.80 (s, 1H), 7.62 - 7.49 (m, 3H), 7.43 (m, 1H), 6.73 (br s, 2H), 5.33 (s, 2H), 4.65 (t, J = 5.6 Hz, 1H), 4.35 (br s, 2H), 4.27 - 4.21 (m, 2H), 3.55 - 3.44 (m, 2H), 2.61 (br t, J = 6.8 Hz, 2H), 2.28 (s, 3H), 1.95 (s, 3H), 1.31 (t, J = 7.2 Hz, 3H). MS (ES^+) m/e 511 $(M+H)^+$.

[0481] **Example 64** (M101)

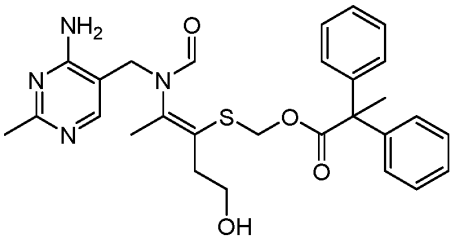
[0482] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl 2,2-dimethyl-3-(naphthalen-2-yl)propanoate



[0483] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.90 (s, 1H), 7.88 - 7.78 (m, 4H), 7.64 - 7.58 (m, 1H), 7.51 - 7.44 (m, 2H), 7.28 - 7.19 (m, 1H), 6.78 (m, 2H), 4.95 (br s, 2H), 4.66 (t, *J* = 5.6 Hz, 1H), 4.33 (br s, 2H), 3.50 - 3.41 (m, 2H), 3.01 - 2.92 (m, 2H), 2.48 (br s, 2H), 2.28 (m, 3H), 1.91 (s, 3H), 1.16 - 1.12 (m, 6H). MS (ES$^+$) m/e 523 (M+H)$^+$.

[0484] **Example 65** (M103)

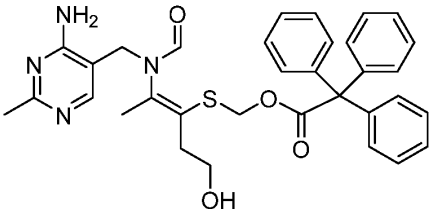
[0485] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl 2,2-diphenylpropanoate



[0486] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.78 (s, 1H), 7.64 (s, 1H), 7.28 - 7.34 (m, 6H), 7.12 - 7.36 (m, 4H), 6.72 - 6.70 (m, 2H), 5.04 (br s, 2H), 4.60 (t, *J* = 5.6 Hz, 1H), 4.27 (br s, 2H), 3.28 - 3.32 (m, 2H), 2.38 - 2.35 (m, 2H), 2.27 (s, 3H), 1.88 - 1.85 (m, 6H). MS (ES$^+$) m/e 521 (M+H)$^+$.

[0487] **Example 66** (M105)

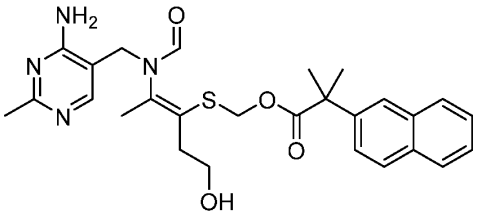
[0488] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl 2,2,2-triphenylacetate



[0489] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.89 (s, 1H), 7.80 (s, 1H), 7.34 - 7.28 (m, 11H), 7.11 - 7.09 (m, 7H), 6.77 - 6.72 (m, 2H), 5.39 (br s, 2H), 4.56 (t, *J* = 5.6 Hz, 1H), 4.33 (br s, 2H), 3.20 - 3.19 (m, 2H), 2.28 (s, 3H), 2.08 - 1.99 (m, 2H), 1.89 (s, 3H). MS (ES$^+$) m/e 583 (M+H)$^+$.

[0490] **Example 67** (M107)

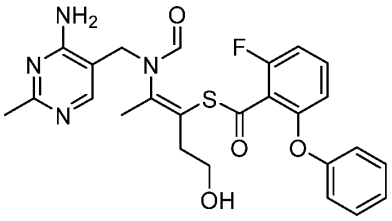
[0491] (Z)-((2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl)thio)methyl 2-methyl-2-(naphthalen-2-yl)propanoate



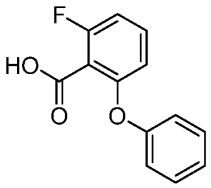
[0492] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.92 - 7.82 (m, 4H), 7.76 (s, 1H), 7.70 - 7.64 (m, 1H), 7.53 - 7.46 (m, 2H), 7.44 - 7.37 (m, 1H), 6.69 (br d, J = 12.4 Hz, 2H), 4.96 (s, 2H), 4.59 (t, J = 5.6 Hz, 1H), 4.35 - 4.12 (m, 2H), 3.32 - 3.25 (m, 2H), 2.34 (br t, J = 6.8 Hz, 2H), 2.29 - 2.20 (m, 3H), 1.77 (s, 3H), 1.64 - 1.57 (m, 6H). MS (ES$^+$) m/e 509 $(M+H)^+$.

[0493] **Example 68** (M126)

[0494] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-fluoro-6-phenoxybenzothioate



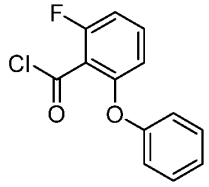
[0495] Step 1. 2-Fluoro-6-phenoxybenzoic acid



[0496] To a solution of 2-bromo-6-fluorobenzoic acid (5.00 g, 22.8 mmol, 1.00 eq) and phenol (3.87 g, 41.1 mmol, 3.61 mL, 1.80 eq) in DMF (100 mL) was added CuI (434 mg, 2.28 mmol, 0.10 eq) and Cs_2CO_3 (22.3 g, 68.5 mmol, 3.00 eq). The mixture was stirred at 100 °C for 10 hrs under N_2, cooled to rt, diluted with water (300 mL) and washed with ethyl acetate (200 mL x 2). The aqueous phase was adjusted pH ~ 1with HCl (2 M) and extracted with ethyl acetate (300 mL x 2). The combined organic phase was washed with brine (100 mL x 2), dried over Na_2SO_4, filtered and concentrated under vacuum to give crude product which was purified by reversed-phase HPLC (5% ~ 55% acetonitrile in water, 0.1% HCl) and concentrated under vacuum to remove acetonitrile. The residue was extracted with ethyl acetate (300 mL x 2). The combined organic phase was washed with brine (100 mL x 2), dried over Na_2SO_4, filtered and concentrated under vacuum to give the title compound (1.10

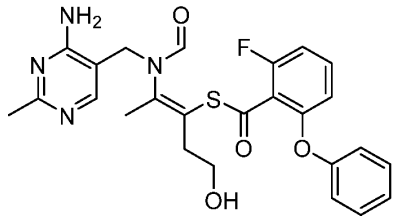
g, 4.74 mmol, 21% yield) as yellow solid. ^{1}H NMR (400 MHz, DMSO-d6) δ 7.47 - 7.36 (m, 3H), 7.20 - 7.16 (m, 1H), 7.13 - 7.06 (m, 1H), 7.03 (dd, J_1 = 0.80 Hz, J_2 = 8.80 Hz, 2H), 6.74 (d, J = 8.40 Hz, 1H). MS (ES$^+$) m/e 233 (M+H)$^+$.

