

Rational Design, Re-engineering and Characterisation of Tetrahydrofolate Riboswitches in Bacteria

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences

September 2015

James Leigh

Manchester Institute of Biotechnology
School of Chemistry

Contents

Abstract	16
Declaration	17
Copyright Statement	17
Acknowledgements.....	18
Abbreviations.....	19
1.1 RNA structure and function	22
1.2 Transcription: Mechanisms and regulation	25
1.3 Translation: Ribosome structure, mechanisms and regulation	28
1.4 Riboswitches.....	29
1.4.1 Riboswitches overview	29
1.4.2 Riboswitch structure	30
1.4.3 Riboswitch classes and ligands	30
1.4.4 Mechanisms of riboswitch-mediated control of gene expression	32
1.4.4.1 Overview of transcriptional and translational riboswitch mechanisms.....	32
1.4.4.2 Kinetic and thermodynamic control of riboswitch-mediated gene expression	33
1.4.5 The adenine-responsive add A-riboswitch.....	35
1.4.6 The preQ riboswitch	36
1.4.7 Tandem riboswitch architectures	38
1.5 Existing Small-Molecule Inducible Gene Expression Systems	40
1.5.1 Small-molecule inducible repressor/operator systems.....	40
1.5.2 The <i>lac</i> operon	40
1.5.3 The ‘all-or-nothing’ phenomenon: The <i>ara</i> system.....	42
1.5.4 Summary of the limitations of current small-molecule inducible expression systems.....	42
1.6 Engineered riboswitches as tools for synthetic biology.....	44
1.6.1 Generating RNA aptamers in vitro: SELEX.....	44
1.6.2 Exploitation of riboswitch domain modularity to create chimeric expression systems.....	46
1.6.2.1 The SAM-I riboswitch and chimeric transcriptional terminator devices	48
1.6.2 Other approaches to riboswitch engineering and the creation of more complex regulatory devices	49
1.6.3 Engineered riboswitches in the Micklefield laboratory.....	51
1.6.3.1 Re-engineering an orthogonally selective <i>add</i> A-riboswitch	52

1.6.3.2 Re-engineering an orthogonally selective preQ riboswitch	54
1.7 The Tetrahydrofolate (THF) Riboswitch	56
1.7.1 Folate biosynthesis.....	56
1.7.2 The role of THF and reduced folates in one-carbon metabolism.....	58
1.7.3 THF riboswitch overview and phylogeny	58
1.7.4 THF aptamer domain structure and ligand recognition	60
1.7.5 THF riboswitch ligand binding specificity and regulatory function in vitro	62
1.7.6 THF riboswitch ligand binding site cooperativity	65
1.7.7 In vivo characterisation of the THF riboswitch	67
1.8 Aims of Project.....	68
2.1 Chapter Introduction	71
2.1.1 Overview of the synthesis of folate analogues	71
2.1.2 Aims and Objectives.....	75
2.2 Results and Discussion.....	76
2.2.1 Riboswitch binding sites and ligand design.....	76
2.2.2 Synthesis of 5,6,7,8-tetrahydropteridines	80
2.2.3 Synthesis of 5,6,7,8-tetrahydropyrido[2,3- <i>d</i>]pyrimidines	82
2.2.4 Synthesis of C6-alkyl-5,6,7,8-tetrahydropyrido[2,3- <i>d</i>]pyrimidines	83
2.2.5 Synthesis of α -alkylphenyl-substituted acrylaldehydes	85
2.2.6 Synthesis of C6-alkylphenyl-substituted pyrido[2,3- <i>d</i>]pyrimidines.....	86
2.2.7 Synthesis of 6-alkylphenyl-5,6,7,8-tetrahydropyrido[2,3- <i>d</i>]pyrimidines.....	88
2.2.8 Synthesis of N-((2-amino-4-oxo-3,4,5,6,7,8-hexahydropyrido[2,3- <i>d</i>] pyrimidin-6-yl)methyl)-N-phenylformamide	90
2.3 Chapter conclusions	93
3.1 Chapter Introduction	96
3.1.1 Aims and Objectives.....	96
3.2 Results and Discussion.....	99
3.2.1 Engineering and characterisation of a chimeric <i>folT/metE</i> THF riboswitch <i>in vivo</i>	99
3.2.1.1 Design and construction of the <i>in vivo</i> reporter gene construct	100
3.2.1.2 Design and creation of riboswitch mutants	101
3.2.1.3 Determination of riboswitch basal expression levels	102
3.2.1.4 Riboswitch expression response to a single concentration of folinic acid	103
3.2.1.5 Riboswitch expression response to single concentrations of T0A and T1A	106
3.2.1.6 Riboswitch expression response to diaminopterin analogues	107

3.2.2 Engineering and characterisation of native THF riboswitches <i>in vivo</i>	108
3.2.2.1 Selection of candidate THF riboswitches	108
3.2.2.2 Construct design and creation.....	111
3.2.2.3 Dose response profiles of three candidate riboswitches using folinic acid	111
3.2.2.4 Creation of an <i>L. casei</i> riboswitch mutant library and determination of basal expression levels	114
3.2.2.5 <i>L. casei</i> riboswitch expression response at a single concentration of folinic acid.....	116
3.2.2.6 Dose response profiles using folinic acid	118
3.2.2.7 Riboswitch dose-response profiles using T0A.....	119
3.2.2.8 <i>L. casei</i> riboswitch library expression response to a single concentration of T0A.....	121
3.2.2.9 Exploiting the effect of T0A to improve THF riboswitch expression parameters	123
3.2.2.10 Investigating the regulatory effect of T0A analogues.....	125
3.2.2.11 Screening T1A with the <i>L. casei</i> mutant library	130
3.4 Chapter Conclusions.....	134
4.1 Chapter Introduction	138
4.1.1 Overview of isothermal titration calorimetry (ITC)	138
4.1.2 <i>In vitro</i> Transcription Assays.....	141
4.1.3 In-line probing analysis.....	142
4.1.4 <i>In vitro</i> characterisation of the THF riboswitch	143
4.1.5 Aims and Objectives.....	144
4.2 Results and Discussion.....	146
4.2.1 Synthesis and purification of RNA for ITC analysis	146
4.2.1.1 <i>In vitro</i> transcription of <i>S. mutans</i> THF aptamer domains.....	147
4.2.1.2 FPLC purification of <i>S. mutans</i> RNA transcription products	148
4.2.1.3 <i>In vitro</i> transcription of <i>L. casei</i> THF aptamer domains	149
4.2.1.4 FPLC purification of <i>L. casei</i> RNA transcription products	151
4.2.1.5 Preparative-scale denaturing polyacrylamide gel electrophoresis (PAGE) purification of <i>L. casei</i> RNA transcription products	153
4.2.2 <i>In vitro</i> analysis of THF aptamer domains using isothermal titration calorimetry (ITC).....	154
4.2.2.1 ITC analysis of the <i>Streptococcus mutans</i> wild-type THF aptamer domain	154
4.2.2.2 ITC analysis of the <i>Streptococcus mutans</i> M1 THF aptamer domain	156
4.2.2.3 ITC analysis of the <i>Streptococcus mutans</i> M2 aptamer domain	158

4.2.2.4 ITC analysis of the <i>Lactobacillus casei</i> wild-type aptamer domain	160
4.2.2.5 ITC analysis of the <i>Lactobacillus casei</i> M1 aptamer domain	162
4.2.3 <i>in vitro</i> transcription termination assays using the <i>Smu folT/Bsu metE</i> chimeric riboswitch	162
4.2.3.1 Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis	165
4.2.3.2 <i>In vitro</i> transcription assays utilising the <i>Smu folT/Bsu metE</i> chimeric riboswitch	166
4.3 Chapter conclusions	169
5. Project conclusions and future work.....	171
5.1 Conclusions: Characterisation of the wild-type THF riboswitch <i>in vivo</i>	171
5.2 Conclusions: Re-engineering of the THF riboswitch.....	172
5.3 Future work.....	173
6.1 Overview of experimental apparatus and techniques	177
6.2 Methods – <i>in vivo</i> reporter gene assays.....	179
6.2.1 Plasmid extractions, sequencing and glycerol cell stocks	179
6.2.2 Site-directed mutagenesis	179
6.2.3 Preparation of chemically competent <i>E. coli</i> DH5 α cells	181
6.2.4 Heat-shock transformation of <i>E. coli</i> DH5 α cells with mutagenic plasmids ...	181
6.2.5 Preparation of electro-competent <i>E. coli</i> DH5 α pLOI707HE/slr0642 cells.....	181
6.2.6 Transformation of <i>E. coli</i> DH5 α pLOI707HE/slr0642 with mutagenic plasmids by electroporation.....	182
6.2.7 <i>E. coli</i> GFP- <i>lv</i> reporter gene expression assay	182
6.3 Methods - Enzyme over-expression and purification	184
6.3.1 Expression of recombinant T7 RNA Polymerase (T7-RNAP).....	184
6.3.2 Purification of T7 RNA Polymerase (T7-RNAP).....	184
6.3.3 SDS-Polyacrylamide Gel Electrophoresis (PAGE) analysis.....	185
6.4 Methods - RNA synthesis and purification.....	186
6.4.1 <i>In vitro</i> transcription reactions.....	186
6.4.2 Urea PAGE analysis.....	187
6.4.3 RNA purification using size exclusion FPLC	187
6.4.4 Urea PAGE preparative scale purification.....	188
6.5 Methods - Isothermal Titration Calorimetry (ITC)	189
6.6 Chemical Synthesis	191
6.7 Methods - Buffers and Media Recipes	210
6.8 Sequences of primers and <i>in vivo</i> expression constructs	214
6.8.1 <i>In vivo</i> expression assay construct sequences.....	214

6.8.1.1 <i>S. mutans</i> THF/metE chimera GFP expression construct	214
6.8.1.2 <i>S. mutans</i> THF wild-type GFP expression construct	214
6.8.1.3 <i>L. casei</i> THF wild-type GFP expression construct	215
7. References	217

Table of Figures

Figure 1 – Overview of RNA structure, base-pair classification and nomenclature . A) Interacting faces of ribonucleotides; B) Relative glycosidic bond orientations; C) Mechanism of RNA intramolecular hydrolysis.	22
Figure 2 – Non-canonical RNA base pairings. A) Example structures and H-bonding motifs of six common ribonucleoside interactions; B) Exhaustive list of all possible canonical and non-canonical RNA base pairings.	24
Figure 3 – Overview of the three stages of transcription in <i>E. coli</i> . A) Initiation – RNA polymerase (RNAP) binds to the DNA promoter region, facilitated by transcription factors, to form an initiation complex; B) Elongation – procession of the transcription complex along the DNA template catalyses the formation of the RNA transcript; C) Termination (intrinsic shown) – formation of an RNA terminator stem loop succeeded by a poly-uridylate tract destabilises the transcription complex resulting in dissociation of the RNAP from the template strand.	26
Table 1 – Overview of known riboswitch classes including cognate ligand and original discovery paper	31
Figure 4 – General mechanisms of riboswitch mediated control of gene expression. A) Transcriptional control. In this example, a ligand binding event induces the formation of a downstream terminator stem in the riboswitch expression platform causing the RNAP to disengage. B) Translational control. In this example, a ligand binding event causes the release of the Shine Dalgarno sequence from a stem loop, allowing translation initiation to occur.....	32
Figure 5 – X-ray crystallography structure of the <i>add A</i> -riboswitch from <i>Vibrio vulnificus</i> ³² . A) Structure of the <i>add A</i> aptamer domain showing overall tertiary architecture; B) Structure of the <i>add A</i> ligand binding site showing adenine recognition by four unpaired uridine residues.....	35
Figure 6 – X-ray crystallography structure of the PreQ class I riboswitch from <i>Bacillus subtilis</i> ⁶⁵ . A) Structure of the PreQ aptamer domain showing tertiary architecture; B) structure of the ligand binding site showing the PreQ1 binding motif.....	37
Figure 7 – Overview of the <i>lac</i> repressor system. A) Mechanism of genetic regulation by <i>lacI</i> . Left: the <i>lac</i> repressor (LacI) binds the operator region thereby blockading transcription. Right: IPTG binds to the <i>lac</i> repressor, inducing a conformational change which releases LacI from the DNA template, allowing transcription by the endogenous RNA polymerase (RNAP). B) Structures of the natural inducer of the <i>lac</i> operon, allolactose (left), and the synthetic analogue isopropyl- β -D-1-thiogalactopyranoside (IPTG; right).....	41

Figure 8 – Overview of the ‘all-or-nothing’ effect using the example of the *ara* system (modified from Kealsing *et al.*⁶⁹). A) In the absence of arabinose the P_{BAD} promoter the expression construct is repressed by the regulatory protein AraC, blockading transcription of the gene of interest. B) Addition of arabinose allows the compound to bind AraC to de-repress P_{BAD}, as well as inducing the expression of the arabinose transporter *araE* from the genomic DNA. C) Representation of the relationship between expression output observed at the population level and the expression output of individual cells. 43

Figure 9 – Overview of a typical SELEX procedure (redrawn from Ellington *et al.*⁹⁸)... 45

Figure 1 – Overview of encoded co-transcriptional folding. A) Folding of the aptamer domain begins during transcription by RNA polymerase (RNAP). B) A ligand binding event stabilises the aptamer domain P1 stem before the downstream expression platform has been transcribed. C) P1 stem formation favours only one expression platform conformation – in this case, a terminator stem.....46

Figure 11 – A) X-ray crystallography structure of the M6'' riboswitch binding PPDA⁸¹. B) Ligand binding site showing the PPDA recognition motif and the mutated residues (shown in green in 2D representation). C) Ligand binding site of the wild-type *addA* riboswitch..... 53

Figure 12 – A) Ligand binding motif for the preQ riboswitch binding preQ₁ obtained by x-ray crystallography⁶⁵. B) Proposed ligand binding motif for the preQ M1 (C17U) mutant riboswitch binding DPQ₀..... 54

Figure 13 – Structure and numbering scheme of THF. The three key moieties are highlighted. 56

Figure 14 – Folate biosynthesis pathway. A) Biosynthetic pathway showing structures of intermediates and genes encoding enzymes involved in the catalytic steps. B) Overview of folate biosynthesis and transport genes in bacteria showing the frequency with which they are found to be controlled by THF-responsive riboswitches. 57

Figure 15 – Tetrahydrofolate and related compounds in bacterial single carbon metabolism. Genes involved in each transformation are shown next to the appropriate arrow. For reactions whereby a different gene is involved in each catalytic direction - two genes are listed. Genes for reactions in the top to bottom direction are shown on the left of the arrow; genes for reactions in the right to left direction are shown above the arrow. 59

Figure 16 – Structure of the *S. mutans* THF riboswitch aptamer domain. A) X-ray crystallography structure of the THF aptamer binding folinic acid¹⁰⁴. B) Secondary structure of the THF aptamer domain based on x-ray crystallography data. Structural features are labelled and colouring is consistent across both images. 60

Figure 17 – THF riboswitch binding sites and folinic acid recognition motifs¹⁰⁴. Top: 3WJ site; Bottom: PK site. 61

Table 2 – Summary of <i>in vitro</i> ligand binding/regulatory parameters for the THF riboswitch	63
Figure 18 – Structures of compounds which are known to bind the THF riboswitch ^{57, 104, 110}	64
Figure 19 – Structures of folic acid and examples of DHFR inhibitors currently used in the clinic	71
Figure 20 – Overview of synthetic routes to 6-substituted 5-deazapteridines; a) 2-methoxyethanol, 120 °C, 24 h (DeGraw <i>et al</i>) ¹³⁹ ; b) NaEtOH, EtOH, reflux, 12 h; c) Raney Ni, HCO ₂ H 80 °C, 2 h; d) NaBH ₄ , MeOH, rt, 2h; e) Ph ₃ PBr ₂ Me ₂ NAC, rt, 2h; f) Me ₂ NAC, rt 24 h (Piper <i>et al</i>) ¹⁴² and Graffner-Nordberg <i>et al</i>) ¹⁴⁵ ; g) NaOH, H ₂ O, reflux 4h (Piper <i>et al</i>) ¹⁴² and Temple <i>et al</i>) ¹⁴¹).	73
Figure 21 – A) Structures, IUPAC nomenclature and short-hand labelling for pyrido[2,3d]pyrimidines. B) Overview of condensation reactions utilised in the synthesis of 6-substituted pyrido[2,3d]pyrimidines; a) water, piperidine acetate, reflux (Stark <i>et al</i>) ¹⁴⁷ ; b) aqueous AcOH (60%), reflux 12 h (Kuwada <i>et al</i>) ¹⁴⁶ ; AcOH, reflux 5 h (Troschutz <i>et al</i>) ¹⁴⁸).	75
Figure 22 – Overview of known and predicted ligand binding motifs for the wild-type THF riboswitch. A) Folinic acid recognition by the wild-type THF riboswitch at the 3WJ and PK ligand binding sites (obtained by x-ray crystallography ¹⁰⁴); B) Predicted ligand binding motifs for the THF riboswitch and synthetic folate analogues; C) Secondary structure of the THF riboswitch aptamer domain showing the 3WJ and PK ligand binding sites.....	77
Figure 23 - Overview of known and predicted ligand binding motifs for the wild-type, M1 (U25C/C53U) and M2 (U35C/U42C) THF riboswitches. A) Folinic acid recognition by the wild-type THF riboswitch at the 3WJ and PK ligand binding sites (obtained by x-ray crystallography ¹⁰⁴); B) Predicted ligand binding motifs for THF riboswitch mutants and synthetic folate analogues. Top: Predicted ligand binding motif for the M1 (U25C/C53U) riboswitch and ‘di-keto’ THF analogues. Bottom: Predicted ligand binding model for the M2 (U35C/U42C) riboswitch and ‘di-amino’ THF analogues. C) Secondary structure of the THF riboswitch aptamer domain showing the 3WJ and PK ligand binding sites. ...	80
Figure 24 – Synthetic strategy for the preparation of 5,6,7,8-tetrahydropteridines; a) NaHCO ₃ , H ₂ O, 70 °C, 3 h; b) KOH, H ₂ O, RT, 24 h; c) H ₂ , PtO ₂ , H ₂ O, RT, 24 h; d) H ₂ , PtO ₂ , TFA, RT, 24 h.....	81
Figure 25 – Synthetic strategy for the preparation of 5,6,7,8-tetrahydropyrido[2,3-d]pyrimidine analogues. a) AcOH, H ₂ O, reflux 100 °C, 3.5 h; b) AcOH, reflux 120 °C, 10 h; c) No solvent, reflux 150 °C, 12 h; d) H ₂ , PtO ₂ , TFA, RT, 24 h; e) NaOMe, MeOH, microwave, 140 °C, 10 mins.	82

- Figure 26 – Overview of the synthetic strategy for the synthesis of 6-alkyl-2,4-disubstituted-5,6,7,8-tetrahydropyridopyrimidine analogues. a) AcOH, reflux, 12 h; b) H₂, PtO₂, TFA, 24 h; c) TFA, reflux, 12 h..... 84
- Figure 27 – Overview of the synthesis of 2-alkylphenyl-substituted acrylaldehydes. a) DMSO, (COCl)₂, DCM, -78 °C, then Et₃N; b) pyrrolidine, N,N-dimethyl-4-aminobenzoic acid, HCHO, DCM, 40 °C..... 85
- Figure 28 – Condensation reactions for the synthesis of 6-alkylphenyl-substituted pyrido[2,3-d]pyrimidines; a) AcOH, reflux 6 h; b) TFA, reflux 6 h..... 87
- Figure 29 – Overview of the synthesis of 6-alkylphenyl-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidines. a) H₂, PtO₂, TFA; b) H₂, Pd/C, TFA. 90
- Figure 30 – Overview of the synthesis of N-((2-amino-4-oxo-3,4,5,6,7,8-hexahydropyrido[2,3-d]pyrimidin-6-yl)methyl)-N-phenylformamide. a) CuCl, KOH, RT, 12 h; b) Boc₂O, EtOH, RT, 24 h; c) (COCl)₂, DMSO, DCM, then Et₃N; d) pyrrolidine, N,N-dimethyl-4-aminobenzoic acid, HCHO, DCM, 40 °C; e) AcOH, reflux, 12 h; f) HCOOH, Ac₂O, RT, 3 h; g) H₂, PtO₂, TFA, rt, 48 h..... 92
- Figure 31 – Compounds successfully synthesised and carried forward for testing in the *in vivo* assay 93
- Figure 32 – Known and predicted binding motifs for THF riboswitch ligands based on x-ray crystallography data¹⁰⁴. A) Known (THF/folinic acid) and predicted (all others) binding motifs for ‘keto-amino-faced’ compounds with the wild-type riboswitch at both binding sites; B) predicted ligand binding motif for ‘diketo-faced’ compounds with the M1 (U25C/C53U) mutant riboswitch 3WJ site; C) predicted ligand binding motif for ‘diamino-faced’ compounds with the M2 (U35C/U42C) mutant riboswitch PK site. 98
- Figure 33 – A) Putative switching mechanism for the *Smu folT/Bsu metE* chimeric riboswitch. The *S. mutans folT* aptamer domain is shown in blue and the *B. subtilis metE* expression platform in orange. Left: In the ligand-unbound conformation the expression platform forms an anti-terminator stem, allowing transcription of the ORF. Right: A ligand binding event stabilises the P3 stem and induces pseudoknot and P1 stem formation, favouring the establishment of a downstream terminator stem and the premature abortion of transcription. B) Chimeric reporter gene construct used in the GFP expression assays 99
- Figure 34 – A) Basal GFP expression levels for wild-type (WT) *Smu folT/Bsu metE* and mutant library. The bar height indicates the mean normalised fluorescence of three experiments and the error bars represent the standard deviation. B) Secondary structure of the *S. mutans* THF aptamer domain. Key residues targeted for mutagenesis are highlighted in red. C) Summary of riboswitch mutants. 102
- Figure 35 – A) Relative GFP expression levels for *S. mutans/metE* THF riboswitch mutants with and without folinic acid. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM

folinic acid. Bar height indicates the mean of three experiments and error bars show the standard deviation. B) Secondary structure of the *S. mutans* THF aptamer domain. Key residues targeted for mutagenesis are highlighted in red. C) Summary of riboswitch mutants. D) Structure of folinic acid..... 104

Figure 36 – A) Relative *GFP* expression levels for *S. mutans/metE* THF riboswitch mutants with and without T0A. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM T0A. Bar height indicates the mean of three independent experiments and error bars show the standard deviation. B) Structure of T0A. C) Relative *GFP* expression levels with and without 2 mM T1A. D) Structure of T1A. E) Summary of riboswitch mutants. 106

Figure 37 – Structures of synthesised compounds and commercially available pyrimidines screened with the M2 riboswitch 107

Figure 39 – A) Secondary structure and consensus sequence alignment of THF riboswitch aptamer domains from *Streptococcus mutans*, *Alkaliphilus metalliredigens*, and *Lactobacillus casei* in the ligand-bound state. Residues conserved across all three riboswitches are shown in blue and ligand binding site residues are shown in red with the nucleotide numbering taken from the *S. mutans* sequence. (Note: *L. casei* riboswitch actually contains the U35C mutation, but residue is highlighted for clarity). B) Ligand binding motifs for the *S. mutans* THF riboswitch, obtained by x-ray crystallography¹⁰⁴..... 108

Figure 40 – Putative switching mechanisms for two translational-‘OFF’ THF riboswitches. A) *Streptococcus mutans*; B) *Lactobacillus casei*. In both cases, it is proposed that a ligand binding event strengthens the P3 stem, stabilises pseudoknot formation, and causes sequestration of the Shine-Dalgarno sequence within a downstream stem loop. Diagrams were created using Mfold¹⁶⁴ and Vienna¹⁶⁹ web server secondary structure predictions. 110

Figure 41 – Fully native construct design used in GFP reporter gene assays 111

Figure 42 – Dose-responsive control of *GFP* expression by THF riboswitches. A) *Lactobacillus casei*; B) *Streptococcus mutans*; C) *Streptococcus mutans/metE* chimera; D) *Alkaliphilus metalliredigens* (data from Dr. Helen Vincent); E) No riboswitch control strain; F) Structure of folinic acid; G) Expression parameters. Data points indicate the mean of three experiments and error bars show standard deviation. 112

Figure 43 – A) Basal *GFP* expression levels for *L. casei* THF riboswitch mutant library. Bar height indicates the mean fluorescence normalised for OD based on three experiments. Error bars show the standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Overview of riboswitch mutants..... 115

Figure 44 – A) Relative *GFP* expression levels for *L. casei* THF riboswitch mutants with and without folinic acid. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM folinic acid. Bar height

indicates the mean of three independent experiments and error bars show the standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Overview of riboswitch mutants..... 116

Table 3 – Comparison of gene expression output for the *Smu folT/Bsu metE* and *L. casei* riboswitches..... 117

Figure 45 – Dose-responsive control of gene expression with the *L. casei* binding site knockout mutant riboswitches. A) M3 (U7C, PK site knockout); B) M4 (U26C, 3WJ site knockout); C) M5 (U7C/U26C, double binding site knockout); D) structure of folinic acid; E) expression parameters for the mutant riboswitches with wild-type data shown for comparison..... 118

Figure 46 – Dose-response experiments for riboswitch reporter constructs with increasing concentrations of T0A. A) *L. casei*; B) *S. mutans* fully native; C) *S. mutans* chimera; D) no riboswitch; E) expression parameters; F) structure of T0A..... 120

Figure 47 – A) Relative *GFP* expression levels for *L. cas* THF riboswitch mutants with and without T0A. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM T0A. Bar height indicates the mean of three experiments and error bars show the standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Structure of T0A. D) Overview of riboswitch mutants..... 121

Figure 48 – Absolute fluorescence levels for the wild-type and mutant THF riboswitches in the presence (light blue) and absence (dark blue) of 2 mM T0A. Bar height indicates the mean of three experiments with error bars showing standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Overview of riboswitch mutants..... 123

Figure 49 – OD-normalised *GFP* expression mediated by the *L. cas* wild-type THF riboswitch in response to addition of folinic acid. Black circles indicate OD-normalised fluorescence in response to the addition of folinic acid in the presence of 2 mM T0A; white circles indicate response to addition of folinic acid only. Data points represent the mean of three experiments with error bars showing standard deviation. 124

Figure 50 – A) X-ray crystallography structure of the *S. mutans* THF aptamer domain bound to folinic acid¹⁰⁴ with expanded PK site shown. At both ligand binding sites folinic acid adopts a high energy pseudo-boat conformation, positioning the pABA side chain in close proximity to the backbone RNA. B) Comparison of energy minimised free solution conformation of folinic acid (pink) and pseudo-boat conformation observed at the THF riboswitch ligand binding sites (red). 126

Figure 51 - Relative *GFP* expression levels for *L. cas* THF riboswitch mutants in response to T0A analogues. Dark blue bars indicate normalised basal fluorescence;

pastel coloured bars indicate relative fluorescence in the presence of 2 mM ligand. Bar height indicates the mean of three independent experiments and error bars show the standard deviation. A) T0B; B) T0C; C) T0F; D) Structures of relevant compounds. E) Overview of riboswitch mutants. 128

Figure 52 – A) Relative *GFP* expression levels for the *L. cas* THF riboswitch in response to the commercial compound lometrexol. Bar height indicates the mean of three experiments and error bars show the standard deviation (data from Dr. Helen Vincent). B) Structures of folinic acid and lometrexol. 129

Figure 53 - Relative *GFP* expression levels for *L. cas* THF riboswitch mutants with and without T1A. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM T1A. Bar height indicates the mean of three experiments and error bars show the standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Structure of T1A. D) Overview of riboswitch mutants..... 131

Figure 54 – *GFP-IV* expression response by the *L. cas* M6 (U7C/U26C/C55U) riboswitch to increasing T1A concentrations. Data points are the mean of three experiments and error bars indicate the standard deviation. 132

Figure 55 - Dose-responsive repression of *GFP* expression by the *L. cas* THF-M1 riboswitch. (A) Dose response curves showing normalised gene expression levels at varying folinic acid concentrations. Black circles show varying folinic acid concentrations only and white circles show varying folinic acid concentrations at fixed 2 mM T1A. (B) Dose response curves showing normalised gene expression levels at varying T1A concentrations. Black circles show varying T1A concentrations only and white circles show varying T1A concentrations at fixed 2 mM T0A. (C) Structures of folinic acid, T0A and T1A. (D) M1 riboswitch expression parameters..... 133

Figure 56 - Diagram showing the apparatus of a typical isothermal titration calorimeter. 138

Figure 57 – Equations for determining the binding parameters of a molecular interaction. The change in enthalpy (ΔH) and binding constant (K_A) are directly determined from the ITC experiment and the remaining parameters are calculated using these equations. 139

Figure 58 – Urea PAGE analysis of *in vitro* transcription reactions for THF aptamer domains. A) *S. mutans*. B) and C) *L. casei*. See table D) for the identity of the gel lanes and descriptions of the aptamer domain sequences..... 146

Figure 59 – A) UV absorbance elution profile for the FPLC size exclusion purification of the *S. mutans* wild-type THF aptamer domain showing absorbance at 260 (red) and 280 (blue) nm. The separation protocol utilised a GE healthcare Superdex 75 Hiload 26/600 pg gel filtration column run isocratically in FPLC buffer at 2 ml/min on an AKTA purifier FPLC system. B) Urea PAGE analysis of the indicated FPLC fractions..... 149

Figure 60 – Comparison of THF aptamer domain secondary structures. A) Ligand-bound *S. mutans* secondary structure as determined by x-ray crystallography¹⁰⁴. B) Predicted ligand bound *L. casei* secondary structure based on *S. mutans* sequence alignment and minimal energy conformation determined using the Mfold web server. 150

Figure 61 – A) UV absorbance elution profile for the FPLC size exclusion purification of the *L. casei* wild-type THF aptamer domain showing absorbance at 260 (red) and 280 (blue) nm. The separation protocol utilised a GE Healthcare Superdex HiLoad 26/600 pg gel filtration column run isocratically in FPLC buffer at 2 ml/min on an AKTA purifier FPLC system. B) Urea PAGE analysis of the indicated FPLC fractions. 152

Figure 62 – ITC analysis of ligand binding by the *S. mutans* wild-type THF aptamer domain. A) Binding isotherm for the titration of folinic acid. B) Binding isotherm for the titration of T0A. C) Binding isotherm for the titration of T1A. D) Structures of folinic acid, T0A and T1A E) Thermodynamic parameters calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0. 155

Figure 63 - ITC analysis of ligand binding by the *S. mutans* M1 THF aptamer domain. A) Binding isotherm for the titration of T1A. B) Binding isotherm for the titration of 7-deazaxanthine. C) Binding isotherm for the titration of folinic acid. D) Structures of T1A, 7-deazaxanthine, 7-deazaguanine and folinic acid. E) Thermodynamic parameters calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0. 157

Figure 64 - ITC analysis of ligand binding by the *S. mutans* M2 THF aptamer domain (data provided by Dr. Chris Robinson). A) Binding isotherm for the titration of T2A. B) Binding isotherm for the titration of T2C. C) Binding isotherm for the titration of folinic acid. D) Structures of T2A, T2B and folinic acid. E) Thermodynamic parameters calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0. 159

Figure 65 - ITC analysis of ligand binding by the *L. casei* wild-type THF aptamer domain. A) Binding isotherm for the titration of folinic acid with the aptamer domain purified under native conditions. B) Binding isotherms for the titration of T0A with the aptamer domain purified under native conditions. C) Binding isotherms for the titration of T1A with the aptamer domain purified under native conditions. D) Binding isotherms for the titration of folinic acid with the aptamer domain purified under denaturing conditions. D) Structures of T0A, T1A and folinic acid. E) Thermodynamic parameters for the interactions observed using the aptamer domain purified under native conditions. Parameters were calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0. 161

Figure 66 – ITC analysis of ligand binding by the *L. casei* M1 (U26C/C55U) aptamer domain. A) Binding isotherm for the titration of T1A with the aptamer domain. B) Structure of T1A. C) Thermodynamic parameters for the interactions observed using the aptamer domain purified under native conditions. Parameters were calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0. 162

Figure 67 – Proposed secondary structure, ligand binding and transcriptional termination model for the <i>folT/metE</i> chimeric riboswitch. The <i>folT</i> aptamer domain is shown in blue and the <i>metE</i> expression platform is shown in orange. Ligand binding induces the formation of a downstream intrinsic transcriptional terminator stem in the expression platform causing the RNA polymerase to abort transcription.....	163
Figure 68 – Overview of the single-turnover <i>in vitro</i> transcription assay. A) The DNA template containing the T7 promoter is pre-incubated with T7-RNAP and radiolabelled UTP; B) Transcription is initiated by the addition of a mixture of the remaining NTPs, the ligand of interest and an excess of the polyanionic saccharide heparin; C) Upon transcription termination the T7-RNAP binds to heparin, preventing further rounds of transcription. The relative levels of read-through and truncated transcripts can be assessed by urea PAGE.....	164
Figure 69 – Overview of TBIO PCR-based gene synthesis. See body text for details.	165
Figure 70 - Agarose gel of PCR products; lanes (L->R) Initial TBIO PCR product; re-amplified PCR product; ladder.	166
Figure 71 - Urea PAGE analysis of <i>in vitro</i> transcription assays. A) Multi-round assays, conducted in the absence of ligand or in the presence of THF or folinic acid; B) Single-turnover assay conducted in the presence or absence of folinic acid. Multi-round assays were stained using SYBR gold (Life Technologies), whilst the single-turnover assay was imaged by autoradiography through incorporation of ³² P in the RNA transcript from ³² P-UTP.....	167
Figure 72 – Structures of compounds required for structure-activity relationship studies of the THF riboswitch.....	174
Table 5: Standard parameters used for performing ITC experiments with the MicroCal VP-ITC microcalorimeter.....	189
Table 6: Standard parameters used for performing ITC experiments with the MicroCal PEAQ-ITC automated microcalorimeter.....	190

Abstract

The ability to independently control the expression of multiple genes in response to small-molecule inducers is highly desirable in the study of metabolic pathways and has numerous applications in the fields of drug discovery and synthetic biology. The existing small-molecule-inducible gene expression tools such as the *lac*, *ara* and *tet* systems have a number of limitations including poor titratability and a tendency to promote cell population heterogeneity. To address these shortcomings, the Micklefield laboratory has developed a novel method for controlling gene expression by re-engineering naturally occurring regulatory elements known as riboswitches. Riboswitches are mRNA-based, small-molecule-responsive genetic switches found in the 5'-UTR of many bacterial genes. We have previously shown that it is possible to re-engineer natural riboswitches, such that they are no longer activated by their cognate metabolite, but instead respond to a non-natural synthetic analogue not ordinarily present in the cell.

Our most recent target for re-engineering was the tetrahydrofolate (THF)-responsive riboswitch found to control the *folT* gene in Firmicutes. The THF riboswitch is unique in that it contains two distinct ligand binding sites within the same aptamer domain, and was considered a candidate for the engineering of a dual-input, repressive gene expression system. To date, the THF riboswitch has not been characterised *in vivo* and therefore we also sought to further understand the ligand specificity and the role of the two binding sites in eliciting the riboswitch regulatory response. We established an *in vivo* reporter gene assay in *E. coli* and demonstrated that the THF riboswitch is able to dose-dependently repress gene expression in response to the THF analogue folinic acid. Ligand binding to the aptamer domain PK site was found to be the primary mechanism by which the regulatory response is mediated *in vivo*. We designed, synthesised and screened *in vivo* a series of THF analogues, which identified a potent inhibitor of the wild-type riboswitch, and was confirmed *in vitro* using ITC. Moreover, the de-repressive activity of the inhibitor compound was exploited to optimise the expression parameters of the wild-type riboswitch. In parallel, we successfully re-engineered one of the two aptamer domain ligand binding sites, such that it selectively bound a non-natural THF analogue whilst rejecting folinic acid as demonstrated *in vitro*. Additionally, this mutant riboswitch-ligand pair was shown to dose-dependently repress gene expression *in vivo*. This represents a significant step towards the engineering of a dual-input THF riboswitch gene expression system.

Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.
- ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.
- iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy, in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations and in The University's policy on Presentation of Theses.

Acknowledgements

I would like to thank Prof. Jason Micklefield for giving me the opportunity to undertake this project as well as guidance throughout. I am very grateful to Dr. Helen Vincent and Dr. Chris Robinson for their support, training and advice during my PhD. Thanks also goes to Dr. Neil Dixon for his guidance, particularly in the early days.

Finally, a thank you to my family for all their support and patience.

Abbreviations

2AP	2-aminopurine
3WJ	Three-way junction
AdoCbl	Adenosylcobalamin
APS	Ammonium persulfate
ATP	Adenosine triphosphate
bp	Base pair
c-di-GMP	Cyclic-di-guanosinemonophosphate
CTP	Cytidine triphosphate
DAP	2,6-diaminopurine
DEPC	Diethylpyrocarbonate
DHF	7,8-dihydrofolate
DHFR	Dihydrofolate reductase
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraaceticacid
ES	Electrospray
FMN	Flavin mononucleotide
FPLC	Fast Protein Liquid Chromatography
GFP	Green Fluorescent Protein
GTP	Guanosine triphosphate
HPLC	High Pressure Liquid Chromatography
HR-MS	High Resolution Mass Spectrometry
IPTG	Isopropyl β -(D)-1-thiogalactopyranoside
IR	Infrared Spectroscopy
ITC	Isothermal Titration Calorimetry
K_D	Dissociation constant
LB	Lysogeny broth
mRNA	Messenger ribonucleic acid
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
nt	Nucleotide

NTP	Ribonucleotide triphosphate
NTPs	A mixture comprising equal concentrations of ATP, GTP, UTP and CTP
OD	Optical density
PAGE	Poly-Acrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PK	Pseudoknot
ppm	Parts Per Million
preQ₀	7-cyano 7-deazaguanine
preQ₁	7-aminoethyl 7-deazaguanine
RBS	Ribosomal binding site
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNase	Ribonuclease
rRNA	Ribosomal RNA
SAM	S-adenosylmethionine
SD	Shine-Dalgarno
SDS	Sodium Dodecyl Sulphate
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SHAPE	Selective 2'-Hydroxyl Acylation analysed by Primer Extension
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TEMED	Tetramethylethylenediamine
THF	5,6,7,8-tetrahydrofolic acid
TLC	Thin Layer Chromatography
TMS	Trimethylsilane
TPP	Thiamine pyrophosphate
TRIS	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
UTP	Uridine triphosphate
UTR	Untranslated region
UV	Ultraviolet
WT	Wild-type

Chapter 1

Introduction

1.1 RNA structure and function

RNA plays a pivotal and varied role in cellular function, particularly in the regulation of gene expression. Historically, RNA was considered a passive messenger molecule, acting merely as a vehicle for the transfer of genetic information from DNA into proteins. However, RNA is now known to influence every step of gene expression, through a variety of mechanisms largely attributable to its ability to form intricate secondary and tertiary structures¹.

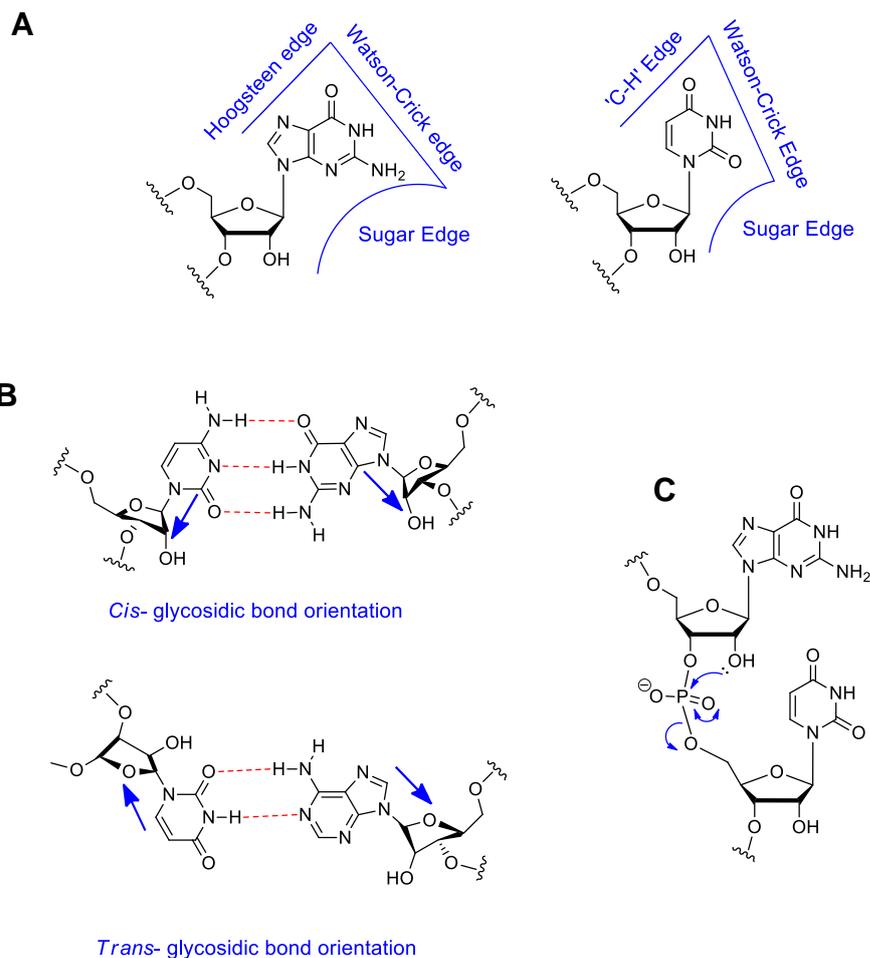


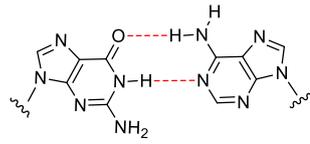
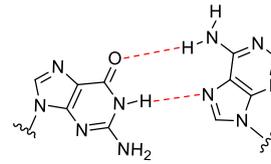
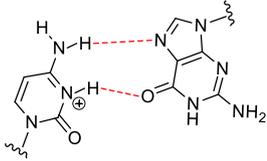
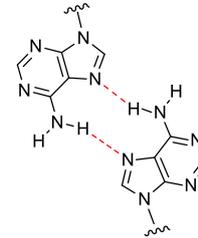
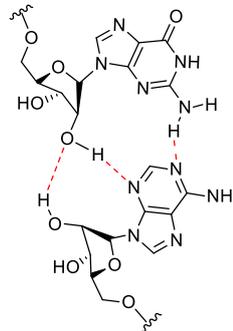
Figure 2 – Overview of RNA structure, base-pair classification and nomenclature . A) Interacting faces of ribonucleotides; B) Relative glycosidic bond orientations; C) Mechanism of RNA intramolecular hydrolysis.

The interaction between nucleobases within nucleic acids occurs in two key ways. Firstly, via π - π stacking to adjacent bases on the same strand, which is the major impetus for nucleic acid folding. Secondly, through edge-to-edge interactions via hydrogen bonding; this can determine strand directionality. The canonical Watson-

Crick base pairings (GC and A-T/U) are unique in that in the case of DNA they give rise to the B-form double helix and substitution of any of the four combinations at any position does not perturb the tertiary helical structure². However, the structural complexity of RNA is a consequence of its ability to form a variety of base pairings outside of the canonical Watson-Crick interactions. Indeed, only around 60% of ribonucleotides within structured RNAs participate in canonical base-pairing². Furthermore, the structure of ribonucleotides does not favour the formation of stable, anti-parallel B-form RNA double helices, which are the hallmark of DNA structure. The presence of the RNA 2'-OH moiety means that the C2'-endo pucker conformation adopted by DNA in the B-form helix is unfavourable for RNA. The structure of ribonucleotides allows for hydrogen bonding interactions between three distinct faces, as shown in **Figure 1A**. These are the 'Watson-Crick edge', the 'Hoogsteen' edge (purines) and 'C-H' edge (pyrimidines), and the sugar edge which includes the 2'-OH of the ribose ring. In addition to the three edges, each pair of interacting ribonucleotides can also be classified by the relative orientations of the glycosidic bonds³, which can be either *cis*- or *trans*- (**Figure 1B**). This gives a total of 12 possible base pair geometries, a summary and examples of which can be found in **Figure 2**.

The diversity of possible ribonucleotide interactions gives rise to the large structural variety exhibited by RNA molecules. Common RNA secondary structural motifs include stem loops, pseudoknots, K-turns and junctions. RNA stems comprise a double-stranded helical structure of base-paired residues at the end of which is a region of unpaired nucleotides known as a loop⁴. One frequently encountered structural feature is the base-pairing of two complementary loops located at distal regions of the RNA – termed a pseudoknot. Additionally, unpaired nucleotides are often encountered at regions which serve to join several stems loops, known as junctions¹. The extent to which RNA nucleotides are engaged in base-pairing interactions is the basis for a number of different *in vitro* analytical techniques for probing RNA structure, including SHAPE⁵ and in-line probing⁶. Despite its inherent flexibility, RNA has also been well studied using x-ray crystallography⁷ and there are many notable examples of complex RNA structures being solved in this way, famously including the entire 30S ribosomal subunit in 1999 by Ramakrishnan⁸.