[0497] Step 2. 2-Fluoro-6-phenoxybenzoyl chloride



[0498] To a solution of 2-fluoro-6-phenoxybenzoic acid (500 mg, 2.15 mmol, 1.00 eq) in DCM (10.0 mL) and DMF (1.57 mg, 21.5 umol, 1.66 uL, 0.01 eq) was added $(COCl)_2$ (328 mg, 2.58 mmol, 226 uL, 1.20 eq) drop-wise at 25 °C. The reaction mixture was stirred at 25 °C for 0.5 hr. The reaction was quenched with anhydrous MeOH and concentrated under reduced pressure to give the title compound (540 mg, crude) as yellow oil was directly used for next step without further purification.

[0499] Step 3. (Z)-S-(2-(N-((4-Amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-fluoro-6-phenoxybenzothioate

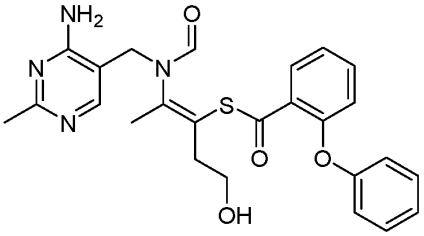


[0500] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.93 (s, 1H), 7.83 (s, 1H), 7.58 - 7.48 (m, 1H), 7.46 - 7. 39 (m, 1H), 7.24 - 7.17 (m, 1H), 7.13 - 7.02 (m, 1H), 6.78 (d, J = 8.40 Hz, 1H), 6.42 (s, 2H), 4.47 - 4.25 (m, 1H), 3.47 (br d, J = 5.20 Hz, 1H), 2.58 (br t, J = 6.80 Hz, 1H), 2.29 (s, 3H), 2.05 (s, 1H). MS (ES$^+$) m/e 497 (M+H)$^+$.

[0501] **Examples 69 to 99**: The following compounds were synthesized using essentially the same procedure for the preparation of M126 and appropriate starting material.

[0502] **Example 69** (M113)

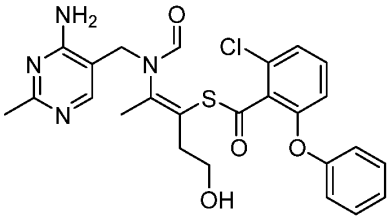
[0503] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-phenoxybenzothioate



[0504] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.83 (s, 1H), 7.79 (s, 1H), 7.66 - 7.56 (m, 2H), 7.43 - 7.36 (m, 2H), 7.27 (dt, J_1 = 0.8, J_2 = 7.6 Hz, 1H), 7.19 - 7.12 (m, 1H), 7.01 - 6.95 (m, 3H), 6.78 - 6.59 (m, 2H), 4.57 (t, J = 5.6 Hz, 1H), 4.34 (br s, 2H), 3.36 (br d, J = 6.0 Hz, 2H), 2.43 (t, J = 6.4 Hz, 2H), 2.16 (s, 3H), 2.08 (s, 3H). MS (ES$^+$) m/e 479 (M+H)$^+$.

[0505] **Example 70** (M127)

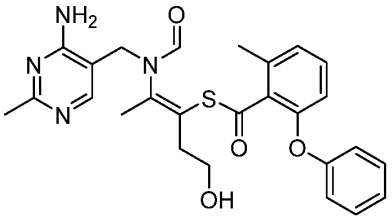
[0506] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-chloro-6-phenoxybenzothioate



[0507] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.91 (s, 1H), 7.81 (s, 1H), 7.53 - 7.45 (m, 1H), 7.44 - 7.37 (m, 2H), 7.33 (br d, J = 8.00 Hz, 1H), 7.19 (br t, J = 8.00 Hz, 1H), 7.08 - 6.97 (m, 2H), 6.90 (br d, J = 8.40 Hz, 1H), 6.83 - 6.58 (m, 2H), 4.62 (br t, J = 5.20 Hz, 1H), 4.35 (br s, 2H), 3.41 - 3.37 (m, 2H), 2.59 - 2.54 (m, 2H), 2.27 (s, 3H), 2.07 (s, 3H). MS (ES$^+$) m/e 514 (M+H)$^+$.

[0508] **Example 71** (M128)

[0509] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-methyl-6-phenoxybenzothioate

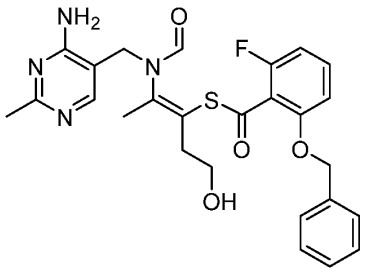


[0510] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.91 (s, 1H), 7.79 (s, 1H), 7.40 - 7.31 (m, 3H), 7.16 - 7.05 (m, 2H), 6.98 - 6.92 (m, 2H), 6.80 - 6.60 (m, 3H), 4.57 (t, J = 5.20 Hz, 1H), 4.34

(br s, 2H), 3.37 - 3.33 (m, 2H), 2.53 - 2.51 (m, 2H), 2.26 (s, 3H), 2.24 (s, 3H), 2.06 (s, 3H). MS (ES^+) m/e 493 $(M+H)^+$.

[0511] **Example 72** (M129)

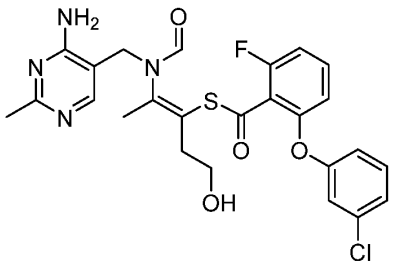
[0512] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(benzyloxy)-6-fluorobenzothioate



[0513] ^1H NMR (400 MHz, DMSO-d6) δ 7.91 (s, 1H), 7.80 (s, 1H), 7.50 - 7.25 (m, 7H), 7.04 (d, J = 8.40 Hz, 1H), 6.89 (t, J = 8.80 Hz, 1H), 6.77 - 6.52 (m, 2H), 5.21 (s, 2H), 4.63 (t, J = 5.60 Hz, 1H), 4.33 (s, 2H), 3.43 - 3.38 (m, 2H), 2.59 - 2.55 (m, 2H), 2.26 (s, 3H), 2.05 (s, 3H). MS (ES^+) m/e 511 $(M+H)^+$.