A

*cis*-Watson-Crick/Watson-Crick*cis*-Watson-Crick/Hoogsteen*Trans*-Watson-Crick/Hoogsteen*Trans*-Hoogsteen/Hoogsteen*Cis*-Sugar Edge/Sugar Edge

B

No.	Glycosidic bond orientation	Interacting Edges	Local Strand Orientation
1	<i>Cis</i>	Watson-Crick/Watson-Crick	Antiparallel
2	<i>Trans</i>	Watson-Crick/Watson-Crick	Parallel
3	<i>Cis</i>	Watson-Crick/Hoogsteen	Parallel
4	<i>Trans</i>	Watson-Crick/Hoogsteen	Antiparallel
5	<i>Cis</i>	Watson-Crick/Sugar Edge	Antiparallel
6	<i>Trans</i>	Watson-Crick/Sugar Edge	Parallel
7	<i>Cis</i>	Hoogsteen/Hoogsteen	Antiparallel
8	<i>Trans</i>	Hoogsteen/Hoogsteen	Parallel
9	<i>Cis</i>	Hoogsteen/Sugar Edge	Parallel
10	<i>Trans</i>	Hoogsteen/Sugar Edge	Antiparallel
11	<i>Cis</i>	Sugar Edge/Sugar Edge	Antiparallel
12	<i>Trans</i>	Sugar Edge/Sugar Edge	Parallel

Figure 3 – Non-canonical RNA base pairings. A) Example structures and H-bonding motifs of six common ribonucleoside interactions; B) Exhaustive list of all possible canonical and non-canonical RNA base pairings.

However, whilst chemical probing techniques have also shown application *in vivo*^{9, 10}, the structure of RNA within the cell has been less thoroughly characterised.

1.2 Transcription: Mechanisms and regulation

Transcription is the process of transferring the genetic information contained within DNA to a molecule of RNA. During transcription, the nucleotide sequence from the template DNA strand is 'copied' into a complementary RNA transcript, catalysed by the enzyme RNA polymerase. Arguably the most well characterised transcriptional machinery is the *E. coli* system¹¹, which will be used as the basis for discussion in this section. The chemical processes which underpin transcription are straightforward, however the variety of regulatory mechanisms deployed by organisms to control the process are diverse and numerous¹². Given that transcription is the foremost process in the transfer of genetic information into functional proteins, it is understandable that this event is a major focal point in the regulation of gene expression¹¹.

Transcription is categorised into three successive events: initiation, elongation and termination; and there are examples of regulatory control mechanisms which operate at each of these stages. During the initiation phase (**Figure 3A**), the RNA polymerase (RNAP) binds to the promoter region of the gene to be transcribed and multiple transcription factors increase the relative availability of the promoter. Transcriptional activation typically occurs via interaction of the promoter-bound RNA polymerase to other DNA-bound proteins, either distal or adjacent to the site of transcription initiation. Interaction between distal DNA-bound proteins and the RNA polymerase typically occurs by controlled DNA looping^{13, 14}. Transcriptional regulation during the initiation phase is primarily an 'inter-operon' process, in that multiple operons compete for a finite amount of free RNA polymerase¹⁵. Competition occurs at both the thermodynamic and kinetic levels¹⁶. The distribution of RNA polymerase on the available promoter regions of the DNA template is directly related to the binding affinity of the RNA polymerase for the promoter sequence^{17, 18}, however overall binding affinity is dependent on both the promoter sequence and the extent to which this is enhanced by transcription factors¹⁹. Kinetically, the rate at which the transcription complex can initiate RNA synthesis and proceed beyond the promoter region into the elongation phase is also important for transcriptional regulation²⁰. Once the nascent RNA transcript has become adequate in length to stabilise the transcription complex such that dissociation from the DNA template becomes unfavourable, transcription initiation is said to be complete. During the elongation phase (**Figure 3B**), the RNA polymerase progresses along the DNA template, catalysing phosphodiester bond formation as the RNA oligonucleotide grows, and also participating in specific pausing events.

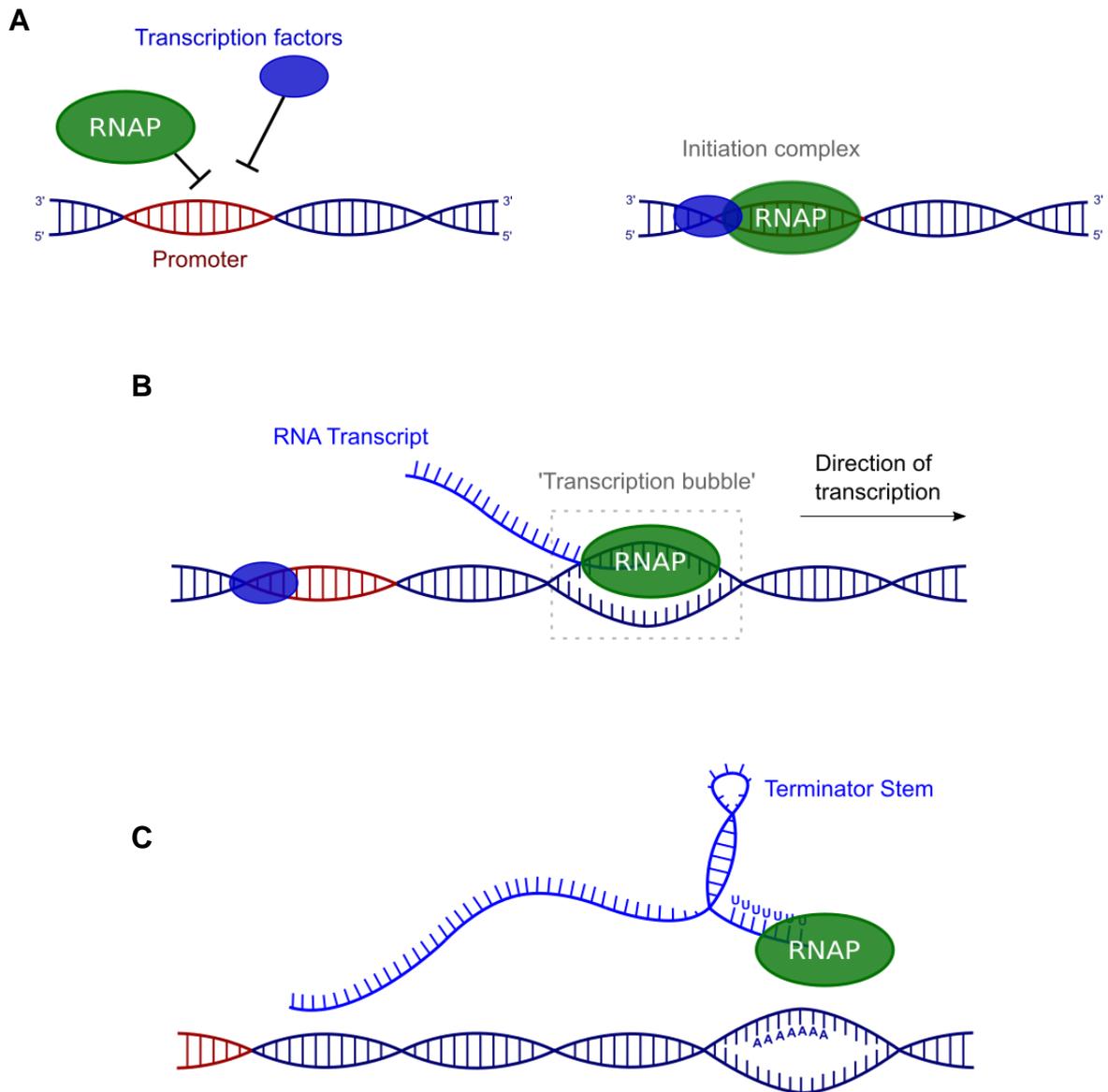


Figure 4 – Overview of the three stages of transcription in *E. coli*. A) Initiation – RNA polymerase (RNAP) binds to the DNA promoter region, facilitated by transcription factors, to form an initiation complex; B) Elongation – procession of the transcription complex along the DNA template catalyses the formation of the RNA transcript; C) Termination (intrinsic shown) – formation of an RNA terminator stem loop succeeded by a poly-uridylate tract destabilises the transcription complex resulting in dissociation of the RNAP from the template strand.

The timing of these events is dependent upon the DNA sequence, interactions with regulatory proteins, and the concentration of free NTPs in solution²¹. Regulation of transcription at the elongation stage is essentially ‘intra-operon’ in that only the control of the relative rates of elongation and termination determines the rate at which synthesis of the specific transcript is completed¹⁵. Following elongation and formation of the RNA transcript, the final step of the process is transcription termination (**Figure**

3C), occurring when the transcription complex encounters a terminator sequence on the DNA template. In *E. coli*, transcription termination is termed either 'intrinsic', whereby termination occurs without the recruitment of additional protein factors, or 'rho-dependent', whereby the transcription termination factor, rho, is required for release of the complex from the template¹¹. Intrinsic terminators account for approximately 50% of termination sites in *E. coli* and are characterised by a DNA sequence which encodes a stable RNA terminator stem-loop followed by a poly-uridylate region at the 3' end of the transcript¹¹. The sequence of dA-rU base-pairs forming the DNA-RNA hybrid structure is particularly unstable relative to its DNA-DNA and RNA-RNA counterparts²² and this effect combined with the formation of the RNA terminator stem causes the release of the transcription complex from the DNA template.

1.3 Translation: Ribosome structure, mechanisms and regulation

Translation is the process by which genetic information contained within mRNA is converted into a functional protein. The process of translation is catalysed by the ribosome, a composite RNA-protein complex, and (in a similar fashion to transcription) is categorised into three phases: initiation, elongation and termination. Structurally, the ribosome contains three key domains: the A site, which serves as the access point for aminoacylated tRNA molecules; the P site, where the amino acid chain is elongated and the E site, through which the de-aminoacylated tRNAs are able to vacate the ribosomal complex. There are a number of differences in the mechanisms which underpin translation in prokaryotes and eukaryotes, with the eukaryotic systems generally being more complex in nature, and this overview will focus on prokaryotic translational mechanisms. In *E. coli*, the initiation phase begins with interaction between the mRNA Shine-Dalgarno (SD) sequence (located approximately 6-9 nucleotides upstream of the start codon) and the corresponding anti-SD sequence at the 3' terminus of the 16S rRNA²³. Following association of the 30S ribosomal subunit, binding of three initiation factors (IF1, IF2, and IF3) facilitates start codon accommodation and enhances the fidelity of the initiation process. The step is completed by association of the larger 50S ribosomal subunit and the binding of the first aminoacylated tRNA within the ribosomal P site. Initiation is generally considered to be the rate-limiting step in protein synthesis and there are a number of regulatory mechanisms involved at the initiation step²³. In prokaryotes, this is typically mediated by blocking access to the SD sequence. This primarily occurs via sequestration of the mRNA ribosome binding site, either by base-pairing interactions within the same mRNA strand, interaction with a small *trans*-acting RNA, or with the assistance of mRNA-binding proteins²⁴. During elongation, the ribosome catalyses peptide bond formation between the amino acids covalently attached to tRNAs. The initiator tRNA bound to the P site during the initiation phase is positioned in close proximity to a second aminoacylated tRNA bound at the ribosomal A site, facilitating peptide bond formation²⁵. The de-aminoacylated tRNA is shuffled from the P site to the E site before vacating the ribosome, whilst the tRNA carrying the growing peptide chain is transferred from the A site to the P site, ready for the next round of elongation. Termination of translation occurs when the ribosomal A site encounters a stop codon, which is one of three possible sequences (UAA, UAG and UGA) and is recognised by two independent release factors (termed RF1 and RF2), binding of which to the ribosomal P site initiates hydrolysis and release of the peptide chain from the P site tRNA²³.

1.4 Riboswitches

1.4.1 Riboswitches overview

One common and crucial feature of all bacteria is the ability to continuously modulate gene expression in response to both chemical and physical fluctuations in their environment. This allows bacterial metabolism to be optimised, avoiding inappropriate physiological responses and unnecessary expenditure of resources²⁶. Historically, this has been thought to be achieved through the use of protein factors, however in the last decade it has become clear that a significant proportion of bacterial genetic regulation is performed by non-coding RNAs. In particular, the *cis*-acting regulatory elements located in the 5'-UTRs of bacterial mRNAs, known as riboswitches, are widespread throughout prokaryotic phylogeny. Indeed, they control ~4% of the genes in *Bacillus subtilis* alone²⁷. Riboswitches respond to physiological stimuli without the aid of protein factors, hence allowing mRNA transcripts to govern their own expression. Whilst there are some examples of riboswitches which respond to physical changes within the cell²⁸, they most commonly respond to changes in cellular concentration of specific metabolites and either induce or repress expression of the downstream gene accordingly. Moreover, the genes under riboswitch control are typically involved in the biosynthesis or transport of the aforementioned metabolite; hence these genetic switches serve to attenuate cellular ligand concentrations when they are perturbed from desirable levels.

Given their structural and mechanistic simplicity, as well as their widespread distribution across the bacterial kingdom, it is possible that modern riboswitches may be descendants of ancient regulatory mechanisms that evolved before the advent of protein based enzymes and genetic factors²⁹. If that is the case, then the fact that riboswitches have persisted in the organisms of today indicates that they must represent an extremely efficient method for regulation of gene expression. Compared with protein-based gene expression systems, riboswitches are primitive and are unable to achieve a paralleled level of sophistication with regard to genetic decision making and it may be for this reason that riboswitches have been super-seeded by protein based regulatory systems in eukaryotic organisms. It is likely that riboswitches are still found in many prokaryotes because selective pressures have led to these organisms retaining a more stream-lined genome.

1.4.2 Riboswitch structure

Riboswitches typically consist of two domains: an aptamer domain, which binds the cognate ligand; and an expression platform which undergoes a conformational change on binding that directly affects expression of the downstream gene. The evolutionary constraints which shape these two domains are significantly different. Components of expression platforms almost always require particular secondary structure elements but show very low sequence conservation across bacterial species³⁰. In contrast, aptamer domains are highly conserved, with each class exhibiting a distinctive arrangement of secondary structures with well conserved residues involved in tertiary interactions as well as ligand binding³⁰. Indeed, riboswitch aptamer domains are likely to represent evolutionarily optimum structures given the significant penalties which would ensue should a mutation arise in one of the key binding-site residues. In order to selectively bind target metabolites amongst the huge variety of small-molecules in the cellular environment, aptamer domains have evolved to interact with a diverse range of ligand functional groups. Moreover, a range of ribonucleotide moieties have been shown to be involved in ligand binding include the Watson-Crick, Hoogsteen and sugar-edges of the nucleobase, as well as the ribose and phosphate groups.

1.4.3 Riboswitch classes and ligands

Riboswitches have evolved to respond to a wide variety of cellular metabolites. To date, at least twenty classes of riboswitches have been identified that recognise a broad range of ligands including: vitamin co-factors³¹; purine nucleobases³²; amino acids^{33, 34}; amino sugars³⁵; metal ions³⁶; enzyme co-factors³⁷ and halogen ions³⁸. An overview of the known riboswitch classes is shown in **Table 1**. The metabolites which are known to activate riboswitches are chemically diverse and do not contain a common recognition functionality. Unsurprisingly, many riboswitch ligands contain nucleobases or moieties which mimic the Watson-Crick face of purines and pyrimidines. Examples include the adenine³⁹ and guanine⁴⁰ responsive purine riboswitches, the thiaminepyrophosphate (TPP) riboswitch⁴¹ and the SAM-sensing riboswitches^{42, 43, 44}. However, riboswitch aptamer domains have also evolved to bind amino acids such as glycine³³ and lysine⁴⁵ as well as magnesium ions⁴⁶. The majority of ligands rely on heteroatoms to form hydrogen bonds and participate in electrostatic interactions with the RNA. Typical hydrogen bonding interactions occur between ligand edges and conserved unpaired residues within the riboswitch aptamer domain in the

same plane⁴⁷. Moreover, aromatic ligand moieties are often engaged in stacking interactions between adjacent RNA nucleobases. Negatively charged ligand functionalities, such as the carboxylate group of amino acids, can be accommodated by riboswitch aptamers by interaction via metal cations such as Mg²⁺ and K⁺. Somewhat surprisingly, riboswitches have also evolved to sense intracellular fluoride ions³⁸, the interaction between which is mediated by both Mg²⁺ and water molecules.

Table 1 – Overview of known riboswitch classes including cognate ligand and original discovery paper

Ligand Category	Riboswitch Class	Effector Ligand(s)	Year of Discovery
Nucleosides/Nucleobases	Purine/G	Guanine	2003 (Breaker <i>et al.</i>) ⁴⁰
	Purine/A	Adenine	2004 (Breaker <i>et al.</i>) ⁴⁸
	Purine/dG	2'-Deoxyguanosine	2007 (Breaker <i>et al.</i>) ⁴⁹
Co-factors	TPP	Thiamine pyrophosphate	2001 (Soberon <i>et al.</i>) ⁵⁰
	FMN	Flavin mononucleotide	2002 (Breaker <i>et al.</i>) ⁵¹
	AdoCbl	Adenosylcobalamin	2002 (Breaker <i>et al.</i>) ⁵²
	SAM (I-V)	S-adenosylmethionine (SAM)	2003 (SAM-I, Breaker <i>et al.</i>) ⁵³
			2005 (SAM-II, Breaker <i>et al.</i>) ⁴³
2006 (SAM-III, Henkin <i>et al.</i>) ⁵⁴			
2008 (SAM-IV, Breaker <i>et al.</i>) ⁵⁵			
2009 (SAM-V, Breaker <i>et al.</i>) ⁵⁶			
THF	Tetrahydrofolic acid	2010 (Breaker <i>et al.</i>) ⁵⁷	
Amino acids	Glycine	Glycine	2004 (Breaker <i>et al.</i>) ⁵⁸
	Lysine	Lysine	2003 (Breaker <i>et al.</i>) ^{59, 60}
	Glutamine	L-Glutamine	2011 (Breaker <i>et al.</i>) ³⁴
Metal/halogen ions	Mg ²⁺ /ykoK	Magnesium	2004 (Breaker <i>et al.</i>) ⁶¹
	F ⁻	Flouride	2012 (Breaker <i>et al.</i>) ³⁸
Bacterial second messengers	c-di-GMP-II	c-di-GMP	2010 (Breaker <i>et al.</i>) ⁶²
	c-di-AMP	ci-di-AMP	2013 (Breaker <i>et al.</i>) ⁶³

1.4.4 Mechanisms of riboswitch-mediated control of gene expression

1.4.4.1 Overview of transcriptional and translational riboswitch mechanisms

Riboswitch expression platforms can typically adopt two mutually exclusive secondary structures, which are inter-converted as a result of ligand binding/dissociation in the aptamer domain⁶⁴. The nature of the expression platform determines the mechanism by which the riboswitch achieves its genetic control. Riboswitches function at both the transcriptional and translational level.

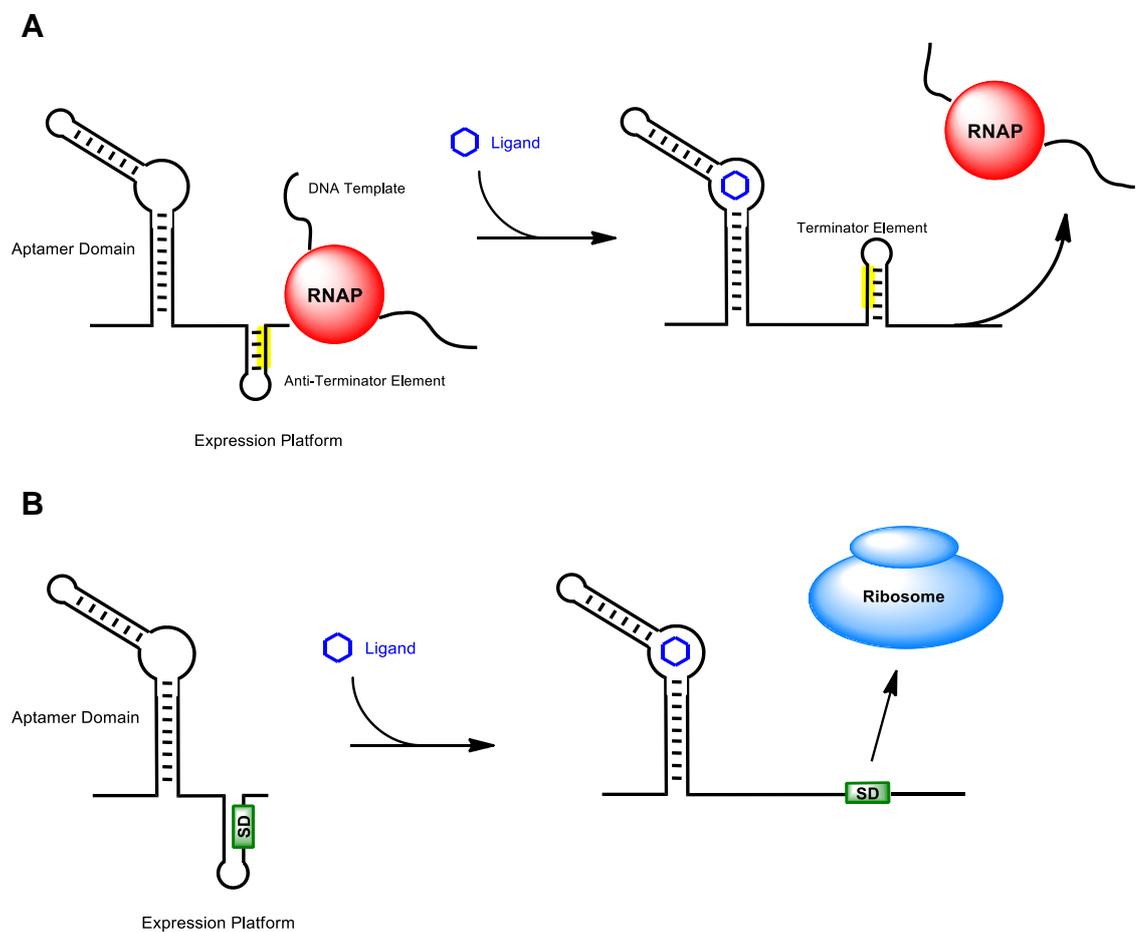


Figure 5 – General mechanisms of riboswitch mediated control of gene expression. A) Transcriptional control. In this example, a ligand binding event induces the formation of a downstream terminator stem in the riboswitch expression platform causing the RNAP to disengage. **B) Translational control.** In this example, a ligand binding event causes the release of the Shine Dalgarno sequence from a stem loop, allowing translation initiation to occur.

Transcriptional riboswitches operate *via* the formation of a terminator stem-loop in the expression platform following either ligand binding or dissociation in the aptamer domain. This terminator element causes the native RNA polymerase to abort transcription and hence expression of the downstream gene is repressed³⁰. In the case of 'ON' transcriptional riboswitches, the ligand-unbound state favours terminator stem formation and ligand binding initiates the sequestration of key terminator stem residues into an anti-terminator element. Conversely, transcriptional-'OFF' riboswitches favour anti-terminator formation in the absence of ligand, and ligand binding results in terminator stem formation and consequent abortion of transcription (**Figure 4A**). This type of genetic control is observed in a number of different riboswitch classes including the PreQ₁⁶⁵, SAM-I⁶⁶ and lysine⁶⁷, riboswitches.

Translational riboswitches function *via* the sequestration or release of the mRNA Shine-Dalgarno sequence (ribosomal binding site) on ligand binding. In the case of translational-'ON' systems, the Shine-Dalgarno (SD) sequence is sequestered in the ligand-unbound conformation, preventing the mRNA transcript from binding the ribosome and hence inhibiting translation (**Figure 4B**). In the ligand-bound state, the SD sequence is released, allowing ribosomal binding to occur and hence expression of the downstream gene is up-regulated³⁰. Examples of riboswitches which operate at the translational level are the adenine³² and SAM-II⁶⁸ riboswitches.

1.4.4.2 Kinetic and thermodynamic control of riboswitch-mediated gene expression

Transcriptional riboswitches are able to influence genetic decisions during the timeframe of their own transcription by the endogenous polymerase. It is therefore crucial that any ligand binding events and subsequent changes in secondary structure occur with sufficient rapidity to influence polymerase activity; transcriptional riboswitches therefore operate under stringent kinetic constraints³⁰. The kinetics of riboswitch folding and ligand recognition has attracted a lot interest, with methods such as single molecule force spectroscopy⁶⁹, fluorescence spectroscopy⁷⁰ and spin-labelling⁷¹ used to probe the time-scales on which transcriptional riboswitches operate. In particular, a study by Block *et al.* used single molecule force spectroscopy to assess the kinetics of RNA folding and ligand binding in the *pbuE* adenine riboswitch from *Bacillus subtilis*. Rates were determined by fully extending the RNA under high force

and then determining the probability of obtaining fully re-folded RNA as a function of time for varying concentrations of adenine⁷². This yielded time constants of: 3s for RNA folding; 4 s for ligand association (at 3 μ M adenine); and ~8 s for ligand dissociation. Since bacterial polymerases have average elongation rates of around 50 nt s⁻¹⁷³, RNA folding and ligand binding occur on a similar time-scale to transcription of the riboswitch by the endogenous polymerase. Hence, it seems that the rate of folding and ligand binding are more important in determining the overall genetic decision than the equilibrium binding constant of the ligand for the aptamer domain. Riboswitches which operate in this way are said to be under 'kinetic control'.

This is exemplified by the observation that ligand concentrations much higher than the apparent dissociation constant, K_D , are needed to effect half-maximal regulatory response (T_{50}) *in vitro* for many classes of transcriptional riboswitch^{45, 66, 70}. Furthermore, the idea of kinetic control offers an explanation as to how bacteria have evolved riboswitches with what appear to be excessively high affinities²⁹. In *E. coli*, a concentration of a single molecule per cell corresponds to values of around the low nanomolar range. However, there are examples of riboswitches which exhibit apparent dissociation constants in the picomolar range^{74, 75, 76}, hence are able to detect metabolites with at least two orders of magnitude greater sensitivity than required for concentrations of only one molecule per cell. It seems therefore, that these transcriptional switches do not reach equilibrium with their cognate ligand, and instead rely on the kinetics of ligand binding to drive the regulation of gene expression. This mechanism could potentially be exploited for the purposes of engineering novel riboswitch based gene control elements. It may be possible to reduce the concentration of ligand required to activate a kinetically driven riboswitch by increasing the length of the linker-region of nucleotides which separate the aptamer from the expression platform. Potentially, this would increase the window of opportunity for the ligand binding event before the RNA polymerase moves beyond the influence of the riboswitch²⁹. In contrast, translational riboswitches rely on the thermodynamics of ligand binding to achieve control of gene expression. Given that the timeframe for activation of the riboswitch switching mechanism is not constrained by the kinetics of transcription, the aptamer domain has sufficient time to reach thermodynamic equilibrium with its cognate ligand. Thus, *in vitro* binding constants obtained for translational riboswitches are generally a more accurate reflection of the cellular ligand concentrations required for their activation.

1.4.5 The adenine-responsive add A-riboswitch

There are a number of different riboswitches, from a variety of organisms that have evolved to modulate gene expression in response to purines^{40, 77}. The *add* A-riboswitch is an adenine-binding riboswitch found in the 5'-UTR of adenine deaminase (*add*) in *Vibrio vulnificus*³⁹. The riboswitch functions via a translational-‘ON’ mechanism, whereby binding of adenine to the aptamer domain induces a conformational change in the expression platform such that the SD sequence is released from sequestration in a stem loop. Thus, when the cellular adenine concentration is elevated above desirable levels, the *add* A-riboswitch induces the expression of a gene responsible for metabolising adenine and the endogenous concentration decreases.

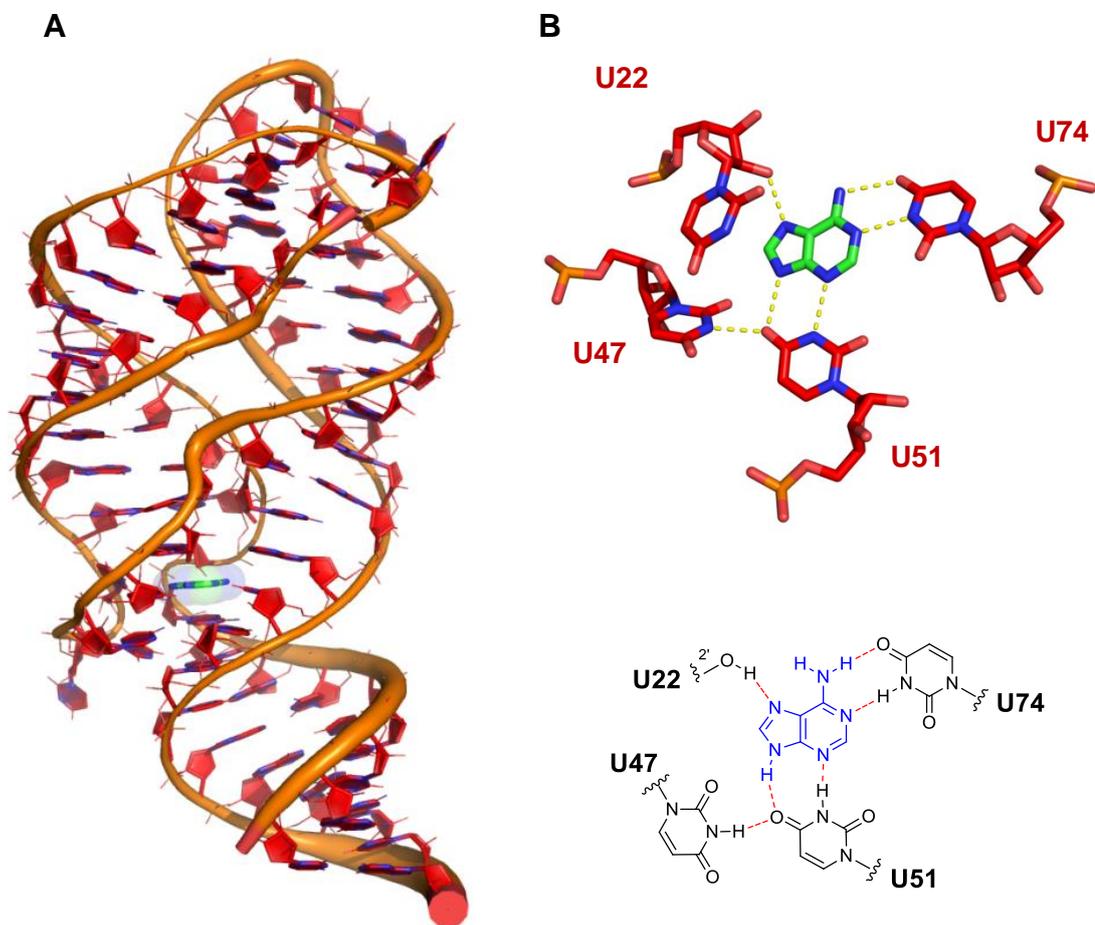


Figure 6 – X-ray crystallography structure of the *add* A-riboswitch from *Vibrio vulnificus*³². A) Structure of the *add* A aptamer domain showing overall tertiary architecture; B) Structure of the *add* A ligand binding site showing adenine recognition by four unpaired uridine residues.

The x-ray crystal structure of the *add* A-riboswitch aptamer domain was solved by Serganov in 2004³² and is shown in **Figure 5**.

The *add A*-riboswitch aptamer domain has a classical ‘tuning-fork’ structure, containing a ligand binding site located at the junction of three stem loops, with a pseudoknot interaction between loops L2 and L3 governing the overall architecture. The ligand binding site consists of four unpaired uridine residues which form a hydrogen bonding network with the ligand, providing the basis of the recognition motif (**Figure 5B**). The Watson-Crick interaction between adenine and U74 provides the basis for the differentiation of adenine from cellular guanine, with homologous guanine-responsive purine riboswitches possessing a cytidine residue at this position³². Moreover, it has also been shown that the U74C mutation confers guanine binding ability to the *add A*-riboswitch⁷⁸. The relative simplicity of the ligand binding motif, and the thorough extent of characterisation both *in vitro* and *in vivo*, has presented the *add A*-riboswitch as an attractive target in the development of novel gene expression systems^{79, 80, 81}.

1.4.6 The preQ riboswitch

The cellular metabolites preQ₁ and preQ₀ (collectively: preQ) are intermediates in the biosynthesis of queuosine, a hypermodified nucleoside used in certain bacterial tRNAs⁸². As such, a number of bacteria have evolved preQ-sensing riboswitches to regulate the cellular concentration of these metabolites. The preQ riboswitch is found to control expression of the *B. subtilis queCDEF* operon, which encodes a series of genes involved in queuosine biosynthesis⁸³. The riboswitch functions as a transcriptional-‘OFF’ switch, whereby binding of preQ₁ or preQ₀ to the aptamer domain induces the formation of a downstream transcriptional terminator stem causing the RNA polymerase to disengage⁸⁴. The x-ray crystallography structure of the preQ riboswitch has revealed insights into the structure of the aptamer domain and the ligand recognition motifs utilised to selectively distinguish the aforementioned metabolites from the cellular milieu⁶⁵. An overview is shown in **Figure 6**. The preQ aptamer domain is the smallest reported in the literature comprising a mere 34 nt. The aptamer secondary structure constitutes a single stem, capped by a loop within which lies the ligand binding site. A tertiary pseudoknot interaction between nucleotides beyond the 3’ end of the P1 stem and unpaired residues in the loop defines the overall riboswitch architecture, and is also the primary feature of the switching mechanism⁸⁴.

The ligand binding motif is shown in **Figure 6B**. Interestingly, all four types of ribonucleotide are involved in ligand recognition.

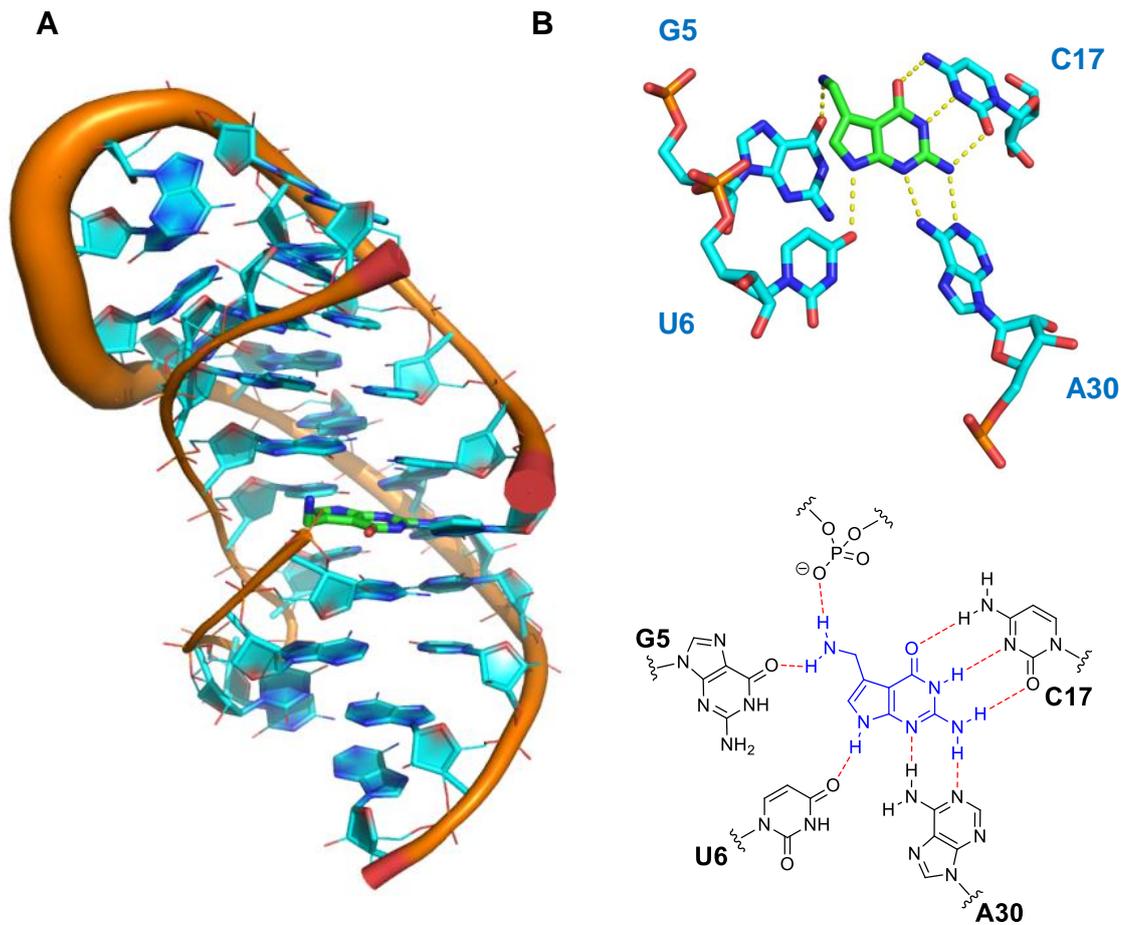


Figure 7 – X-ray crystallography structure of the PreQ class I riboswitch from *Bacillus subtilis*⁶⁵. A) Structure of the PreQ aptamer domain showing tertiary architecture; B) structure of the ligand binding site showing the PreQ₁ binding motif.

As with the majority of riboswitch aptamer domains which have evolved to bind purines and their derivatives, a key feature in the binding of preQ₁ is a Watson-Crick interaction between the ligand and an unpaired pyrimidine within the ligand binding site. In this case, the C17 residue performs this function. Interestingly preQ₁ also forms a hydrogen bond with an adjacent phosphate residue, which demonstrates the breadth of RNA moieties which can contribute to ligand recognition and may also provide the basis for distinguishing preQ₁ from cellular guanine⁸⁵. As with the *add* A-riboswitch, the preQ riboswitch has also been the subject of reengineering by the Micklefield lab⁸⁶, and represents the first example of an ‘OFF’ switch rationally modified in this way.

1.4.7 Tandem riboswitch architectures

The majority of riboswitches occur in isolation, whereby a single metabolite binds a single aptamer domain and elicits a gene regulatory response. However, there are examples of multiple aptamer domains controlling expression of a single gene, allowing for the execution of more sophisticated regulatory decisions. These tandem riboswitch constructs can be two or more aptamers with identical ligand specificities – allowing for enhanced sensitivity to the cellular ligand pool – or they can have differing specificities allowing multiple chemical inputs for a single regulatory output⁷⁶. The most ubiquitous tandem riboswitches are those that bind the amino acid glycine³³. Tandem glycine riboswitches typically contain two aptamer domains with high sequence and secondary structure homology, arranged consecutively and separated by a linker region. Crucially, the regulatory response of both aptamers is effected by a single, downstream expression platform. Moreover, ligand recognition is cooperative – the binding of glycine to one aptamer influences the binding to the other – and consequently the scope of regulatory effects can be provoked over a tighter range of ligand concentrations³³. This more sophisticated form of riboswitch mediated genetic control is likely to have arisen to maintain sufficient cellular glycine levels for protein synthesis.

Another intriguing example of tandem riboswitch architectures is the SAM-II/SAM-V construct, which is the only known example of two different riboswitch classes which respond to the same metabolite being found in a tandem arrangement. The riboswitch, identified from marine metagenomic sequence data, consists of two S-adenosylmethionine-responsive riboswitch subunits, fused together in a sequential arrangement⁵⁶. The 5' end contains the transcriptional-‘OFF’ switch, SAM-II, which is upstream of a translational-‘OFF’ SAM-V riboswitch, allowing the organism control of gene expression at both the transcriptional and translational level. Hence, at the point of translation initiation, the full length mRNA transcript will only be synthesised at low SAM concentrations, since activation of the SAM-II portion of the tandem construct will prematurely terminate transcription. However, if during the lifetime of the mRNA within the cell the SAM concentration increases, then the SAM-V substructure of the riboswitch would respond by sequestering the SD sequence and preventing translation initiation⁵⁶.

Tandem riboswitches containing two or more aptamer domains with different ligand specificities are much less common. The most prominent example of this type of motif is found in *Bacillus clausii*, where two independent riboswitches selective for S-

adenosylmethionine (SAM) and coenzyme B₁₂ (AdoCbl) control the expression of *metE*. This gene encodes an enzyme which catalyses the SAM dependent formation of methionine from homocysteine⁸⁷. Each aptamer has its own expression platform, which forms a transcriptional terminator element in the ligand-bound state; hence the two ligands can independently repress *metE* expression. In this sense, the tandem motif functions as a natural Boolean NOR gate, whereby either of the two chemical inputs yields a single output of gene repression. *Bacillus clausii* also employs another enzyme for the methylation of homocysteine known as MetH, which is more efficient than MetE and utilises the co-enzyme methylcobalamin, a derivative of AdoCbl⁸⁸. MetE activity is therefore only important when the cellular AdoCbl pool is depleted, offering an explanation as to why this less efficient route to methionine biosynthesis is repressed when the cellular concentrations of either AdoCbl or SAM are high³⁰. These examples illustrate the ways in which RNA elements have evolved to make sophisticated genetic decisions without the need of protein factors.

1.5 Existing Small-Molecule Inducible Gene Expression Systems

1.5.1 Small-molecule inducible repressor/operator systems

There are a number of examples of gene expression tools which have been created over the last 40 years for the selective control of heterologous gene expression. The majority of these devices have been constructed in the context of maximising the production of a protein of interest. Indeed, the ability to insert a specific heterologous gene under the control of a native promoter with the goal of producing large quantities of the desired protein has been one of the major driving forces behind the modern biotechnology industry⁸⁹.

In the context of maximising titres in the production of heterologous proteins, the ability to control the timing of the initiation of the desired protein over-expression is also of importance. The significant metabolic flux as a result of the excessive production of heterologous proteins can be detrimental to the host cell and it is desirable to allow the cell culture to reach optimum density before over-expression commences⁹⁰. This temporal control of gene expression in bacteria has been achieved by utilising small molecule-inducible native promoter/repressor systems such as the *lac*, *tet* and *ara* systems⁹¹, with the *E. coli lac* system probably being the most widely utilised in research laboratories^{92, 93}.

1.5.2 The *lac* operon

The *lac* operon contains the genes *lacZ*, *Y* and *A*, with *lacZ* encoding the enzyme β -galactosidase. This enzyme is responsible for the hydrolysis of the di-saccharide lactose into glucose and galactose as well as the production of allolactose – the natural inducer of the *lac* operon. The gene *lacY* encodes a lactose transport protein, responsible for uptake of lactose into the cell⁹⁴, whilst *lacA* encodes a galactoside acetyltransferase with a putative role in detoxification⁹⁵. The gene *lacI* encodes the lactose repressor protein which selectively binds the *lac* operator region thereby inhibiting transcription of the *lac* operon. This mechanism of genetic regulation has been exploited to create an expression system which is responsive to the addition of a synthetic small-molecule analogue of the natural inducer, allolactose. An overview of the mechanistic basis of the expression system is shown in **Figure 7A**. The

heterologous gene of interest is cloned downstream of the *lac* promoter in a suitable expression vector. During the cell culture growth phase (and in the absence of any inducer ligands), LacI blockades transcription of the gene of interest, thereby preventing the placement of deleterious metabolic burden on the host strain. When the cell culture has reached optimum density, isopropyl- β -D-1-thiogalactopyranoside (IPTG) is added to the growth media to induce gene expression. IPTG binds to LacI, inducing conformational changes in the protein causing it to release from the DNA template, thereby allowing transcription of the gene of interest and expression is induced.

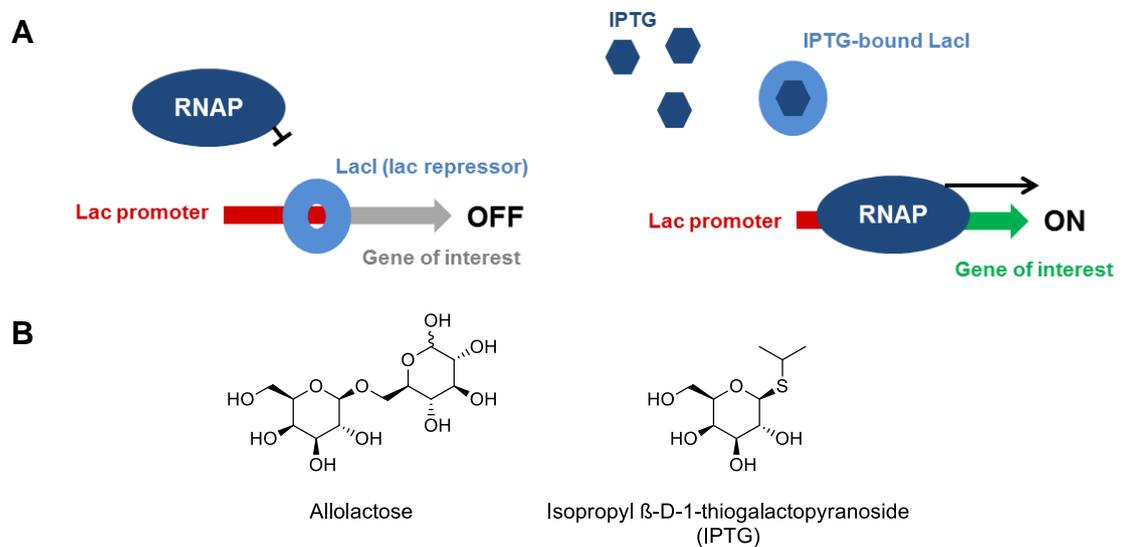


Figure 8 – Overview of the *lac* repressor system. A) Mechanism of genetic regulation by *lacI*. Left: the *lac* repressor (LacI) binds the operator region thereby blockading transcription. Right: IPTG binds to the *lac* repressor, inducing a conformational change which releases LacI from the DNA template, allowing transcription by the endogenous RNA polymerase (RNAP). B) Structures of the natural inducer of the *lac* operon, allolactose (left), and the synthetic analogue isopropyl- β -D-1-thiogalactopyranoside (IPTG; right).

The *lac* system has been widely utilised to maximise the production of therapeutic proteins in the biotechnology industry, as well as finding widespread application in biochemistry research laboratories. Indeed, the system is very well suited to the over-expression of one or two proteins for further characterisation. However, in the context of metabolic engineering and the manipulation of entire gene networks to maximise the production of a small-molecule product, the *lac* system is much less appropriate, as the fine tuning of expression levels is not achievable⁸⁹.

1.5.3 The ‘all-or-nothing’ phenomenon: The *ara* system

Using the related arabinose system (*ara*), it is possible to tune the gene expression output as a function of ligand concentration. However, there are fundamental problems with utilising the arabinose system in this context. In the engineering of metabolic pathways, it is desirable to have small-molecule responsive expression systems which induce gene expression across the entire bacterial population in a homogenous manner. Some of the most advantageous inducible promoters (including the *ara* and *lac* systems) display ‘all-or-nothing’ expression, whereby the number of cells within the culture which are induced varies as a function of inducer concentration. Hence, at the population level, it can appear that the gene expression output varies with inducer ligand concentration, but this is actually a consequence of a heterogeneous mixture of cells exhibiting either ‘ON’ or ‘OFF’ expression outputs (see **Figure 8**)⁸⁹. There are a number of problems with this. The ‘all-or-none’ effect could have detrimental consequences for the host, as cells expressing genes at maximal level would be face a more substantial metabolic burden – and hence grow more slowly – than uninduced cells. This in turn has consequences for product homogeneity, especially in the context of engineering primary metabolic pathways^{89, 96}.

1.5.4 Summary of the limitations of current small-molecule inducible expression systems

The existing small-molecule inducible expression systems have found widespread utility across recombinant protein production and gene functional analysis⁹⁰, however they have a number of limitations. The lack of titratability and exact control over expression outputs, as well as a tendency to promote population heterogeneity makes them unsuitable for a number of applications. In particular, in the engineering of bacterial biosynthetic pathways for the production of desired small-molecules, the ability to finely tune the expression of the biosynthetic genes is of paramount importance⁹⁶. Existing expression systems such as the *lac* and *ara* systems are very effective at maximising expression of heterologous proteins, but are poor choices when more precise control of expression output is required. As such, over the last decade, there has been significant effort to develop new expression systems which allow dose-dependent modulation of gene expression in response to exogenous small-molecules. In particular, riboswitches have emerged as prime engineering candidates⁹⁷.

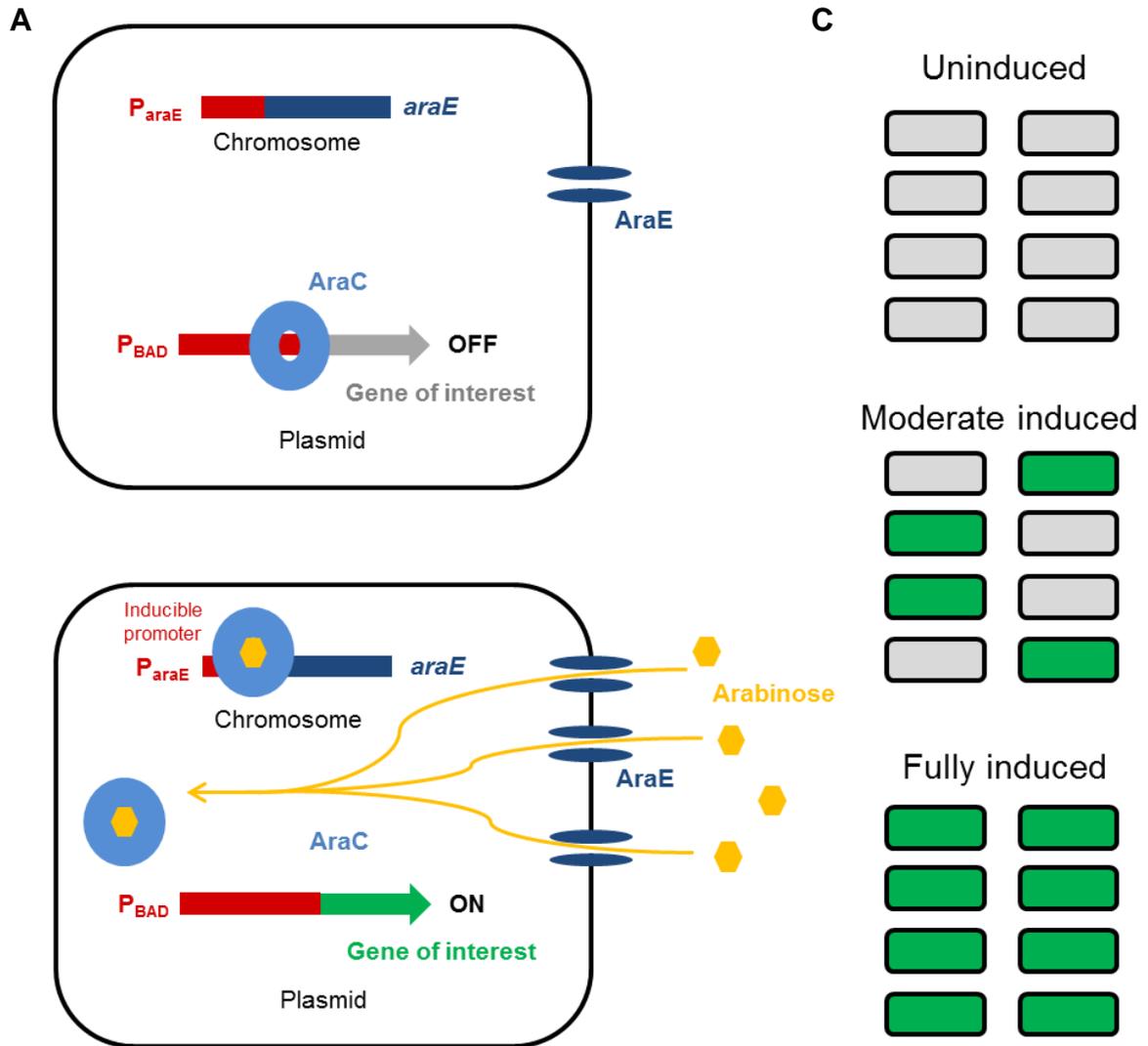


Figure 9 – Overview of the ‘all-or-nothing’ effect using the example of the *ara* system (modified from Kealsing *et al*⁸⁹). A) In the absence of arabinose the P_{BAD} promoter the expression construct is repressed by the regulatory protein AraC, blocking transcription of the gene of interest. B) Addition of arabinose allows the compound to bind AraC to de-repress P_{BAD} , as well as inducing the expression of the arabinose transporter *araE* from the genomic DNA. C) Representation of the relationship between expression output observed at the population level and the expression output of individual cells.

1.6 Engineered riboswitches as tools for synthetic biology

Compared with promoter/repressor-based expression systems, the mechanisms by which riboswitches exert their genetic control are relatively simple. Riboswitches are able to respond to the binding of small molecules to dose-dependently modulate gene expression, in a protein-free manner, utilising a region of non-coding RNA consisting of only around 50-200 nucleotides. This makes them particularly attractive targets for re-engineering⁷⁹ and the methodologies deployed to achieve this have been diverse. Strategies include the *in vitro* selection of aptamer domains for a ligand of choice^{98, 99}, and the combining of these aptamer domains with native expression platforms¹⁰⁰. Other approaches to proceed from *in vitro* aptamer to functional riboswitch have involved high throughput screening of large pools of RNAs containing randomised sequences, with the goal of identifying a functional expression platform¹⁰¹. The performance of riboswitches as gene expression tools *in vivo* has been optimised by rationally strengthening and weakening particular secondary structure elements, thus modulating the riboswitch ON/OFF state equilibrium¹⁰². Additionally, it has also been shown that some natural riboswitch domains are modular in nature, and that aptamers and expression platforms from distinct organisms and riboswitch classes can be combined to produce chimeric systems with new ligand specificities and mechanisms of action^{100, 103, 104}. Finally, the Micklefield laboratory has shown that it is possible to re-engineer naturally occurring riboswitches, such that they no longer respond to their cognate metabolite but are instead activated by the addition of a non-natural ligand not ordinarily present within the cell^{79, 81, 86}.

1.6.1 Generating RNA aptamers in vitro: SELEX

The generation of RNA aptamers *in vitro* has been explored extensively since the first reports of the successful application of this methodology by Ellington *et al.*⁹⁸ and Tuerk *et al.*⁹⁹ in 1990. The Systematic Evolution of Ligands by Exponential Enrichment (SELEX) was originally used to generate RNA aptamers for proteins and small-molecule organic dyes, but has since been shown to have broad application in the generation of high-affinity binding domains for a variety of small molecules, cells and microorganisms¹⁰⁵. An overview of a typical SELEX procedure is shown in **Figure 9**. The procedure utilises a library of DNA oligonucleotide templates followed by transcription and multiple rounds of selection to generate RNA aptamers which bind the

ligand of choice. The chemically synthesised ssDNA template library contains up to 60 nucleotides in a randomised region, flanked by short constant regions containing a sequence to which primers are annealed during the subsequent PCR step, as well as the appropriate promoter for the initiation of transcription (typically the T7 bacteriophage RNA polymerase is used, requiring the corresponding T7 promoter region). The RNA sequences are generated by *in vitro* transcription and purified, most commonly by urea PAGE, to ensure strand length homogeneity. This procedure generates approximately $10^{12} - 10^{15}$ unique RNA sequences with approximately 40-100 copies of each RNA oligonucleotide¹⁰⁵. The RNA library is then counter selected against the target ligand which has been immobilised on a matrix, and for this step affinity chromatography is most often the procedure of choice.

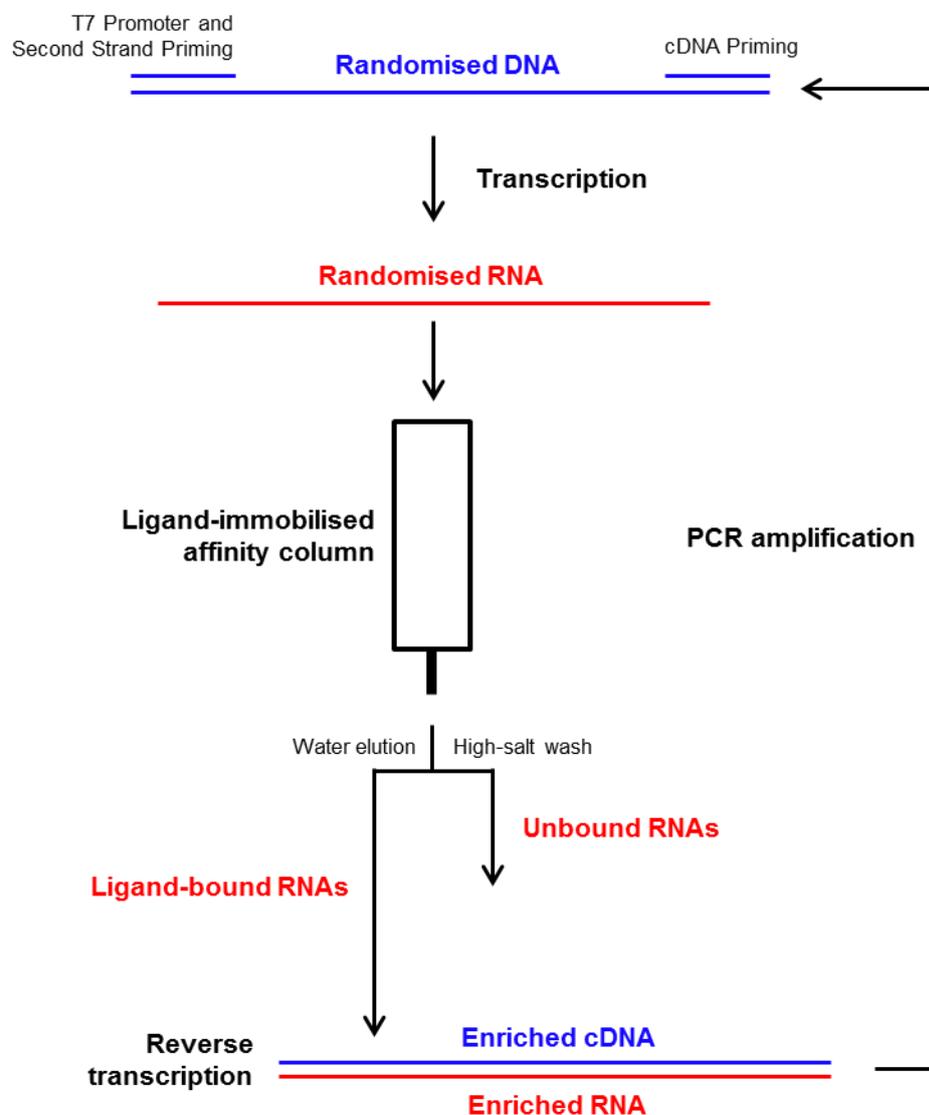


Figure 10 – Overview of a typical SELEX procedure (redrawn from Ellington *et al.*⁹⁸)

Using a buffer with a high salt concentration, the RNA library is applied to an affinity column onto which the ligand of choice has been immobilised. Non-binding species are removed by washing with high-salt buffer before elution of the ligand-bound RNAs using water. The water-eluted RNAs are then reverse transcribed using reverse transcriptase to generate the corresponding cDNAs which are then PCR amplified. The procedure is then repeated by treating the DNA with T7-RNA polymerase to generate an enriched RNA library which can be subjected to further rounds of selection. The cDNA of the ultimate enriched RNA pool is then finally cloned and sequenced. Whilst the SELEX procedure has been fruitful in the production of a large number of RNA aptamers which can bind a variety of small molecule (and biomolecular) ligands¹⁰⁶, the use of SELEX-derived aptamers as the basis for functional riboswitches is relatively limited. Of the numerous small-molecule binding SELEX aptamers reported in the literature, only a handful have found applications as riboswitches. The few examples include the theophylline¹⁰⁷, tryptophan¹⁰⁸, and 2,4-dinitrotoluene¹⁰⁹ SELEX-derived aptamers. One possible reason why so few of the known examples have been shown to function effectively *in vivo* is that the ligand binding sites of *in vitro*-derived aptamers are typically on the surface of the RNA. Consequently, the scope of ligand functional groups which are involved in RNA binding is usually relatively narrow and this could explain the poor ligand specificity *in vivo*. Moreover, whilst the modularity of natural riboswitch expression platforms has been demonstrated^{100, 110}, their ability to combine with SELEX-derived aptamer domains has thus far been limited.

1.6.2 Exploitation of riboswitch domain modularity to create chimeric expression systems

The modularity of riboswitch aptamer domains and expression platforms has been demonstrated by Batey *et al.*, who have created a series of chimeric riboswitch constructs, responsive to a variety of small-molecules. These gene expression devices are fusions of transcriptional expression platforms with a variety of natural and SELEX-derived aptamer domains, responsive to a number of ligands including lysine, FMN and theophylline¹⁰⁰ and folinic acid¹⁰⁴. The group demonstrated that this methodology does not require the selection of device-specific linker sequences to allow the transmission of the ligand binding event to the expression platform. This allowed the rapid engineering of novel aptamer-expression platform fusions, without the need for extensive optimisation.

These chimeric, transcriptional-‘OFF’ switches were shown to effect transcription termination *in vitro*, although only two of the examples were shown to repress the expression of *gfpuv* in *E. coli*¹⁰⁰.

It is not immediately apparent how these expression platforms could be directed by such a variety of distinct aptamer domains. One model for the folding of RNA secondary structures, known as “encoded co-transcriptional folding”, proposes that as well as the relative stability of secondary structure elements, the 5’ to 3’ sequence order also dictates the overall folding of the RNA¹¹¹. In other words, folding of the riboswitch occurs as it is being transcribed, and nucleotides in the 5’ region (aptamer domain) form secondary structure elements whilst transcription of the remaining sequence (expression platform) is still occurring.

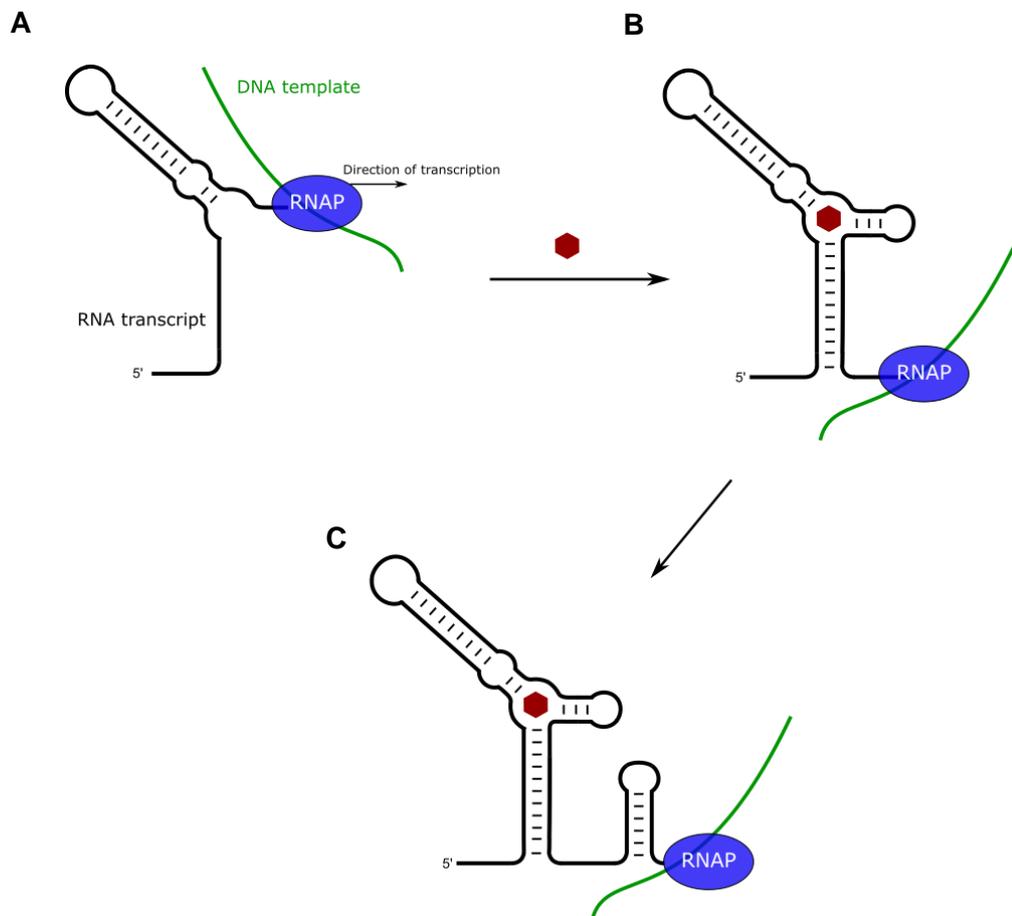


Figure 11 – Overview of encoded co-transcriptional folding. A) Folding of the aptamer domain begins during transcription by RNA polymerase (RNAP). B) A ligand binding event stabilises the aptamer domain P1 stem before the downstream expression platform has been transcribed. C) P1 stem formation favours only one expression platform conformation – in this case, a terminator stem.

Triggered by ligand binding, switching occurs by changing between two mutually exclusive RNA conformations. In this case, one conformation is defined by the aptamer domain P1 stem and the other by the anti-terminator stem of the expression platform. Crucially, formation of the terminator stem itself is merely an inevitable consequence of P1 stem formation (see **Figure 10**). Hence, these transcriptional expression platforms are able to co-operate with aptamer domains with which a ligand binding event disrupts or stabilises formation of a P1 stem¹⁰⁰. It follows therefore, that riboswitch aptamers for which ligand binding events stabilise/disrupt the formation of a pseudo-knot (eg. SAM-II, PreQ), rather than a P1 stem are not thought to be suitable for these types of chimeric switching systems.

1.6.2.1 The SAM-I riboswitch and chimeric transcriptional terminator devices

S-adenosylmethionine (SAM) is a ubiquitous enzyme co-factor, most commonly used as a methyl donor by methyl transferase enzymes, but is also involved in a variety of other reactions including donation of methylene, ribosyl and amino groups¹¹². SAM is intimately involved in the synthesis and recycling of cellular methionine and is consequently implicated in almost all of the cellular metabolic processes which involve sulphur¹¹³. Given the importance of SAM in a variety of metabolic processes, it is vital that its concentration is tightly regulated. Many bacterial species have achieved this level of control through the evolution of SAM-binding riboswitches. SAM sensing riboswitches are found across the breadth of bacterial taxonomy, including firmicutes, fusobacteria, α -, β -, γ -proteobacteria, chlorobi and deinococcus/thermus, as well as appearing in a number of metagenomic sequences¹¹⁴. For most effectors, there is typically only one class of associated riboswitch which has evolved to selectively bind it. However, to date, there are a total of six riboswitch classes which have been shown to bind SAM: SAM I – V and SAM – I/IV, with SAM-I being the most prevalent^{43, 44, 55, 56, 115}. Each class of riboswitch represents an evolutionarily distinct solution to the binding of SAM by RNA. SAM riboswitches generally control expression of genes which are implicated in sulphur metabolism, and in particular genes which are involved in the processing of SAM within the cell. SAM-responsive riboswitches are known to control the production of a number of enzymes including SAM synthetase, methionine synthetase and various methyl transferases¹¹⁶. SAM riboswitches harness a number of different mechanisms including transcriptional termination⁴², inhibition of translation initiation¹¹⁷ and in rare cases through the use of antisense devices¹¹⁸.

The SAM-I riboswitch is a transcriptional-‘OFF’ switch which effects termination of transcription in response to sufficient concentrations of intracellular SAM¹¹⁹. The SAM-I riboswitch found in the 5'-UTR of several *metX* genes in *Bacillus subtilis* has been particularly well characterised, especially with respect to the switching mechanism. The binding of SAM to the SAM-I aptamer domain stabilises P1 stem formation, which favours an expression platform conformation with an intrinsic transcriptional terminator stem, thus causing the RNA polymerase to disengage¹¹⁶. Ligand-induced P1 stem formation is also known to be a key feature of the switching mechanism of other transcriptional riboswitches including the adenine-responsive *pbuE* riboswitch from *B. subtilis*⁷². It has been shown by Batey *et al.* that this mechanism of ligand-induced P1 stem formation/disruption can be exploited to create chimeric riboswitches, whereby aptamer domains and expression platforms from distinct riboswitch classes and bacterial species can be combined to create novel transcriptional terminator devices^{100, 103}. The *B. subtilis* SAM-I riboswitch expression platform has been shown to be particularly conducive to this approach and has been combined with aptamer domains sensing guanine, lysine, 2-aminopurine¹⁰⁰ and THF¹¹⁰ to create chimeric expression systems functioning *in vitro* and in some cases *in vivo*.

1.6.2 Other approaches to riboswitch engineering and the creation of more complex regulatory devices

In addition to the fusion of non-native aptamer domains and expression platforms, entirely synthetic riboswitches comprising SELEX aptamer domains and *de novo* designed expression platforms have been reported recently by Wachsmuth¹²⁰. The group developed an *in silico* pipeline for the rational design of riboswitch expression platforms which function by transcription termination. Fusion of these sequences with a theophylline aptamer enabled the creation of a transcriptional-‘ON’ riboswitch which was shown to control the expression of reporter genes (GFP, *lacZ*) in *E. coli*. However, the resulting induction factors of around 3 are much poorer than those of naturally occurring or otherwise engineered riboswitch systems⁷⁹.

Natural riboswitches have been identified which are able to make complex genetic decisions, comparable with that of evolutionarily more recent protein-based expression systems⁸⁷ and these discoveries have inspired the creation of synthetic riboswitch constructs of rivalled complexity. The Yokobayashi group were able to mimic naturally occurring tandem riboswitch architectures by creating synthetic riboswitch systems,

based on the theophylline and TPP aptamers, which functioned as Boolean AND and NAND logic gates¹²¹. Using a *tetA*-based dual selection method, the group screened for functional systems from a library of clones with randomised regions of ~20 bases linking the theophylline aptamer and existing TPP 'ON' and 'OFF' switches.

Another example of more sophisticated riboswitch-based devices was demonstrated by Yokobayashi *et al.*, who constructed a tandem riboswitch, consisting of 'ON' and 'OFF' sub-units, which when combined produced a synthetic RNA device which acts as a chemical band-pass filter¹²². The two component riboswitches are TPP responsive, and function by transcriptional termination and translational activation respectively. At low intracellular TPP concentrations, the full length mRNA is transcribed, but expression is translationally repressed, whereas at high TPP concentrations transcription termination is effected. Consequently, maximum gene expression output is observed at an intermediate concentration of TPP, at which the two riboswitch sub-units are not completely repressed.

1.6.3 Engineered riboswitches in the Micklefield laboratory

One drawback of using SELEX-derived aptamers for *in vivo* applications is that the ligand binding sites are typically found on the surface of the RNA. Consequently, the scope of ligand functional groups which are involved in aptamer domain binding is usually relatively narrow and this leads to poor ligand specificity *in vivo*. Naturally derived riboswitches however, offer high binding affinities, precise ligand specificities and the metabolites which they recognise are typically small and drug-like. They display a significant degree of universality, in that they have been shown to control the expression of non-native reporter genes *in vivo* and are transferable between species. Moreover, riboswitches are intrinsically dose-dependent⁷⁹ and offer a range of expression levels which are correlated to ligand concentration. For *in vivo* applications as gene expression tools, natural riboswitches will be activated by the endogenous cellular ligand pool and this places a fundamental limit on the level of control which can be achieved using these natural systems. In the case of 'ON' switches, the presence of the cognate metabolite in the cell leads to 'leaky' basal expression levels. Moreover, binding of the natural ligand to 'OFF' switches leads to a background repression of gene expression and low basal levels in the absence of exogenous activator ligand. In order to address this issue, the focus of the Micklefield laboratory has been to re-engineer these natural systems to develop orthogonally selective riboswitches: mutant RNAs which no longer respond to the natural ligand but instead recognise synthetic ligands not found in the endogenous metabolite pool of the host organism.

This strategy has been used by the Micklefield laboratory to generate a number of orthogonally selective mutant riboswitches by re-engineering a natural adenine-sensing riboswitch from the *add* gene of *Vibrio vulnificus*^{79, 80, 81} and a preQ-responsive riboswitch found in *B. subtilis*⁸⁶. These two riboswitches were considered as good candidates for re-engineering for a variety of reasons. Both the *add A*- and preQ riboswitches have been well characterised, both *in vivo* and *in vitro*. Crucially, the x-ray crystallography structures of the riboswitch aptamer domains have been solved, which allowed the rational design of mutant riboswitch-ligand pairs. These two riboswitches were considered to be particularly good candidates because it was envisaged from the x-ray crystallography structures that ligand-binding could be modulated by the reversal of a few key hydrogen bonds.

1.6.3.1 Re-engineering an orthogonally selective *add A*-riboswitch

The *add A*-riboswitch is a translational-‘ON’ riboswitch, which induces expression of the downstream gene by releasing the SD sequence on binding adenine, thus promoting translation initiation¹²³. Through a combination of chemical genetics and genetic selection, the Micklefield laboratory was able to re-engineer the wild-type *add A*-riboswitch to create a series of mutants which respond to non-natural purine analogues.

From the crystal structure of the wild-type riboswitch, four binding site nucleotides are involved in ligand recognition. Of these, residues U47 and U51 were selected for mutagenesis as they were not deemed to be crucial in the folding of the aptamer domain¹²³. Saturation mutagenesis at these two positions produced 15 mutant RNAs which were each cloned upstream of a chloramphenicol resistance gene on a pMOD3 plasmid, transformed into *E. coli* and screened against a library of around 80 heterocyclic compounds. In order to identify mutant-ligand pairs, transformed strains were grown in the presence and absence of each ligand and the number and size of the resulting colonies was observed. This screening process led to the identification of two lead mutant-ligand pairs, which were then further optimised based on mFold computational analysis. The resulting mutant riboswitches M6'' (U28G, G42C, U47C, U51C) and M6C'' (U28G, G42C, U47C, U51C, U74C) were shown to respond to synthetic ligands ammeline and azacytosine but showed no activity at all with the natural ligand adenine. This orthogonality was fully characterised *in vivo* and *in vitro* using an *eGFP* expression assay and isothermal titration calorimetry (ITC) respectively.

One of the problems with the mutant M6''-ammeline expression system is the extremely poor solubility of the ammeline activator ligand. Building on this work, the M6'' system was further optimised with the creation of a second-generation ligand, PPDA, which had both improved solubility and increased binding affinity for the M6'' aptamer⁸¹. The x-ray crystallography structure of the M6''-PPDA complex is shown in **Figure 11A** and **11B**, with the wild-type aptamer domain bound to adenine shown in **Figure 11C** for comparison. Interestingly, the x-ray crystallography structure of M6''-PPDA revealed an interaction between the binding site C51 residue and N1 of PPDA, which would be expected to be unfavourable due to lone-pair repulsion. It was therefore postulated and confirmed by ITC that the interaction is actually mediated by a proton. In addition to aptamer-ligand engineering, it was also shown that the second

generation M6''-PPDA system was able to control the expression of functional genes governing motility and cell morphology in *E. coli* and *B. subtilis*.

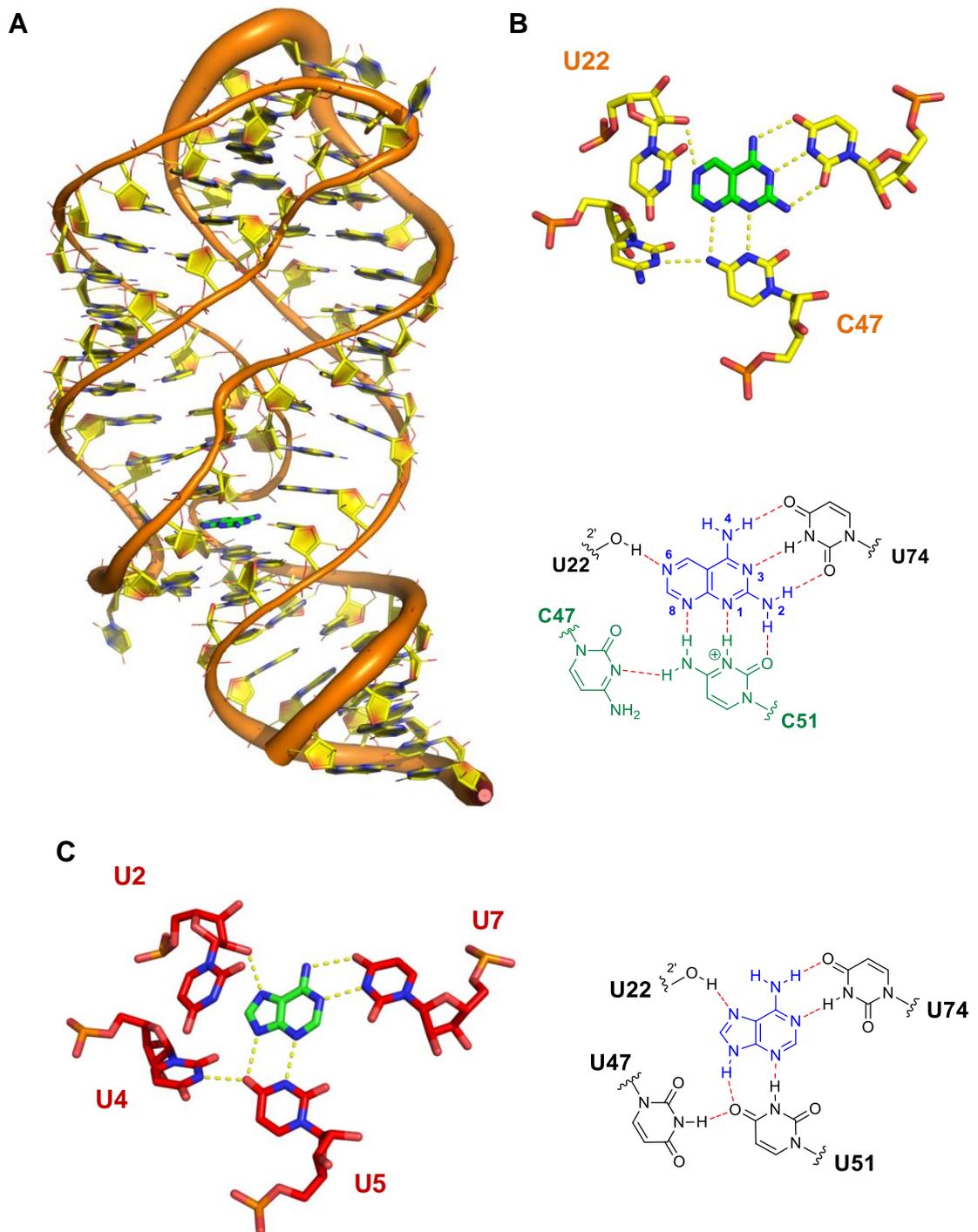


Figure 12 – A) X-ray crystallography structure of the M6'' riboswitch binding PPDA⁸¹. B) Ligand binding site showing the PPDA recognition motif and the mutated residues (shown in green in 2D representation). C) Ligand binding site of the wild-type *add* A-riboswitch.

The modular nature of riboswitch expression platforms was also demonstrated by the creation of a chimeric transcriptional-‘OFF’ switch based on the M6” aptamer domain and the expression platform from the *xpt* G-riboswitch which functions by transcriptional repression.

1.6.3.2 Re-engineering an orthogonally selective preQ riboswitch

To demonstrate the applicability of the orthogonal mutant riboswitch rationale to ‘OFF’ riboswitches as well as ‘ON’ systems, the Micklefield group also targeted the transcriptional-‘OFF’ preQ riboswitch for re-engineering⁸⁶.

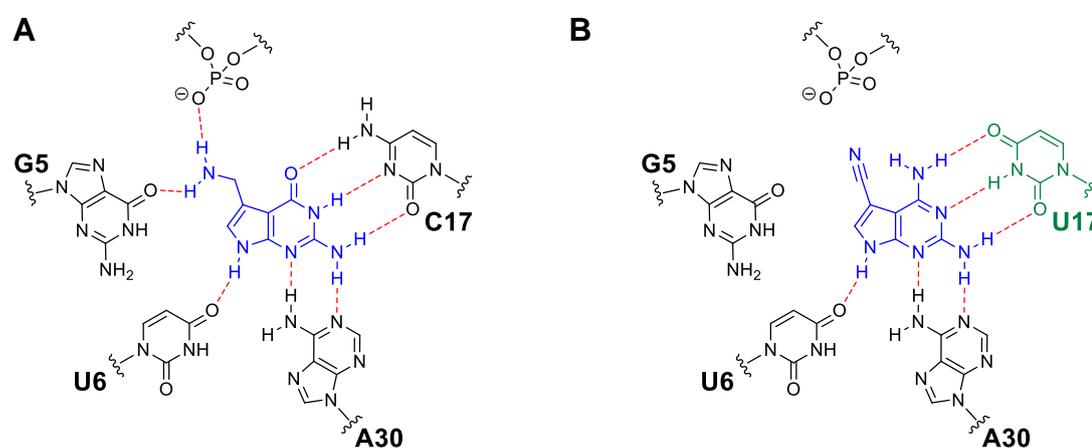


Figure 13 – A) Ligand binding motif for the preQ riboswitch binding preQ₁ obtained by x-ray crystallography⁶⁵. B) Proposed ligand binding motif for the preQ M1 (C17U) mutant riboswitch binding DPQ₀.

However, rather than screening a large number of compounds against an extensive mutant library, a more rational and selective strategy was employed to identify orthogonal mutant-ligand pairs. From the x-ray crystallography structure of the wild-type preQ aptamer domain (**Figure 12A**), the residues G5, U6 and C17 were identified as key nucleotides involved in ligand recognition and these were selected for site-directed mutagenesis to create a small library of preQ riboswitch mutants. To minimise disruption of riboswitch folding and to maintain the size of the ligand binding site, the substitution of pyrimidines for purines and *vice versa* was avoided. This resulted in a library of just seven mutants. It was envisaged that mutants possessing the C17U mutation would likely bind synthetic ‘diamino-faced’ analogues of preQ₁ and thus small

selection of seven analogues was synthesised for screening with the mutant library. Screening was based on a *lacZ* reporter gene assay in *B. subtilis*, which identified the M1-DPQ₀ mutant-ligand pair as a functional repressive expression system. The structure of DPQ₀ and the proposed binding motif with the preQ M1 (C17U) mutant riboswitch is shown in **Figure 12B**. Interestingly, the wild-type system proved to be almost entirely repressed by the endogenous preQ pool, further highlighting the need for re-engineering of this system. In addition to the reporter gene assay, the M1-DPQ₀ system was shown to be able to control the expression of the *B. subtilis* cell morphology gene *mreB*. Microscopy was used to show that the morphological phenotype could be modified when *mreB* is under the conditional control of the mutant riboswitch. Finally, the aptamer-ligand binding affinities were assessed using ITC.

1.7 The Tetrahydrofolate (THF) Riboswitch

1.7.1 Folate biosynthesis

5,6,7,8-tetrahydrofolate (THF) is a ubiquitous enzyme cofactor which participates in one-carbon transfer reactions¹²⁴ and is utilised across all domains of life. It is involved in the biosynthesis of a number of important biomolecules including purines, amino acids, modified tRNA nucleotides and vitamin cofactors¹²⁵. The structure and numbering scheme for THF is shown in **Figure 13**. The compound is comprised of three key moieties: a reduced pterin ring, derived from the guanine function of GTP; a *p*-aminobenzoyl group, produced by the shikimate pathway¹²⁶; and a glutamyl moiety. THF accepts and donates one-carbon units at varying levels of oxidation at the N5 and N10 positions. Bacteria produce THF *de novo* in a seven-step metabolic pathway from GTP, however higher organisms lack several key enzymes and therefore require dietary folate¹²⁷. This distinction has been exploited to produce antifolate antibiotics which inhibit key steps in the bacterial folate pathway¹²⁸.

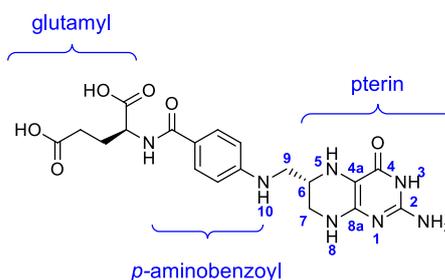
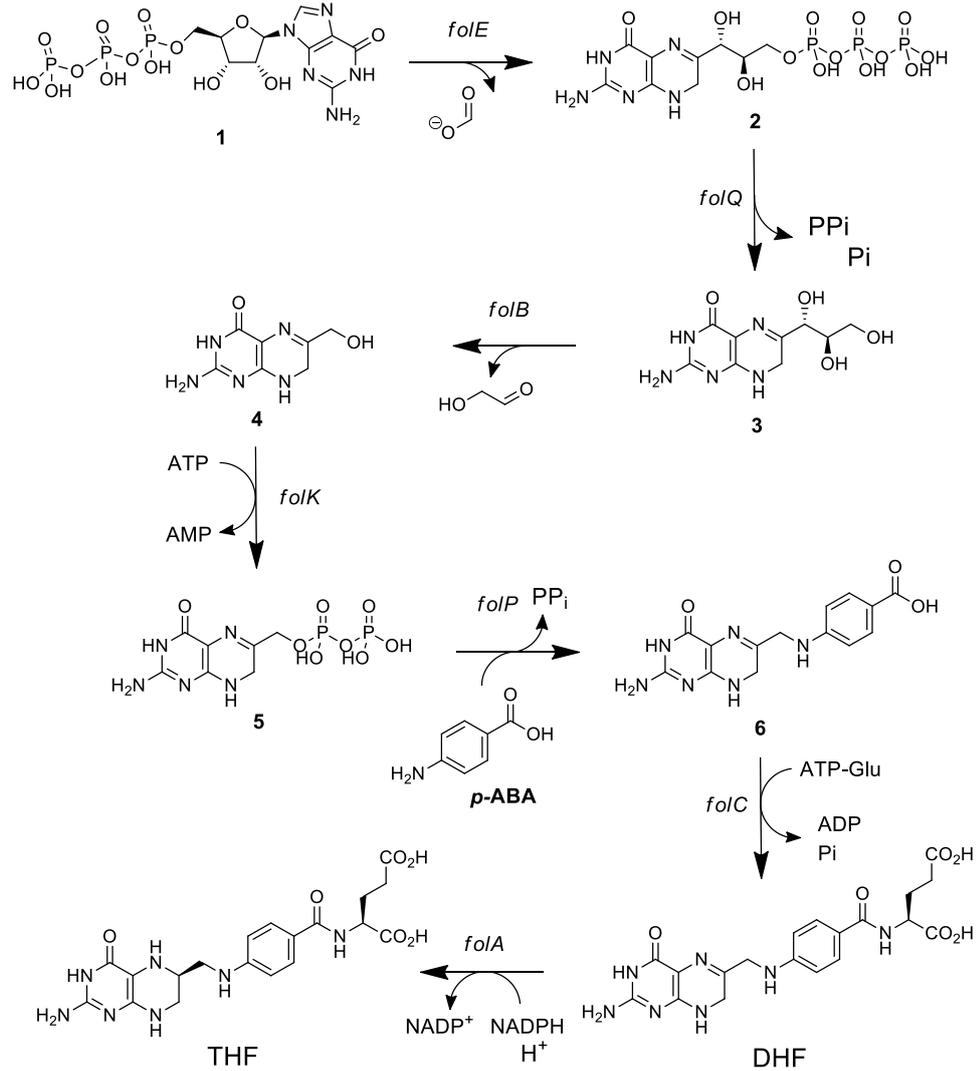


Figure 13 – Structure and numbering scheme of THF. The three key moieties are highlighted.

An overview of bacterial THF biosynthesis is shown in **Figure 14A**. The first step is the hydrolysis of GTP **1** to 7,8-dihydroneopterin triphosphate **2**, catalysed by GTP cyclohydrolase (encoded by *folE*) with the loss of formate. Subsequent dephosphorylation by the *folQ*-encoded phosphatase gives 7,8-dihydroneopterin **3**, from which the neopterin aldolase (encoded by *folB*)-catalysed loss of glycoaldehyde leads to the formation of 6-hydroxymethyl-7,8-dihydropterin **4**. This intermediate is then phosphorylated by the corresponding *folK*-encoded kinase to give 6-hydroxymethyl-7,8-dihydroneopterinpyrophosphate **5**.

A



B

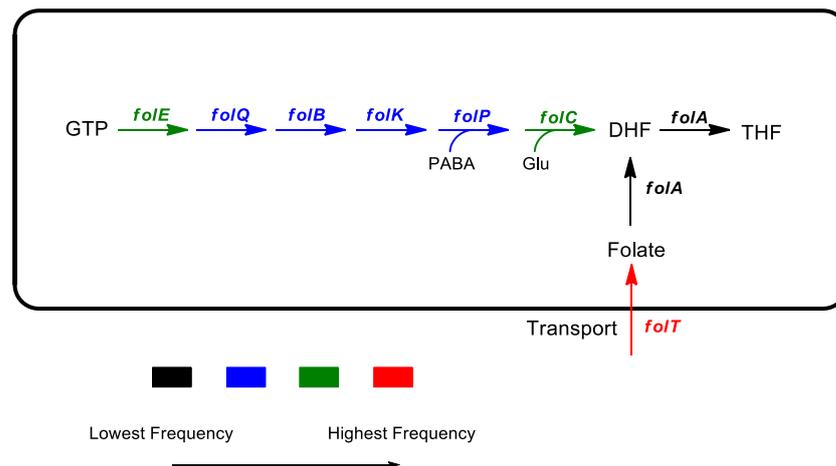


Figure 14 – Folate biosynthesis pathway. A) Biosynthetic pathway showing structures of intermediates and genes encoding enzymes involved in the catalytic steps. B) Overview of folate biosynthesis and transport genes in bacteria showing the frequency with which they are found to be controlled by THF-responsive riboswitches.