[0514] **Example 73** (M130)

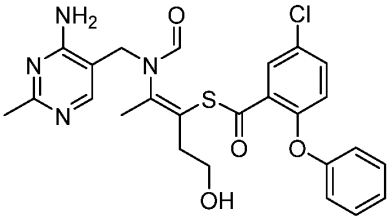
[0515] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(3-chlorophenoxy)-6-fluorobenzothioate



[0516] ^1H NMR (400 MHz, DMSO-d6) δ 7.86 (s, 1H), 7.80 (s, 1H), 7.63 - 7.55 (m, 1H), 7.46 - 7.40 (m, 1H), 7.28 - 7.18 (m, 2H), 7.13 (t, J = 2.00 Hz, 1H), 6.98 (dd, J_1 = 2.00 Hz, J_2 = 8.40 Hz, 1H), 6.92 (d, J = 8.40 Hz, 1H), 6.73 - 6.57 (m, 2H), 4.64 (t, J = 5.60 Hz, 1H), 4.34 (s, 2H), 3.40 (br s, 2H), 2.60 - 2.58 (m, 2H), 2.26 (s, 3H), 2.06 (s, 3H). MS (ES^+) m/e 532 $(M+H)^+$.

[0517] **Example 74** (M131)

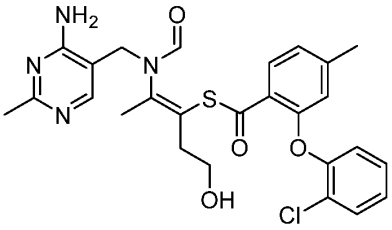
[0518] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 5-chloro-2-phenoxybenzothioate



[0519] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.85 (s, 1H), 7.80 (s, 1H), 7.61 (dd, J = 12.0 Hz, 1H), 7.58 (d, J = 4.00 Hz, 1H), 7.45 - 7.39 (m, 2H), 7.22 - 7.16 (m, 1H), 7.04 (d, J = 7.60 Hz, 2H), 6.97 (d, J = 8.80 Hz, 1H), 6.76 - 6.73 (m, 2H), 4.61 - 4.56 (m, 1H), 4.35 (s, 2H), 3.43 - 3.38 (m, 2H), 2.49 - 2.45 (m, 2H), 2.17 (s, 3H), 2.10 (s, 3H). MS (ES$^+$) m/e 514 (M+H)$^+$.

[0520] **Example 75** (M133)

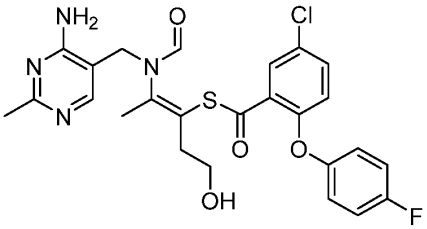
[0521] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(2-chlorophenoxy)-4-methylbenzothioate



[0522] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.82 (s, 1H), 7.78 (s, 1H), 7.60 - 7.54 (m, 2H), 7.35 (dt, J_1 = 1.60 Hz, J_2 = 7.8 Hz, 1H), 7.21 (dt, J = 1.60 Hz, J_2 = 8.00 Hz, 1H), 7.10 (d, J = 8.00 Hz, 1H), 6.99 (dd, J_1 = 1.60 Hz, J_2 = 8.00 Hz, 1H), 6.70 - 6.60 (m, 3H), 4.57 (t, J = 5.60 Hz, 1H), 4.33 (br s, 2H), 3.43 - 3.37 (m, 2H), 2.45 - 2.41 (m, 2H), 2.28 (s, 3H), 2.16 (s, 3H), 2.06 (s, 3H). MS (ES$^+$) m/e 528 (M+H)$^+$.

[0523] **Example 76** (M134)

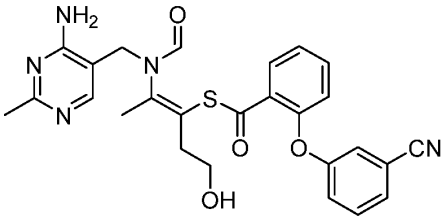
[0524] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 5-chloro-2-(4-fluorophenoxy)benzothioate



[0525] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.84 (s, 1H), 7.79 (s, 1H), 7.65 - 7.63 (m, 2H), 7.29 - 7.24 (m, 2H), 7.13 - 7.11 (m, 2H), 6.66 (d, J = 7.20 Hz, 1H), 6.64(br s, 2H), 5.60 (t, J = 5.60 Hz, 1H), 4.35 (s, 2H), 3.42 - 3.39 (m, 2H), 2.56 - 2.55 (m, 2H), 2.15 (s, 3H), 2.11 (s, 3H). MS (ES$^+$) m/e 532 (M+H)$^+$.

[0526] **Example 77** (M135)

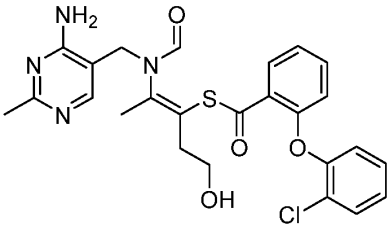
[0527] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(3-cyanophenoxy)benzothioate



[0528] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.81 (s, 1H), 7.78 (s, 1H), 7.72 - 7.62 (m, 2H), 7.61 - 7.54 (m, 2H), 7.46 - 7.35 (m, 2H), 7.28 - 7.20 (m, 1H), 7.14 (d, J = 7.60 Hz, 1H), 6.69 (br s, 2H), 4.58 (t, J = 5.60 Hz, 1H), 4.34 (br s, 2H), 3.45-3.38 (m, 2H), 2.42 - 2.35 (m, 2H), 2.15 (s, 3H), 2.08 (s, 3H). MS (ES$^+$) m/e 504 (M+H)$^+$.

[0529] **Example 78** (M136)

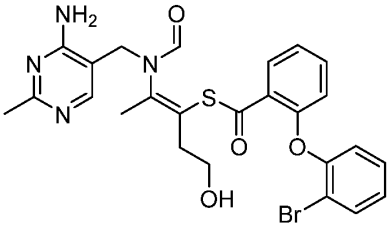
[0530] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(2-chlorophenoxy)benzothioate



[0531] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.85 (s, 1H), 7.80 (s, 1H), 7.64 (dd, J_1 = 1.60 Hz, J_2 = 7.60 Hz, 1H), 7.61 - 7.55 (m, 2H), 7.40 - 7.33 (m, 1H), 7.31 - 7.25 (m, 1H), 7.24 - 7.19 (m, 1H), 7.06 - 7.00 (m, 1H), 6.86 (d, J = 7.60 Hz, 1H), 6.67 (br s, 2H), 4.59 (t, J = 5.60 Hz, 1H), 4.35 (br s, 2H), 3.42 - 3.35 (m, 2H), 2.46 - 2.44 (m, 2H), 2.16 (s, 3H), 2.08 (s, 3H). MS (ES$^+$) m/e 514 (M+H)$^+$.