At this point *p*-aminobenzoic acid (*p*A_{BA}), a product of the shikimate pathway¹²⁶, is utilised in a nucleophilic substitution reaction catalysed by 7,8-dihydropteroate synthase (*folP*), yielding 7,8-dihydropteroate **6**. Subsequent glutamylation by dihydrofolate synthase (*folC*) produces 7,8-dihydrofolate (DHF), which is reduced by dihydrofolate reductase (DHFR) to give 5,6,7,8-tetrahydrofolate (THF).

1.7.2 The role of THF and reduced folates in one-carbon metabolism

THF and other reduced folates are a family of enzyme cofactors which act as a vehicle for the transfer of one carbon units from the N5 and N10 positions. These molecules are interconverted by a network of reactions known collectively as one-carbon metabolism. A summary of cellular THF derivatives and the genes responsible for their interconversion is shown in **Figure 15**. THF serves as the primary scaffold onto which single carbon units are transferred temporarily, for their use in the biosynthesis of a number of important metabolites. Whilst the variety of cellular THF analogues can be interconverted enzymatically, each cofactor form is specific to a particular biosynthetic pathway¹²⁹. For instance, the formyl group of 10-formylTHF **9** is utilised in purine biosynthesis and is incorporated at the C2 and C8 positions of the purine ring¹³⁰. The one-carbon moiety of 5,10-methyleneTHF **8** is required in nucleotide biosynthesis for the conversion of uridylate to thymidylate and similarly, the N5 methyl group of 5-methylTHF **7** is utilised in the re-methylation of homocysteine to methionine during amino acid biosynthesis¹³¹. Folinic acid **11** (5-formylTHF) acts only as an intermediate in the interconversion of THF and 5,10-methenylTHF **10** and is not used as cofactor in folate-dependent one-carbon transfer reactions.

1.7.3 THF riboswitch overview and phylogeny

THF riboswitches have been identified in around 60 different genomic sequences, the majority of which are from organisms within the Firmicutes phylum⁵⁷. They are predominantly found in the 5'-UTR of the folate transporter gene *folT*, although there are also examples of THF riboswitches controlling folate biosynthesis genes including *folC* and the *folQBPK* operon. An overview of folate biosynthesis and transport along with the frequency with which these genes are found to be controlled by THF-responsive riboswitches is shown in **Figure 14B**. The precise mechanism of genetic

regulation by THF riboswitches has not been unequivocally established, however it is known that they function as 'OFF' switches, repressing expression of the downstream gene in response to increasing THF concentrations. Given the known sequence and predicted secondary structures of the THF riboswitch expression platforms, it is thought likely that they operate at the translational level through RBS sequestration on ligand binding⁵⁷.

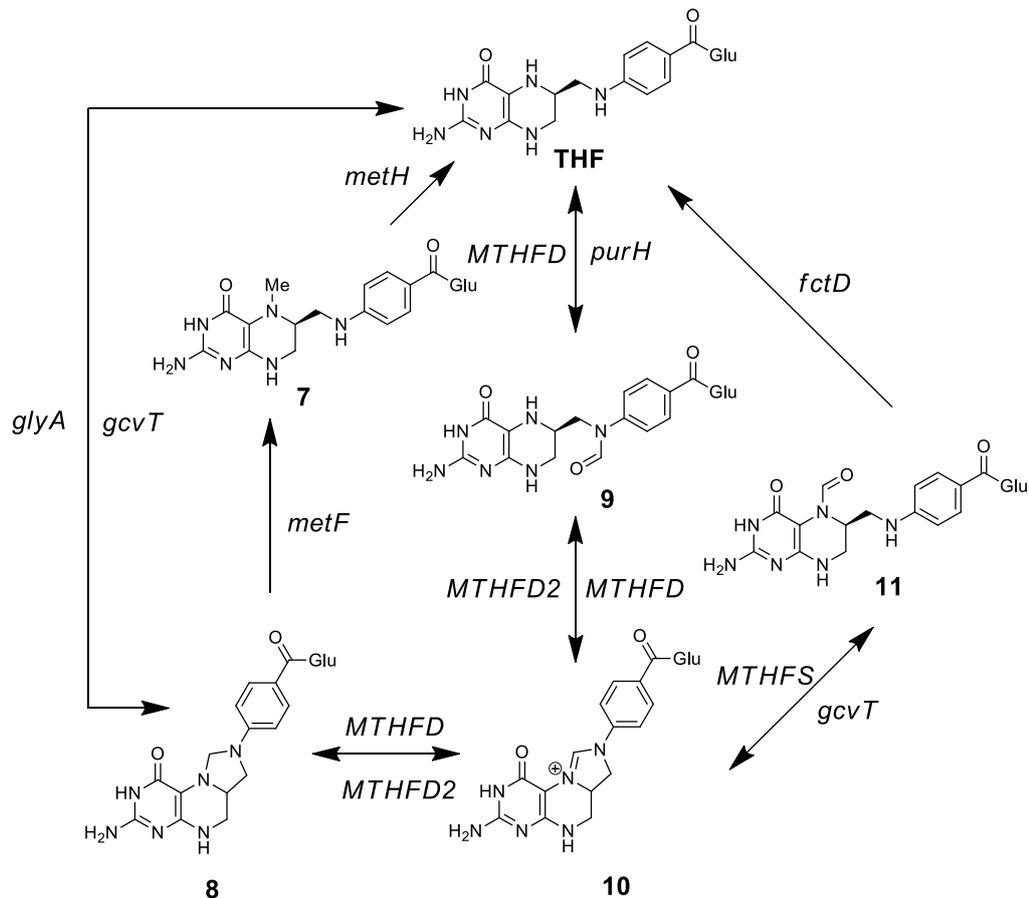


Figure 15 – Tetrahydrofolate and related compounds in bacterial single carbon metabolism. Genes involved in each transformation are shown next to the appropriate arrow. For reactions whereby a different gene is involved in each catalytic direction - two genes are listed. Genes for reactions in the top to bottom direction are shown on the left of the arrow; genes for reactions in the right to left direction are shown above the arrow.

1.7.4 THF aptamer domain structure and ligand recognition

The x-ray crystallography structure of the ligand-bound THF riboswitch aptamer domain from the *folT* gene of *Streptococcus mutans* has been solved by Batey *et al*¹⁰⁴ and is shown in **Figure 16**. Uniquely, the THF aptamer contains two distinct ligand binding sites and is the only published example of this occurrence in the literature. THF itself is unstable in air¹³² and the structure was solved with the THF analogue folinic acid bound at each site.

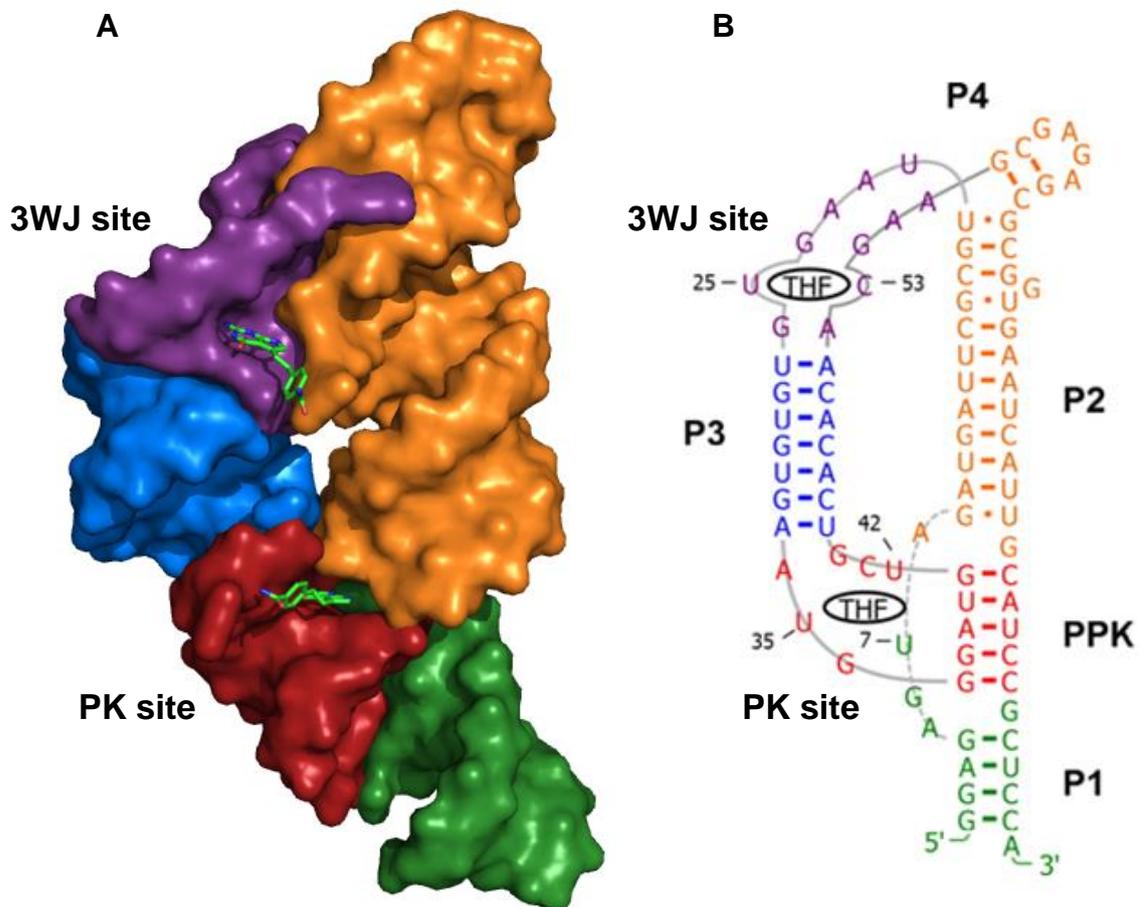
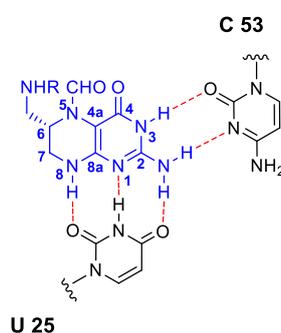
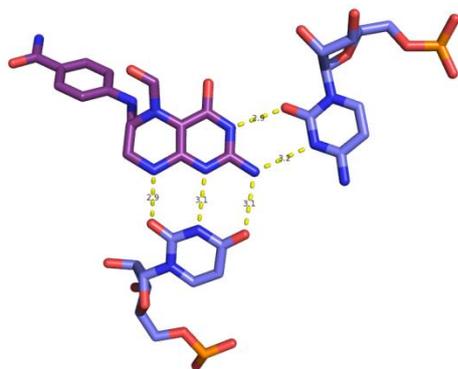


Figure 16 – Structure of the *S. mutans* THF riboswitch aptamer domain. A) X-ray crystallography structure of the THF aptamer binding folinic acid¹⁰⁴. B) Secondary structure of the THF aptamer domain based on x-ray crystallography data. Structural features are labelled and colouring is consistent across both images.

The THF aptamer is formed of two pairs of coaxially stacked helices (P1/P3 and P2/P4), positioned via the formation of a pseudoknot and a three-way junction. The pseudoknot forms one of the two ligand binding sites (denoted the 'PK site'), whilst the other site (denoted '3WJ site') lies at the base of the P3 stem, adjacent to the 3-way junction which joins the P2, P3 and P4 stems. The pseudoknot is constituted of five Watson-Crick base pair interactions between the nucleotides in the L3 loop and the unpaired region separating the P1 and P2 stems. Ligand recognition at the two binding sites occurs via two similar but distinct binding motifs and is shown in **Figure 17**.

3WJ site



PK site

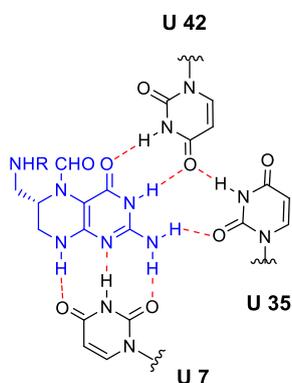
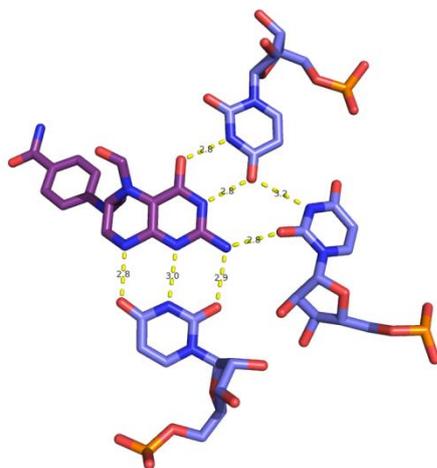


Figure 17 – THF riboswitch binding sites and folinic acid recognition motifs¹⁰⁴. Top: 3WJ site; Bottom: PK site.

Both the 3WJ and PK binding sites only have clear hydrogen bonding interactions with the pterin moiety of folinic acid. In the case of the 3WJ binding site, the folinic acid pterin functionality is recognised by the unpaired U25 and C53 residues. U25 accepts hydrogen bonds from the folinic acid C2-amino group and N8, whilst donating a

hydrogen bond to the sp²-hybridised N1 atom. C53, meanwhile, interacts with the Watson-Crick face of the bound ligand, forming hydrogen bonds with the C2-amino group and N3. The x-ray crystallography structure also reveals a stacking interaction between the *p*ABA moiety of folinic acid and the G68 residue of the THF aptamer domain.

At the PK binding site, folinic acid is bound by hydrogen bonding interactions between three uridine residues U7, U35 and U42. Similarly to U25 at the 3WJ site, the PK site U7 residue accepts hydrogen bonds from the folinic acid C2-amino group and N8, whilst donating a hydrogen bond to N1. Additionally, and analogously to the 3WJ site C53 residue, the PK site U42 forms two hydrogen bonds with the Watson-Crick face of folinic acid. Finally, the third uridine residue U35 serves to position U42 whilst also forming a hydrogen bond with the folinic acid C2-amino group. It has been reported that the *p*ABA moiety interacts with the backbone RNA at the PK site, and that this could be an important feature in the overall regulatory function of the THF riboswitch¹¹⁰. However, the specific mode of interaction and how this might influence the riboswitch regulatory response has not been at all defined.

1.7.5 THF riboswitch ligand binding specificity and regulatory function *in vitro*

The ligand binding specificity of the THF riboswitch has been extensively explored *in vitro*, with more than 20 compounds identified which bind the aptamer domain^{57, 104, 110}. A summary can be found in **Table 2**. In-line probing and ITC analysis by Breaker *et al.*⁵⁷ and Batey *et al.*¹⁰⁴ respectively has shown that THF riboswitches selectively bind reduced forms of folate such as THF and the partially reduced dihydrofolate (DHF) whilst rejecting folic acid⁵⁷. X-ray crystallography studies discussed in the previous section have revealed the importance of a reduced C7-N8 bond in allowing hydrogen bond donation from N8 to the adjacent U7 and U25 residues of the two ligand binding sites, and this is thought to be the mechanism by which the THF aptamer distinguishes the reduced forms of folate.

Table 2 – Summary of *in vitro* ligand binding/regulatory parameters for the THF riboswitch

Ligand	Riboswitch	Method(s)	K _D (μM)	T ₅₀ (μM)	Reference(s)
6S-THF	<i>S. mutans</i>	ITC, transcription assay	13 ± 1	32 ± 1	Batey <i>et al</i> ¹¹⁰
6R-THF	<i>S. mutans</i>	ITC, transcription assay	19 ± 2	110 ± 10	Batey <i>et al</i> ¹¹⁰
6S-folinic acid	<i>S. mutans</i>	ITC, transcription assay	9.6 ± 0.3	19 ± 1	Batey <i>et al</i> ¹¹⁰
Dihydrofolate	<i>S. mutans</i>	ITC, transcription assay	23 ± 1	89 ± 6	Batey <i>et al</i> ¹¹⁰
Tetrahydrobiopterin	<i>S. mutans</i>	ITC, transcription assay	18 ± 3	190 ± 20	Batey <i>et al</i> ¹¹⁰
6,7-dimethyl-tetrahydrobiopterin	<i>S. mutans</i>	ITC, transcription assay	57 ± 3	590 ± 30	Batey <i>et al</i> ¹¹⁰
7-deazaguanine	<i>S. mutans</i>	ITC, transcription assay	66 ± 1	690 ± 20	Batey <i>et al</i> ¹¹⁰
Pemetrexed	<i>S. mutans</i>	ITC, transcription assay	110 ± 10	>10,000	Batey <i>et al</i> ¹¹⁰
Lometrexol	<i>S. mutans</i>	ITC, transcription assay	37 ± 2	190 ± 20	Batey <i>et al</i> ¹¹⁰
2,6-diaminopurine	<i>S. mutans</i>	ITC, transcription assay	0.5 ± 0.1	130 ± 10	Batey <i>et al</i> ¹¹⁰
Adenine	<i>S. mutans</i>	ITC, transcription assay	8 ± 1	2500 ± 800	Batey <i>et al</i> ¹¹⁰
2-aminopurine	<i>S. mutans</i>	ITC, transcription assay	26 ± 3	>10,000	Batey <i>et al</i> ¹¹⁰
8-aminoadenine	<i>S. mutans</i>	ITC, transcription assay	160 ± 20	>10,000	Batey <i>et al</i> ¹¹⁰
6-hydroxy-2,4,5-triaminopyrimidine	<i>S. mutans</i>	ITC, transcription assay	84 ± 5	1400 ± 200	Batey <i>et al</i> ¹¹⁰
THF	<i>A. metalliredigens</i>	In-line probing	~0.07	N/A	Breaker <i>et al</i> ⁵⁷
Dihydrofolate	<i>A. metalliredigens</i>	In-line probing	~0.3	N/A	Breaker <i>et al</i> ⁵⁷
Tetrahydrobiopterin	<i>A. metalliredigens</i>	In-line probing	~0.3	N/A	Breaker <i>et al</i> ⁵⁷
5-methyl THF	Various	In-line probing	Binding detected*	N/A	Breaker <i>et al</i> ⁵⁷
10-formyl THF	Various	In-line probing	Binding detected*	N/A	Breaker <i>et al</i> ⁵⁷
5,10-methenyl THF	Various	In-line probing	Binding detected*	N/A	Breaker <i>et al</i> ⁵⁷
5,10-methylene THF	Various	In-line probing	Binding detected*	N/A	Breaker <i>et al</i> ⁵⁷

* It was reported that binding was detected with “several other representatives from this RNA class”, but neither the identity of the native organism nor the binding parameters were published.

THF riboswitches have been shown to tolerate ligand functionalization at both the N5 and N10 positions with binding having been observed for folinic acid (5-formyl THF), 5-methyl THF, 10-formyl THF, 5,10-methenyl THF and 5,10-methylene THF⁵⁷. These folate derivatives are known cellular metabolites involved in one-carbon transfer reactions and this suggests that the THF riboswitch may be able to sample a large pool of related metabolites in the determination of a regulatory response. An overview of reduced folate metabolism and the structures of cellular N5 and N10-functionalised THF analogues is shown in **Figure 15**.

The *Streptococcus mutans* THF riboswitch was investigated *in vitro* by Batey *et al.* utilising an *in vitro* transcription assay. The THF aptamer domain was fused to the expression platform of a *B. subtilis* SAM-I riboswitch, thus creating a chimeric system which functions by transcriptional termination on ligand binding¹¹⁰ (for further details on this chimeric riboswitch, see **Section 1.6.2.1**). Interestingly, this analysis revealed a discrepancy between ligands which are able to bind the THF riboswitch aptamer domain, versus ligands which are able to both bind and elicit a regulatory response. A variety of natural and non-natural folate analogues have been shown by ITC to bind the THF aptamer domain including: tetrahydrobiopterin; 6,7-dimethyl-tetrahydrobiopterin; 7-deazaguanine; and pemetrexed. The structures of these compounds can be found in **Figure 18**.

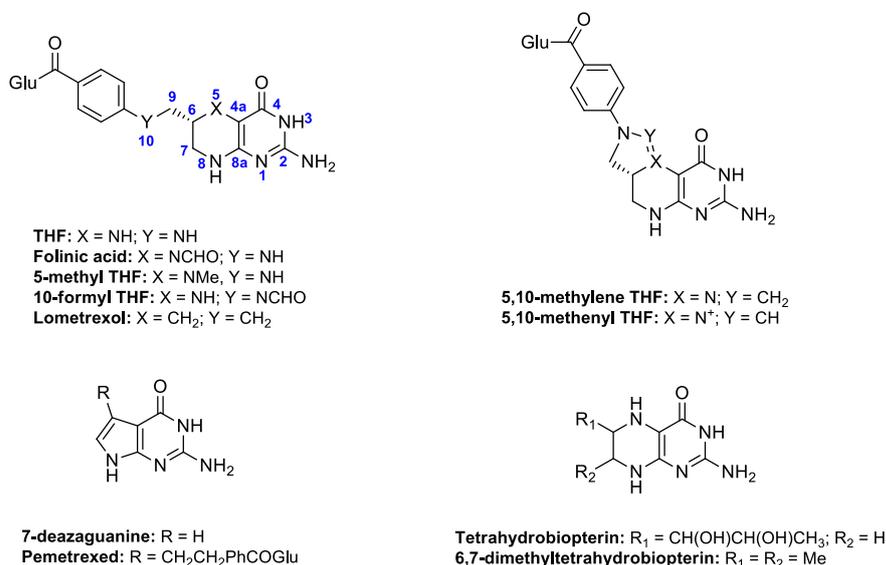


Figure 18 – Structures of compounds which are known to bind the THF riboswitch^{57, 104, 110}

However, whilst all of these compounds exhibit binding affinities within the same order of magnitude as THF/folinic acid, they show a significantly diminished ability to induce the riboswitch regulatory response *in vitro*. The working hypothesis of Batey *et al.* is that a specific conformation of the THF *p*ABA side chain, such that it interacts with the backbone RNA, is required to mediate the riboswitch regulatory response. This is supported by the fact that the aforementioned ligands all either lack this structural feature, or (in the case of pemetrexed) possess a planar pterin functional analogue which positions the side-chain in an alternative orientation. Moreover, whilst the binding affinities of the 6*S* and 6*R* diastereoisomers of THF are almost identical, the 6*R* isomer is more than 3-fold less effective in eliciting the riboswitch regulatory response¹¹⁰. This again suggests that the correct orientation of the *p*ABA side chain within the THF

binding site is an important requirement for regulating gene expression. This phenomenon is further exemplified by the observation that adenine and 2,6-diaminopurine are able to bind the THF PK site via a 'retro binding' motif but have a significantly reduced regulatory ability relative to their binding affinity.

1.7.6 THF riboswitch ligand binding site cooperativity

Binding site cooperativity is a well explored phenomenon in biochemistry. In this context, 'cooperativity' refers to the extent to which ligand binding at one site influences binding at one or more other sites in a multi-binding site scenario. Cooperativity is typically determined by fitting an *in vitro* binding curve to a Hill Equation and then solving to obtain the Hill coefficient. A Hill coefficient of greater than one indicates a positively cooperative system, whereby binding to one site increases the likelihood of binding to another. Conversely, a Hill coefficient of less than one implies a negatively cooperative system, with binding at one site reducing the binding affinity at another. A coefficient of exactly one indicates that binding at the different sites is entirely independent¹³³. With respect to riboswitches, the cooperativity of tandem riboswitch systems such as the glycine³³ and SAM-II/SAM-V⁵⁶ has been investigated. Additionally, Batey *et al.* have used an *in vitro* transcription assay to study the cooperativity of the two ligand binding sites within the THF aptamer^{104, 110}. This revealed a positive binding cooperativity, with a Hill coefficient of 1.8. Additionally, in order to investigate the effects of the two ligand binding sites on the riboswitch regulatory response, the group used the solved x-ray crystallography structure to postulate mutations which would abrogate binding at each of the two binding sites. Two mutations were identified which were then confirmed to disrupt binding by ITC and SHAPE analysis. The U7C mutation was shown to prohibit the binding of folinic acid to the PK site whilst U25C ablated binding to the 3WJ site. The double mutant U7C/U25C was shown to not recognise folinic acid at either of the two ligand binding sites¹⁰⁴.

These mutant aptamer domains were then fused to the *metE* expression platform from *B. subtilis* (which functions as a transcriptional terminator) and used in an *in vitro* transcription assay to assess the relative importance of the two ligand binding sites in the riboswitch regulatory function (see **Section 1.6.2.1** for further details). The level of transcription termination in response to increasing ligand concentrations was determined by measuring the relative levels of full-length versus terminated transcripts

by urea PAGE¹¹⁰. At physiologically relevant magnesium concentrations (3.5 μM), the U7C mutation (PK site binding disrupted) was shown to completely negate the regulatory ability of the riboswitch. However, the U25C mutation (3WJ binding site disrupted) had a much more moderate effect on riboswitch regulatory ability, with a T_{50} (concentration of effector ligand required for 50% transcriptional termination) of 91 μM compared with 19 μM for the wild-type. As expected, the double mutant U7C/U25C showed no increase in transcriptional termination at folinic acid concentrations up to 3 mM. These data indicate that, in an *in vitro* setting, and in the context of the chimeric system used in this study, the PK binding site is much more important in governing the overall riboswitch regulatory response than the 3WJ site¹¹⁰. This could be rationalised structurally, where it is clear that ligand binding at the PK site is crucial for pseudoknot formation which is the key structural feature which defines the aptamer domain tertiary architecture in the ligand-bound state. However, this conclusion should come with a number of caveats. Firstly, the *in vitro* study utilised a chimeric riboswitch comprising the THF aptamer domain and the expression platform of a transcriptional-‘OFF’ SAM-I riboswitch. It is therefore not clear if the observed roles of the two ligand binding sites are representative of native THF riboswitches, or are merely an artefact of the chimeric system. Indeed, the folding of the SAM-I expression platform is known to be governed by P1 stem formation/disruption in the aptamer domain¹¹⁹, whilst the structure and switching mechanism of the THF riboswitch expression platform has not been investigated. If anything, the use of the SAM-I *metE* expression platform is likely to overstate the role of P1 stem formation and hence overstate the role of the PK site in inducing the riboswitch regulatory effect. Secondly, the NTP concentrations used in the assay were an order of magnitude lower than the known concentration of free NTPs in bacteria such as *E. coli*^{134, 135}. Since both the rate of transcriptional elongation and the efficiency of intrinsic transcriptional termination are known to depend on NTP concentration, the physiological relevance of the *in vitro* transcription assay as described¹¹⁰ should be questioned.

1.7.7 In vivo characterisation of the THF riboswitch

Whilst the THF riboswitch aptamer domain has been investigated *in vitro*^{57, 104, 110}, there have to date been no reports of the characterisation of the THF riboswitch in an *in vivo* setting. One possible explanation for this is transport of THF across bacterial cell membranes. Riboswitches are typically studied in common laboratory organisms such as *E. coli* and *B. subtilis*, neither of which have native transporter proteins to facilitate the uptake of THF/folinic acid into the cell. The absence of any *in vivo* characterisation as well as the shortcomings of the *in vitro* transcription assay from which inferences about the riboswitch regulatory function have been made¹¹⁰ (see **Section 1.7.6**), makes the establishment of an *in vivo* reporter system a high priority for investigations for the THF riboswitch.

1.8 Aims of Project

Since its discovery in 2010⁵⁷, there have been a number of published studies on the *in vitro* characterisation of the THF riboswitch^{104, 110}. These studies have utilised techniques such as in-line probing and ITC to explore the riboswitch ligand binding specificity, as well as a transcription termination assay to model the riboswitch regulatory function and probe the role of the two ligand binding sites. However, to date there has been no published characterisation of the THF riboswitch in an *in vivo* setting. Moreover, the existing *in vitro* studies of ligand binding specificity and riboswitch regulatory function have a number of limitations. THF riboswitch ligand binding specificity has been limited to the study of commercially available THF analogues, which restricts the diversity of compounds for testing. Furthermore, the *in vitro* transcription assay which was used to model the riboswitch gene regulatory function may not be a faithful representation of its behaviour in the cell. The assay was performed using a chimeric THF riboswitch containing a non-native expression platform - which appears to operate by a different switching mechanism - using assay conditions which differ from those of the cellular environment by an order of magnitude.

Thus, the first objective of this project is the characterisation of the wild-type THF riboswitch *in vivo* by establishing a reporter gene assay in *E. coli*. The riboswitch ligand binding and regulatory specificity will be investigated through the synthesis and *in vivo* screening of suitable THF analogues identified from x-ray crystallography data. Any 'hits' will be confirmed *in vitro* using isothermal titration calorimetry (ITC). Additionally, the role of the two riboswitch ligand binding sites in mediating the gene regulatory response will be investigated *in vivo* by creating a series of riboswitch mutants which have previously been identified by Batey *et al.* to abrogate binding to each site *in vitro*¹⁰⁴.

Riboswitches have evolved to respond to fluctuating concentrations of important intracellular metabolites. However, the presence of these metabolites within the cellular environment limits the use of natural riboswitches as gene expression tools. In the case of 'ON' riboswitches, binding of the cognate metabolite leads to 'leaky' expression, whilst the basal expression levels of 'OFF' riboswitches are usually too low because of background repression by endogenous ligands. The Micklefield group have demonstrated that it is possible to re-engineer naturally occurring riboswitches such that they no longer respond to their cognate ligands present within the cell, but instead are activated by non-natural, synthetic compounds. These orthogonal riboswitches were

used to dose-dependently control the expression of reporter and functional genes in *E.coli* and *B. subtilis*^{79, 80, 81, 86}. These engineered riboswitches function as single-input/single-output regulatory systems, whereby the addition of a single synthetic ligand induces (in the case of the *add A*-riboswitch^{79, 81}) or represses (in the case of the preQ riboswitch⁸⁶) the expression of a single gene. The THF riboswitch is unique in that it contains two ligand binding sites within the same aptamer domain¹⁰⁴. This riboswitch therefore presents an opportunity for the engineering of a dual-input regulatory system, whereby the addition of two distinct synthetic ligands can be used to repress the expression of the downstream gene.

Therefore, the second objective of this project is the re-engineering of the THF riboswitch to create orthogonal mutant riboswitch-ligand pairs, following the same strategy utilised in the creation of orthogonal *add A*-⁸¹ and preQ riboswitches⁸⁶. Using the published x-ray crystallography data¹⁰⁴, mutations will be identified at each of the two ligand binding sites, which will provide the reversal of hydrogen bond directionality necessary for the binding of a synthetic ligand known to be accessible through relatively straightforward synthesis. The strategy will be to create two independent THF riboswitch mutants, each with one binding site separately re-engineered, before combining the two sets of mutations to create a second generation riboswitch with two engineered binding sites. If it is only possible to re-engineer one of the two ligand binding sites, a fall-back strategy will be to create a second generation mutant with one site successfully re-engineered and the other site knocked out through known mutations¹⁰⁴. Mutant riboswitches and synthetic ligands will be screened using an *in vivo* reporter gene assay in *E. coli* and any identified 'hits' will be confirmed *in vitro* using ITC.

Chapter 2

Synthesis of ligands to probe the THF riboswitch

2.1 Chapter Introduction

2.1.1 Overview of the synthesis of folate analogues

Folic acid and its reduced derivatives dihydrofolic acid (DHF) and tetrahydrofolic acid (THF) are important cellular metabolites and are involved in a variety of metabolic processes. Reduced folates participate in the transfer of one-carbon units of varying oxidation states from the N5 and N10 positions and are key co-factors in the biosynthesis of nucleic acids¹²⁴. As such, folates are of particular importance during phases of frequent cell division and growth and this has been exploited in the development of therapeutic agents which inhibit folate biosynthesis. This strategy in drug development has driven the exploration of synthetic folate analogues and the structural variety of known compounds in this class is diverse¹³⁶.

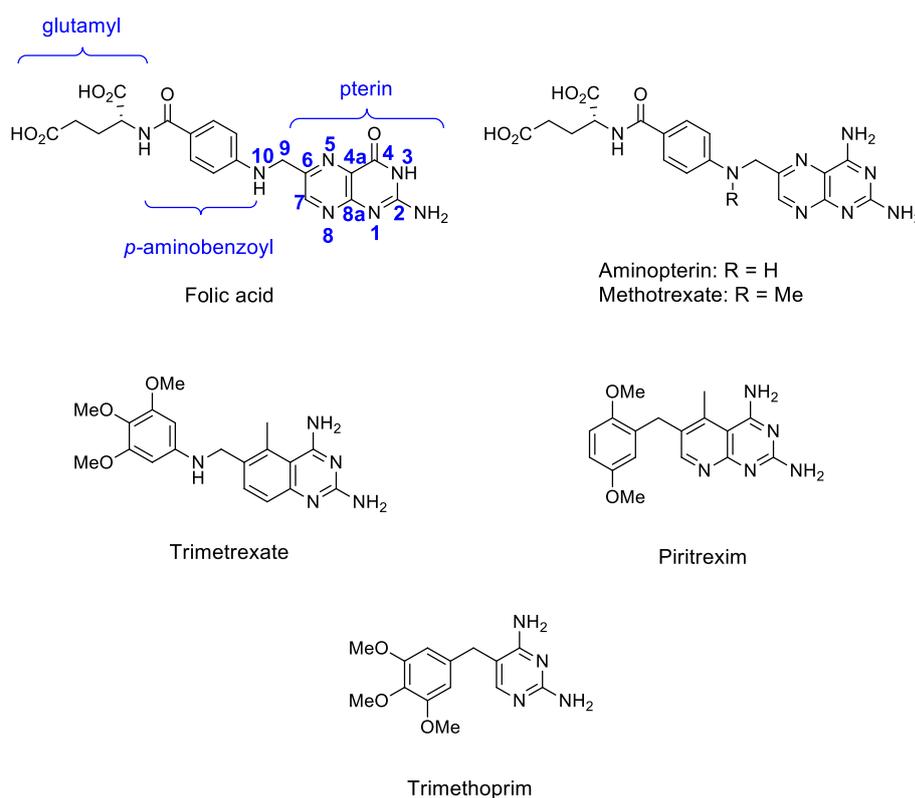


Figure 19 – Structures of folic acid and examples of DHFR inhibitors currently used in the clinic

The most common therapeutic target in the development of inhibitors of folate biosynthesis is the enzyme dihydrofolate reductase (DHFR). Folate-derived antagonists of DHFR have been used in a variety of clinical settings and have been shown to exhibit anti-neoplastic, anti-bacterial, anti-inflammatory and

immunosuppressive properties¹³⁷. Examples of DHFR inhibitors (also known as anti-folates) which are currently utilised in the clinic can be found in **Figure 19**. The scaffolds of known anti-folates are diverse, but typically contain a few common structural motifs. Compounds of this class feature a bi-cyclic fused pyrimidine ring system which serves to mimic the pterin moiety of folic acid, such as the 2,4-diaminopteridine group of aminopterin and methotrexate or the pyrido[2,3-*d*]pyrimidine moiety of piritrexim (see **Figure 19**). This bicyclic group is usually functionalised at the C6-position with a side chain which mimics the folate *p*-aminobenzoyl (*p*ABA) group. The *de novo* synthesis of these types of compounds has been extensively explored, with analogues bearing a number of different functionalities at the C6-position having been synthesised^{138, 139, 140, 141}. An overview of the known synthetic strategies relevant to this project is shown in **Figures 20** and **21**.

Synthetic routes towards folic acid analogues can be generally categorised by the stage of the synthesis at which the *p*ABA mimic is installed at the C6-position of the pterin or pyrido[2,3-*d*]pyrimidine system. Strategies developed by DeGraw¹³⁹ involve the synthesis of 2-amino-3-cyanopyridines bearing the desired *p*ABA derivative at the 5-position, **12**. These intermediates are then subjected to an annulation reaction with guanidine to give the desired C6-functionalised 2,4-diaminopyrido[2,3-*d*]pyrimidines **13**. Other strategies, such as those deployed by Piper¹⁴² and Gangjee¹⁴⁰, involve the generation of 2-amino-3,5-dicyanopyridine or 2-amino-3,5-dicyanopyrazine, **14**, as substrates for the guanidine annulation reaction. The resulting 6-cyano-2,4-diaminopteridines and 6-cyano-2,4-diaminopyrido[2,3-*d*]pyrimidines, **15**, are then converted in three steps to a key methylbromo intermediate, **18**. This compound is then used in nucleophilic substitution reactions with an amine carrying the appropriate functionality, and there are numerous examples of amine nucleophiles which will participate^{136, 140}. This strategy is attractive in that the acquisition of the methylbromo compounds, **18**, allows the facile generation of a library of 2,4-diaminopterins and 2,4-diamino-5-deazapterins, **19**, with a variety of functionalities installed at the C6-position. It has also been shown by Piper¹⁴² and others¹⁴¹ that it is possible to obtain the 2-amino-4-hydroxypteridine equivalents, **20**, by refluxing the corresponding 2,4-diaminopteridines, **19**, in 1 M NaOH. Alternative routes have been described by Yang¹⁴³ and Rosowsky¹⁴⁴ which proceed via reductive amination of the appropriate pteridine-6-carboxyaldehyde with an amine coupling partner. However this approach is somewhat more laborious since the pterin C2-amino group must be protected in order to prevent self-condensation.

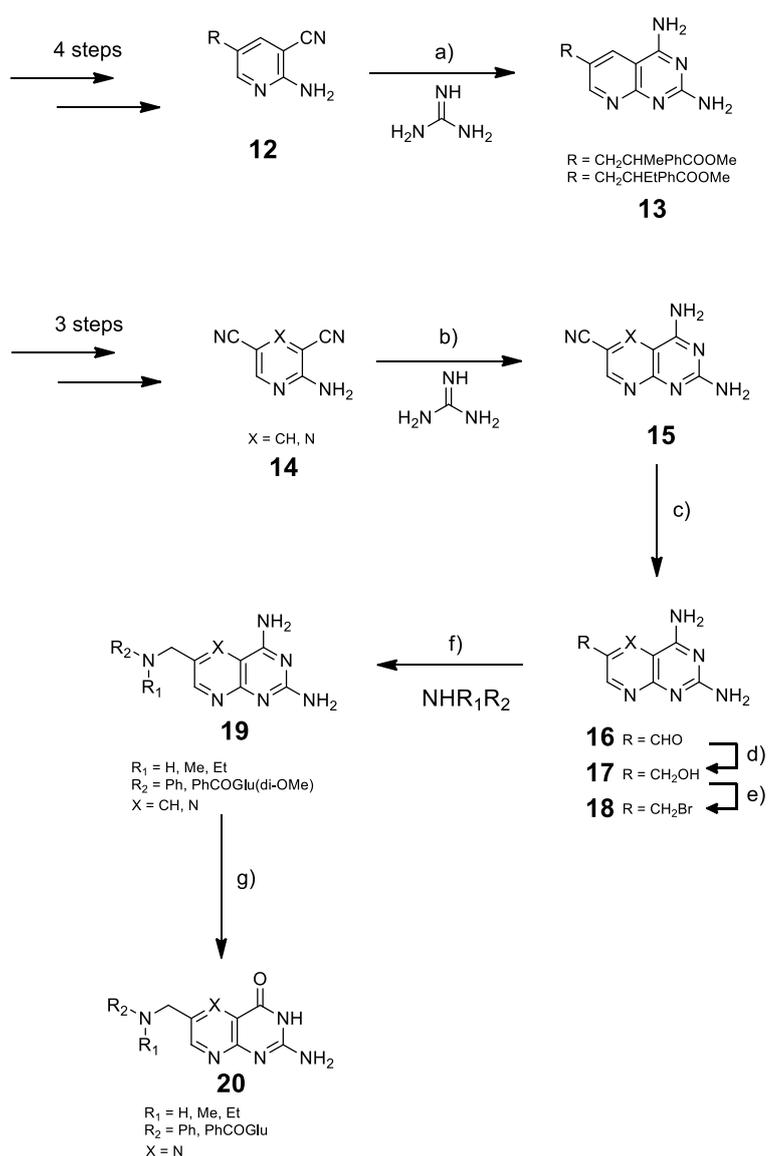


Figure 20 – Overview of synthetic routes to 6-substituted 5-deazapteridines; a) 2-methoxyethanol, 120 °C, 24 h (Sirotnak *et al*)¹³⁹; b) NaEtOH, EtOH, reflux, 12 h; c) Raney Ni, HCO₂H 80 °C, 2 h; d) NaBH₄, MeOH, rt, 2h; e) Ph₃PBr₂ Me₂NAC, rt, 2h; f) Me₂NAC, rt 24 h (Sirotnak *et al*)¹⁴² and Hallberg *et al*)¹⁴⁵; g) NaOH, H₂O, reflux 4h (Sirotnak *et al*)¹⁴² and Montgomery *et al*)¹⁴¹).

One of the major drawbacks of all these strategies is the requirement of a nucleophilic amine to participate in either nucleophilic substitution or reductive amination, since this sets the requirement that the resulting pteridine or pyrido[2,3-*d*]pyrimidine products contain a nitrogen atom at the 10-position. For folate analogues bearing *pABA* mimics without the N10 functionality, an alternative strategy is required.

The synthesis of 6-substituted pyrido[2,3-*d*]pyrimidines can be achieved through condensation of α -substituted acroleins with an appropriate pyrimidine^{146, 147, 148}. These

reactions typically require forcing conditions, usually the reflux of the two condensation partners in water or acetic acid^{147, 148}. An overview of known strategies involving condensation reactions is shown in **Figure 21B**. One potential advantage of utilising condensation reactions to build scaffolds of this type is that once the desired acrolein condensation partner has been generated, it can be reacted with a variety of pyrimidine compounds to generate ‘keto-amino-faced’, ‘diketo-faced’ or ‘diamino-faced’ pyrido[2,3-*d*]pyrimidines (**Figure 21A**). The use of acroleins bearing leaving groups such as β -amino or β -methylsulphonyl moieties has been explored by Stark¹⁴⁷ and Kuwada¹⁴⁶, and these compounds have been utilised in the synthesis of ‘keto-amino-faced’ pyrido[2,3-*d*]pyrimidines (**22 & 24**) and ‘diketo-faced’ pyrido[2,3-*d*]pyrimidines (**21 & 23**) bearing aliphatic and aromatic substituents at the C6-position. However, the introduction of the β -amino and β -methylsulphonyl moieties introduces an additional two steps to the synthesis. Alternatively, it has been shown by Troschutz¹⁴⁸ that the condensation-annulation reaction will proceed in neat acetic acid without the use of a leaving group and this was demonstrated for the synthesis of **25**. However, the substrate scope of this reaction has not been properly established, and the extent to which the pyrimidine condensation partner and the acrolein α -substituent can be varied is not clear.

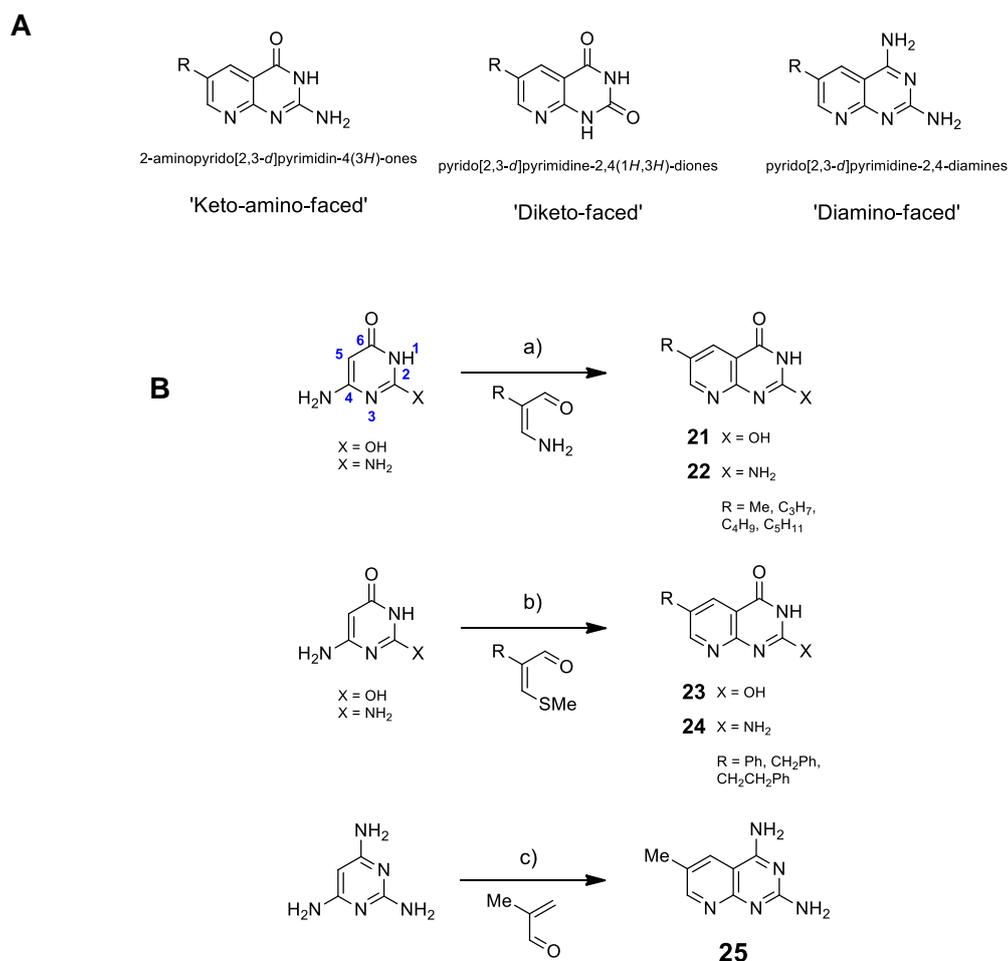


Figure 21 – A) Structures, IUPAC nomenclature and short-hand labelling for pyrido[2,3*d*]pyrimidines. B) Overview of condensation reactions utilised in the synthesis of 6-substituted pyrido[2,3*d*]pyrimidines; a) water, piperidine acetate, reflux (Stark *et al*)¹⁴⁷; b) aqueous AcOH (60%), reflux 12 h (Nobuhiro *et al*)¹⁴⁶; AcOH, reflux 5 h (Troschutz *et al*)¹⁴⁸.