[0532] **Example 79** (M137)

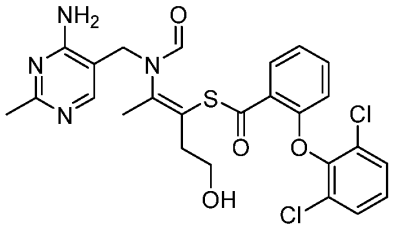
[0533] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(2-bromophenoxy)benzothioate



[0534] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.85 (s, 1H), 7.80 (s, 1H), 7.73 (dd, J_1 = 1.60 Hz, J_2 = 8.00 Hz, 1H), 7.64 (dd, J_1 = 1.60 Hz, J_2 = 8.00 Hz, 1H), 7.60 - 7.54 (m, 1H), 7.44 - 7.37 (m, 1H), 7.28 (t, J = 7.60 Hz, 1H), 7.18 - 7.12 (m, 1H), 7.00 (dd, J_1 = 1.60 Hz, , J_2 = 8.00 Hz, 1H), 6.84 (d, J = 8.40 Hz, 1H), 6.75 - 6.60 (m, 2H), 4.59 (t, J = 6.00 Hz, 1H), 4.35 (br s, 2H), 3.40 (q, J = 6.80 Hz, 2H), 2.47 - 2.43 (m, 2H), 2.16 (s, 3H), 2.08 (s, 3H). MS (ES$^+$) m/e 558 (M+H)$^+$.

[0535] **Example 80** (M138)

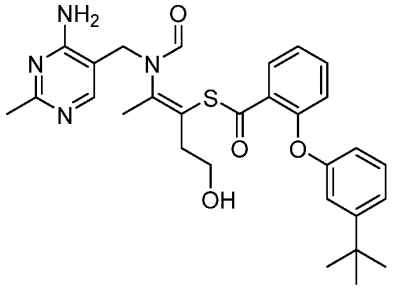
[0536] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(2,6-dichlorophenoxy)benzothioate



[0537] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.30 - 8.25 (m, 1H), 8.13 (dd, J_1 = 3.20 Hz, J_2 = 6.4 Hz, 1H), 7.98 (s, 1H), 7.88 - 7.80 (m, 2H), 7.69 - 7.64 (m, 2H), 7.47 (d, J = 7.60 Hz, 1H), 6.71 (br s, 2H), 4.69 (br t, J = 5.60 Hz, 1H), 4.43 (br s, 2H), 3.55 - 3.47 (m, 2H), 2.72 (s, 3H), 2.66 (br t, J = 6.80 Hz, 2H), 2.18 (s, 3H), 2.13 (s, 3H). MS (ES$^+$) m/e 457 (M+H)$^+$.

[0538] **Example 81** (M139)

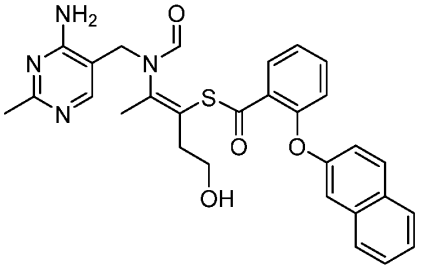
[0539] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(3-(tert-butyl)phenoxy)benzothioate



[0540] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.86 (s, 1H), 7.80 (s, 1H), 7.66 - 7.54 (m, 2H), 7.34 - 7.23 (m, 2H), 7.19 (br d, J = 8.00 Hz, 1H), 7.05 - 6.93 (m, 2H), 6.79 - 6.56 (m, 3H), 4.58 (t, J = 5.60 Hz, 1H), 4.34 (br s, 2H), 3.40 - 3.37 (m, 2H), 2.44 (br t, J = 6.80 Hz, 2H), 2.17 (s, 3H), 2.07 (s, 3H), 1.26 (s, 9H). MS (ES$^+$) m/e 535 (M+H)$^+$.

[0541] **Example 82** (M140)

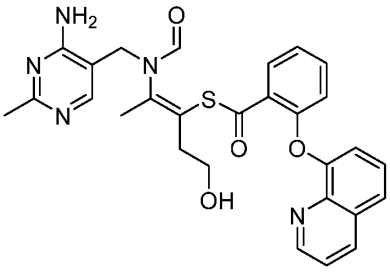
[0542] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(naphthalen-2-yloxy)benzothioate



[0543] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.97 (d, J = 8.80 Hz, 1H), 7.91 (d, J = 7.60 Hz, 1H), 7.86 - 7.81 (m, 2H), 7.77 (s, 1H), 7.72 - 7.65 (m, 1H), 7.65 - 7.58 (m, 1H), 7.54 - 7.41 (m, 2H), 7.37 - 7.29 (m, 2H), 7.26 (dd, J_1 = 2.40 Hz, J_2 = 8.80 Hz, 1H), 7.10 (d, J = 8.40 Hz, 1H), 6.70 (br s, 2H), 4.50 (t, J = 5.60 Hz, 1H), 4.32 (br s, 2H), 3.23 (q, J = 6.40 Hz, 2H), 2.35 (br t, J = 6.40 Hz, 2H), 2.14 (s, 3H), 2.05 (s, 3H). MS (ES$^+$) m/e 529 (M+H)$^+$.

[0544] **Example 83** (M141)

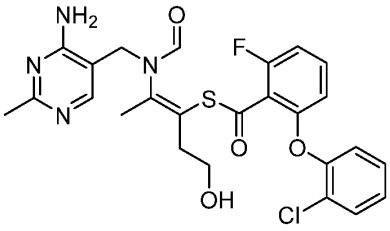
[0545] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(quinolin-8-yloxy)benzothioate



[0546] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.86 (s, 1H), 7.84 - 7.72 (m, 2H), 7.68 - 7.59 (m, 1H), 7.49 - 7.32 (m, 3H), 7.32 - 7.17 (m, 2H), 7.11 - 6.84 (m, 1H), 6.83 - 6.47 (m, 3H), 6.44 - 6.06 (m, 1H), 4.64 - 4.40 (m, 2H), 4.13 - 3.95 (m, 1H), 3.56 - 3.48 (m, 2H), 2.80 - 2.70 (m, 2H), 2.28 (s, 3H), 1.83 (br s, 3H). MS (ES$^+$) m/e 530 (M+H)$^+$.

[0547] **Example 84** (M142)

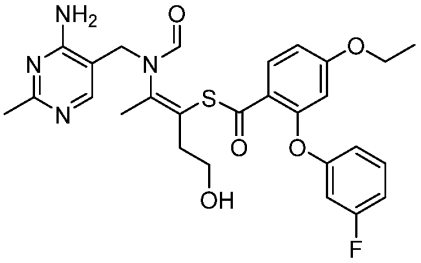
[0548] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(2-chlorophenoxy)-6-fluorobenzothioate



[0549] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.89 (s, 1H), 7.80 (s, 1H), 7.58 (dd, J_1 = 1.60 Hz, J_2 = 8.00 Hz, 1H), 7.51 - 7.45 (m, 1H), 7.41 - 7.36 (m, 1H), 7.29 - 7.23 (m, 1H), 7.18 - 7.08 (m, 2H), 6.87 - 6.61 (m, 2H), 6.55 (d, J = 8.40 Hz, 1H), 4.61 (t, J = 5.60 Hz, 1H), 4.33 (s, 2H), 3.42 - 3.38 (m, 2H), 2.57 - 2.55 (m, 2H), 2.27 (s, 3H), 2.06 (d, J = 1.60 Hz, 3H). MS (ES$^+$) m/e 532 (M+H)$^+$.