2.1.2 Aims and Objectives

The objectives of this project were divided into two strands (see **Section 1.8**). Firstly, the characterisation of the THF riboswitch *in vivo* by exploring both the wild-type ligand binding specificity and the role of the two ligand binding sites in mediating the regulatory response. Secondly, the re-engineering of the THF riboswitch to develop novel gene expression tools, based on orthogonal mutant riboswitch-ligand pairs. Both of these aims require the synthesis of THF analogues, the strategies for which are outlined in this chapter.

2.2 Results and Discussion

2.2.1 Riboswitch binding sites and ligand design

It was determined that the roles of the two binding sites in the riboswitch regulatory function could be elucidated using a series of riboswitch mutants and the commercially available compound folinic acid (a stable, close structural analogue of THF). This study is discussed in Chapter 3. However, for the exploration of ligand binding specificity, it would be necessary to synthesise a selection of THF analogues. These compounds were designed based on the available x-ray crystallography data¹⁰⁴, which identified key ligand structural features which participate in hydrogen bonding to the aptamer domain, as well as other ligand regions which were considered as candidates for modification. An overview of the folinic acid ligand binding motifs and the analogues designed for this study are shown in **Figure 22**. One of the desired outcomes of the ligand specificity study was the synthesis of a THF analogue which would both activate the wild-type riboswitch and also permeate the bacterial cell membrane without the use of a heterologously expressed transporter protein. With this in mind, compounds were designed without the highly negatively charged glutamate moiety, which was also not believed to be involved in interaction with the RNA based on the available structural data¹¹⁰. Other THF functionalities targeted for modification were the N5 position and the *p*-aminobenzoyl (*p*ABA) side chain. The N5 atom is known not to participate in binding to the THF aptamer domain, and replacement with a methylene group is known to impart a greater stability of the resulting compound towards aerial oxidation¹⁴⁹. Additionally, it has been reported by Batey *et al.* that whilst the *p*ABA moiety of THF/folinic acid does not appear to impact ligand binding affinity, its presence does appear to be a requirement to mediate the riboswitch regulatory effect¹¹⁰. Thus, analogues with a variety of side-chains at the C6-position were targeted for synthesis to explore this observation further (**Figure 22B**). The available *in vitro* data indicates that the difference in binding affinity and regulatory ability of the 6*R*- and 6*S*-diastereoisomers of THF is moderate¹¹⁰. Therefore, in order to simplify the synthesis of the desired analogues, the racemic mixtures of the corresponding 6*R*- and 6*S*-enantiomers were prepared.

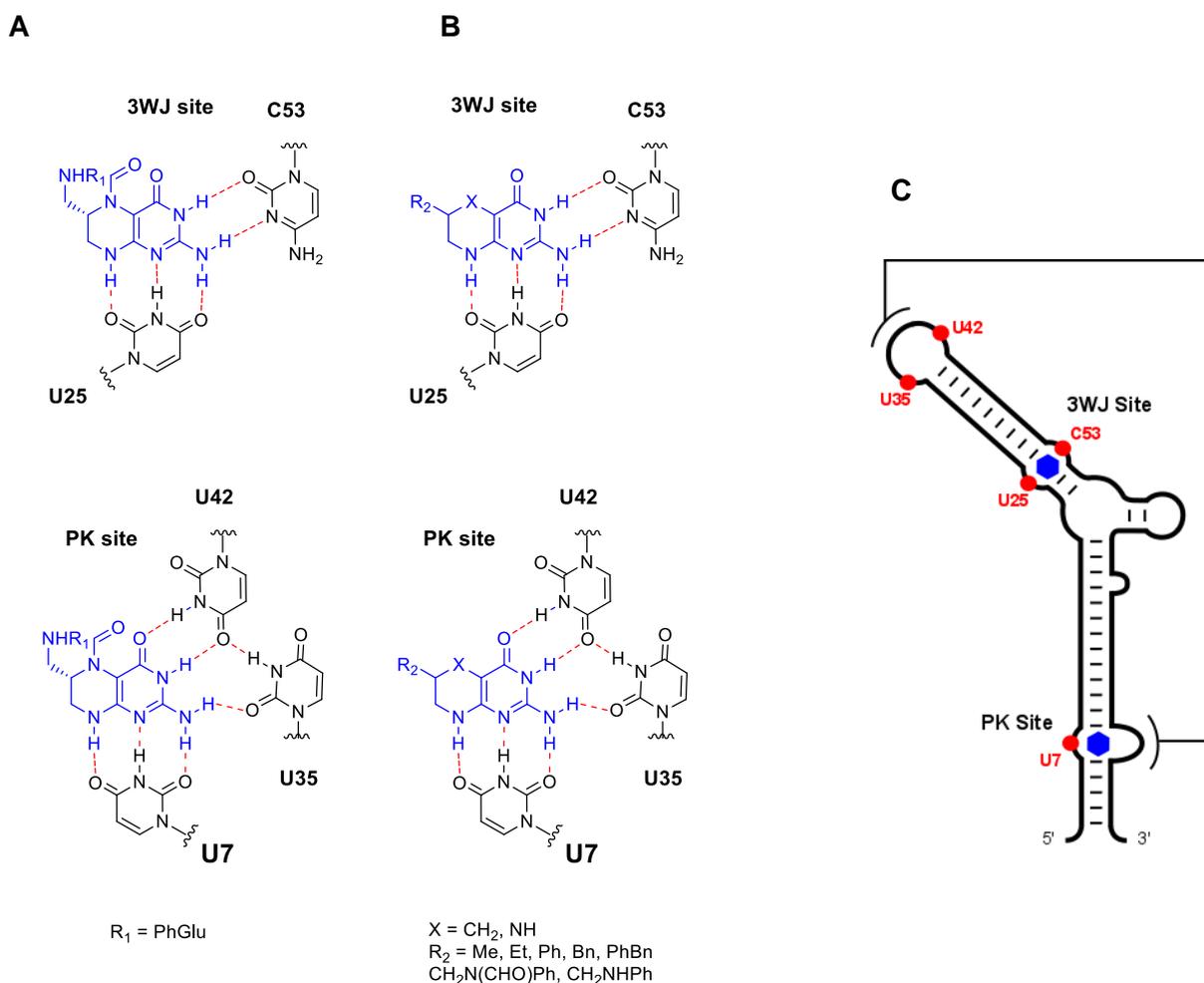


Figure 22 – Overview of known and predicted ligand binding motifs for the wild-type THF riboswitch. **A)** Folinic acid recognition by the wild-type THF riboswitch at the 3WJ and PK ligand binding sites (obtained by x-ray crystallography¹⁰⁴); **B)** Predicted ligand binding motifs for the THF riboswitch and synthetic folate analogues; **C)** Secondary structure of the THF riboswitch aptamer domain showing the 3WJ and PK ligand binding sites.

With respect to the second goal of this project – the re-engineering of the THF riboswitch to generate an orthogonal gene expression system – the available x-ray crystallography data was again used to govern the overall strategy, which is summarised in **Figure 23**. An orthogonal riboswitch system should respond to a non-natural ligand not present in the cell, whilst rejecting the natural compound which almost always would be present. The Micklefield laboratory has demonstrated with a number of riboswitch classes that by rationally introducing mutations into the ligand binding site it is possible to re-engineer them such that they reject their cognate ligand and instead respond to a non-natural synthetic compound. Since the THF riboswitch has two ligand binding sites which bind THF/folinic acid via two distinct motifs, it was considered that it may be possible to re-engineer the riboswitch to respond to two

distinct ligands (moreover, it was not possible to rationalise mutations which would allow a single synthetic ligand to bind each mutated site). An overview of this strategy is shown in **Figure 23**.

Two sets of mutant-ligand pairs were designed. Firstly, the double mutant **M1** containing the U25C and C53U mutations at the 3WJ site was designed to bind 'di-keto' analogues whilst rejecting THF/folinic acid (**Figure 23B, top**). The **M1** mutant was designed such that the PK site was unmodified. Secondly, the double mutant **M2** containing the U35C and U42C mutations at the PK site was designed to bind 'di-amino' analogues whilst rejecting THF/folinic acid (**Figure 23B, bottom**). The **M2** mutant was designed such that the 3WJ site was unmodified. It was envisaged that if this strategy was successful, the two sets of mutations could eventually be combined to create a quadruple mutant with two re-engineered binding sites which would no longer respond to reduced folates, but instead be activated by two distinct synthetic ligands. As a fall-back strategy, if re-engineering was only successful at one of the two sites, then a mutation could be introduced at the non-engineered site to abrogate binding to THF/folinic acid, thereby creating an orthogonal mutant riboswitch which is activated by a single synthetic ligand^{79, 81, 86}. As with the THF analogues designed to bind the wild-type riboswitch described in **Figure 22B**, the two series of compounds were designed with a variety of side chains at the C6-position as well as both N5 and 5-deaza functionalities (**Figure 23B**). Again, racemic mixtures of the two possible enantiomers were used since enantiopurity has previously been shown to only moderately impact ligand binding and regulatory ability¹¹⁰ and a more straightforward synthesis was desired.

A**B**

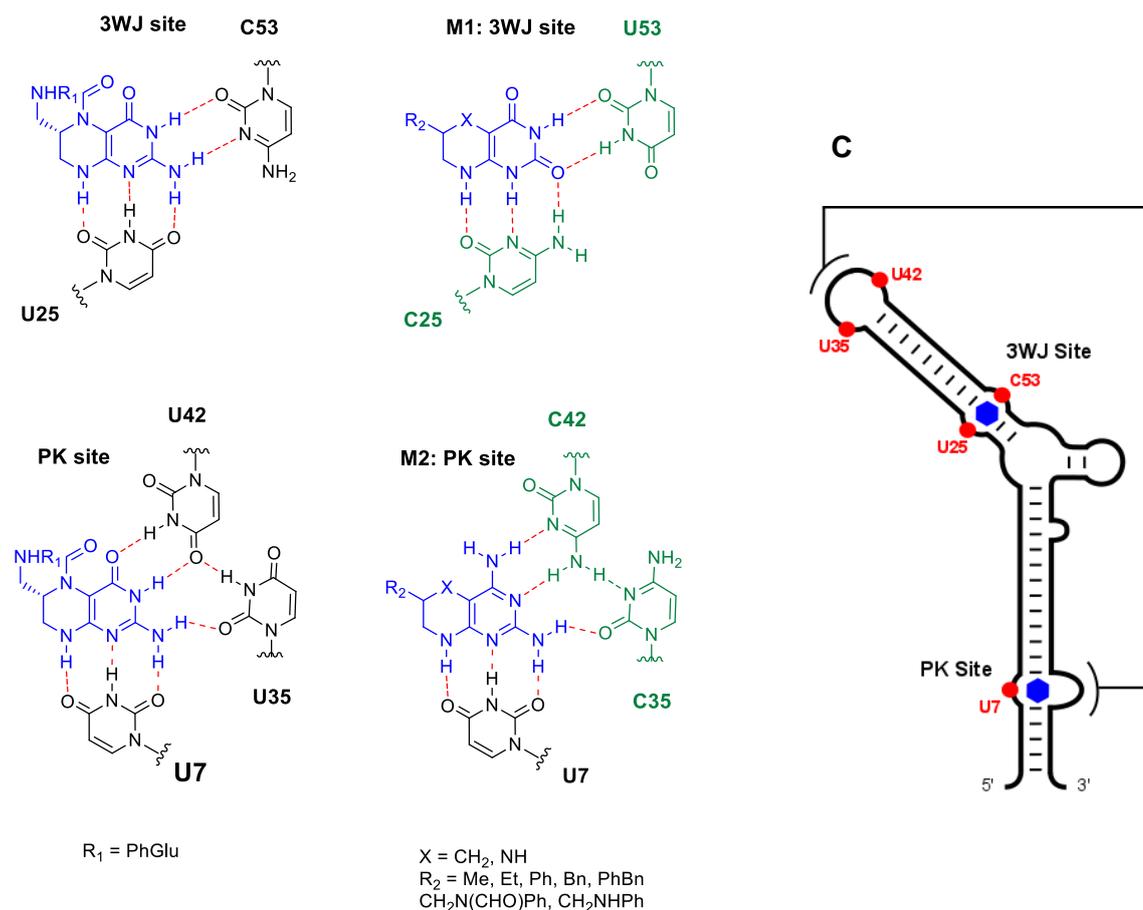


Figure 23 - Overview of known and predicted ligand binding motifs for the wild-type, M1 (U25C/C53U) and M2 (U35C/U42C) THF riboswitches. A) Folinic acid recognition by the wild-type THF riboswitch at the 3WJ and PK ligand binding sites (obtained by x-ray crystallography¹⁰⁴); B) Predicted ligand binding motifs for THF riboswitch mutants and synthetic folate analogues. Top: Predicted ligand binding motif for the M1 (U25C/C53U) riboswitch and ‘di-keto’ THF analogues. Bottom: Predicted ligand binding model for the M2 (U35C/U42C) riboswitch and ‘di-amino’ THF analogues. C) Secondary structure of the THF riboswitch aptamer domain showing the 3WJ and PK ligand binding sites.

2.2.2 Synthesis of 5,6,7,8-tetrahydropteridines

Synthesis of the 5,6,7,8-tetrahydropteridines was attempted using a two-step strategy, combining syntheses previously outlined by Palanki¹⁵⁰, Konrad¹⁵¹ and Kenyon¹⁵². An overview is shown in **Figure 24**. The strategy involved the synthesis of pteridine compounds via known condensation-annulation reactions from the appropriate pyrimidine, followed by attempted hydrogenation with the goal of obtaining the 5,6,7,8-

tetrahydro- derivatives. Accordingly, pteridines **28** and **29** were synthesised by condensation of the appropriate tetrasubstituted pyrimidines (**26** and **27**) with glyoxal in the presence of base^{150, 152}. Lumazine, **30**, was commercially available and thus not synthesised in this manner. The reduced pteridine **31** was successfully synthesised in good yield by hydrogenation with H₂ and Adam's catalyst as a suspension in water over 24 hours¹⁵¹. Pterin **29** and lumazine **30** were also hydrogenated under similar conditions also outlined by Konrad¹⁵¹, using initially water, and then TFA as a solvent due to very poor substrate solubility.

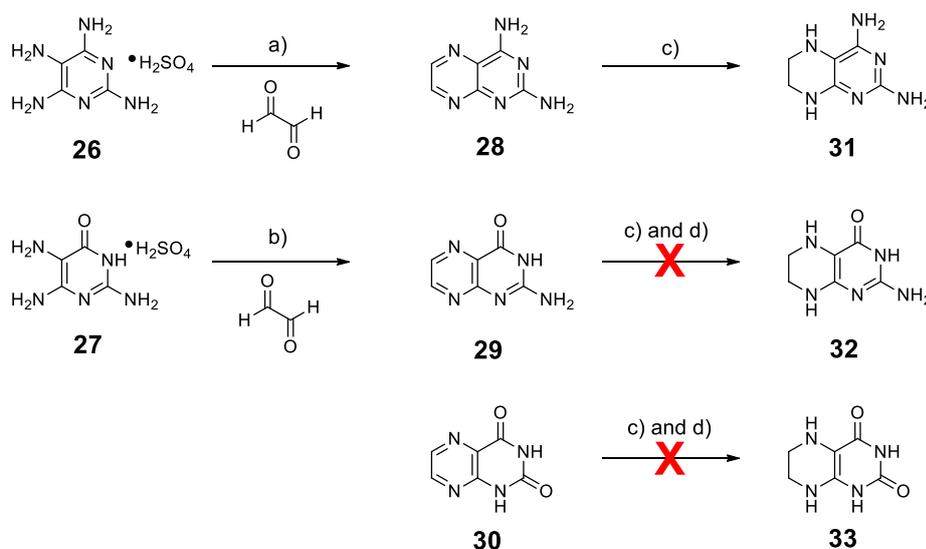


Figure 24 – Synthetic strategy for the preparation of 5,6,7,8-tetrahydropteridines; a) NaHCO₃, H₂O, 70 °C, 3 h; b) KOH, H₂O, RT, 24 h; c) H₂, PtO₂, H₂O, RT, 24 h; d) H₂, PtO₂, TFA, RT, 24 h.

However, it was not possible to isolate the desired reduced products **32** and **33** following the published protocols. Monitoring these reactions by TLC indicated new product formation and complete consumption of starting material in both cases, however TLC, mass spectrometry and proton NMR analysis of the crude reaction residue following the removal of the solvent indicated the presence of starting material only. It is assumed that this is a result of rapid aerial oxidation of the reduced compounds, which has been previously reported¹⁵³. Attempts to isolate **32** and **33** under inert gas conditions were unsuccessful, as were attempts to trap the reduced products by N5-acetylation with acetic anhydride as reported by Pfeleiderer¹⁵⁴. Following this, focus was switched to N5-deazapteridine analogues (5,6,7,8-tetrahydropyrido[2,3-d]pyrimidines, see **Figure 25**), which are reportedly much more stable and are not susceptible to rapid oxidation back to the fully aromatic compound¹⁵⁵.

2.2.3 Synthesis of 5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines

The synthesis of 5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines was accomplished using a similar two-step strategy to the 5,6,7,8-tetrahydropteridines, whereby condensation-annulation reactions were utilised to generate aromatic precursor compounds, which were then reduced by hydrogenation.

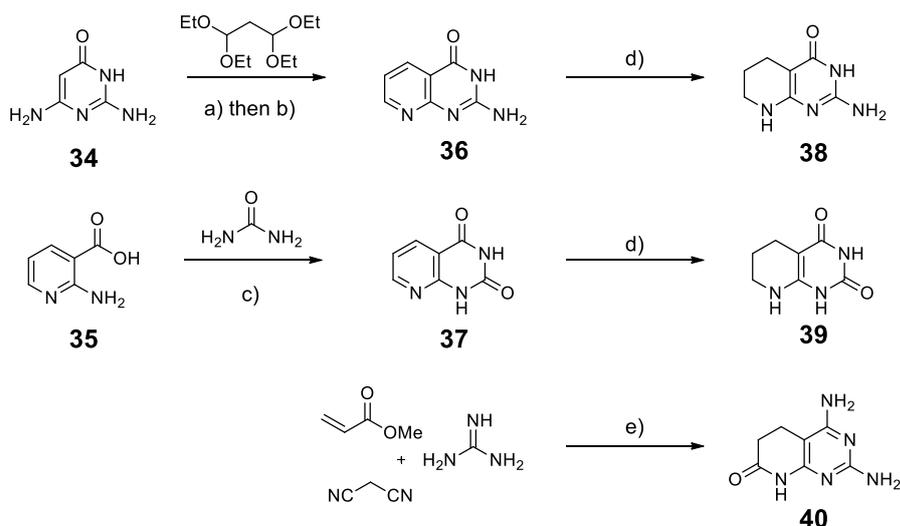


Figure 25 – Synthetic strategy for the preparation of 5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidine analogues. a) AcOH, H₂O, reflux 100 °C, 3.5 h; b) AcOH, reflux 120 °C, 10 h; c) No solvent, reflux 150 °C, 12 h; d) H₂, PtO₂, TFA, RT, 24 h; e) NaOMe, MeOH, microwave, 140 °C, 10 mins.

A summary of the synthesis is found in **Figure 25**. Compounds **36** and **37** were synthesised using reactions outlined in the literature by Pfeleiderer¹⁵⁶ and Li¹⁵⁷ respectively. Reflux of pyrimidine **34** and 1,1,3,3-tetraethoxypropane in water/acetic acid effected successive nucleophilic substitution and elimination reactions to furnish **36**. Synthesis of **37** was achieved by refluxing 2-aminonicotinic acid, **35**, in liquid urea at 150 °C. Compounds **36** and **37** were both successfully hydrogenated to their 5,6,7,8-tetrahydro- derivatives **38** and **39** using a modification of the procedure patented by Kubota¹⁵⁸. The published protocol was modified to increase the loading of Adam's catalyst to 0.2 molar equivalents and the reaction time was increased to 24 h. Compound **40** was accessible via a three-component, one-pot condensation reaction reported by Mont¹⁵⁹. A solution of guanidine in methanol was prepared by dissolving guanidine carbonate in NaOMe/MeOH and removing the resulting Na₂CO₃ precipitate

by filtration. Upon addition of methyl acrylate and malonitrile the solution was subjected to microwave radiation which afforded compound **40**. Attempts to remove the carbonyl moiety from compound **40** by reduction with $\text{BH}_3\cdot\text{THF}$ as previously reported for similar compounds by Teixido¹⁴⁹ were not successful.

2.2.4 Synthesis of C6-alkyl-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines

The synthesis of the C6-alkyl-substituted 5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines was achieved following a similar two-step strategy as outlined in the previous sections, and is summarised in **Figure 26**. Synthesis of the pyrido[2,3-*d*]pyrimidines **43-46** was achieved by modification and substrate scope expansion of the method reported by Troschutz¹⁴⁸. The Troschutz procedure outlines the synthesis of **43** by condensation of 2,4,6-triaminopyrimidine with α -methylacrolein. This methodology was extended to include pyrimidines **34** and **42** and modified to accommodate their differing reactivities and solubilities, initially using commercially available 2-methyl and 2-ethylacroleins as the condensation partners. These acroleins were refluxed with the appropriate pyrimidine in AcOH, or (in the case of 4-amino-2,6-dihydroxypyrimidine **42**) TFA due to poor substrate solubility.

The isolation and purification of the resulting compounds **43-46** was achieved through a number of different approaches, with poor compound solubility being the most significant hurdle. Compound **43** was isolated by removing the reaction solvent *in vacuo* and washing the resulting crude residue with copious amounts of various solvents. Attempts to re-crystallise the crude product from DMF were not successful, but the compound proved to be sufficiently soluble to attempt purification using C18 reverse phase chromatography in water/MeOH. Although this did not completely remove the 2,4,6-triaminopyrimidine starting material from the sample, the procedure yielded a product of sufficient purity such that the presence of **43** could be confirmed by ¹H-NMR and mass spectrometry analysis. This partially purified product was taken forward to the next step.

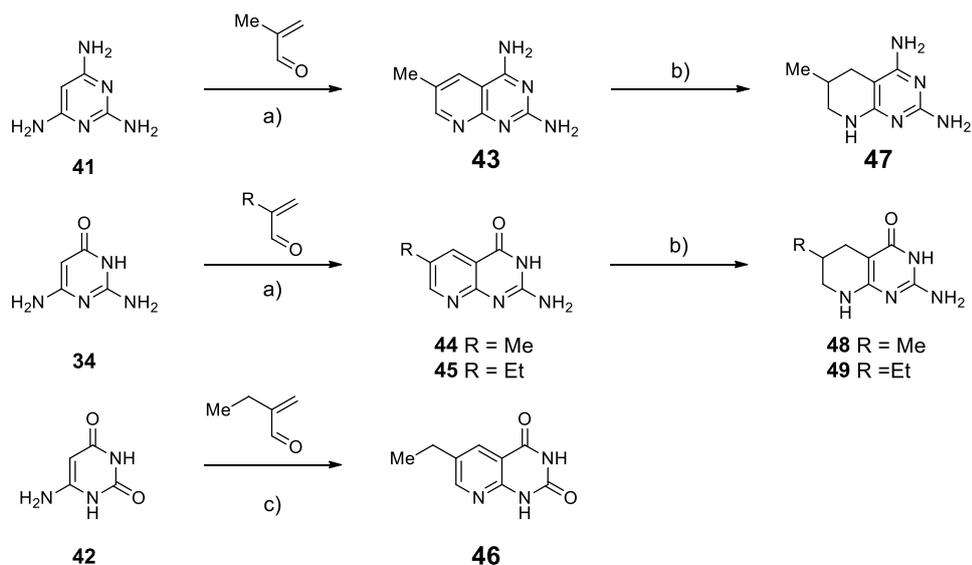


Figure 26 – Overview of the synthetic strategy for the synthesis of 6-alkyl-2,4-disubstituted-5,6,7,8-tetrahydropyridopyrimidine analogues. a) AcOH, reflux, 12 h; b) H₂, PtO₂, TFA, 24 h; c) TFA, reflux, 12 h.

In the case of compounds **44** and **45**, the products precipitated from refluxing AcOH and this precipitate was then isolated by centrifugation and washed with myriad solvents to remove impurities. This procedure gave analytically pure samples of **44** and **45** as confirmed by NMR and mass spectrometry. Compound **46** was precipitated from the reaction mixture by adding diethylether, washing the precipitate with copious amounts of diethylether, MeOH and DCM and then recrystallizing from DMF to furnish **46** in high purity. Reactions of α -ethylacrolein with **41** and α -methylacrolein with **42** yielded complex, inseparable mixtures of products and starting materials, and consequently these compounds were not pursued further.

Following isolation and purification of intermediates **43-46**, the compounds were taken forward for hydrogenation using Adam's catalyst in TFA to yield the desired reduced compounds **47-49**. Reactions were monitored by TLC (5:1, CHCl₃/MeOH) for consumption of starting material before centrifugation to remove the catalyst and removal of the solvent *in vacuo*. Analytically pure samples were obtained by recrystallising the crude reaction residues from MeOH. Due to time constraints, the hydrogenation of **46** was not successfully completed.

2.2.5 Synthesis of α -alkylphenyl-substituted acrylaldehydes

The diversification of the Troschutz methodology was further explored by synthesising a library of α -substituted acroleins as potential pyrimidine condensation partners. An overview of this synthesis is shown in **Figure 27**. The synthetic strategy was built around a procedure previously reported by Erkilla¹⁶⁰, who had demonstrated that an α -methenyl group can be installed on a variety of aliphatic aldehydes using an aldol/mannich reaction with formaldehyde. This methodology was first applied to the two commercially available aldehydes **48** and **49** to give the corresponding 2-substituted acrylaldehydes **51** and **52**. Reactions were monitored for the consumption of the aldehyde by TLC (10:1, Hex/EtOAc) and worked up by extraction into ethyl acetate. Compound **52** was then successfully purified by flash chromatography and confirmed by NMR and mass spectrometry analysis. The formation of **51** was confirmed by NMR and mass spectrometry analysis of the crude product; however this compound proved to be unstable (previously noted in the literature¹⁶¹) and decomposed during attempts at flash chromatography. Due to its instability, crude **51** was carried forward into the subsequent condensation reaction without further purification, and was freshly prepared immediately prior to use.

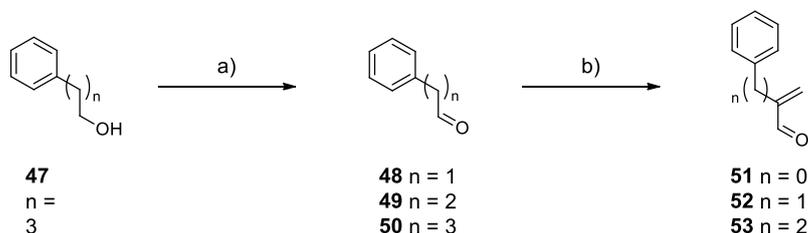


Figure 27 –Overview of the synthesis of 2-alkylphenyl-substituted acrylaldehydes. a) DMSO, (COCl)₂, DCM, -78 °C, then Et₃N; b) pyrrolidine, N,N-dimethyl-4-aminobenzoic acid, HCHO, DCM, 40 °C.

Aldehyde **50** was accessible via Swern oxidation of the corresponding alcohol **47** using standard conditions. Reactions were monitored by TLC (10:1, Hex/EtOAc) for the consumption of the alcohol and on completion were extracted with ether and the solvent removed *in vacuo*. The presence of **50** in the crude reaction residue was confirmed by mass spectrometry and proton NMR analysis and integration of the aldehyde proton resonance indicated a crude yield of 93%. The crude aldehyde product was then used in the α -methenylation reaction without further purification. Indeed, it was possible to extend the Erkilla methodology to install the α -methenyl

group on **50** to give **53**, which was purified by flash chromatography to give an isolated yield of 39% over the two steps.

2.2.6 Synthesis of C6-alkylphenyl-substituted pyrido[2,3-*d*]pyrimidines

Condensation of the synthesised α -alkylphenylacrylaldehydes **51-53** with the pyrimidines of interest was then extensively investigated. As with the aliphatic condensations described in the previous section, the reactivity and solubility of the pyrimidines as well as the solubility of the condensation products defined the optimisation process. In particular, one of the challenges in optimising this procedure was the very poor solubility of the condensation products, which prohibited the use of flash chromatography or indeed recrystallization for isolation and purification. Indeed, the large number of hydrogen bond donors/acceptors as well as the potential for π - π stacking interactions is thought to lead to aggregation – and hence poor solubility – of many of the compounds discussed in this chapter. The synthetic strategy employed in the preparation of these compounds is outlined in **Figure 28**. Initial investigations into the condensation of **34** with **51** using the previously employed 1:1 molar stoichiometry of the condensation partners gave the desired product as determined by mass spectrometry and the appearance of new heteroaromatic resonances on the proton NMR. However, isolation of the product **54** proved challenging using these conditions. A 1:1 stoichiometry did not allow the complete consumption of the pyrimidine starting material **34**. Additionally, multiple additional side products were detected. Molecular masses which would correspond to imines formed by the condensation of the aldehyde moiety of **51** with the C2-amino and C4-amino groups of pyrimidine **34** were detectable by mass spectrometry. It was not possible to separate these multiple products upon evaporation of the solvent, again due to very limited solubility. Two modifications were developed to address these issues.

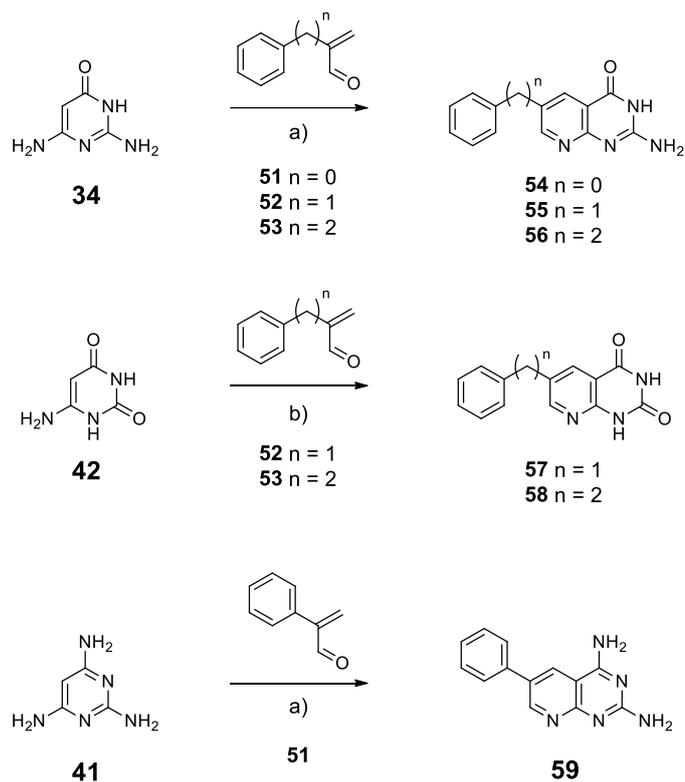


Figure 28 – Condensation reactions for the synthesis of 6-alkylphenyl-substituted pyrido[2,3-d]pyrimidines; a) AcOH, reflux 6 h; b) TFA, reflux 6 h.

Firstly, a slight excess (1.1 equivalents) of **51** was used to promote complete consumption of the pyrimidine starting material. Secondly, the insolubility of the desired product was exploited to facilitate isolation. Dissolving **34** in minimum refluxing AcOH before addition of the **51** encouraged the precipitation of the desired product **54** from refluxing solution, which was then isolated and washed extensively with multiple solvents to remove any residual impurities. This approach also proved successful for the synthesis of the **55** and **56**. Similar modifications were made for the synthesis of compounds **57-59**. The pyrimidine **42** is very poorly soluble - even in refluxing AcOH - and instead TFA was used as a solvent. The use of TFA did, however, prohibit the precipitation of the desired products in this case. To allow the isolation of compounds **57-58**, the reaction was monitored by TLC until the pyrimidine was completely consumed. The reaction mixture was then cooled to room temperature and diethylether was added dropwise until a precipitate formed, which was isolated by centrifugation and washed with multiple solvents to remove excess impurities. Finally, the synthesis of **59** was achieved following the same procedure as described for compounds **54-56**, however in order to isolate and purify this compound, it was necessary to remove all

the reaction solvent *in vacuo* and dissolve the crude residue in EtOH. The insoluble particles were then isolated by centrifugation and washed repeatedly with multiple solvents.

Condensations of **42** and **51**, **41** and **52**, and **41** and **53** were not successful using the aforementioned methodology. In each case, isolation of the precipitate formed during the reaction yielded a highly insoluble mixture of products that were not separable. In the case of the condensations involving **41**, one possibility is that additional condensations between the C2- and C6-amino groups and the acroleins **52** and **53** resulted in multiple additional side-products. The reaction procedure was modified further, in an attempt to isolate the desired pyrido[2,3-*d*]pyrimidines. Firstly, the reaction times were increased to up to 48 h in an attempt to drive the reaction through to the formation of the bi-cyclic aromatic product. This was not successful. Secondly, the volume of AcOH used in the reaction was sequentially increased, with the goal of encouraging the selective precipitation of the desired pyrido[2,3-*d*]pyrimidines whilst retaining the unwanted side-products in solution. This was again not met with success and compounds **54-59** were prioritised and taken forward to the final stage of the synthesis.

2.2.7 Synthesis of 6-alkylphenyl-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines

The reductions of compounds **54-56** were first investigated using the hydrogenation conditions that had proved to be successful in the synthesis of the 5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines described in the previous section, using platinum dioxide and hydrogen gas in TFA (see **Figure 29**). Reactions were monitored by TLC, which indicated the formation of two new products after approximately 4 hours. The reaction was continued until the starting material was completely consumed and at this point only one additional product was detected by TLC. The crude product was isolated and analysed by mass spectrometry to reveal a compound with a molecular weight of six mass units larger than expected. Re-crystallisation from methanol allowed for proton NMR analysis which revealed a number of unexpected aliphatic resonances and a lack of aromatic protons which would be expected for **54-56**. These data indicate that the hydrogenation was not selective and that the phenyl ring of the side-chain was reduced to cyclohexyl under these conditions. To investigate this further, the hydrogenation of **56** was monitored over time by mass spectrometry with the goal of

identifying a suitable endpoint such that the reaction could be halted before the reduction of the phenyl ring occurred (assuming that the reduction of the pyridine moiety proceeded more rapidly). Aliquots were taken every 30 minutes for a total of four hours and samples were analysed using mass spectrometry with electrospray ionisation in positive mode. This revealed the formation of both the desired 5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidine, **60**, and the unwanted side product containing the reduced phenyl ring, **61**, after only 30 minutes. In addition to monitoring over time, reduced PtO₂ loadings of as low as 0.1% were also explored, with no success. Attempted separation of the two products also proved problematic, with compound solubility again proving to be a major stumbling block. Separation was attempted using flash chromatography (CHCl₃/MeOH), C18 reverse phase column chromatography (water/MeOH) and finally C18 reverse phase HPLC, without success.

Alternative catalysts were explored including the use of Pd/C at 1-10% loadings. At 10% Pd/C loading, the reaction was monitored for the consumption of the starting material as indicated by TLC, at which point the reaction was worked up. Analysis by proton NMR estimated the ratio of **60:61** in the reaction mixture at 3:2. Purification of this crude mixture was again attempted by flash chromatography, C18 reverse phase column chromatography and C18 reverse phase HPLC, without success. At this point, the discontinuation of this synthetic route was considered for a number of reasons. Firstly, the solubility of the products of the hydrogenation reaction in water, DMSO and 1 M NaOH was very limited (< 1 mM), which limits their use in the *in vivo* assay. Secondly, the inability to separate the desired products **60** from side-products **61** would prevent meaningful conclusions being drawn about the structure-activity relationship between these compounds and the THF riboswitch. Finally, the reduction reactions were also attempted using **54**, **55** and **57-59** as substrates, giving the same mixture of products which were not able to be separated using the aforementioned techniques. The synthesis of this compound series was not pursued further, however the 3:2 crude mixture of **60** and **61** was taken forward for testing *in vivo*.

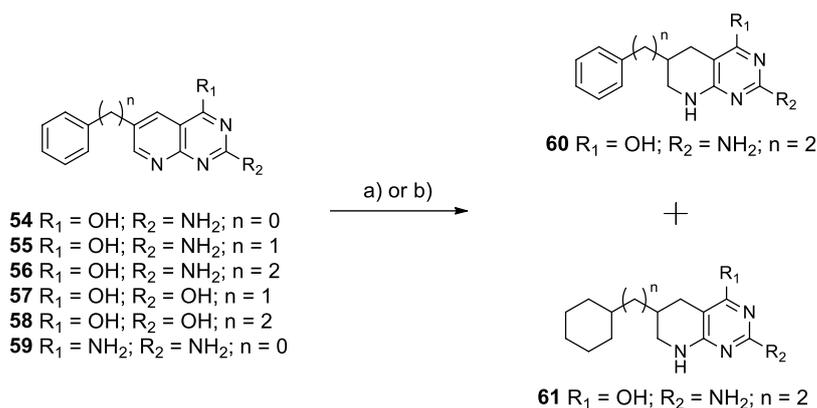


Figure 29 – Overview of the synthesis of 6-alkylphenyl-5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines. a) H_2 , PtO_2 , TFA; b) H_2 , Pd/C, TFA.

2.2.8 Synthesis of N-((2-amino-4-oxo-3,4,5,6,7,8-hexahydropyrido[2,3-*d*]pyrimidin-6-yl)methyl)-N-phenylformamide

The synthetic strategy for the preparation of **70** was designed around the condensation reaction which had proven successful in the synthesis of the 6-substituted pyrido[2,3-*d*]pyrimidines described in the previous section. The synthetic strategy can be found in **Figure 30**. It was envisaged that condensation would be the crucial step in the reaction series and therefore the synthesis of gram quantities of **66** was prioritised, to allow the condensation reaction to be optimised. **66** was accessible in four steps from commercially available starting materials. In the first step, copper-catalysed *N*-arylation of 3-amino-1-propanol following a procedure reported by Yin¹⁶², effected the formation of 3-anilino-1-propanol **62**. This compound was then subjected to the Swern oxidation and α -methenylation procedure described for the previous compound series to yield **66**. However, when this compound was subsequently subjected to the condensation conditions (reflux in AcOH) with the pyrimidine **34**, the reaction mixture rapidly turned dark and viscous. Although no analysis of this mixture was possible, it is assumed that this was a result of polymerisation, with the unprotected amine reacting with the electrophilic aldehyde either directly or via Michael addition. To alleviate this, it was first attempted to boc-protect **66** directly, however this produced a complex and inseparable mixture of products as detected by TLC. Instead, compound **62** was selectively *N*-boc protected by stirring overnight with Boc anhydride in ethanol without the addition of base to give **63**. This was then subjected to Swern oxidation to furnish **65** and α -

methenylation to give **67**, which was purified by flash chromatography and fully characterised by NMR and mass spectrometry.

Condensation of **67** with pyrimidine **34** was achieved by dissolving **34** in minimum refluxing acetic acid, adding 1.1 equivalents of **67** and refluxing for 18 hours. As with the previous condensation reactions, the condensation product precipitated from solution after approximately 1 hour, however isolation at this point resulted in an inseparable mixture of *N*-boc and boc-protected condensation products. Extending the reflux time to 18 hours encouraged complete boc-deprotection to give a precipitate of **68** in 29% yield. The very limited solubility of this compound again posed challenges for purification, with standard chromatography methods not being possible. Re-precipitation from sodium hydroxide solution by neutralisation was not successful. A slightly higher purity was obtained, by re-crystallising from DMF, although this resulted in the loss of approximately 90% of material so this procedure was not utilised at scale. The best compromise of yield and purity was reached by washing the crude precipitate with copious amounts of a number of solvents of varying polarity, including acetic acid, water, methanol, dichloromethane, acetone, ethyl acetate and diethylether.

The corresponding condensation reaction of **67** and pyrimidine **42** was also attempted using minimum TFA as a solvent due to the even poorer solubility of **42**. However, this resulted in the formation of a dark and viscous solution characteristic of polymerisation. It is possible that the more rapid boc-deprotection of **67** in TFA (compared with acetic acid) lead to a greater concentration of **66** in the reaction mixture and polymerisation ensued.

Exposure of **68** to the previously utilised hydrogenation conditions, using Adam's catalyst in TFA, lead to the benzylic cleavage of C9-N10 bond to generate compounds **72** and **73**, which were identified by mass spectrometry. Hydrogenolysis is a well-established method of removing *N*-benzyl protecting groups and benzylic cleavage has been reported for compounds similar to **68** by Singh¹⁶³. In the same study, the group also reported a procedure to circumvent this problem and this was utilised in this synthesis. *N*10-formylation was effected by stirring **68** in a formic acid/acetic anhydride mixture for three hours and then precipitating the resulting product **69** with cold ether. Hydrogenation of **69** was achieved using Adam's catalyst and H₂ gas with TFA as a solvent. After 18 hours, TLC analysis indicated the consumption of the starting material and proton NMR and mass spectrometry analysis of the crude material confirmed the presence of the desired compound **70**. A small amount of over-reduced compound

(containing a cyclohexyl side-chain) was detected by mass spectrometry and was estimated to constitute approximately 10% of the sample by NMR peak integration. The compound was purified by C18 reverse phase HPLC and fully characterised by NMR and mass spectrometry. Due to the lack of material available (~5 mg) it was decided not to proceed with the hydrolysis of the *N*-10 formyl group and compound **70** was used in the *in vivo* studies.