[0550] **Example 85** (M143)

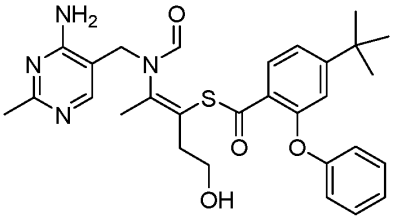
[0551] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-ethoxy-2-(3-fluorophenoxy)benzothioate



[0552] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.79 (s, 1H), 7.77 (s, 1H), 7.69 (d, J = 8.80 Hz, 1H), 7.44 - 7.36 (m, 1H), 6.98 (dt, J_1 = 2.40 Hz, J_2 = 8.40 Hz, 1H), 6.91 - 6.85 (m, 1H), 6.76 (dd, J_1 = 2.00 Hz, J_2 = 8.00 Hz, 1H), 6.65 (br d, J = 2.00 Hz, 2H), 6.54 (d, J = 2.40 Hz, 1H), 4.55 (t, J = 5.60 Hz, 1H), 4.31 (br s, 2H), 4.06 (q, J = 7.20 Hz, 2H), 3.37 - 3.33 (m, 2H), 2.40 (br t, J = 6.80 Hz, 1H), 2.17 (s, 3H), 2.06 (s, 3H), 1.30 (t, J = 6.80 Hz, 3H). MS (ES$^+$) m/e 541 (M+H)$^+$.

[0553] **Example 86** (M145)

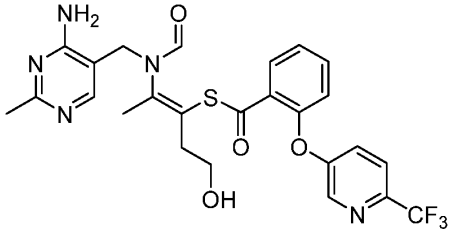
[0554] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-(tert-butyl)-2-phenoxybenzothioate



[0555] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.89 (s, 1H), 7.83 (s, 1H), 7.64 (br d, J = 8.40 Hz, 1H), 7.40 - 7.36 (m, 2H), 7.32 (br d, J = 7.20 Hz, 1H), 7.19 - 7.08 (m, 1H), 7.01 (s, 1H), 6.96 (d, J = 8.40 Hz, 2H), 6.30 (br s, 2H), 4.35 (s, 2H), 4.21 - 4.06 (m, 1H), 3.50 - 3.43 (m, 2H), 2.55 (br s, 2H), 2.25 (s, 3H), 2.06 (s, 3H), 1.26 (s, 9H). MS (ES$^+$) m/e 535 (M+H)$^+$.

[0556] **Example 87** (M146)

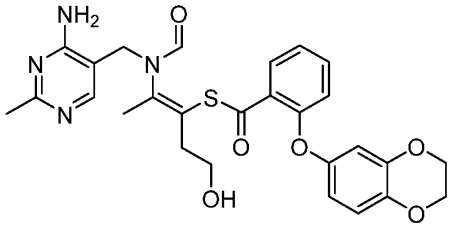
[0557] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-((6-(trifluoromethyl)pyridin-3-yl)oxy)benzothioate



[0558] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.50 (d, J = 2.40 Hz, 1H), 7.89 (d, J = 8.80 Hz, 1H), 7.85 - 7.63 (m, 4H), 7.50 - 7.40 (m, 2H), 7.39 - 7.27 (m, 1H), 6.71 (br s, 2H), 4.59 (t, J = 5.60 Hz, 1H), 4.35 (br s, 2H), 2.59 (br d, J = 6.80 Hz, 2H), 2.40 (br t, J = 6.40 Hz, 2H), 2.16 (s, 3H), 2.09 (s, 3H). MS (ES$^+$) m/e 548 (M+H)$^+$.

[0559] **Example 88** (M147)

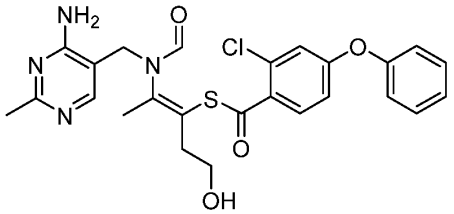
[0560] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-((2,3-dihydrobenzo[b][1,4]dioxin-6-yl)oxy)benzothioate



[0561] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.85 (s, 1H), 7.80 (s, 1H), 7.65 - 7.51 (m, 2H), 7.22 (t, J = 7.60 Hz, 1H), 6.93 (d, J = 8.40 Hz, 1H), 6.87 (d, J = 8.80 Hz, 1H), 6.67 (br s, 2H), 6.56 (d, J = 2.40 Hz, 1H), 6.48 (dd, J_1 = 2.80 Hz, J_2 = 8.80 Hz, 1H), 4.67 - 4.54 (m, 1H), 4.35 (br s, 2H), 4.25 - 4.23 (br d, J = 2.40 Hz, 4H), 3.41 (q, J = 6.40 Hz, 2H), 2.48 - 4.46 (m, 2H), 2.16 (s, 3H), 2.09 (s, 3H). MS (ES$^+$) m/e 537 (M+H)$^+$.

[0562] **Example 89** (M149)

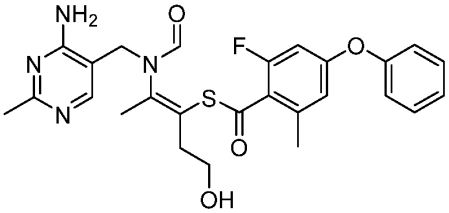
[0563] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-chloro-4-phenoxybenzothioate



[0564] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.88 (s, 1H), 7.84 (s, 1H), 7.61 (d, J = 8.80 Hz, 1H), 7.50 (t, J = 8.00 Hz, 2H), 7.33 - 7.28 (m, 1H), 7.22 - 7.15 (m, 3H), 6.98 (dd, J_1 = 2.40 Hz, J_2 = 8.80 Hz, 1H), 6.79 - 6.68 (m, 2H), 4.67 (t, J = 5.60 Hz, 1H), 4.45 - 4.41 (m, 2H), 3.49 - 3.45 (m, 2H), 2.60 - 2.58 (m, 2H), 2.20 (s, 3H), 2.15 (s, 3H). MS (ES$^+$) m/e 514 (M+H)$^+$.

[0565] **Example 90** (M150)

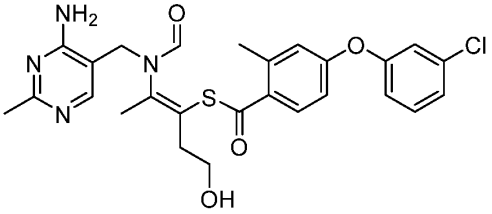
[0566] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-fluoro-6-methyl-4-phenoxybenzothioate



[0567] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.91 (s, 1H), 7.83 (s, 1H), 7.49 - 7.42 (m, 2H), 7.29 - 7.22 (m, 1H), 7.14 (d, J = 8.00 Hz, 2H), 6.79 - 6.57 (m, 4H), 4.71 (t, J = 5.20 Hz, 1H), 4.39 (br s, 2H), 3.49 - 3.43 (m, 2H), 2.60 (br s, 2H), 2.26 (m, 3H), 2.19 (s, 3H), 2.12 (s, 3H). MS (ES$^+$) m/e 511 (M+H)$^+$.