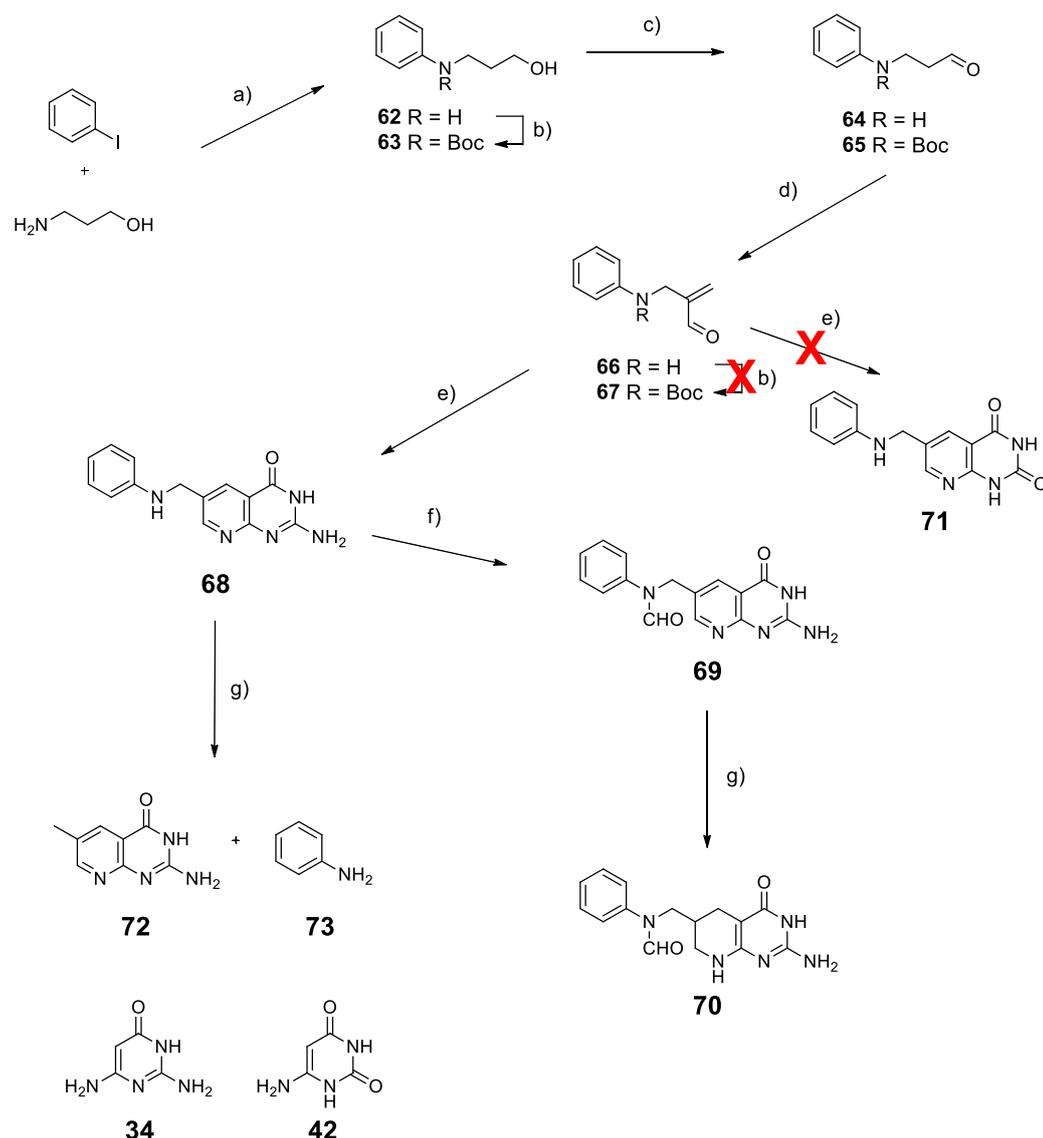


Figure 30 – Overview of the synthesis of *N*-((2-amino-4-oxo-3,4,5,6,7,8-hexahydropyrido[2,3-d]pyrimidin-6-yl)methyl)-*N*-phenylformamide. a) CuCl, KOH, RT, 12 h; b) Boc₂O, EtOH, RT, 24 h; c) (COCl)₂, DMSO, DCM, then Et₃N; d) pyrrolidine, *N,N*-dimethyl-4-aminobenzoic acid, HCHO, DCM, 40 °C; e) AcOH, reflux, 12 h; f) HCOOH, Ac₂O, RT, 3 h; g) H₂, PtO₂, TFA, rt, 48 h.

2.3 Chapter conclusions

This chapter has described the design and synthesis of a series of THF analogues to be screened using an *in vivo* reporter gene assay outlined in Chapter 3. Three classes of compounds were designed and representatives from each were successfully synthesised. The structures of these compounds are shown in **Figure 31**, as well as the nomenclature carried forward for use in the *in vivo* and *in vitro* studies.

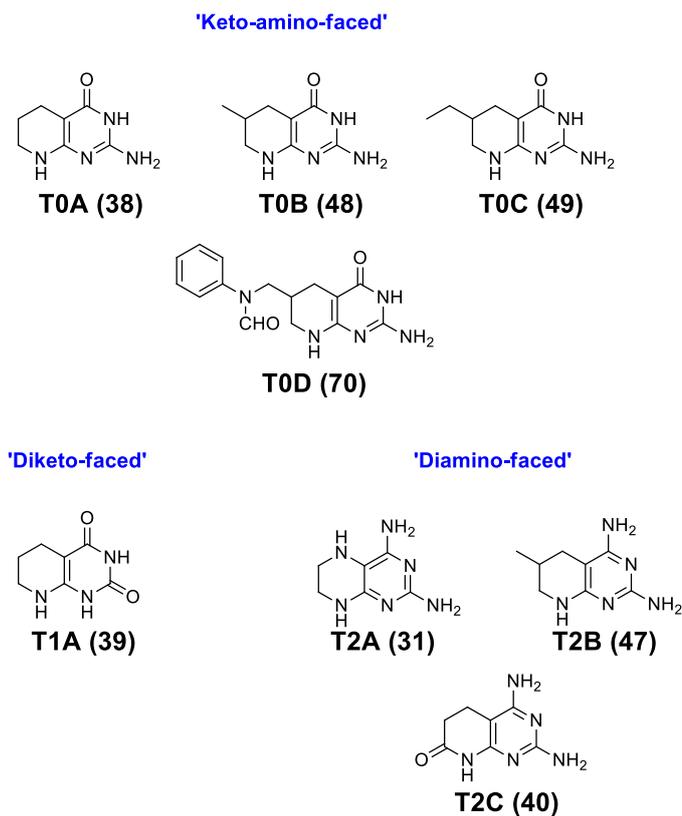


Figure 31 – Compounds successfully synthesised and carried forward for testing in the *in vivo* assay

Early investigations into the synthesis of 5,6,7,8-tetrahydropteridines were successful for the synthesis of 'diamino-faced' compound **T2A**, but due to susceptibility to aerial oxidation, the equivalent 'di-keto' and 'keto-amino-faced' compounds were not accessible. The synthesis of pyrido[2,3-*d*]pyrimidines was extensively explored, which in the case of the C6-substituted compounds pivoted around condensation reactions between a variety of acroleins and pyrimidines. The work described here represents an expansion of the substrate scope of this reaction¹⁴⁸. Commercially available α -methyl

and α -ethyl acroleins were first investigated as condensation partners, which established that the reaction could be applied to pyrimidines **34** and **42** as well as 2,4,6-triaminopyrimidine **41** which was known to participate¹⁴⁸. Additional acroleins were then synthesised with a number of different substituents at the α -position, including phenyl, benzyl and ethylphenyl and these were shown to undergo condensation with all three pyrimidines to give the corresponding pyrido[2,3-*d*]pyrimidines. The lack of selectivity of the subsequent hydrogenation and an inability to separate product mixtures due to poor solubility meant that the corresponding 5,6,7,8-tetrahydropyrido[2,3-*d*]pyrimidines were inaccessible.

Compound **T0D** was successfully synthesised in eight steps from commercially available starting materials. The pivotal step of the synthesis proved to be the condensation reaction, which was successful following the generation of the N-boc α -methylanilinoacrolein **67**. Hydrogenation of the resulting compound led to benzylic cleavage of the C9-N10 bond, however this problem was circumvented by installing a formyl protecting group on N10. **T0D** is a particularly interesting compound for testing in the *in vivo* assay since it allows the determination of the importance of the N10 atom in THF riboswitch activation.

Chapter 3

Engineering and characterisation of THF riboswitches *in vivo*

3.1 Chapter Introduction

The discovery of the THF riboswitch was first reported in 2010⁵⁷ and since then the published studies have focused on the *in vitro* characterisation of this system. Techniques such as in-line probing⁵⁷ and ITC^{104, 110} have been utilised to investigate the ligand binding specificity, whilst x-ray crystallography has been employed to explore aptamer domain structure and ligand binding motifs. Additionally, an *in vitro* transcription assay has been used to model the riboswitch regulatory effects¹¹⁰.

However there have so far been no reports of the *in vivo* characterisation of THF riboswitches. Furthermore, the published *in vitro* data leaves many questions about the riboswitch unanswered, not least around the discrepancy between ligands which are able to bind the THF aptamer domain (as determined by ITC) versus ligands which are able to effect a regulatory response *in vitro* (as determined by a transcription assay)¹¹⁰. For further details of these experiments, see **Section 1.7.5**. It has not yet been determined whether or not this discrepancy is a result of the chosen *in vitro* assay conditions or whether this represents a biologically relevant phenomenon. Additionally, the role of the two ligand binding sites within the THF aptamer domain has been established *in vitro*, but these effects have yet to be established in an *in vivo* setting and it is not yet clear if both sites do indeed contribute to riboswitch regulatory function *in vivo*.

3.1.1 Aims and Objectives

The objectives of this project were divided into two main goals (see **Section 1.8**). Firstly, the characterisation of the wild-type THF riboswitch *in vivo* by establishing a reporter gene assay in *E. coli*. The aim was to use THF analogues synthesised in Chapter 2 to probe the ligand binding and regulatory specificity of the wild-type riboswitch. It was also desired to investigate the role of the two THF ligand binding sites in mediating the regulatory function of the riboswitch. This was to be achieved by testing a series of mutations (identified by Batey *et al.*¹⁰⁴ as disrupting ligand binding at each of the two sites) in the *E. coli* reporter gene assay. The second main objective of the project was the re-engineering the THF riboswitch to develop an orthogonal, small-molecule-controlled gene expression system. Specifically, to identify mutations at each of the two riboswitch ligand binding sites which would provide the reversal of hydrogen bond directionality required to allow the binding of a distinct, synthetic compound at

each of the two binding sites. This was to be achieved by screening the rationally designed mutants in an *E. coli* reporter gene assay with a series of compounds synthesised in Chapter 2.

The structures of the compounds synthesised as described in Chapter 2 which were used in *in vivo* testing are shown in **Figure 31**. These compounds can be categorised into three groups based on their structure and intended application. Firstly, the 'keto-amino-faced' compounds (series **T0X**) were designed to probe the ligand binding and regulatory specificity of the wild-type THF riboswitch. It was hypothesised by Batey *et al.* that a structural requirement for ligands to induce a regulatory effect in the THF riboswitch is the presence of a *p*-aminobenzoyl (*p*ABA) or related moiety at the C6-position. Thus the **T0X** series was designed to comprise compounds bearing C6-position side-chains with increasing structural similarity to a *p*ABA group. An overview of their predicted binding motifs is shown in **Figure 32A**. Secondly, the 'diketo-faced' compound **T1A** was designed to bind to the mutated 3WJ site of the **M1** (U25C/C53U) THF riboswitch, for which the postulated binding motif can be found in **Figure 32B**. Finally, the 'diamino-faced' compounds (series **T2X**) were designed with the purpose of binding the mutated PK site of the **M2** (U35C/U42C) THF riboswitch. The posited binding motifs for these potential interactions are shown in **Figure 32C**.

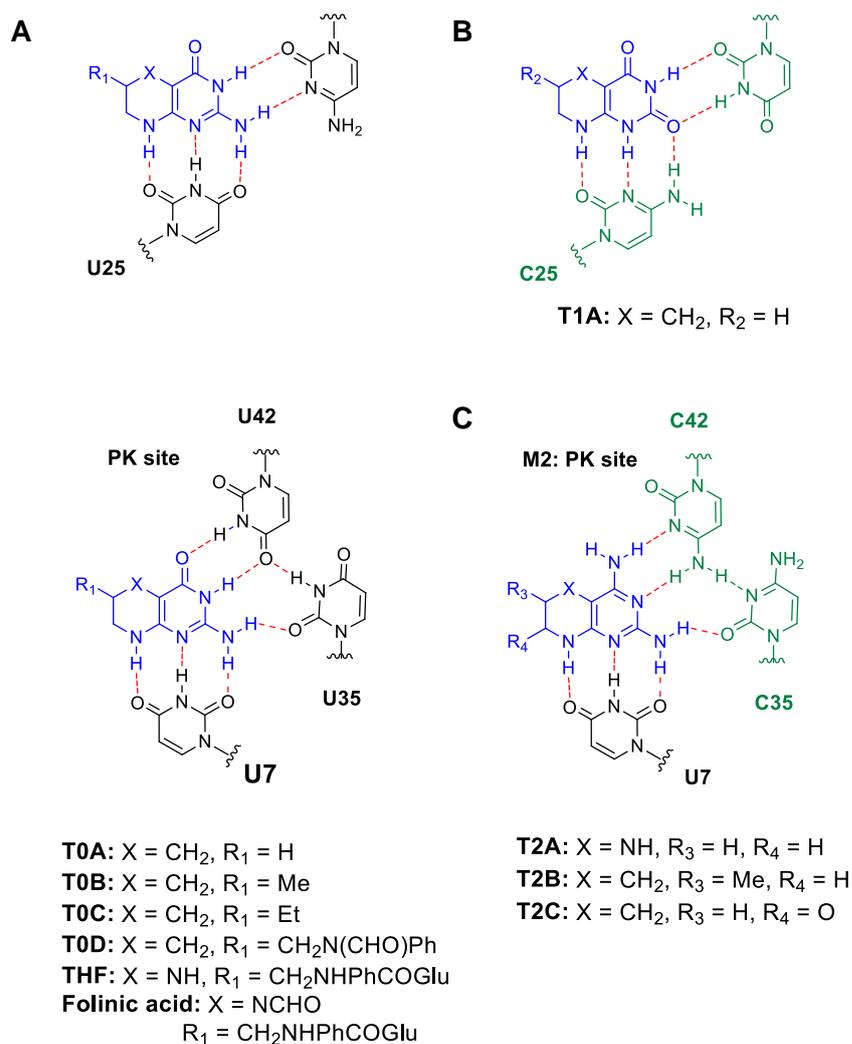


Figure 32 – Known and predicted binding motifs for THF riboswitch ligands based on x-ray crystallography data¹⁰⁴. A) Known (THF/folinic acid) and predicted (all others) binding motifs for ‘keto-amino-faced’ compounds with the wild-type riboswitch at both binding sites; B) predicted ligand binding motif for ‘diketo-faced’ compounds with the M1 (U25C/C53U) mutant riboswitch 3WJ site; C) predicted ligand binding motif for ‘diamino-faced’ compounds with the M2 (U35C/U42C) mutant riboswitch PK site.

3.2 Results and Discussion

3.2.1 Engineering and characterisation of a chimeric *folT/metE* THF riboswitch *in vivo*

The *in vitro* transcription assay, which has been used to explore the ligand specificity of the THF riboswitch by Batey *et al.*, utilised a chimeric riboswitch system whereby the aptamer domain of the *Streptococcus mutans* THF riboswitch (found in the 5'-UTR of the *folT* gene) was fused to the expression platform of the *Bacillus subtilis* SAM-I riboswitch (found in the 5'-UTR of the *metE* gene). The resulting folinic acid (a stable THF analogue) responsive chimeric riboswitch is anticipated to function via a transcriptional-‘OFF’ mechanism.

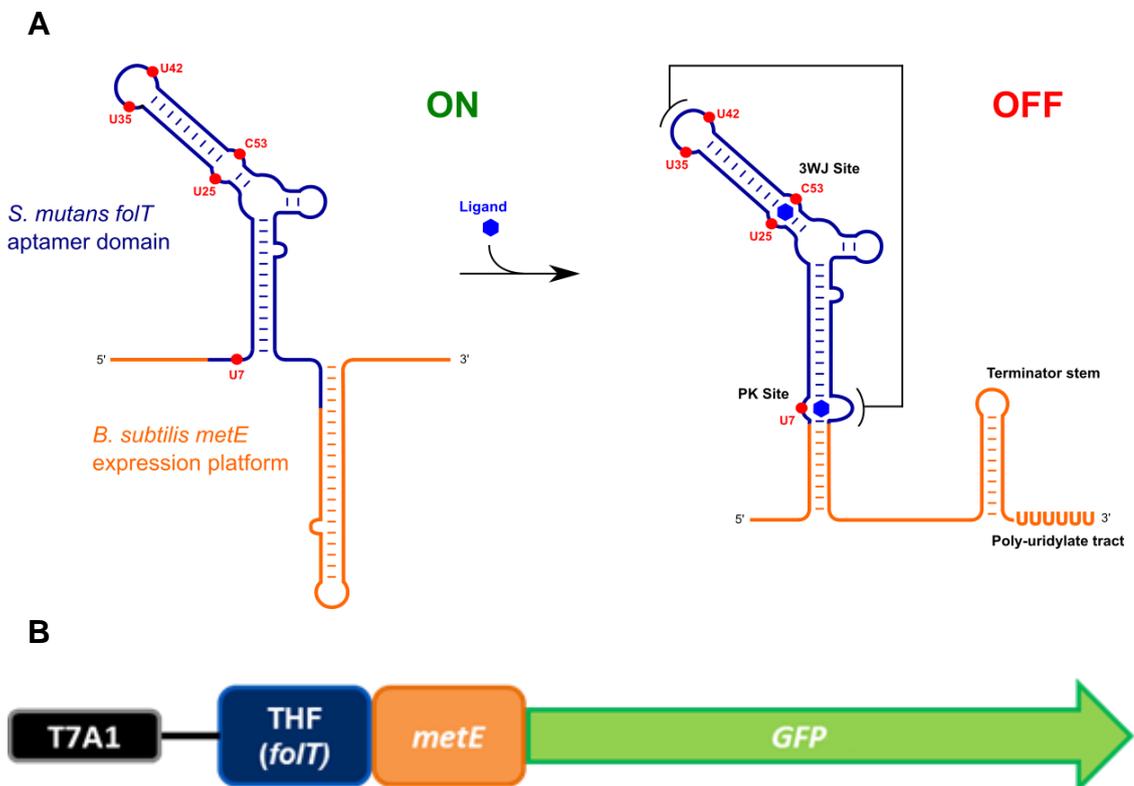


Figure 33 – A) Putative switching mechanism for the *Smu folT/Bsu metE* chimeric riboswitch. The *S. mutans folT* aptamer domain is shown in blue and the *B. subtilis metE* expression platform in orange. Left: In the ligand-unbound conformation the expression platform forms an anti-terminator stem, allowing transcription of the ORF. Right: A ligand binding event stabilises the P3 stem and induces pseudoknot and P1 stem formation, favouring the establishment of a downstream terminator stem and the premature abortion of transcription. B) Chimeric reporter gene construct used in the GFP expression assays

Since this system is already well characterised *in vitro*^{104, 110} and since it has also been shown that chimeric riboswitches containing the *B. subtilis metE* expression platform also function in *E. coli*¹⁰⁰, the chimeric construct was used as a starting point for the *in vivo* characterisation of the THF riboswitch.

An overview of the proposed switching mechanism for the *Smu folT/Bsu metE* chimera is shown in **Figure 33A**. This was created using the Mfold web server¹⁶⁴ in combination with the ligand-bound aptamer domain structure and putative expression platform conformation proposed by Batey *et al.*¹⁰⁴. In the ligand-unbound state, the expression platform forms an anti-terminator stem and the ORF is transcribed (gene ON). Ligand binding at the PK site stabilises the P1 stem and induces pseudoknot formation, causing a conformational change in the expression platform such that a downstream terminator stem is formed, causing premature abortion of transcription (gene OFF). It is assumed that ligand binding at the 3WJ site strengthens the P3 stem and facilitates PK site formation and indeed positively cooperative binding has been observed *in vitro*¹⁰⁴.

3.2.1.1 Design and construction of the *in vivo* reporter gene construct

A riboswitch-controlled GFP reporter construct was designed and cloned by colleague Dr. Helen Vincent. The chimeric *Smu folT/Bsu metE* riboswitch was cloned upstream of the *GFP* reporter gene expressed constitutively from a strong T7A1 promoter in a pACYC-Duet1 vector (see **Section 6.8**). This was used to transform *E. coli* DH5 α pLOI707HE/slr0642, containing the *Synechocystis slr0642* folate transporter (known to function in *E. coli*, plasmid provided by Dr. Andrew Hanson of the University of Florida)¹⁶⁵ under the control of the *lac* promoter. This gave the *E. coli* DH5 α pLOI707HE/slr0642, pACYC-Duet1-T7A1-S.mutTHFwt/metE-GFP strain, whereby expression of the folate transporter was IPTG inducible and ligands which were able to activate the riboswitch were expected to cause a reduction in GFP expression. The folate transporter was required since folates are not able to permeate bacterial cell membranes and *E. coli* does not have a native transporter. Subsequent mutant library creation and *in vivo* assays were carried out by the author, unless otherwise stated.

3.2.1.2 Design and creation of riboswitch mutants

In addition to the strain containing the wild-type *S. mutans* THF aptamer, a selection of mutants was prepared using site-directed mutagenesis, utilising the aforementioned pACYC-Duet1-T7A1-*S.mut*THFwt/*metE*-GFP-lv plasmid as the template. A total of five mutants were created, of which a summary can be found in **Figure 34C**. The mutants were designed for two main purposes. Firstly, to identify the role of the two ligand binding sites in inducing the riboswitch regulatory response. The single mutants **M3** (U7C) and **M4** (U25C) contain mutations which have been shown *in vitro* to abrogate binding of THF/folinic acid at the PK and 3WJ sites respectively. The double mutant **M5** (U7C/U25C) combines these two mutations¹⁰⁴. By disrupting binding to each of the two binding sites independently (**M3** and **M4**) and concurrently (**M5**), then measuring the effect of these mutations on gene expression output, it was anticipated that the significance of the two binding sites could be established for the first time *in vivo*.

A second set of mutants was designed to rationally reconstruct each of the two ligand binding sites independently such that each site no longer recognises the natural ligand but instead binds a non-natural synthetic analogue. The overall goal of this strategy being to identify mutant riboswitch-ligand pairs which could be utilised as orthogonal gene expression systems. With this in mind, two mutants were designed. The double mutant **M1** (U25C/C53U) contains a modified 3WJ binding site with a reversal of H-bond directionality, designed to accept 'diketo-faced' THF analogues and reject THF/folinic acid (see **Figure 32B**). Similarly, the double mutant **M2** (U35C/U42C) has a modified PK binding site, intended to accept 'diamino-faced' THF analogues whilst rejecting THF/folinic acid (see **Figure 32C**).

Site-directed PCR mutagenesis was carried out on plasmid DNA harvested by mini-prep, with primers designed to contain the desired mutation and to anneal to the target sequence on opposite strands of the template plasmid, using a standard protocol. These cells were grown in overnight cultures, the mutagenic plasmids harvested by mini-prep and the mutations were confirmed by sequencing (GATC Biotech). The mutagenic plasmids isolated by mini-prep were then used to transform electro-competent *E. coli* DH5 α pLOI707HE/slr0642 cells (containing the folate transporter) using a standard electroporation protocol.

3.2.1.3 Determination of riboswitch basal expression levels

Following the creation of the riboswitch mutant-containing strains, the basal expression level of GFP in each strain was measured. The goal of this experiment was to check for any potential misfolding of the riboswitch (that could lead to repressed expression in the absence of exogenous activator ligand) and assess the potential impact cellular metabolites may have on the assay. Normalised *GFP* expression was determined by measuring fluorescence intensity and dividing this figure by the OD_{600} to normalise for cell growth. The results of this initial analysis are shown in **Figure 34A**. With the exception of mutant **M2**, basal *GFP* expression levels for the wild-type and mutant riboswitch sequences were consistent, indicating that in the absence of appropriate effector ligands, the mutations themselves do not intrinsically influence the levels of transcription termination.

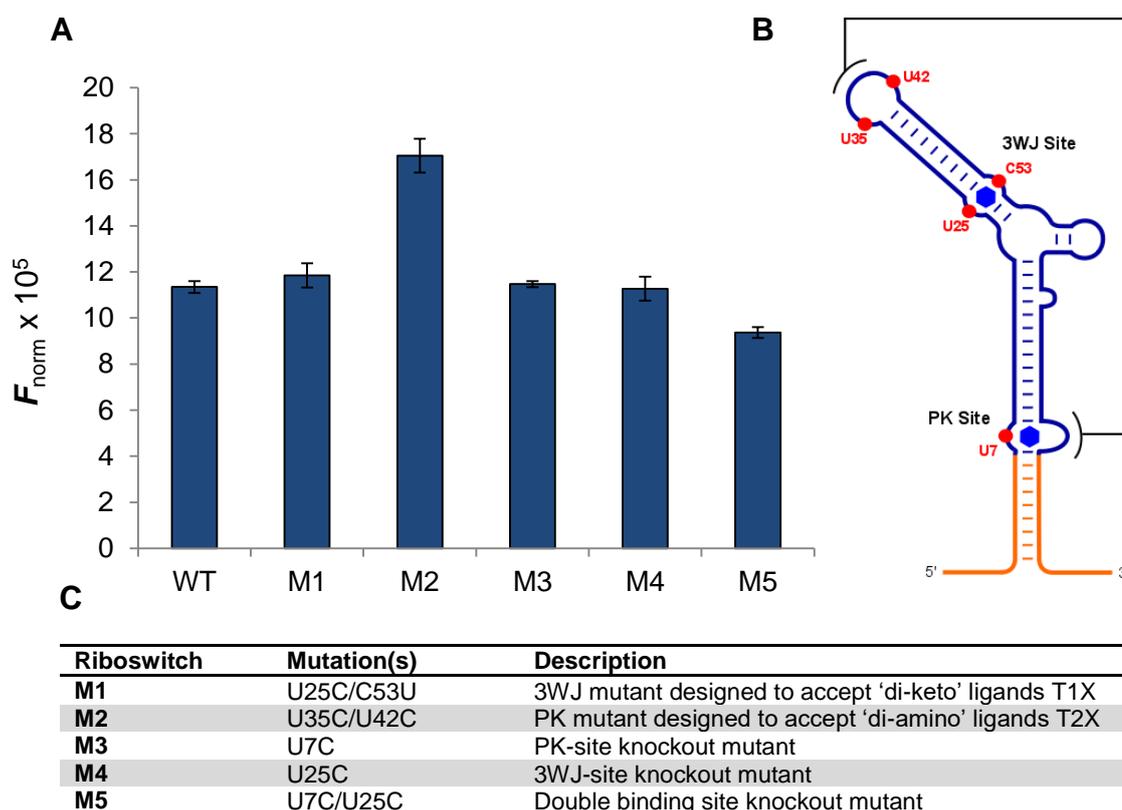


Figure 34 – A) Basal *GFP* expression levels for wild-type (WT) *Smu folT/Bsu metE* and mutant library. The bar height indicates the mean normalised fluorescence of three experiments and the error bars represent the standard deviation. B) Secondary structure of the *S. mutans* THF aptamer domain. Key residues targeted for mutagenesis are highlighted in red. C) Summary of riboswitch mutants.

Additionally, since the basal-level expression for the wild-type and the double binding site knockout mutant **M5** are similar, it can be assumed the cellular levels of THF and related metabolites are not sufficient cause significant background repression of the wild-type riboswitch in this chimeric system. Background repression by endogenous metabolites is an established pitfall in the use of riboswitches as gene expression tools, and has previously been reported by the Micklefield group in the development of orthogonal PreQ riboswitches⁸⁶. However, in this case, the use of a strong promoter and multi-copy plasmid was anticipated to overcome this potential problem.

3.2.1.4 Riboswitch expression response to a single concentration of folinic acid

In order to make a preliminary assessment of the ability of the chimeric THF riboswitches to function as ligand-responsive transcriptional terminators *in vivo*, the wild-type and mutant selection was subjected to a single concentration of folinic acid. Cultures were grown overnight in the presence of IPTG to express the folate transporter, with either 2 mM folinic acid or water added. As previously mentioned, folinic acid is utilised since THF is unstable in air, making it unsuitable for use in aerobic bacterial growth conditions. From the published *in vitro* characterisation of the THF riboswitch¹¹⁰ it was anticipated that mutations at each of the two binding sites would have a deleterious effect on the regulatory response to folinic acid. Specifically, it was expected that mutations which disrupt binding at the PK site would be more deleterious than those which ablate binding at the 3WJ site. Therefore, mutants **M1-5** were all expected to repress expression in response to folinic acid to a lesser extent than the wild-type. Furthermore, the magnitude of this reduction in performance was expected to be greater for mutants **M2**, **M3** and **M5**, which all contain mutations at the PK site.

The results of this experiment are shown in **Figure 35A**. For ease of comparison, *GFP* expression levels for each strain in the absence of folinic acid have been normalised to 1. In the wild-type strain, addition of 2 mM folinic acid was able to effect up to 70% repression of *GFP* expression, indicating that the chimeric riboswitch does indeed function *in vivo*. However the *Smu folT/Bsu metE* chimeric system has been shown to effect up to 95% transcription termination *in vitro* at folinic acid concentrations as low as 250 μM ¹¹⁰. Even concentrations of folinic acid up to 5 mM were only able to cause 74% repression in this *in vivo* assay (see **Section 3.3.3**). This reduced effectiveness *in vivo* may be due to a number of reasons. Firstly, the efficiency of the folate transporter

is not documented, and the concentration of folinic acid added may not reflect the actual intracellular concentration. Secondly, the inherent stability of GFP¹⁶⁶ may limit its application as a study of repressive expression systems. Finally, the efficiency of transcription termination is known to be highly dependent on NTP concentration¹⁶⁷, with higher concentrations leading to reduced termination efficiency.

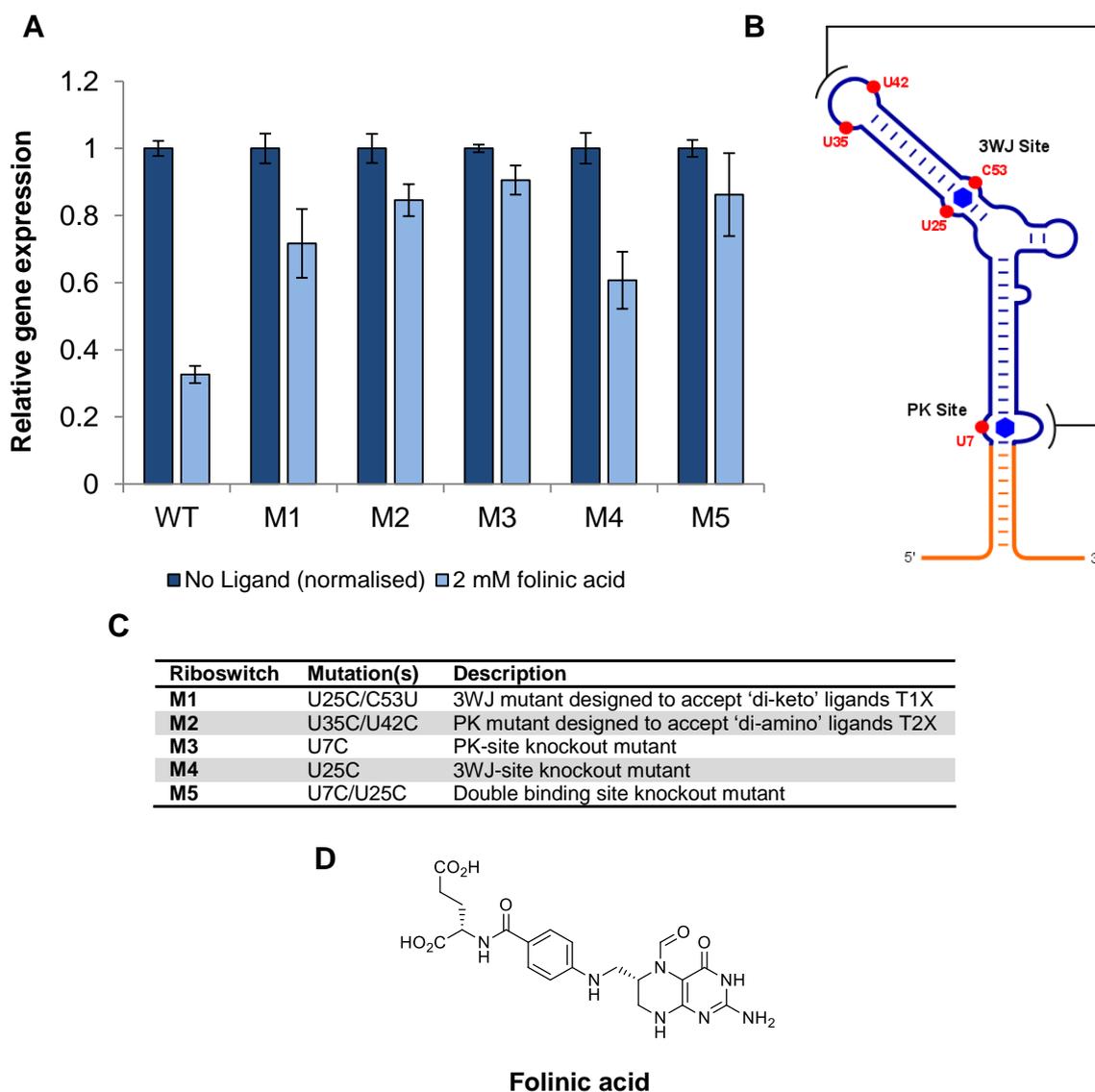


Figure 35 – A) Relative *GFP* expression levels for *S. mutans/metE* THF riboswitch mutants with and without folinic acid. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM folinic acid. Bar height indicates the mean of three experiments and error bars show the standard deviation. **B)** Secondary structure of the *S. mutans* THF aptamer domain. Key residues targeted for mutagenesis are highlighted in red. **C)** Summary of riboswitch mutants. **D)** Structure of folinic acid.

Since the published *in vitro* assays were carried out at NTP concentrations at least one order of magnitude lower than those in *E. coli*¹³⁴, this may explain the discrepancy between the two figures. The folinic acid-induced regulatory effect was significantly diminished in **M3**, with less than 10% repression observed at 2 mM ligand concentration. This is consistent with previous investigations, which have shown that the U7C mutation abrogates ligand binding to the PK site and drastically reduces the ability of the riboswitch to effect transcription termination *in vitro*^{104, 110}. In the same study, disruption of ligand binding at the 3WJ site via the U25C mutation was reported to have only marginal impact (~10% reduction) on transcriptional termination ability¹¹⁰. However, in this *in vivo* study the regulatory response of **M4** (U25C) was only 41% – almost half as effective as the wild-type – indicating that ligand binding to the 3WJ site is important for the regulatory effect *in vivo* in this chimeric system. The complex cellular milieu and vast number of potential interactions means that there are many possible reasons for this observed discrepancy between the *in vitro* and *in vivo* function of the 3WJ binding site. One potential explanation for the difference is the nucleotide concentration and concomitant rate of transcription.

The published *in vitro* transcription assay was carried out at NTP concentrations of 75 μM ¹¹⁰, whilst the concentration of NTPs in *E. coli* is known to lie within the range of 0.5 – 3.0 mM^{134, 135}, at least an order of magnitude greater than that used in the *in vitro* assay. Moreover, the concentration of NTPs is known to affect the rate of transcription elongation, with higher concentrations leading to higher transcription rates¹¹. If there were a riboswitch folding rate enhancement as a result of binding to the 3WJ site (which is transcribed first), this could result in an increased transcriptional termination efficiency which is not observable at the low NTP concentrations (and hence slower transcription elongation rates) employed in the *in vitro* assay¹¹⁰, but is observable at the higher NTP concentrations (faster elongation rates) within the *E. coli* cellular environment. As expected, **M5**, which contains mutations to abrogate folinic acid both binding sites also exhibited a significantly diminished regulatory response. Riboswitch **M1**, which contains a mutated 3WJ site, effected repression at a comparable magnitude (within error) to the known 3WJ site knockout, **M4**. Analogously, the mutant **M2**, comprising a modified PK site was shown to exhibit a diminished repressive ability comparable to the known PK site knockout mutant, **M3**. These data suggest that the **M1** and **M2** mutations abrogate binding to folinic acid at the 3WJ and PK sites respectively, which is desirable in the study of binding site co-operativity and the development of orthogonal riboswitch systems. This hypothesis is confirmed *in vitro* in Chapter 4.

3.2.1.5 Riboswitch expression response to single concentrations of T0A and T1A

Compound **T0A** was screened against the riboswitch mutant library, at 2 mM concentration. The structural similarity of **T0A** to THF/folinic acid along with the available x-ray crystallography data indicated that **T0A** was likely to bind to the wild-type THF riboswitch. However, since it was discovered that compounds without C6-position side-chains do not appear to elicit the riboswitch regulatory effect *in vitro*¹¹⁰, it was unclear whether or not an *in vivo* response would be observed.

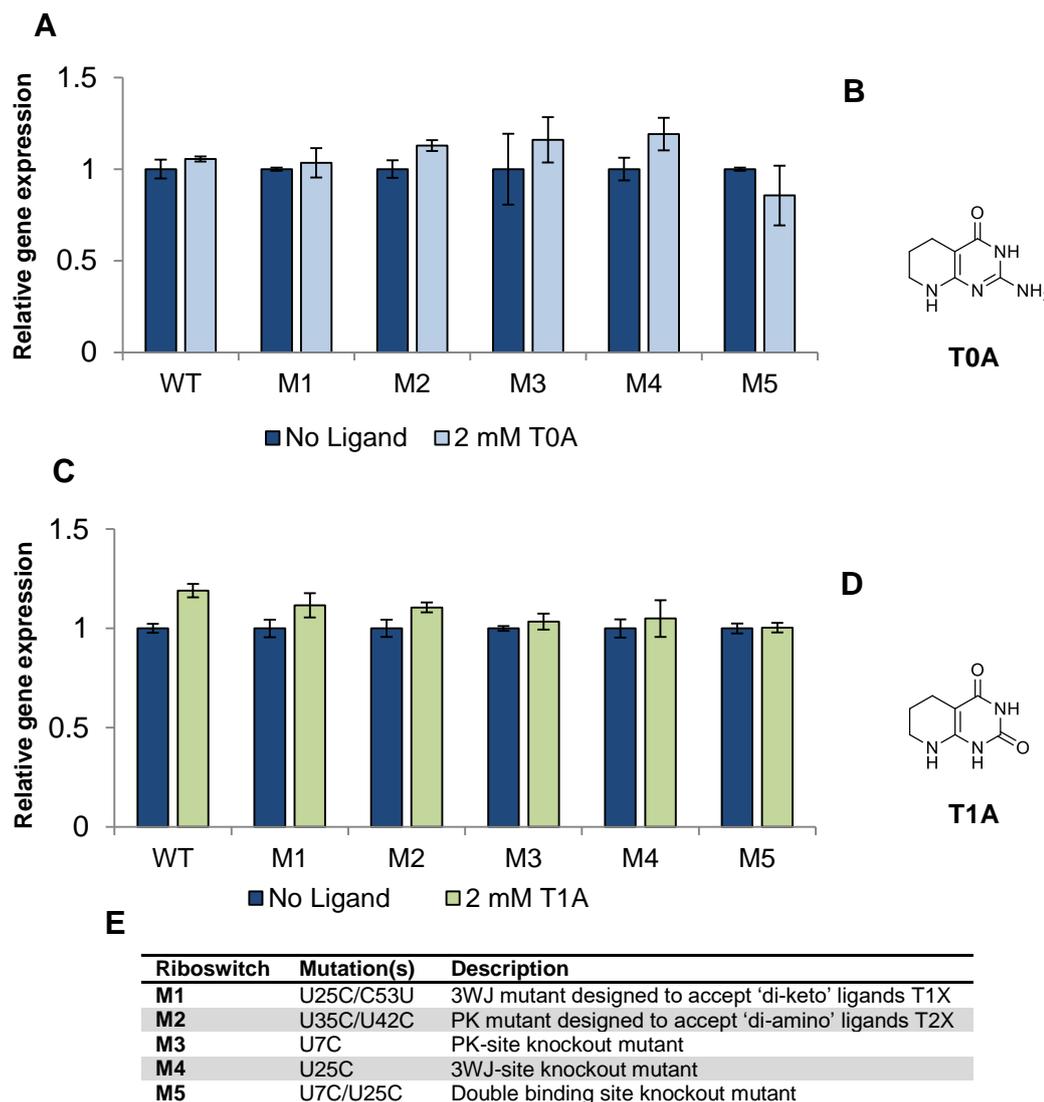


Figure 36 – A) Relative *GFP* expression levels for *S. mutans/metE* THF riboswitch mutants with and without T0A. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM T0A. Bar height indicates the mean of three independent experiments and error bars show the standard deviation. B) Structure of T0A. C) Relative *GFP* expression levels with and without 2 mM T1A. D) Structure of T1A. E) Summary of riboswitch mutants.

T0A had little effect on the expression output of any of the riboswitch-containing strains (**Figure 36A**). Compound **T1A** (**Figure 36D**) was designed to bind and activate the mutant riboswitch **M1**. However, when the riboswitch mutant library was screened with this compound at 2 mM concentration, there was no overall effect on gene expression across any of the strains (**Figure 36C**). The lack of activity of both compounds **T0A** and **T1A** with their target riboswitches (**WT** and **M1** respectively) could be attributed a lack of binding to the aptamer domain, or due to a binding event which does not trigger a regulatory response. An ITC analysis of this hypothesis is presented in Chapter 4.

3.2.1.6 Riboswitch expression response to diaminopterins analogues

The chimeric riboswitch library was also screened against single concentrations (2 mM) of a variety of diaminopterins analogues, which were anticipated to bind the **M2** riboswitch (designed to accept compounds of this type at a modified PK site). The structures of the compounds which were synthesised for this purpose are shown in **Figure 37**.

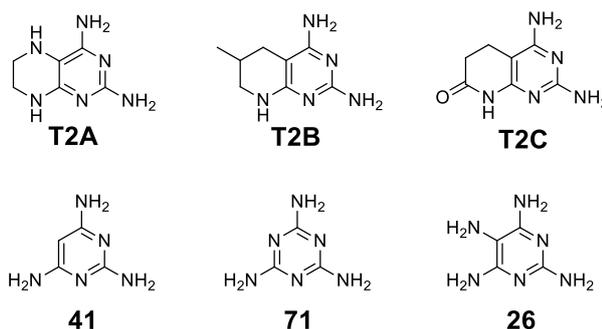


Figure 37 – Structures of synthesised compounds and commercially available pyrimidines screened with the M2 riboswitch

Compound **T2A** proved to be insufficiently stable for analysis in the overnight assay, whilst **T2C** was poorly soluble, and failed to show any effect at saturating concentrations of approximately 1 mM. Finally, **T2B** proved to be potently cytotoxic, causing up to 95% reduction in cell growth and was not deemed suitable for further investigation. In addition, a number of monocyclic, commercially available analogues of these compounds were also tested including 2,4,6-triaminopyrimidine **41**, 2,4,6-triaminotriazine, **71** and 2,4,5,6-tetraaminopyrimidine, **26** (structures shown in **Figure 37**), none of which showed any effect.

3.2.2 Engineering and characterisation of native THF riboswitches *in vivo*

3.2.2.1 Selection of candidate THF riboswitches

The chimeric *Smu folT/Bsu metE* THF riboswitch served as a useful initial model for studying the functionality of THF riboswitches *in vivo*, however the use of this system has a number of limitations. It is not clear whether or not the lack of *in vivo* regulatory activity observed with compounds known to bind *in vitro* is a consequence of the non-native expression platform, or an intrinsic property of the THF riboswitch itself.

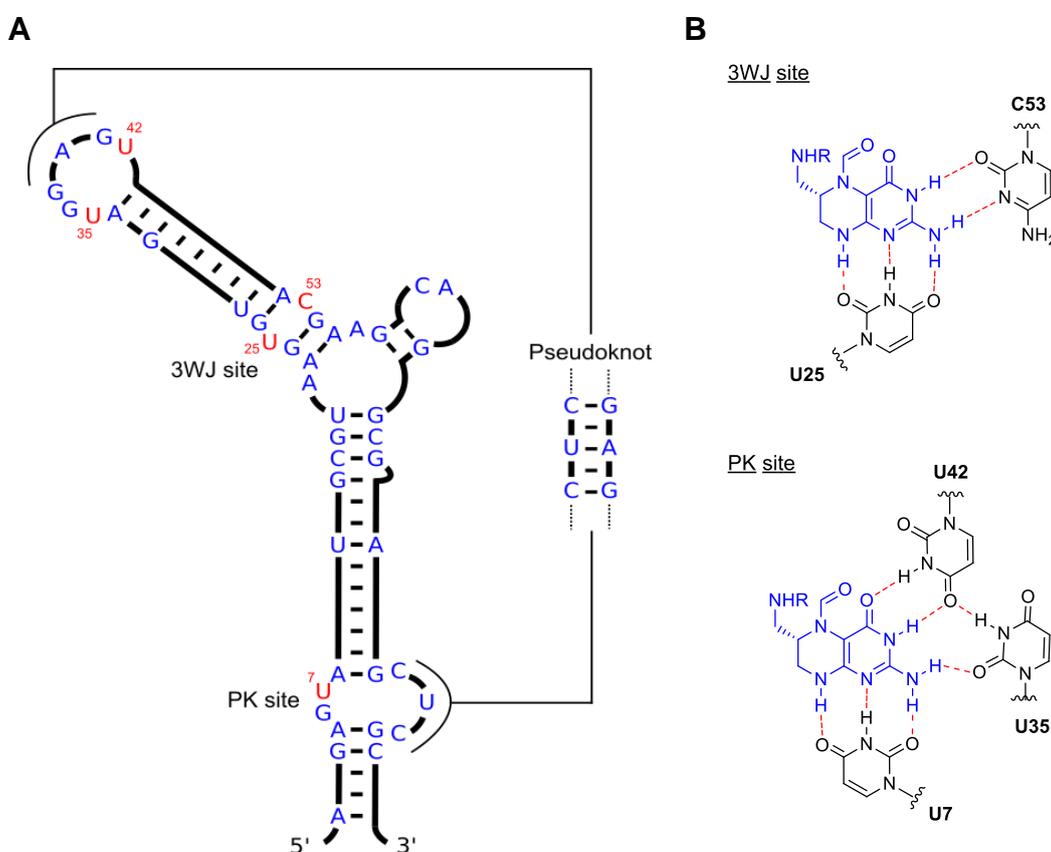


Figure 39 – A) Secondary structure and consensus sequence alignment of THF riboswitch aptamer domains from *Streptococcus mutans*, *Alkaliphilus metalliredigens*, and *Lactobacillus casei* in the ligand-bound state. Residues conserved across all three riboswitches are shown in blue and ligand binding site residues are shown in red with the nucleotide numbering taken from the *S. mutans* sequence. (Note: *L. casei* riboswitch actually contains the U35C mutation, but residue is highlighted for clarity). B) Ligand binding motifs for the *S. mutans* THF riboswitch, obtained by x-ray crystallography¹⁰⁴.