[0568] **Example 91** (M151)

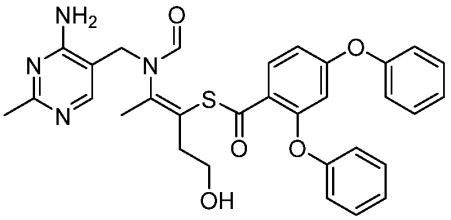
[0569] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-(3-chlorophenoxy)-2-methylbenzothioate



[0570] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.88 (s, 1H), 7.83 (s, 1H), 7.67 (d, J = 8.40 Hz, 1H), 7.47 (t, J = 8.00 Hz, 1H), 7.30 (dd, J_1 = 0.80 Hz, J_2 = 8.00 Hz, 1H), 7.23 (t, J = 2.00 Hz, 1H), 7.09 (dd, J_1 = 2.00 Hz, J_2 = 8.00 Hz, 1H), 7.00 (br d, J = 2.00 Hz, 1H), 6.90 (br dd, J = 2.40 Hz, J_2 = 8.80 Hz, 1H), 6.82 - 6.59 (m, 2H), 4.65 (br t, J = 5.60 Hz, 1H), 4.40 (br s, 2H), 3.46 (q, J = 6.40 Hz, 2H), 2.58 - 2.55 (m, 2H), 2.32 (s, 3H), 2.19 (s, 3H), 2.14 (s, 3H). MS (ES$^+$) m/e 528 (M+H)$^+$.

[0571] **Example 92** (M152)

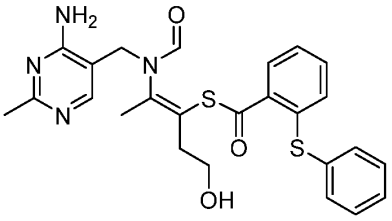
[0572] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2,4-diphenoxybenzothioate



[0573] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.13 (s, 1H), 7.93 (s, 1H), 7.71 (d, J = 8.80 Hz, 1H), 7.43 - 7.41 (m, 5H), 7.28 - 7.18 (m, 2H), 7.09 (m, 5H), 6.79 - 6.77 (m, 1H), 6.38 (d, J = 2.40 Hz, 1H), 4.70 (s, 1H), 4.48 (s, 2H), 3.42 (t, J = 6.00 Hz, 2H), 2.448 - 2.44 (m, 2H), 2.20 (s, 3H), 2.16 (s, 3H). MS (ES$^+$) m/e 571 (M+H)$^+$.

[0574] **Example 93** (M153)

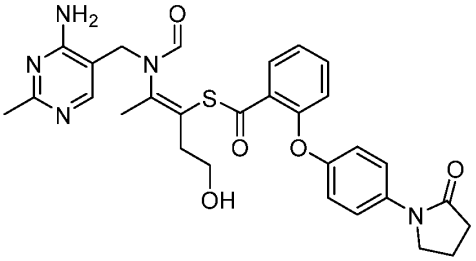
[0575] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(phenylthio)benzothioate



[0576] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.92 (s, 1H), 7.85 (s, 1H), 7.69 (dd, J_1 = 1.20 Hz, J_2 = 8.00 Hz, 1H), 7.53 - 7.40 (m, 7H), 7.34 - 7.28 (m, 1H), 6.90 (d, J = 8.00 Hz, 1H), 6.86 - 6.60 (m, 2H), 4.67 (t, J = 5.60 Hz, 1H), 4.44 - 4.33 (m, 2H), 3.49 - 3.44 (m, 2H), 2.62 - 2.55 (m, 2H), 2.20 (s, 3H), 2.14 (s, 3H). MS (ES$^+$) m/e 495 (M+H)$^+$.

[0577] **Example 94** (M155)

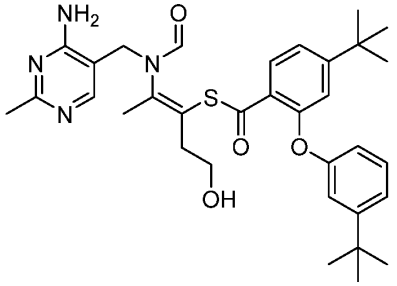
[0578] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(4-(2-oxopyrrolidin-1-yl)phenoxy)benzothioate



[0579] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.84 (s, 1H), 7.79 (s, 1H), 7.69 - 7.63 (m, 3H), 7.62 - 7.54 (m, 1H), 7.31 - 7.22 (m, 1H), 7.04 - 6.99 (m, 2H), 6.96 (d, J = 8.40 Hz, 1H), 6.69 (br s, 2H), 4.59 (t, J = 5.60 Hz, 1H), 4.34 (br s, 2H), 3.82 (t, J = 7.20 Hz, 2H), 3.42 - 3.37 (m, 2H), 2.49 - 2.44 (m, 4H), 2.17 (s, 3H), 2.11 - 2.02 (m, 5H).. MS (ES$^+$) m/e 562 (M+H)$^+$.

[0580] **Example 95** (M156)

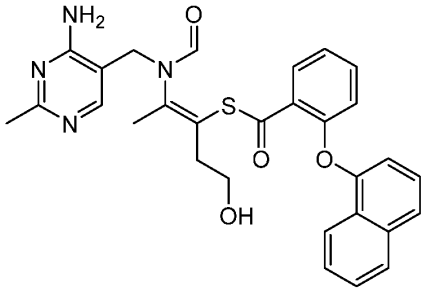
[0581] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 4-(tert-butyl)-2-(3-(tert-butyl)phenoxy)benzothioate



[0582] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.83 (s, 1H), 7.79 (s, 1H), 7.60 (d, J = 8.00 Hz, 1H), 7.39 - 7.30 (m, 2H), 7.16 (d, J = 7.20 Hz, 1H), 6.99 (s, 2H), 6.74 - 6.66 (m, 3H), 4.62 - 4.54 (m, 1H), 4.42 - 4.39 (s, 2H), 2.45 - 2.40 (m, 3H), 2.17 (s, 3H), 2.06 (s, 3H), 1.24 (s, 9H), 1.21 (s, 9H). MS (ES$^+$) m/e 591 (M+H)$^+$.

[0583] **Example 96** (M157)

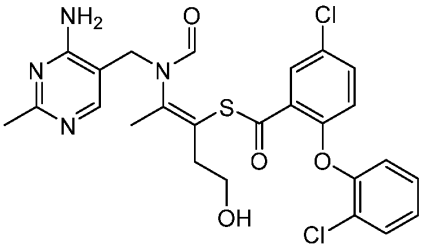
[0584] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(naphthalen-1-yloxy)benzothioate



[0585] ^{1}H NMR (400 MHz, DMSO-d6) δ 8.09 - 8.04 (m, 1H), 8.01 - 7.96 (m, 1H), 7.84 (s, 1H), 7.80 - 7.71 (m, 2H), 7.68 (dd, J_1 = 1.60 Hz, J_2 = 8.00 Hz, 1H), 7.62 - 7.52 (m, 3H), 7.47 (t, J = 8.00 Hz, 1H), 7.32 - 7.24 (m, 1H), 6.97 - 6.88 (m, 2H), 6.69 (br s, 2H), 4.52 (t, J = 5.60 Hz, 1H), 4.33 (br s, 2H), 3.29 - 3.26 (m, 2H), 2.39 (br t, J = 6.40 Hz, 2H), 2.15 (s, 3H), 2.05 (s, 3H). MS (ES$^+$) m/e 529 (M+H)$^+$.