Moreover, the role of the two ligand binding sites in effecting the riboswitch regulatory response may not be faithfully modelled in the chimeric system. Fully native THF riboswitches would represent a more accurate representation of the natural system and may also offer improved expression parameters in the context of developing gene expression devices.

Accordingly, alternative THF riboswitches which have been identified in the literature⁵⁷ were assessed as potential candidates for engineering and characterisation *in vivo*. Candidates were selected based on two factors. Firstly, their sequence homology and the conservation of key residues with the *Streptococcus mutans* THF riboswitch aptamer domain, the x-ray crystallography structure¹⁰⁴ of which forms the basis for the rational mutant design. Secondly, whether or not the candidate riboswitches had an identifiable expression platform that would be anticipated to function in *E. coli*. Riboswitch secondary structure predictions using the Mfold¹⁶⁴ and Vienna¹⁶⁹ web servers were used to postulate expression platform conformations, revealing that the fully native *Streptococcus mutans* riboswitch likely functioned as a translational repressor and also featured a Shine-Dalgarno (SD) sequence compatible with the *E. coli* consensus¹⁷⁰. In addition, two THF riboswitch candidates which appeared to operate at the translational level and contained *E. coli*-compatible SD sequences were also identified from the organisms *Alkaliphilus metalliredigens* and *Lactobacillus casei*.

In addition to the retention of key binding site residues across all three sequences, sequence alignment also revealed a high level of conservation around the structurally important regions of the aptamer domain (**Figure 39A**). Nucleotides forming the pseudoknot and associated PK ligand binding site, as well as those surrounding the three-way junction and adjacent 3WJ binding site were highly conserved across all three strains. Whilst the nucleotides forming the stem regions were generally less well conserved, the overall stem lengths and predicted strengths were consistent. It was therefore taken with some confidence that the mutations rationalised for the *S. mutans* THF riboswitch could be applied the *L. casei* and *A. metalliredigens* sequences to similar effect. All three riboswitches control the expression of the folate transporter *folT* in their native setting.

An overview of the posited switching mechanisms for the native *S. mutans* and *L. casei* riboswitches is shown in **Figure 40**. In both cases, it is thought that the ligand-unbound conformation favours the formation of a downstream stem loop whereby the SD sequence is free to bind the ribosome, allowing translation of the mRNA transcript. Ligand binding induces a conformational change in the expression platform such that

the SD sequence is sequestered within an alternative stem-loop, thus inhibiting translation initiation.

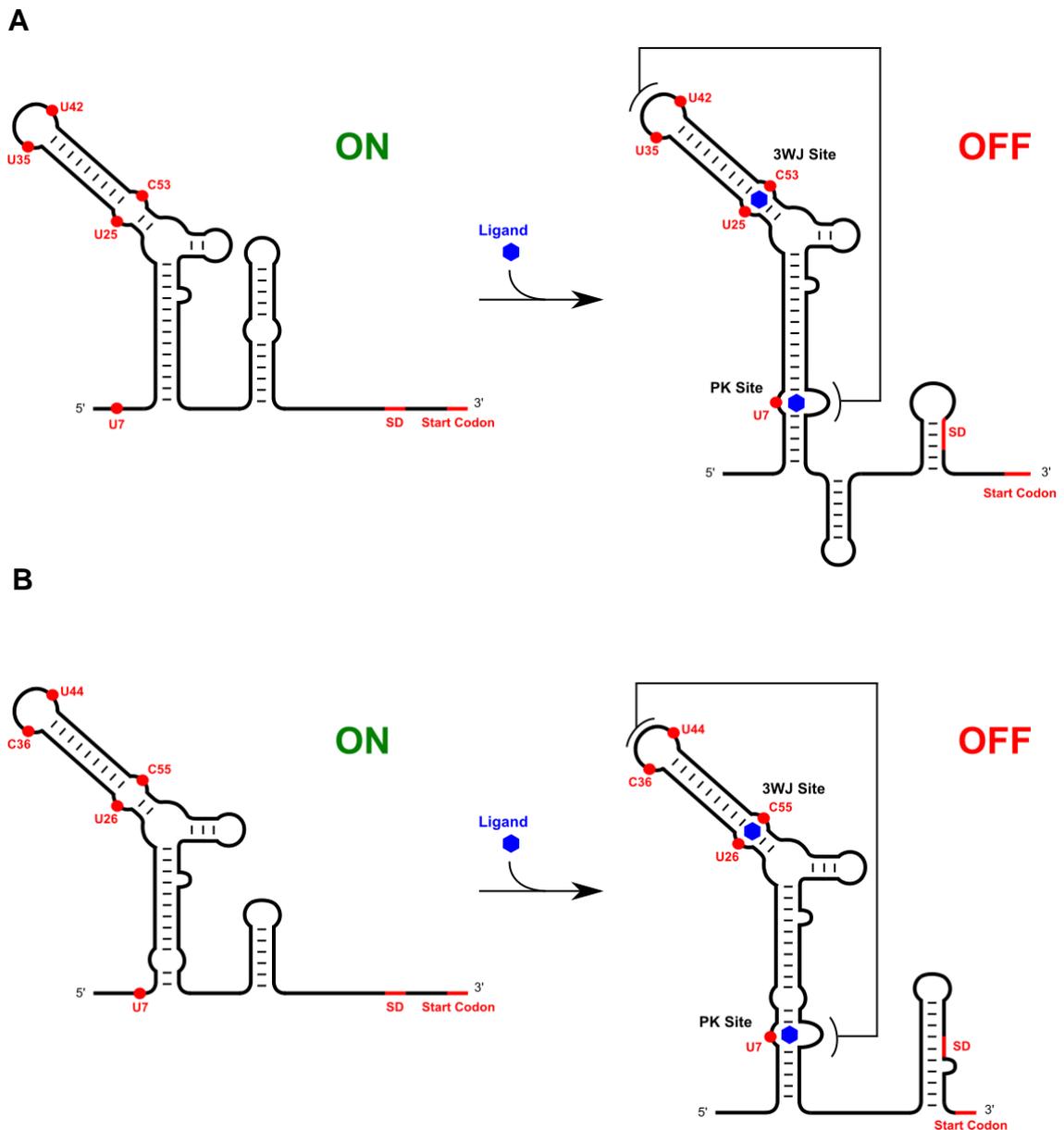


Figure 40 – Putative switching mechanisms for two translational-‘OFF’ THF riboswitches. A) *Streptococcus mutans*; B) *Lactobacillus casei*. In both cases, it is proposed that a ligand binding event strengthens the P3 stem, stabilises pseudoknot formation, and causes sequestration of the Shine-Dalgarno sequence within a downstream stem loop. Diagrams were created using Mfold¹⁶⁴ and Vienna¹⁶⁹ web server secondary structure predictions.

3.2.2.2 Construct design and creation

Reporter gene constructs containing the three riboswitch candidates were designed and cloned by colleague Dr. Helen Vincent as well as a control strain containing the same reporter construct with the riboswitch removed (No-RS). The constructs were designed to closely resemble those used for the *Smu folT/Bsu metE* chimeric THF system with the candidate riboswitch cloned immediately upstream of the *GFP* reporter gene, under the control of the T7A1 promoter in a pACYC-Duet1 vector (**Figure 41**). These constructs were similarly used to transform *E. coli* DH5 α pLOI707HE/slr0642 cells, expressing the *Synechocystis slr0642* folate transporter as described in **Section 3.2**. Unless otherwise stated, all subsequent assays were performed by the author.



Figure 41 – Fully native construct design used in GFP reporter gene assays

3.2.2.3 Dose response profiles of three candidate riboswitches using folic acid

The gene expression responses of the three candidate riboswitch constructs (as well as the *Smu folT/Bsu metE* chimera described in the previous section) were investigated using a broad range of folic acid concentrations from 0 to 5 mM. This data was then used to fit dose response curves using a four parameter logistic function (Sigma Plot 12.0) to generate a series of expression parameters used to characterise and compare riboswitch function *in vivo* (**Figure 42G**). Maximum repression is defined as the percentage decrease in gene expression output at 5 mM folic acid versus 0 mM; T_{50} is the concentration of folic acid required to effect 50% repression of gene expression; and the dynamic range is the concentration of folic acid at 90% of maximum effect divided by the concentration at 10% maximum effect. The dynamic range is an indication of the range of ligand concentrations over which dose-dependent control of gene expression can be achieved. For the development of novel gene expression tools, desirable riboswitch characteristics include a high maximum repression and a low T_{50} – thereby minimizing the amount of effector ligand required to achieve the maximal regulatory effect.

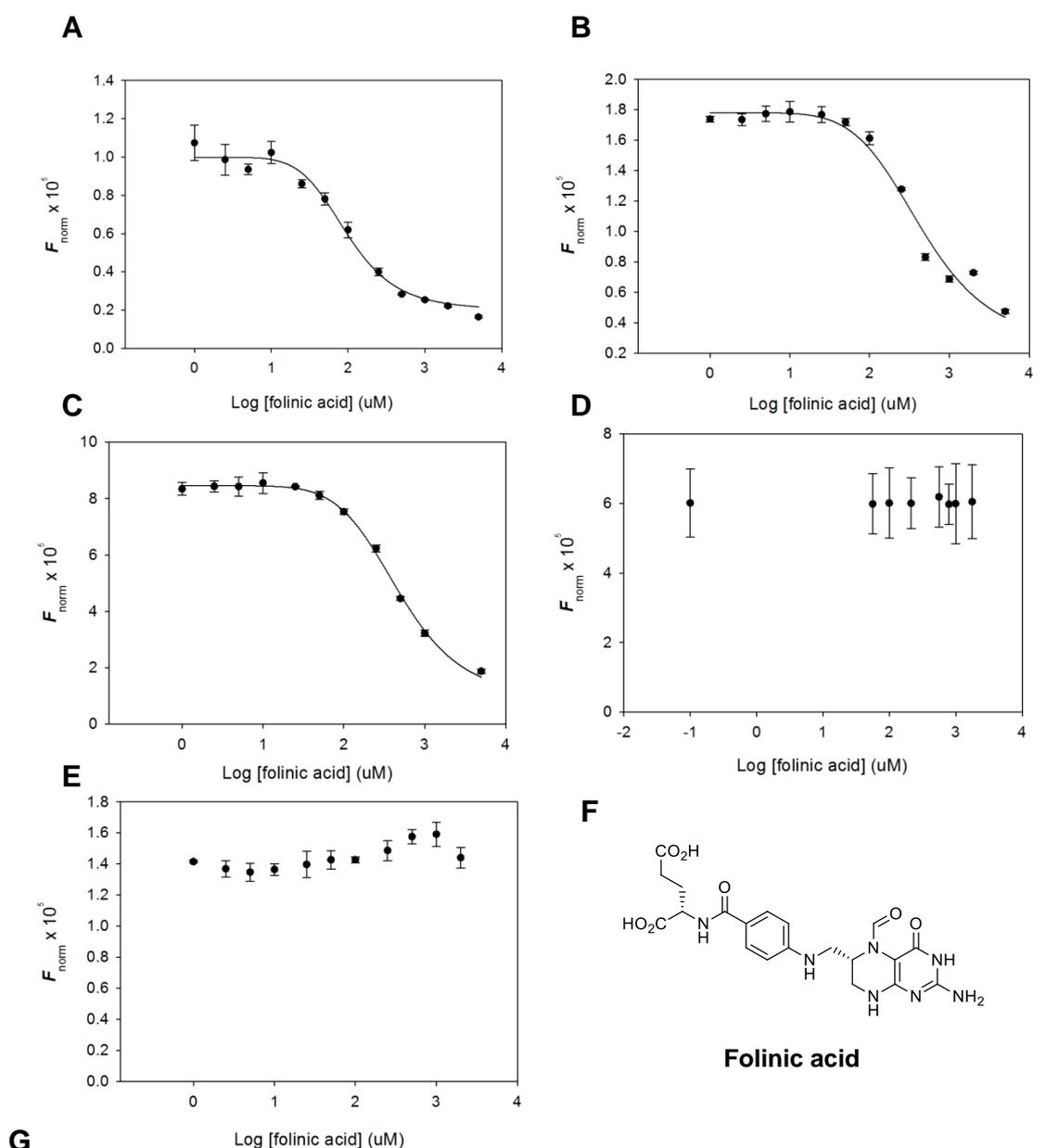


Figure 42 – Dose-responsive control of *GFP* expression by THF riboswitches. A) *Lactobacillus casei*; B) *Streptococcus mutans*; C) *Streptococcus mutans/metE* chimera; D) *Alkaliphilus metalliredigens* (data from Dr. Helen Vincent); E) No riboswitch control strain; F) Structure of folinic acid; G) Expression parameters. Data points indicate the mean of three experiments and error bars show standard deviation.

Of the four riboswitches investigated, three were shown to repress gene expression in a dose-responsive manner on the addition of folinic acid. The *L. casei* THF riboswitch was able to effect the greatest overall repression of gene expression at 84%. The fully native *S. mutans* riboswitch and the chimeric *Smu folT/Bsu metE* riboswitch were slightly less effective, with maximum repression of 72% and 73% respectively. These data indicate that the aptamer domain appears to be more important in determining the maximum repression response, since both the *metE* and native THF expression platforms were similarly effective. The *A. met* riboswitch was not functional at all. The *L. casei* riboswitch also exhibited the lowest T_{50} value, which was 4 to 5-fold lower than the native and chimeric *S. mutans* riboswitches. The *L. casei* riboswitch also gave the broadest dynamic range, which was 3.5- and 7-fold larger than the *S. mutans* chimera and fully native riboswitches respectively.

Comparing basal expression levels, the *L. casei* riboswitch (putative translational mechanism) exhibited the lowest level of expression in the 'ON' state at low folinic acid concentrations. Basal expression for the fully native *S. mutans* riboswitch (putative translational mechanism) was approximately 1.8-fold higher than *L. casei*, whilst the *Smu folT/Bsu metE* chimera (known transcriptional mechanism) exhibited by far the highest 'ON'-state expression, more than four times that of the fully native *S. mutans* riboswitch and eight times that of *L. casei*. There are a number of possible explanations for the observed differences in basal levels of GFP expression. It could be due to a difference in ligand binding affinity by the respective riboswitch aptamer domains. Riboswitches which have higher affinities for THF or other intracellular reduced folates would be regulated to a greater degree by the cellular ligand pool and thus GFP expression in the absence of exogenous ligand would be lower. However, ITC analysis conducted as part of this project revealed that the *L. casei* aptamer domain has a lower affinity for folinic acid than that of the *S. mutans* aptamer (for further details, see **Sections 4.2.4.1** and **4.3.4.4**). Moreover, this does not explain the difference in basal expression levels between the *S. mutans* native and *S. mutans* chimeric THF riboswitches, which contain identical aptamer domains. A further consideration is the difference in the regulatory mechanisms of the respective riboswitches. The transcriptional-'OFF' *Smu folT/Bsu metE* chimeric riboswitch was found to have at least four-fold higher basal GFP expression levels than the native *L. casei* and *S. mutans* native translational-'OFF' riboswitches. The kinetics of riboswitch folding are known to be important in transcriptional riboswitch regulatory function (see **Section 1.4.4.2**) and this could offer an explanation for the apparent lower level of background repression of gene expression by intracellular reduced folates in the transcriptional system.

In the context of developing novel gene expression systems, under certain circumstances the *S. mutans* chimeric THF riboswitch may be a more suitable expression system of choice. For instance, where a higher ‘ON’-state level of gene expression is desired in studies such as gene functional analysis⁸⁶. However, despite having the lowest basal expression level, the *L. casei* riboswitch remains the most attractive choice as the starting point in the development of novel gene expression tools; combining the maximum regulatory response, a low T_{50} and a relatively broad dynamic range. On this basis, the *L. casei* THF riboswitch was selected to be taken forward for mutagenesis studies towards novel gene expression tools and to further characterise the role of the two ligand binding sites in regulating expression.

3.2.2.4 Creation of an *L. casei* riboswitch mutant library and determination of basal expression levels

A small mutant library was designed utilising the same rationale, and created using the same protocols as those outlined for the *S. mutans folT/metE* chimeric riboswitch outlined in **Section 3.2.1**. The appropriate mutations were identified using the aforementioned sequence alignment (see **Section 3.2.2.1**) and it is assumed that these mutations confer the same properties as have been demonstrated for the *S. mutans* riboswitch aptamer¹¹⁰. An initial round of assays was performed to assess the viability of the gene expression constructs and to compare the basal expression levels (‘ON’-state expression levels) across the mutant strains. The results of these assays are shown in **Figure 43A**.

In contrast to the *Smu folT/Bsu metE* riboswitch mutants, the *L. casei* mutant riboswitches showed a much greater degree of variability in the basal level of expression. The **WT**, **M1** and **M4** riboswitches were equivalent within error and exhibited the lowest ‘ON’-state expression levels. Basal expression levels for the PK-site knockout mutant **M3** and double binding site knockout **M5** were more than 3-fold higher than this. This effect is attributed to the presence of intracellular THF (and derivatives) which likely bind to the riboswitch and repress gene expression even without the addition of exogenous folinic acid – thereby lowering the ‘ON’-state expression level. This effect has been reported previously by the Micklefield group in the study of PreQ-responsive riboswitches⁸⁶. For **WT**, **M1** and **M4** this background repressive effect is the most pronounced since all three have intact PK sites and the

PK site has been shown to be the more important of the two in mediating the regulatory effect¹¹⁰.

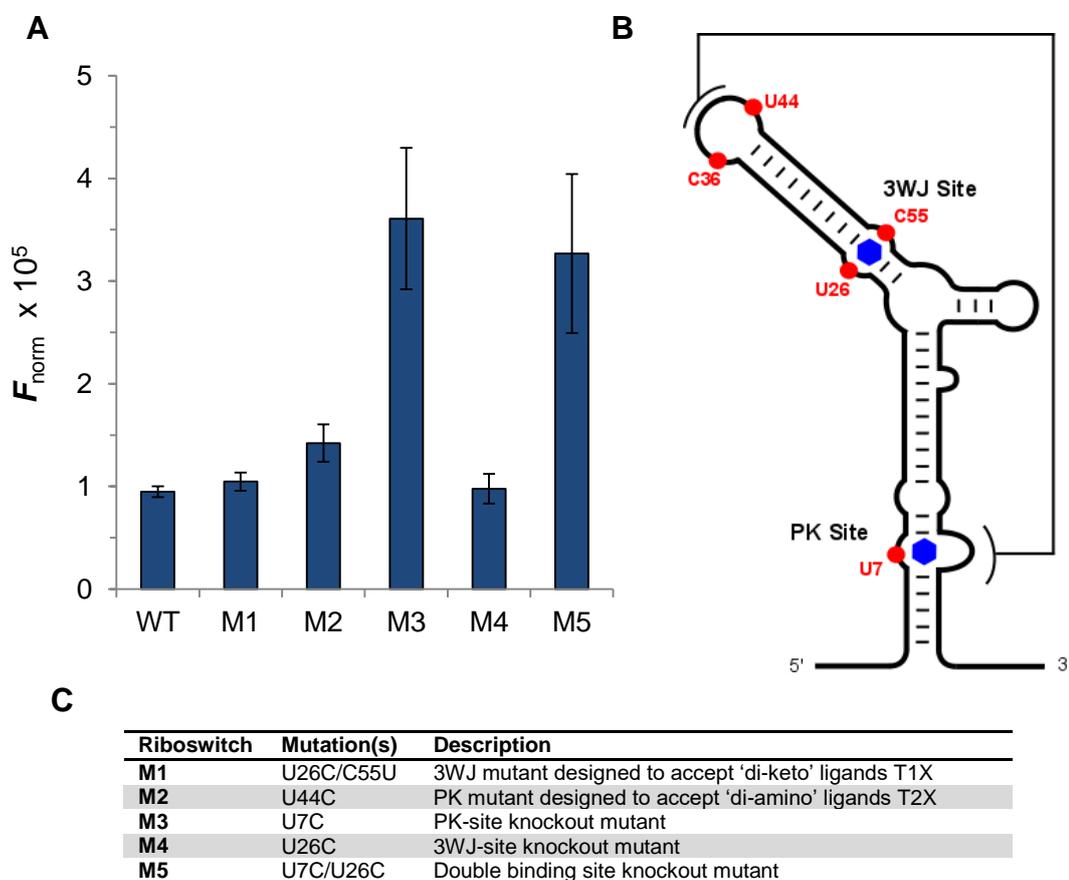


Figure 43 – A) Basal *GFP* expression levels for *L. casei* THF riboswitch mutant library. Bar height indicates the mean fluorescence normalised for OD based on three experiments. Error bars show the standard deviation. **B)** Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. **C)** Overview of riboswitch mutants.

Whereas, abrogating binding of intracellular folate to the PK site via the U7C mutation (which is present in **M3** and **M5**, but not **WT**, **M1** or **M4**) significantly inhibits this background repressive effect and consequently the 'ON'-state expression levels of **M3** and **M5** are far higher. Based on the K_D of serine hydroxymethyltransferase, for which THF is a substrate, the minimal intracellular concentration of THF is estimated to be $\sim 25 \mu\text{M}$ ^{171, 172}. At this concentration, folinic acid has been shown to effect up to 50% termination of transcription *in vitro*¹¹⁰, supporting this hypothesis. Finally, the PK-site mutant **M2**, designed to accept 'diamino' compounds and reject folinic acid, exhibited basal expression levels comparable to **WT**, **M1** and **M4**. This suggests that at

physiological concentrations the U44C mutation does not disrupt abrogate THF binding to the PK site, which is undesirable in the development of an orthogonal system.

3.2.2.5 *L. casei* riboswitch expression response at a single concentration of folinic acid

A preliminary assessment of the regulatory ability of the *L. casei* THF riboswitches was made by challenging each strain with folinic acid at a single concentration of 2 mM and the results are shown in **Figure 44A**. For clarity, the basal expression levels have been normalised to 1 and expression in the presence of 2 mM folinic acid is given relative to the basal expression.

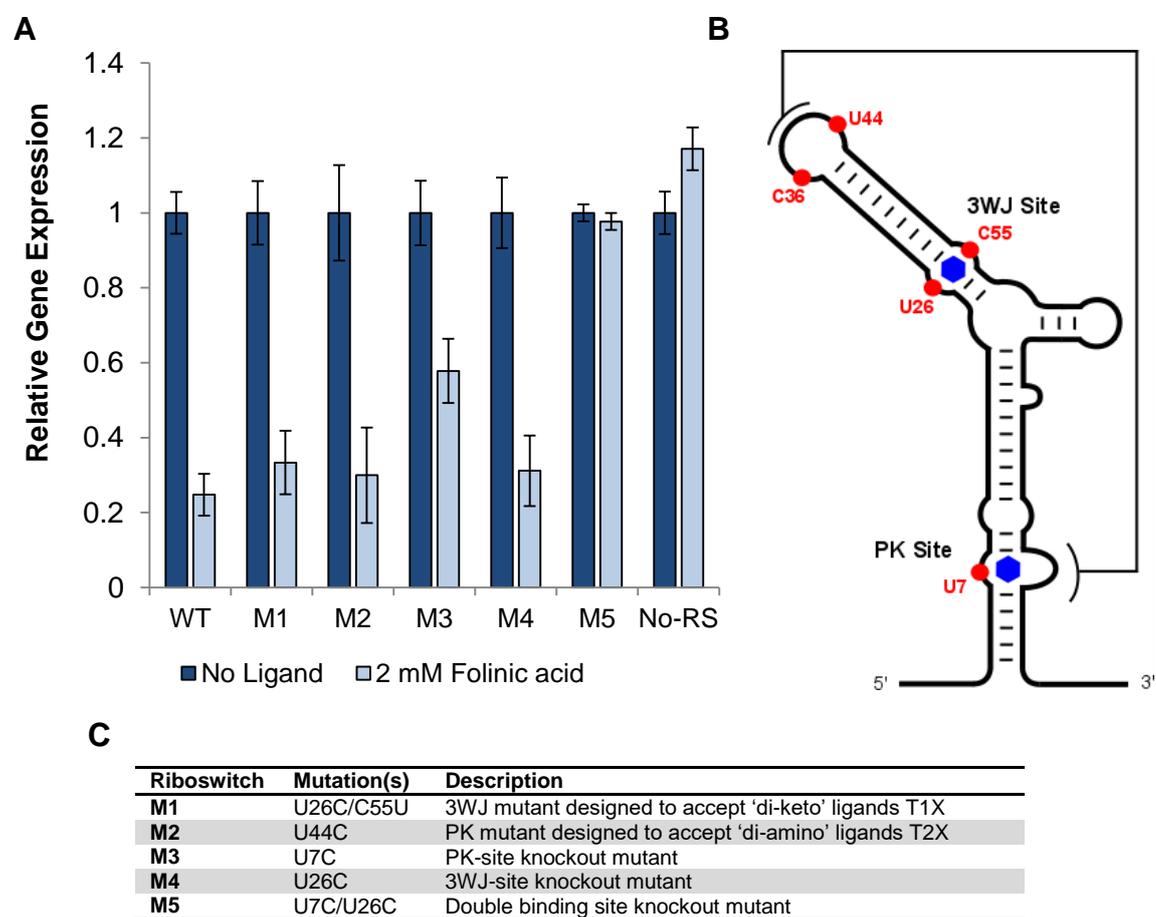


Figure 44 – A) Relative *GFP* expression levels for *L. casei* THF riboswitch mutants with and without folinic acid. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM folinic acid. Bar height indicates the mean of three independent experiments and error bars show the standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Overview of riboswitch mutants.

At this concentration, folinic acid was able to effect up to 78% repression of gene expression with the **WT** riboswitch, a modest improvement on, but comparable to the chimeric system (70%). However, given that the basal expression level of the *L. casei* system is around four times lower than the for *Smu folT/Bsu metE* chimeric riboswitch, the absolute expression levels in the ‘OFF’ state are much lower for *L. casei*. There are also differences in the effects of binding site knockout between the two systems. If the *L. casei* system were to behave in the same fashion as the chimeric riboswitch, it would be expected that PK site knockout would almost entirely negate regulatory function, whereas 3WJ site knockout would reduce performance by around 50%. A summary of the effects of binding site knockout can be found in **Table 3**. In both systems, disrupting ligand binding at both binding sites concurrently essentially negated the regulatory activity. However, compared to the chimeric system, individual binding site knockout was less deleterious to the function of the *L. casei* riboswitch. Knocking out the PK site (**M3**) reduced activity by only 44% for the *L. casei* riboswitch, compared with 86% in the chimeric system. Additionally, disruption of binding at the 3WJ site (**M4**) reduced regulatory activity by only 12% in the *L. casei* system, compared with 41% in *Smu folT/Bsu metE*. Given the lack of available structural information about the *L. casei* riboswitch it is difficult to draw conclusions about this observed difference. However, it does appear that unlike the chimeric system (where PK site knockout effectively negates function), the *L. casei* riboswitch is able regulate expression via ligand binding to the 3WJ site only. In order to investigate this effect further, dose response experiments were performed at a range of folinic acid concentrations

Table 3 – Comparison of gene expression output for the *Smu folT/Bsu metE* and *L. casei* riboswitches

Riboswitch	Mutation	Repression with 2 mM folinic acid (%)	Activity reduction versus wild-type (%)
<i>Smu folT/Bsu metE</i>	Wild-type	70 ± 3	-
	U7C (M3)	10 ± 4	86 ± 6
	U25C (M4)	41 ± 6	41 ± 11
	U7C/U25C (M5)	12 ± 10	86 ± 15
<i>L. casei</i>	Wild-type	78 ± 5	-
	U7C (M3)	44 ± 5	44 ± 10
	U26C (M4)	69 ± 6	12 ± 13
	U7C/U26C (M5)	2 ± 2	97 ± 3

3.2.2.6 Dose response profiles using folinic acid

Dose response profiles for *L. casei* THF riboswitch binding site knockout mutants **M3**, **M4** and **M5** were obtained to further analyse the effects of each binding site on the gene expression output (see **Figure 45**).

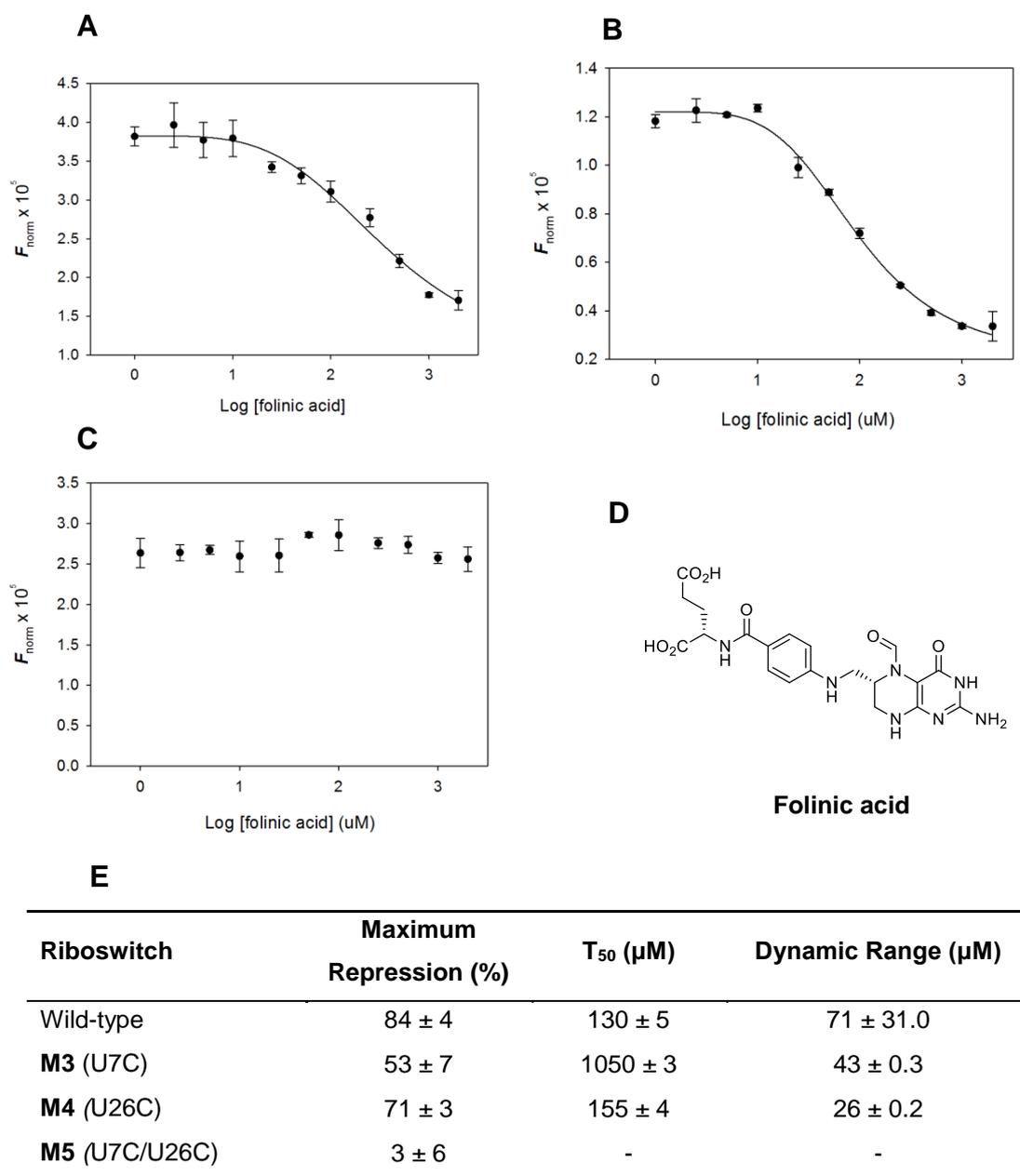


Figure 45 – Dose-responsive control of gene expression with the *L. casei* binding site knockout mutant riboswitches. **A**) M3 (U7C, PK site knockout); **B**) M4 (U26C, 3WJ site knockout); **C**) M5 (U7C/U26C, double binding site knockout); **D**) structure of folinic acid; **E**) expression parameters for the mutant riboswitches with wild-type data shown for comparison.

As well as the previously noted reduction in the magnitude of maximum repression, PK site knockout riboswitch **M3** also displayed a T_{50} around 8-fold higher than the wild-type. Meanwhile, 3WJ site knockout **M4**, whilst being overall slightly less effective than the wild-type in terms of maximum repression, had a comparable T_{50} value of 155 μM . The regulatory response of the two riboswitches containing unmodified PK sites (**WT** and **M4**) therefore occurs at much lower folinic acid concentrations than the mutant with a disrupted PK site (**M3**).

3.2.2.7 Riboswitch dose-response profiles using **T0A**

Compound **T0A** was tested *in vivo* using dose-response experiments with the three reporter systems containing the *L. casei* and *S. mutans* fully native riboswitches as well as the *S. mutans* chimera. The structural similarity of **T0A** to THF/folinic acid along with the available x-ray crystallography data indicated that **T0A** was likely to bind to the wild-type THF riboswitch. However, since it was discovered by Batey *et al.* that compounds without C6-position side-chains do not appear to elicit the riboswitch regulatory effect *in vitro*¹¹⁰, it was unclear whether or not an *in vivo* response would be observed.

The results of these assays are shown in **Figure 46**. Surprisingly, the addition of **T0A** to the *L. casei* THF riboswitch reporter system caused a significant increase in the level of *GFP* expression, which was almost 4-fold higher at 5 mM ligand concentration. This effect was observed in the *S. mutans* fully native reporter strain to a much lesser extent, with an increase in normalised fluorescence of around 27% and in the *S. mutans* chimera even less so (8%). However, the No-RS control strain, which contained the same expression construct but without the riboswitch (*GFP* is constitutively expressed) also gave a small increase in expression at high **T0A** concentrations (18%). It is therefore likely that the small increase in expression observed in the *S. mutans* native and chimeric strains is due to off-target, background effects and not a result of an interaction between **T0A** and the riboswitch. The magnitude of the **T0A**-mediated increase in expression with the *L. casei* THF riboswitch is significantly larger than the background effect and the data can be fitted to a standard dose-response curve. It can therefore be inferred with confidence that the action of **T0A** on the *L. casei* THF riboswitch is causing the observed increase in expression levels. A possible explanation for this is that **T0A** is able to bind the

riboswitch aptamer domain, but does not induce the required conformational changes in the expression platform to elicit the regulatory response.

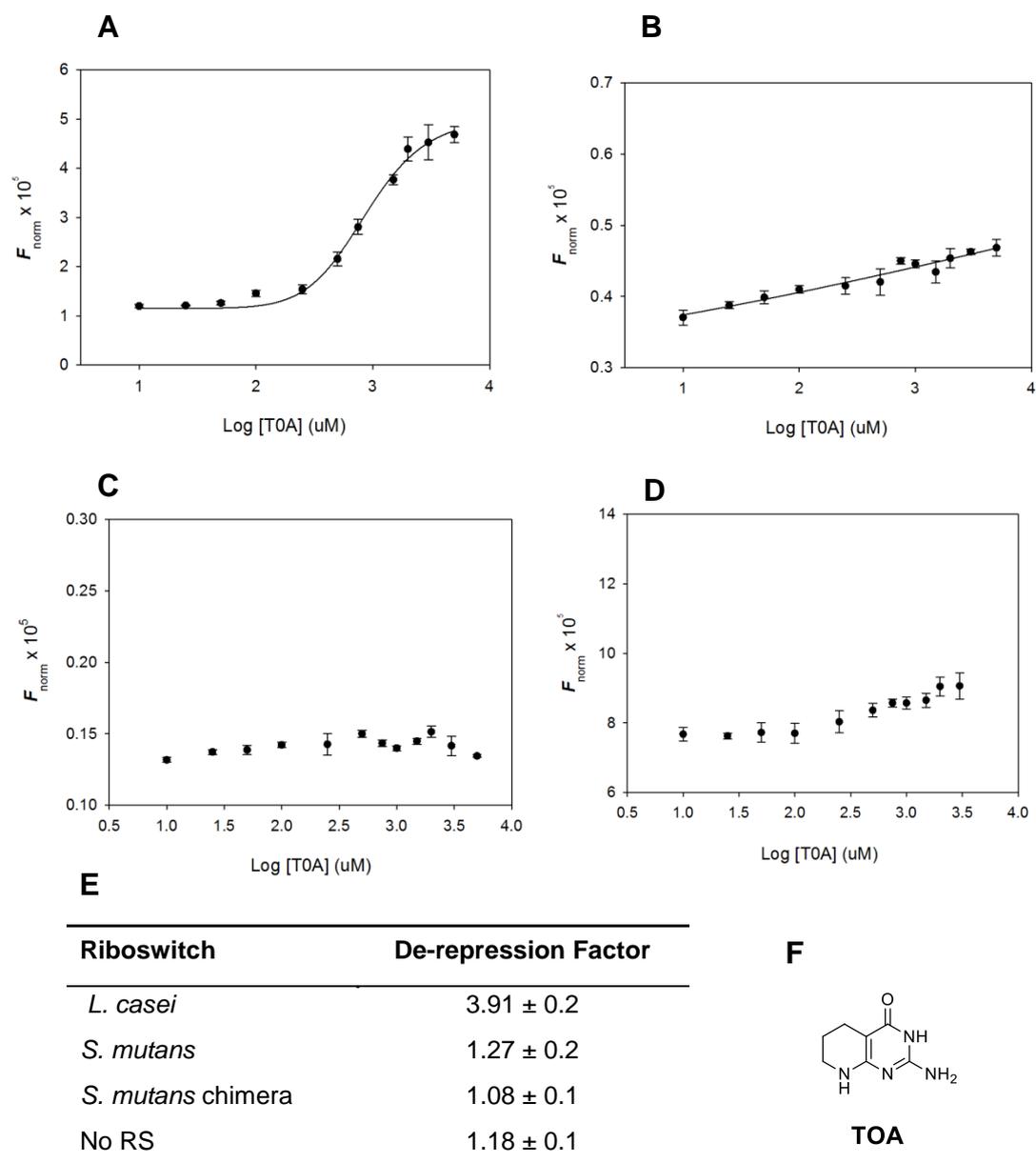


Figure 46 – Dose-response experiments for riboswitch reporter constructs with increasing concentrations of TOA. A) *L. casei*; B) *S. mutans* fully native; C) *S. mutans* chimera; D) no riboswitch; E) expression parameters; F) structure of TOA

Thus, **TOA** can be viewed as a competitive inhibitor for THF and related compounds. This effect has previously been reported for similar compounds *in vitro*¹¹⁰. On this assumption, the observed increase in gene expression can therefore be attributed the addition of **TOA** leading to the displacement of intracellular folates bound to the

riboswitch - thereby *de-repressing* gene expression. It is not clear why this effect only occurs for the *L. casei* THF riboswitch and is absent for the *S. mutans* native and chimeric systems. The binding affinities of **T0A** for the *L. casei* and *S. mutans* aptamer domains are discussed in Chapter 4.

3.2.2.8 *L. casei* riboswitch library expression response to a single concentration of T0A

Screening of **T0A** with the *L. casei* THF mutant library was performed at a single concentration of 2 mM **T0A** and the relative fluorescence of each mutant strain in the presence and absence of the ligand was determined (**Figure 47A**).

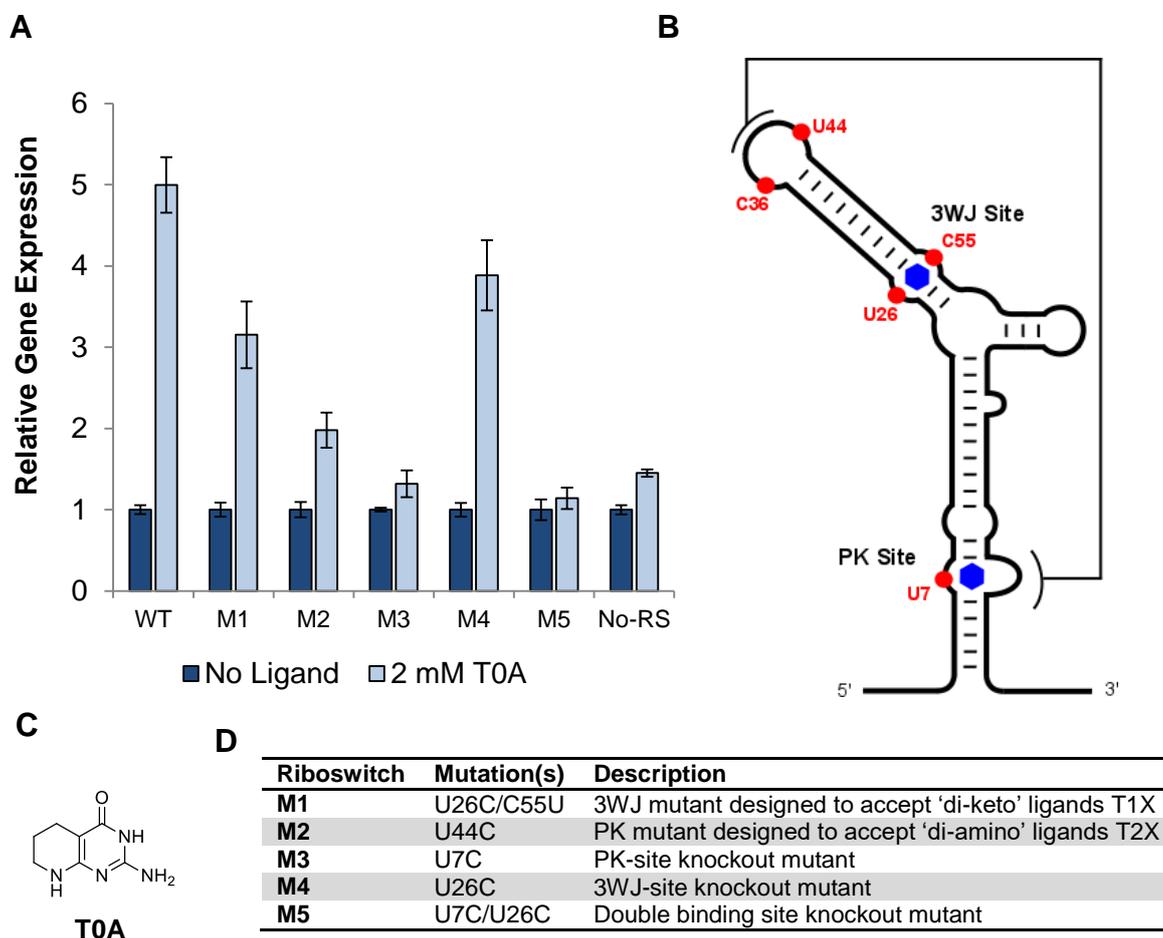


Figure 47 – A) Relative *GFP* expression levels for *L. cas* THF riboswitch mutants with and without T0A. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM T0A. Bar height indicates the mean of three experiments and error bars show the standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Structure of T0A. D) Overview of riboswitch mutants.

The goal of this series of experiments was to understand if the de-repressive effect of **T0A** was mediated through interaction at both of the ligand binding sites equally, or at one of the binding sites preferentially as observed with the wild-type riboswitch and folinic acid (see **Section 3.2.2.5**). To explain the observed gene expression responses, the riboswitch mutants can be segmented into two distinct groups based on the position of the mutations and the putative roles of the two binding sites. In addition to the wild-type (**WT**), compound **T0A** induced significant de-repression in the **M1** and **M4** riboswitch systems, giving approximately 3- and 4-fold increases in *GFP* expression respectively. These three riboswitches either have no mutations at all (wild-type) or have mutations at the 3WJ site only (**M1** and **M4**), thus leaving the PK site unmodified. Conversely, the magnitude of the de-repressive effect was significantly diminished with mutants **M2**, **M3** and **M5** – all of which contain mutated PK sites - with **T0A** inducing gene expression increases of around 1.9-, 1.3- and 1.1-fold respectively in these systems. Moreover, **T0A** also consistently caused a ~1.35-fold increase in *GFP* expression in the control strain (No-RS; containing the same construct minus the riboswitch) indicating that a proportion of the observed de-repression across all the strains can be attributed to off-target effects. Taking this in to account, **T0A** has no overall effect on riboswitches **M3** and **M5**, both of which have PK site mutations which are known to disrupt binding to the structurally analogous folinic acid. Therefore, analogous to folinic acid, the action of **T0A** at the PK site appears to be more crucial for the observed *in vivo* expression response than binding at the 3WJ site.