[0586] **Example 97** (M158)

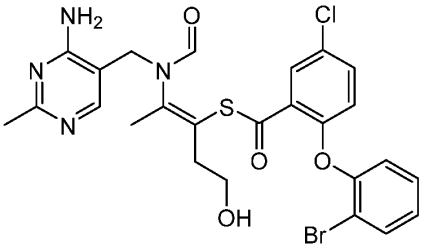
[0587] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 5-chloro-2-(2-chlorophenoxy)benzothioate



[0588] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.87 (s, 1H), 7.81 (s, 1H), 7.64 - 7.58 (m, 3H), 7.39 (t, *J* = 7.20 Hz, 1H), 7.27 (t, *J* = 7.20 Hz, 1H), 7.14 (t, *J* = 8.00 Hz, 1H), 6.86 (d, *J* = 8.40 Hz, 1H), 6.63 (s, 2H), 4.63 - 4.59 (m, 1H), 4.35 (s, 2H), 3.46 - 3.41 (m, 2H), 2.46 - 2.44 (m, 2H), 2.17 (s, 3H), 2.10 (s, 3H). MS (ES$^+$) m/e 548 (M+H)$^+$.

[0589] **Example 98** (M159)

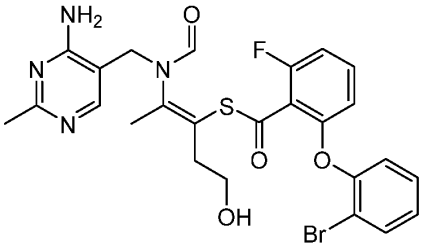
[0590] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(2-bromophenoxy)-5-chlorobenzothioate



[0591] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.87 (s, 1H), 7.81 (s, 1H), 7.78 - 7.74 (m, 1H), 7.65 - 7.58 (m, 2H), 7.48 - 7.42 (m, 1H), 7.25 - 7.18 (m, 1H), 7.16 - 7.13 (m, 1H), 6.86 (d, *J* = 8.40 Hz, 1H), 6.63 (s, 2H), 4.62 - 4.58 (m, 1H), 4.35 (s, 2H), 3.45 - 3.40 (m, 2H), 2.46 - 2.42 (s, 2H), 2.17 (s, 3H), 2.10 (s, 3H). MS (ES$^+$) m/e 592 (M+H)$^+$.

[0592] **Example 99** (M160)

[0593] (Z)-S-(2-(N-((4-amino-2-methylpyrimidin-5-yl)methyl)formamido)-5-hydroxypent-2-en-3-yl) 2-(2-bromophenoxy)-6-fluorobenzothioate



[0594] ^{1}H NMR (400 MHz, DMSO-d6) δ 7.89 (s, 1H), 7.81 (s, 1H), 7.75 (dd, J_1 = 1.20 Hz, J_2 = 7.60 Hz, 1H), 7.56 - 7.41 (m, 2H), 7.23 - 7.19 (m, 1H), 7.18 - 7.08 (m, 2H), 6.63 -

6.61 (m, 2H), 6.56 (d, J = 8.80 Hz, 1H), 4.64 - 4.61 (m, 1H), 4.35 (s, 2H), 3.47 - 3.36 (m, 2H), 2.06 (s, 3H). MS (ES^+) m/e 576 $(M+H)^+$.

[0595] Example 100: Results

[0596] To identify compounds that are more potent in turning on these thiamine analog riboswitches in mammalian cells, and/or have other characteristics such as improved pharmacokenic parameters, additional compounds (e.g., compounds of Formulas IV-VIII) were made and tested for their activity in regulating thiamine analog riboswitches using a luciferase construct harboring 15D10 riboswitch (Luci-15D10, SEQ ID NO:93).

[0597] SEQ ID NO: 93 was obtained by inserting the 15D10 riboswitch into the luciferase reporter gene. Capital letters indicate the luciferase encoding sequence. Lower case letters indicate the intron/alternative exon/intron and riboswitch sequence. The 15D10 aptamer encoding sequence (SEQ ID NO:26) is underlined.

[0598] SEQ ID NO:93:

[0599] ATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATCCG
CTGGAAGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCC
CTGGTTCCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGGACATCACTT
ACGCTGAGTACTTCGAAATGTCCGTTCGGTTGGCAGAAGCTATGAAACGATATGG
GCTGAATACAAATCACAGAATCGTCGTATGCAGTGAAAACTCTCTTCAATTCTTT
ATGCCGGTGTTGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCCCGCGAACGACA
TTTATAATGAACGTGAATTGCTCAACAGTATGGGCATTTCGCAGCCTACCGTGGT
GTTCGTTTCCAAAAAGGGGTTGCAAAAAATTTTGAACGTGCAAAAAAAGCTCCC
AATCATCCAAAAAATTATTATCATGGATTCTAAAACGGATTACCAGGGATTTCAG
TCGATGTACACGTTCGTCACATCTCATCTACCTCCCGGTTTTAATGAATACGATTT
TGTGCCAGAGTCCTTCGATAGGGACAAGACAATTGCACTGATCATGAACTCCTCT
GGATCTACTGGTCTGCCTAAAGGTGTCGCTCTGCCTCATAGAACTGCCTGCGTGA
GATTCTCGCATGCCAGAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGC
GATTTTAAGTGTTGTTCCATTCCATCACGGTTTTGGAATGTTTACTACACTCGGAT
ATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAAGAGCTGTT
TCTGAGGAGCCTTCAGGATTACAAGATTCAAAGTGCGCTGCTGGTGCCAACCCTA
TTCTCCTTCTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTTACA
CGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGAAGCGGTTGCC
AAGAGGTTCCATCTGCCAGGTATCAGGgtgagtctatgggacccttgatgttttctttccccttctttttctatggtt
aagttcatgtcataggaaggggagaagtaacagggtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttct

tcttttaatatacttttttgtttatcttatttctaatactttccctaatctctttctttcagggcaataatgatacaatgtatcatgccgagtaacgctg
tttctctaacttgtaggaatgaattcagatatttccagagaatgaaaaaaaatcttcagtagaaggtaatgtacaggggtccggccttttcat
ttggcaccggtgagaacatacccttcggacctgttcacggataatgccgctgcagggagtacattacgcaccattctaaagaataacag
tgataatttctgggttaaggcaatagcaatatttctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttcatattgctaata
gcagctacaatccagctaccattctgcttttattttatggttgggataaggctggattattctgagtccaagctaggccctttgctaatcatg
ttcatacctcttatcttcctcccacagCAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTG
ATTACACCCGAGGGGGATGATAAACCGGGCGCGGTCGGTAAAGTTGTTCCATTTT
TTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAAA
GAGGCGAACTGTGTGTGAGAGGTCCTATGATTATGTCCGGTTATGTAAACAATCC
GGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTCTGGAGACAT
AGCTTACTGGGACGAAGACGAACACTTCTTCATCGTTGACCGCCTGAAGTCTCTG
ATTAAGTACAAAGGCTATCAGGTGGCTCCCGCTGAATTGGAATCCATCTTGCTCC
AACACCCCAACATCTTCGACGCAGGTGTCGCAGGTCTTCCCGACGATGACGCCGG
TGAACTTCCCGCCGCCGTTGTTGTTTTGGAGCACGGAAAGACGATGACGGAAAA
AGAGATCGTGGATTACGTCGCCAGTCAAGTAACAACCGCGAAAAAGTTGCGCGG
AGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACTCGACGC
AAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGATCGCCG
TGTAA

[0600] Compounds of Formula IV were first tested in HEK 293 cells. As shown in **Figure 9A**, luciferase gene expression was increased upon treatment with compound M10, M16, M18, M19 or M21, in a dose-dependent manner in HEK 293 cells. In comparison with the activity of fursultiamine, M19 is more potent in inducing luciferase expression, generating 168.7-fold increase in luciferase expression at 12.5 μM concentration in HEK 293 cells. Further, these new compounds were tested in mouse liver cell line AML 12 cells. As shown in **Figure 9B**, consistent with the observation from HEK 293 cells, all the compounds increased the luciferase expression from construct harboring 15D10 riboswitch in a dose dependent manner. In comparison with the activity of fursultiamine on the riboswitch 15D10, compounds M10, M16, M19 and M21 are more potent than fursultiamine. M19 and M21 are equally potent in inducing riboswitch 15D10 regulated luciferase expression in this type of cell.

[0601] Compounds of Formulas V and VI together with benfotiamine as control compound were tested in both HEK 293 and AML 12 cells. As shown in **Figure 9C**, compound M30, M31 M32 and M33, as well as benfotiamine had no effect on luciferase expression, while M26, M27, M28, M29 and M34 increased luciferase expression from the

Luci-15D10 construct, with M34 being the most potent compound in HEK 293 cells. In contrast, in AML 12 cells, all the compounds except M30 and M33, increased luciferase expression from Luci-15D10 construct. Consistently within these two cell types, M34 is the most potent inducer among these two groups of compounds in turning on the riboswitch (**Figure 9D**).

[0602] The relatively more potent compounds were further tested for their activity in regulating riboswitch in additional human cell lines. As shown in **Figure 9E** in human ARPE-19 cells, both benfotiamine and M32 showed no activity in increasing riboswitch 15D10-regulated luciferase expression, whereas M10, M19 and M21 have equivalent potency to fursultaimine in increasing luciferase expression from Luci-15D10 construct. Similarly, benfotiamine and M32 have no activity in Hep G2 cells, while M19, M21 and M10 have similar potency at 12.5 μM concentration (**Figure 9F**).

[0603] Compounds M34 to M123 and compounds M126 to M160, with fursultiamine as a control compound, were tested in both HEK 293 and AML 12 cells for their activity in regulating thiamine analog riboswitches using a luciferase construct harboring 15D10 riboswitch (Luci-15D10, SEQ ID NO:93). Increases for the luciferase expression from the Luci-15D10 construct in response to these compounds are shown in Tables 2 and 3.

Table 2. Fold increase in riboswitch 15D10-regulated luciferase expression for compounds M34 to M123 at 50 μM (calculated as the ratio of the luciferase signal from cells treated with the identified compound divided by luciferase signal from untreated cells). # = Compound number. HEK = HEK 293 cells. AML = AML12 cells. Ctrl. = fursultiamine.

#	HEK	AML	#	HEK	AML	#	HEK	AML
M19	676.3	329.5	M59	1.1	1.0	M101	88.6	128.0
M21	618.3	228.6	M62	49.0	109.9	M103	0.9	0.5
M34	163.5	290.6	M63	1.7	19.0	M105	0.2	0.2
M37	76.2	134.3	M64	2.0	3.9	M106	1.2	294.4
M38	1.2	1.0	M65	3.2	67.1	M107	26.0	40.2
M39	102.7	127.2	M66	1.1	0.4	M108	75.8	129.2
M40	83.1	206.0	M67	2.1	0.9	M109	22.6	87.3
M43	10.2	95.3	M68	303.0	209.0	M110	65.7	145.6
M44	176.8	314.1	M70	51.0	235.9	M111	180.7	174.7
M45	1.3	2.9	M72	235.9	295.1	M113	212.9	272.5
M46	1.3	6.0	M73	205.1	200.3	M114	50.1	88.8
M47	25.3	85.1	M76	60.7	134.3	M115	19.9	52.1
M49	414.0	202.3	M79	352.8	358.2	M116	1.3	0.8
M50	93.5	141.8	M82	89.2	73.9	M117	0.8	3.4
M51	392.5	188.6	M88	215.7	355.2	M118	9.3	140.3

M52	363.9	195.1		M91	13.1	150.3		M119	0.9	0.7
M53	2.4	57.0		M92	0.9	23.7		M120	0.5	30.5
M54	3.0	13.9		M97	298.9	126.2		M121	0.9	0.5
M55	2.1	9.0		M98	13.6	57.7		M122	1.3	0.5
M57	1.2	24.4		M99	0.2	0.4		M123	0.4	0.6
M58	1.1	1.5		M100	1.1	0.4		Ctrl.	446.2	314.6

Table 3. Fold increase in riboswitch 15D10-regulated luciferase expression for compounds M126 to M160 at 50 µM (calculated as the ratio of the luciferase signal from cells treated with the identified compound divided by luciferase signal from untreated cells. # = Compound number. HEK = HEK 293 cells. AML = AML12 cells.

#	**HEK**	**AML**		#	**HEK**	**AML**
M113	265.00	348.35		M143	138.05	299.80
M126	147.00	305.39		M145	156.06	60.62
M127	0.46	7.39		M146	53.27	96.56
M128	0.56	7.87		M147	193.36	266.56
M129	10.91	167.48		M148	152.89	288.05
M130	110.40	328.65		M149	175.93	382.94
M131	740.64	508.85		M150	12.27	11.96
M133	99.97	273.22		M151	20.51	231.00
M134	556.94	407.79		M152	68.89	36.62
M135	87.55	44.49		M153	441.91	534.59
M136	132.20	262.53		M155	7.71	117.46
M137	139.56	287.66		M156	0.06	10.13
M138	227.96	195.26		M157	117.52	315.57
M139	37.02	188.39		M158	96.94	235.31
M140	174.69	192.16		M159	85.50	205.55
M141	1.61	3.46		M160	79.51	131.03
M142	85.27	183.85				