Comparing the absolute expression levels allows for a revealing analysis. The absolute expression levels of a subset of the riboswitch mutant library are shown in **Figure 48A**. Considering basal expression (dark blue), it can be seen that the wild-type riboswitch and the 3WJ site knockout mutant **M4** have significantly lower expression levels in the absence of any effector ligand. This is attributed to the action of intracellular folates at the intact PK site, triggering the riboswitch and leading to a background repression of gene expression. Mutant riboswitches **M3** and **M5**, which contain the U7C mutation to disrupt ligand binding at the PK site, exhibit basal expression levels 3-4-fold higher than the wild-type. This effect is presumably because THF riboswitch mutants with abrogated PK site binding are not effectively triggered by reduced folates at the concentrations naturally present within *E. coli*. Compound **T0A** is thought to act as a competitive inhibitor, binding the THF riboswitch but not triggering the regulatory response. In the case of the wild-type and **M4** systems (both of which have unmodified PK sites), addition of 2 mM **T0A** can be seen to de-repress expression to a level which is comparable to the basal levels of **M3** and **M5** (both of which have mutations to

disrupt PK site ligand binding). In other words, addition of 2 mM **T0A** causes the same increase in gene expression levels as disrupting the binding of intracellular folates via the PK site mutation U7C. This supports the hypothesis that **T0A** acts to displace endogenous folates from the THF riboswitch binding site(s) without triggering the riboswitch regulatory response.

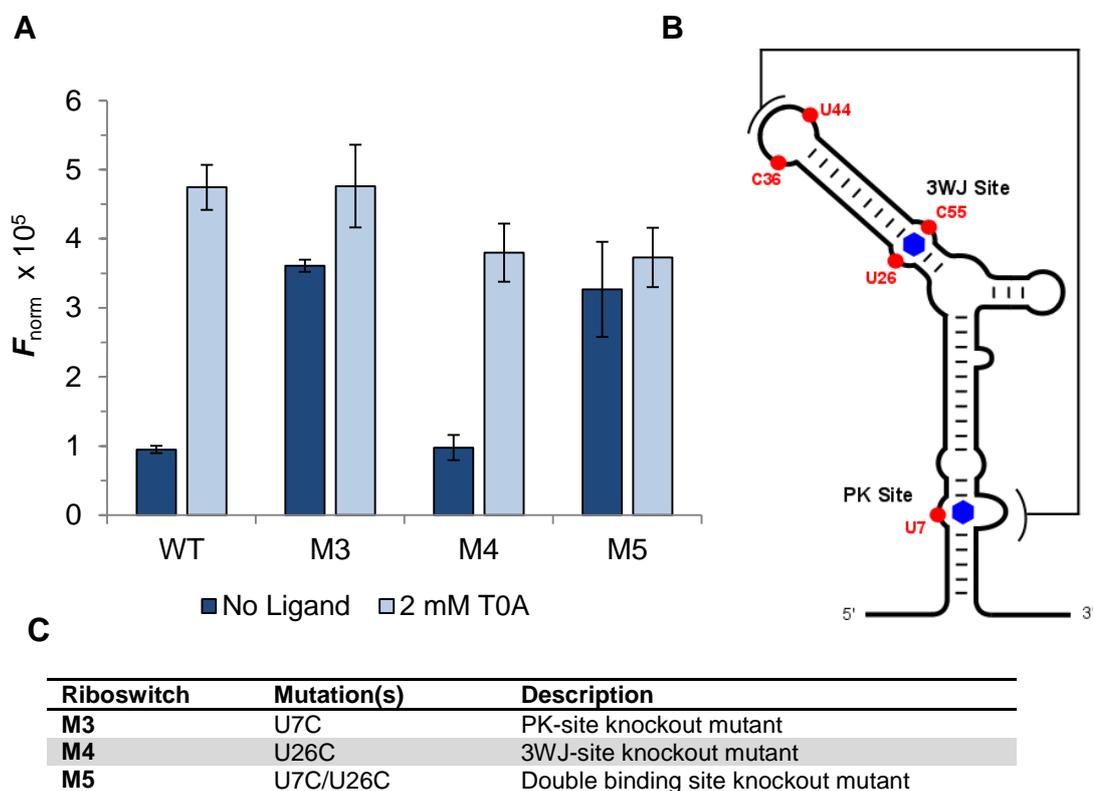
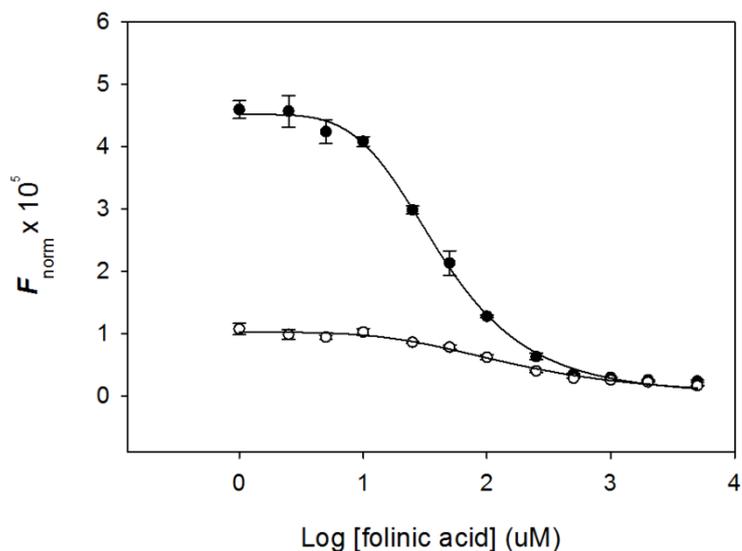


Figure 48 – Absolute fluorescence levels for the wild-type and mutant THF riboswitches in the presence (light blue) and absence (dark blue) of 2 mM T0A. Bar height indicates the mean of three experiments with error bars showing standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Overview of riboswitch mutants

3.2.2.9 Exploiting the effect of T0A to improve THF riboswitch expression parameters

An orthogonal riboswitch-based gene expression system, whereby a mutant riboswitch is successfully engineered to reject its natural compound and accept a synthetic analogue not present in the natural cellular metabolite pool has been shown by the Micklefield group to be a valuable tool in synthetic biology^{79, 80, 86} (see **Section 1.6.3**). One of the main advantages of this strategy is that the orthogonal riboswitch is no

longer triggered by its natural cellular metabolite, thereby eradicating problems with leaky (in the case of ‘ON’ switches) and low basal level (in the case of ‘OFF’ switches) expression. Indeed, low basal level expression with the wild-type system was noted during the development of an orthogonal PreQ riboswitch⁸⁶ and it has already been described for the wild-type *L. casei* THF riboswitch in this thesis (see **Section 3.2.2.3**).



Riboswitch	Maximum Repression (%)	T ₅₀ (μM)	Dynamic Range (μM)
<i>L. casei</i> wild-type	84 ± 4	130 ± 5	71 ± 31.0
<i>L. casei</i> + 2 mM T0A	95 ± 2	40 ± 2	58 ± 0.8

Figure 49 – OD-normalised *GFP* expression mediated by the *L. cas* wild-type THF riboswitch in response to addition of folinic acid. Black circles indicate OD-normalised fluorescence in response to the addition of folinic acid in the presence of 2 mM T0A; white circles indicate response to addition of folinic acid only. Data points represent the mean of three experiments with error bars showing standard deviation.

An alternative strategy to overcome the low basal level expression caused by binding of natural metabolites would be to selectively displace any intracellular ligands from the riboswitch binding site via the addition of a competitive inhibitor compound. The binding affinity of such an inhibitor would need to be sufficiently great to supplant the incumbent activator ligand, but the inhibitor must also be labile enough to be itself displaced on addition of an exogenous activating compound. It was considered that the de-repressive effect of **T0A** could be exploited to improve the performance of the wild-

type riboswitch as a gene expression tool, by significantly increasing the basal expression levels. To investigate this, a dose-response experiment was performed whereby **T0A** was added to the bacterial growth media and then *GFP* expression levels were measured at increasing folinic acid concentrations in the wild-type reporter system. A second dose-response experiment – without **T0A** in the growth media – was carried out in parallel. The results of these two experiments are shown in **Figure 49**.

The de-repressive effect of **T0A** facilitated a higher maximum repression of up to 95% when folinic acid was added, compared with up to 84% repression observed with folinic acid alone. Moreover, the use of the two ligands gave more than a 3-fold reduction in the folinic acid T_{50} and had a minimal effect on the overall dynamic range. Indeed, **Figure 49** illustrates the advantage of the two-ligand expression system clearly: adding **T0A** to the growth media significantly increases the range of expression outputs over which the riboswitch is able to operate.

Interestingly, when the two-ligand experiment is performed in the reverse manner (ie. a fixed concentration of 1 or 2 mM folinic acid is added to the growth media and fluorescence output is measured at increasing **T0A** concentrations), there is no observed response to **T0A** even at concentrations as high as 5 mM. At the relatively low reduced folate concentrations within the cell^{171, 172}, addition of an excess of exogenous **T0A** is able competitively bind (but not activate) the riboswitch, whilst at high folinic acid conditions (when 1 or 2 mM folinic acid is added to the growth media) the competitive inhibition activity of **T0A** is eliminated. This suggests that the folinic acid binding affinity is greater than that of **T0A** and this is explored in Chapter 4.

3.2.2.10 Investigating the regulatory effect of T0A analogues

It is possible that the inhibitory (as opposed to regulatory) effect of **T0A** is due to the lack of the *pABA* side chain at the C6-position, which is present in the regulatory compound folinic acid. It has previously been reported in the literature that there are a number of compounds which are able to bind the THF riboswitch, but are very poor effectors of the regulatory response *in vitro* and this was attributed to the lack of a *pABA* side chain¹¹⁰.

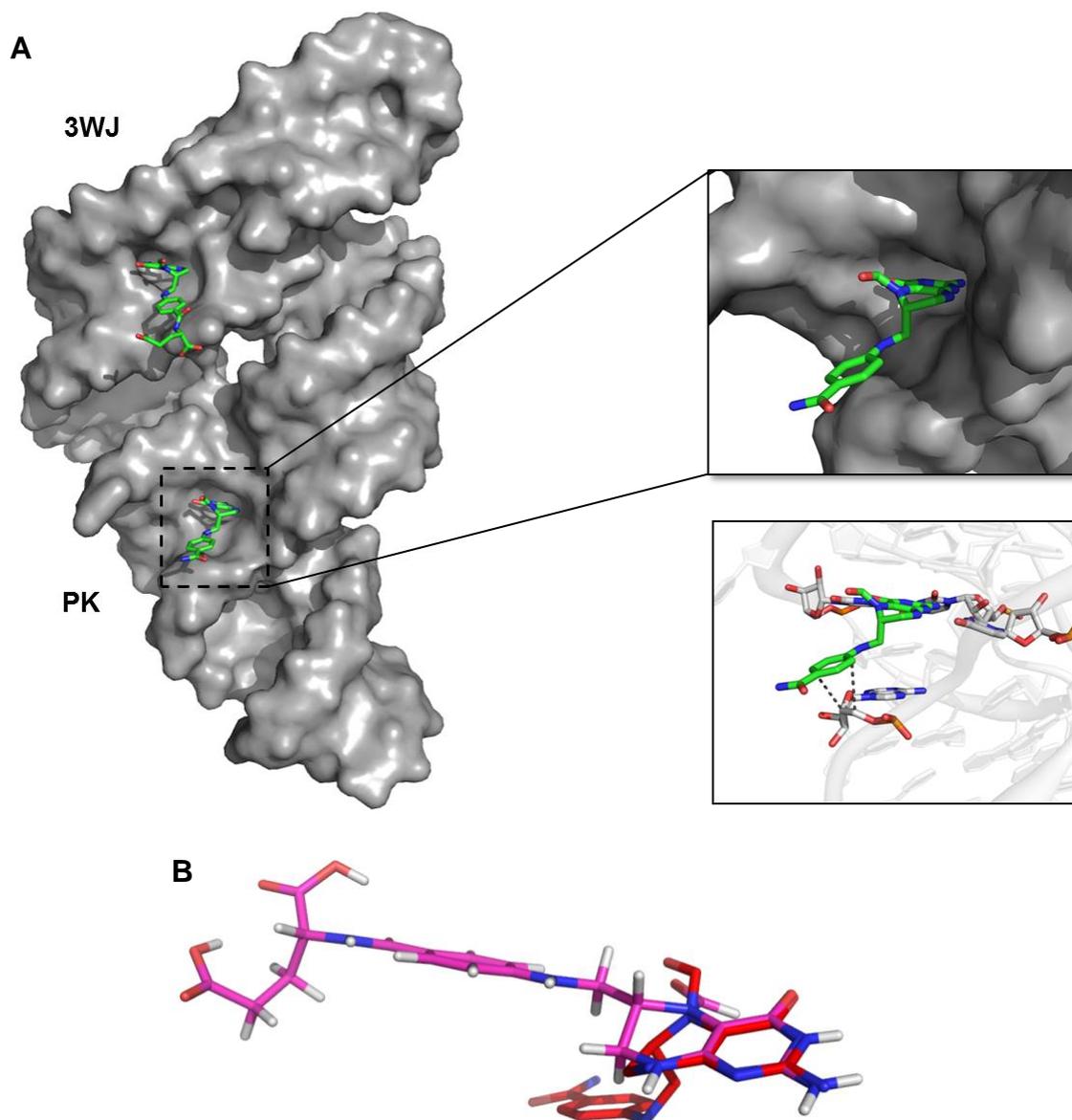


Figure 50 – A) X-ray crystallography structure of the *S. mutans* THF aptamer domain bound to folinic acid¹⁰⁴ with expanded PK site shown. At both ligand binding sites folinic acid adopts a high energy pseudo-boat conformation, positioning the pABA side chain in close proximity to the backbone RNA. **B)** Comparison of energy minimised free solution conformation of folinic acid (pink) and pseudo-boat conformation observed at the THF riboswitch ligand binding sites (red).

In addition, analysis of the x-ray crystallography structure of the THF aptamer domain reveals that at both ligand binding sites, folinic acid adopts a highly strained pseudo-boat conformation which positions the pABA moiety in close proximity to the backbone RNA (**Figure 50A**). At the 3WJ site, this high energy conformation is presumably stabilised by the stacking interaction between the THF pABA group and G68, whilst the pABA moiety appears to interact with the phosphate backbone at the PK site. The lowest energy conformation of THF/folinic acid in free solution places the pABA side-

chain in a pseudo-equatorial position (**Figure 50B**), which implies that the interaction between the *p*ABA moiety and the backbone RNA may be required for the riboswitch to effect a regulatory response.

In order to probe this hypothesis further, a small number of **T0A** analogues functionalised at the C6-position were synthesised as described in Chapter 2. It was anticipated that compounds in this series possessing C6-position side chains most closely resembling the *p*ABA moiety of THF/folinic acid (ie. **T0D**) were more likely to elicit the regulatory effect and repress GFP expression. Conversely, compounds with smaller side-chains that have a greater structural similarity with **T0A** (ie. compounds **T0B** and **T0C**) were considered more likely to be inhibitors and hence an increase GFP expression was expected. These compounds were screened against the mutant library at a single concentration and the results are shown in **Figure 51**.

Compounds **T0B** and **T0C**, bearing C6-methyl and C6-ethyl side-chains respectively, both caused a 2.5-fold increase in gene expression with the wild-type riboswitch. This effect is modest in comparison with **T0A**, which was shown to increase expression levels 5-fold. In addition to the wild-type (and analogous to **T0A**) compounds **T0B** and **T0C** also effected a de-repressive response with riboswitches **M1** and **M4**, both of which contain 3WJ site mutations only and have an in-tact PK site. Moreover, no effect was seen for riboswitches **M2**, **M3** and **M5**, all of which contain mutations at the PK site, indicating that (like **T0A**), compounds **T0B** and **T0C** primarily act at the PK site. Disappointingly, the data obtained for compound **T0D** indicate that this ligand has no overall effect on riboswitch-mediated gene expression. An overall increase in expression levels of 0.5-1.0 fold was seen across all the riboswitch mutants, including the No-RS control strain which lacks the riboswitch sequence entirely. Ruling out intrinsic ligand fluorescence (which is accounted for in the assay design) it is assumed that **T0D**, or a metabolite thereof, has an off-target effect which increases gene expression levels in a manner which is independent of the THF riboswitch construct.

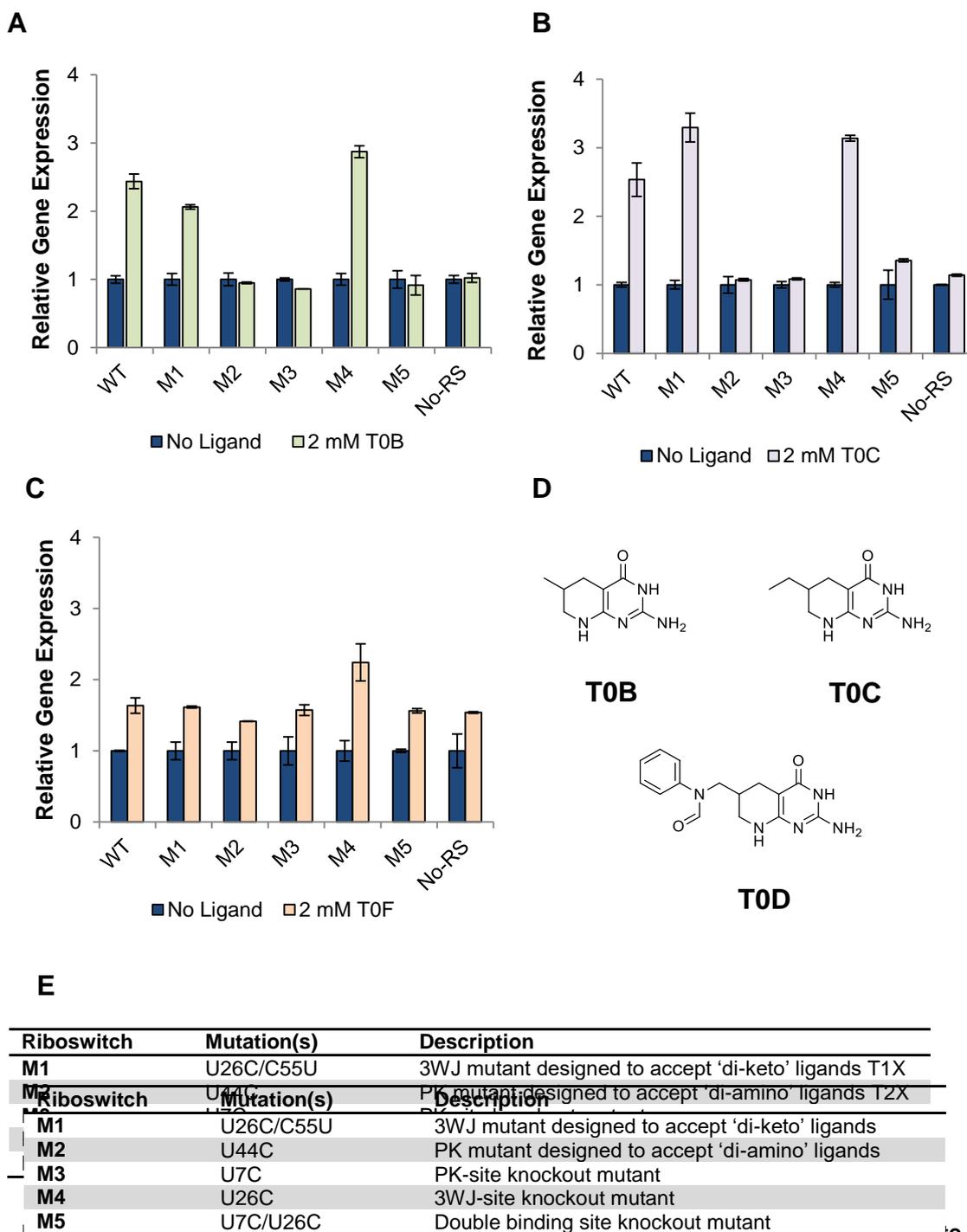


Figure S1. Relative GFP expression levels for *Er. coli* THF riboswitch mutants in response to T0A analogues. Dark blue bars indicate normalised basal fluorescence; pastel coloured bars indicate relative fluorescence in the presence of 2 mM ligand. Bar height indicates the mean of three independent experiments and error bars show the standard deviation. A) T0B; B) T0C; C) T0F; D) Structures of relevant compounds. E) Overview of riboswitch mutants.

The riboswitch was also tested with the commercially available **T0A** analogue lometrexol by colleague Dr. Helen Vincent, which is known to bind the THF riboswitch aptamer with comparable affinity to folinic acid and is able to effect a moderate

regulatory response *in vitro*¹¹⁰. Due to poor compound solubility and the lack of availability of material, lometrexol was only tested against the wild-type and control strains at a single concentration of 1 mM. Being structurally very similar to THF and folinic acid, but possessing a carbon atom at positions 5- and 10- rather than nitrogen, it was anticipated that lometrexol would bind to, and activate the riboswitch leading to a reduction in gene expression. Surprisingly however, the compound caused a moderate increase in gene expression levels with the wild-type riboswitch at this concentration (**Figure 52**). No effect was seen for the control strain, indicating that the observed depression is a result of a direct interaction between the riboswitch and ligand. The fact that folinic acid is able to bind and activate the riboswitch, whilst lometrexol appears to act as an inhibitor, is particularly surprising given the structural similarities between the two compounds. The available x-ray crystallography data indicates that the N5 and N10 atoms of folinic acid do not appear to directly interact with the RNA¹⁰⁴, and this is further supported by the observation that 5,10-methenyl-THF is able to bind the aptamer domain⁵⁷. It is possible that N10 (which is replaced by carbon in lometrexol) is required for the regulatory effect because of the conformational restriction that it places on the adjacent phenyl group.

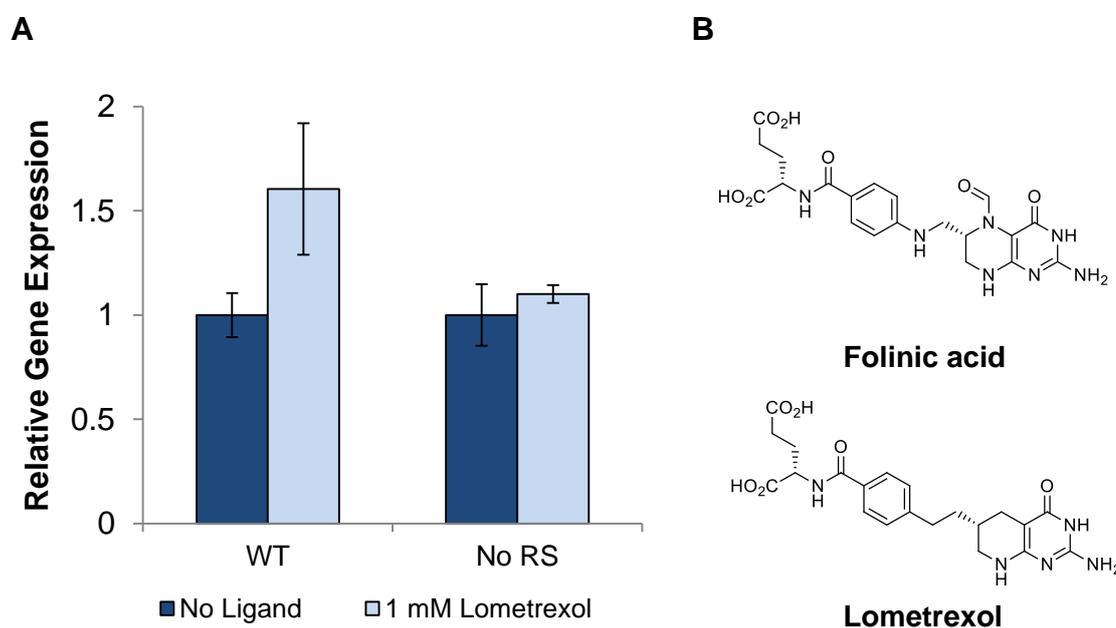


Figure 52 – A) Relative GFP expression levels for the *L. cas* THF riboswitch in response to the commercial compound lometrexol. Bar height indicates the mean of three experiments and error bars show the standard deviation (data from Dr. Helen Vincent). B) Structures of folinic acid and lometrexol.

Since the interaction between the folinic acid/THF *p*ABA group and the backbone RNA appears to be a requirement for the riboswitch regulatory response, it may be that the restricted rotation around the N10-C11 bond, as a result of N10 lone pair delocalisation into the phenyl ring, facilitates the adoption of this active conformation. In the case of lometrexol, where N10 is replaced with carbon, the increased phenyl ring rotational freedom may be a barrier to active conformation adoption. Indeed, given that the de-repressive (or inhibitory) effect of lometrexol is less pronounced than **T0A**, it may be that lometrexol is able to adopt both activating and inhibitory conformations within the ligand binding sites and the net de-repressive effect is a result of the equilibrium favouring the inhibitory state.

3.2.2.11 Screening **T1A** with the *L. casei* mutant library

Compound **T1A** (**Figure 53C**) was screened at a single 2 mM concentration against the *L. casei* riboswitch mutant library and the results are shown in **Figure 53A**. **T1A** was designed to bind the mutant THF riboswitch **M1** at the mutated 3WJ site. Given that no *in vivo* effect had been observed with the *Smu folT/Bsu metE* **M1** riboswitch and **T1A**, it was anticipated that there would be no observed effect *in vivo* with the *L. casei* riboswitch system. However, **T1A** was shown to cause a 40% reduction in gene expression with riboswitch **M1** in contrast to the other riboswitch mutants with which the ligand had no overall effect (within error). This effect is presumably a result of **T1A** binding to the rationally mutated 3WJ site of the **M1** riboswitch. This was surprising, given that the working hypothesis was that riboswitch activating compounds required a C6-substituted *p*ABA moiety analogous to THF/folinic acid. It is not clear if the observed repressive effect is a result of **T1A** binding to the **M1** 3WJ site alone, or if it is caused by a combination of **T1A-M1**(3WJ) binding and intracellular THF/folinic acid acting at the unmodified **M1** PK site. In order to differentiate these possibilities, a new riboswitch mutant was generated, combining the **M1** 3WJ site mutations (U26C/C55U) with PK site knockout (U7C).

Site-directed mutagenesis yielded the triple mutant **M6** (U7C/U26C/C55U), thereby combining the **M1** mutations which modify the 3WJ site to accept **T1A** (U26C/C55U) with the PK site-disruptive mutation U7C. A dose-response *GFP* expression assay with mutant **M6** and increasing concentrations of **T1A** was performed and the results are shown in **Figure 54**. Riboswitch **M6** did not respond at all to **T1A** at concentrations of up to 2 mM, indicating that an intact PK site is required for riboswitch activation by **T1A**.

This suggests that the **M1** 'OFF'-state complex is actually a three component interaction involving **M1**, **T1A** and presumably endogenous THF/folinic acid at the unmodified PK site.

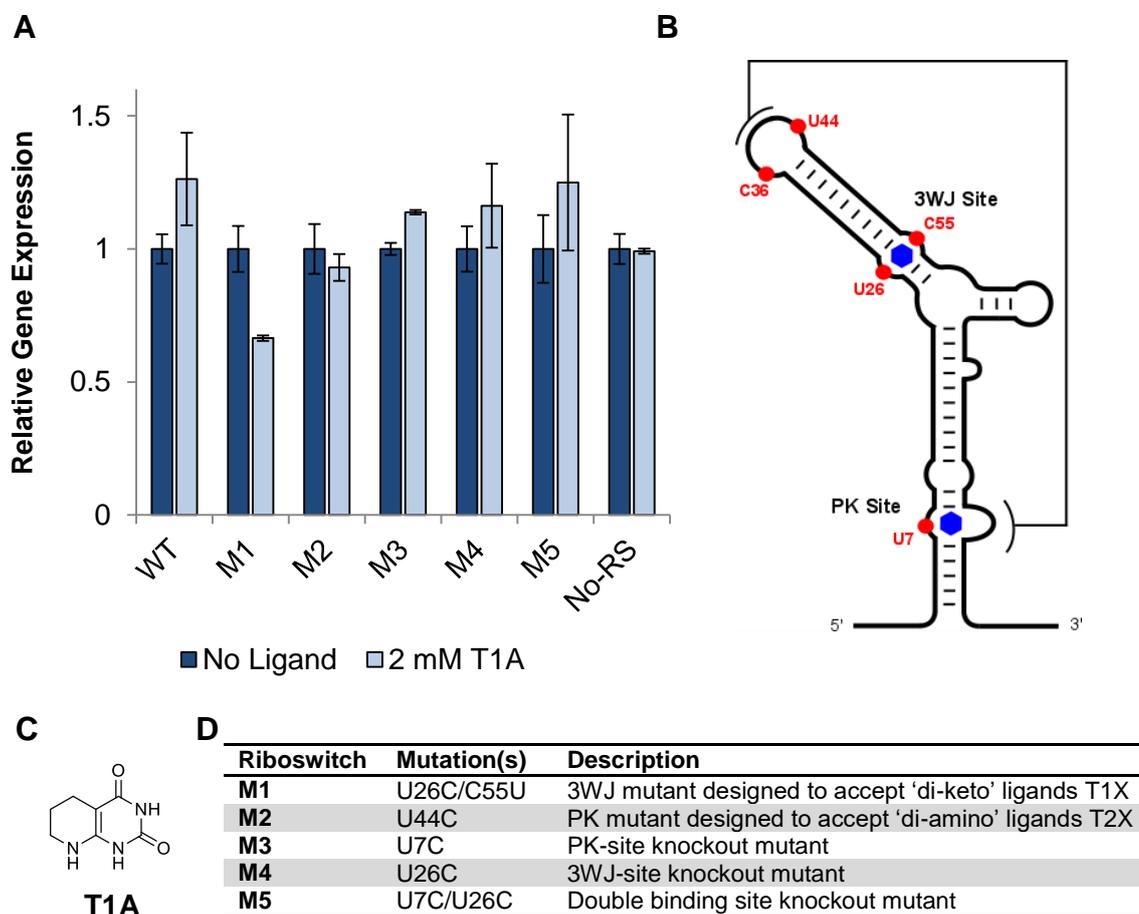


Figure 53 - Relative GFP expression levels for *L. cas* THF riboswitch mutants with and without T1A. Dark blue bars indicate normalised basal fluorescence; light blue bars indicate relative fluorescence in the presence of 2 mM T1A. Bar height indicates the mean of three experiments and error bars show the standard deviation. B) Proposed ligand-bound secondary structure of the *L. casei* THF riboswitch. Key residues targeted for mutagenesis are highlighted in red. C) Structure of T1A. D) Overview of riboswitch mutants.

Following this observation, the **M1** riboswitch was then further explored utilising all three ligand classes known to have regulatory or inhibitory effects. In a similar fashion to the wild-type-T0A-folinic acid system described in **Section 3.3.9**, a series of two ligand combinations were investigated with the goal of improving the expression parameters of the **M1** riboswitch. These results are shown in **Figure 55**. A combination of 2 mM **T1A** and 5 mM folinic acid gave a modest increase in maximum repression levels compared to the wild-type, with the two ligand system effecting up to 78%

repression versus around 70% with folinic acid alone. Additionally, a fixed concentration of 2 mM **T1A** reduced the T_{50} for folinic acid by around 6-fold.

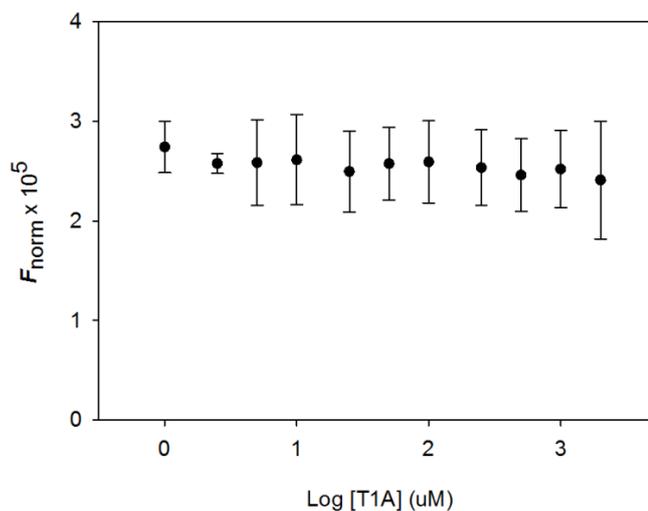


Figure 54 – *GFP-Iv* expression response by the *L. cas* **M6** (U7C/U26C/C55U) riboswitch to increasing **T1A** concentrations. Data points are the mean of three experiments and error bars indicate the standard deviation.

Combining the de-repressor ligand **T0A** with **T1A** and the **M1** riboswitch generated interesting results. As previously shown, **T0A** is able to de-repress the basal expression levels of the **M1** riboswitch, with the working hypothesis being that **T0A** binds and displaces intracellular folates at the PK site, but does induce the conformational changes required for the regulatory response. It was then shown that increasing **T1A** concentrations at a fixed **T0A** concentration of 2 mM, resulted in repression of up to around 50% at 2 mM **T1A** (**Figure 55B**). This effect was surprising. It appears that a combination of two ligands (**T0A** and **T1A**) neither of which have 6-substituted *pABA* side chains thought to be required for riboswitch activation, are able to work in concert to produce a regulatory response in the **M1** riboswitch.

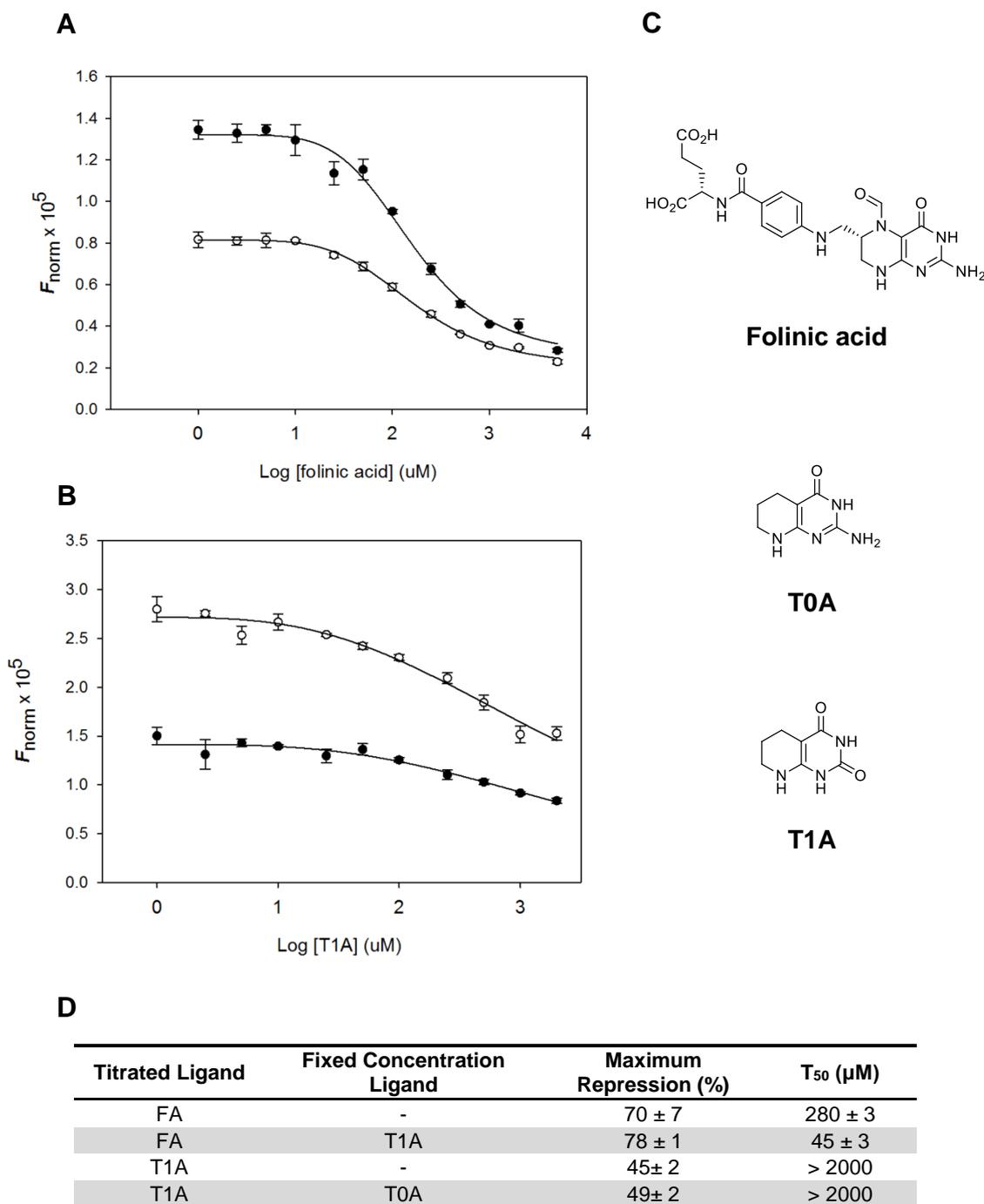


Figure 55 - Dose-responsive repression of *GFP* expression by the *L. cas* THF-M1 riboswitch. (A) Dose response curves showing normalised gene expression levels at varying folinic acid concentrations. Black circles show varying folinic acid concentrations only and white circles show varying folinic acid concentrations at fixed 2 mM T1A. (B) Dose response curves showing normalised gene expression levels at varying T1A concentrations. Black circles show varying T1A concentrations only and white circles show varying T1A concentrations at fixed 2 mM T0A. (C) Structures of folinic acid, T0A and T1A. (D) M1 riboswitch expression parameters.

3.4 Chapter Conclusions

The THF riboswitch has been well characterised *in vitro*, with studies exploring the ligand binding- and regulatory-specificities as well as structural information obtained through x-ray crystallography^{57, 104, 110, 173}. However, until now there have been no reports of the characterisation of the THF riboswitch in an *in vivo* setting. The work presented in this chapter has demonstrated that the THF riboswitch is able to dose-responsively control the expression of the *GFP* reporter gene in a non-native host (*E. coli*). This was shown firstly using the well characterised *Smu folT/Bsu metE* chimeric riboswitch which functions as a transcriptional-‘OFF’ switch (see **Section 3.2.1**), and then exemplified with the native *L. casei* THF riboswitch, thought to repress expression by inhibition of translation initiation (see **Section 3.2.2**). These two systems were then used to identify the role of the two ligand binding sites in mediating the riboswitch regulatory function *in vivo*. This confirmed that ligand binding to the PK site is the primary mechanism responsible for the riboswitch regulatory response, but also that the *L. casei* riboswitch is able to retain significant regulatory ability when binding to the PK site is abrogated – highlighting the importance of the 3WJ site in this system (see **Sections 3.2.1.4** and **3.2.2.5**).

The observation that some compounds are able to bind the THF riboswitch with high affinity but not induce a regulatory effect¹¹⁰ was confirmed *in vivo* and indeed, the THF riboswitch appears to have a very narrow ligand specificity with respect to compounds which activate the switching mechanism. A series of synthetic THF/folinic acid analogues possessing C6-position side-chains were tested in the *in vivo* reporter gene assay (see **Sections 3.2.2.7** and **3.2.2.10**). Surprisingly, compounds **T0A**, **T0B** and **T0C**, bearing C6-H, C6-Me and C6-Et moieties respectively caused an increase in *GFP* expression with the *L. casei* THF riboswitch rather than the expected decrease. In particular, compound **T0A** was shown to have a significant dose-responsive, de-repressive effect, eliciting up to a four-fold increase in *GFP* expression at saturating concentrations. It is hypothesised that this de-repressive effect is a result of competitive inhibition: **T0A** is able to bind to the THF riboswitch and displace endogenous reduced folates which are acting to repress the basal level of gene expression. However, whilst **T0A** is able to competitively bind to the aptamer domain ligand binding sites, it does not elicit the riboswitch regulatory response, causing the observed de-repression of *GFP* expression. This hypothesis is consistent with observations by Batey *et al.* that ligands without C6-position *pABA* moieties are able to bind the THF aptamer but do not elicit

the riboswitch regulatory effect *in vitro*¹¹⁰. In order to support this hypothesis, it is necessary to demonstrate that **T0A** does indeed bind to the THF aptamer domain. An *in vitro* analysis of the **WT-T0A** interaction using ITC is presented in Chapter 4.

The second goal of this project was the re-engineering of the THF riboswitch to create a novel small-molecule-inducible gene expression tool. The major drawback with the use of wild-type riboswitches as gene expression devices is that they are activated by their cognate ligands which are already present within the host cell. In the case of 'ON' switches, this leads to leaky expression and in the case of 'OFF' switches: low basal levels. The Micklefield group has successfully employed a strategy of engineering orthogonal mutant riboswitch-ligand pairs to overcome these pitfalls^{79, 81, 86}. Given that the THF riboswitch contains two ligand binding sites which recognise the ligand via two distinct motifs, the orthogonality strategy was more challenging to implement in this system. However, it was shown that it is possible to re-engineer the THF 3WJ binding site such that the riboswitch regulatory response was triggered by the addition of the synthetic ligand **T1A**. Indeed, the *L. casei* THF riboswitch mutant **M1** (U26C/C55U) was able to effect up to 40% repression of gene expression at saturating **T1A** concentrations (see **Section 3.2.2.11**). Introduction of a mutation known to abrogate binding at the riboswitch PK binding site to create the triple mutant U7C/26C/C55U demonstrated that the **T1A**-elicited regulatory effect is not observed when the PK site is disrupted. It is therefore hypothesised that binding of endogenous reduced folates to the unmodified PK site of **M1** is required for the **M1-T1A** riboswitch regulatory response. In order to confirm that the re-engineered 3WJ site of **M1** does indeed bind **T1A**, an *in vitro* analysis is required, which can be found in Chapter 4.

Whilst re-engineering of the 3WJ binding site was successful, given the secondary role of this site in mediating the riboswitch regulatory response, this was not sufficient to create an effective orthogonal system. Disappointingly, attempts to re-engineer the PK binding site in a similar manner were not successful. Moving forward, the engineering of the PK site should remain a priority given that binding to this site has repeatedly been shown to be a requirement for eliciting the regulatory effect of the THF riboswitch.

As an alternative to the orthogonal mutant riboswitch-ligand pair strategy, it was shown that it was possible to overcome the low basal levels of the wild-type system by exploiting the inhibitory effect of the compound **T0A**. By using **T0A** to de-repress basal gene expression levels the performance of the wild-type *L. casei* THF riboswitch as a gene expression tool was significantly improved (see **Section 3.2.2.9**). Basal

expression levels were increased by up to four-fold without compromising 'OFF'-state expression. This allowed for up to 95% repression of gene expression on addition of folinic acid, with a significantly reduced T_{50} compared with the wild-type system in the absence of **TOA**. However, the use of folinic acid as an effector ligand in a gene expression system is not desirable, given its inability to permeate the bacterial cell membrane without the expression of a heterologous folate transporter. Therefore, one future avenue of study would be the synthesis of a folinic acid analogue, which could penetrate the cell membrane without the use of a non-native transport protein. If this was achieved, the *L. casei* wild-type THF riboswitch – **TOA** expression system could prove to be a valuable resource in the synthetic biology toolkit.

Chapter 4

Characterisation of THF riboswitch- ligand interactions *in vitro*

4.1 Chapter Introduction

4.1.1 Overview of isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is an expedient method for determining the thermodynamics of molecular interactions. It is most commonly used to measure the binding interaction between biological molecules, their small-molecule ligands and/or other biomolecular binding partners. One of the major advantages of ITC is that, given the right conditions, a host of thermodynamic parameters are obtainable from a single experiment¹⁷⁴. This includes ΔG , ΔH , ΔS , the binding affinity, K_A , and the binding stoichiometry, n . These parameters are obtained by directly measuring the change in heat (ΔH_{obs}) as a result of the interaction when one binding partner is injected into a solution of another. A diagram of a typical titration calorimetry apparatus is shown in **Figure 56**. The instrument contains a sample cell, used to contain binding partner A (usually the biomolecule) into which is injected binding partner B (usually the small-molecule) using an injection syringe. A reference cell containing water is also present, and the system functions by maintaining a constant temperature difference for both the sample and reference cells relative to the adiabatic jacket (ΔT_2). The heating of each cell is controlled by an automated electronic feedback system and the instrument monitors the energy required to maintain a constant temperature difference between the two cells over time (ΔT_1).

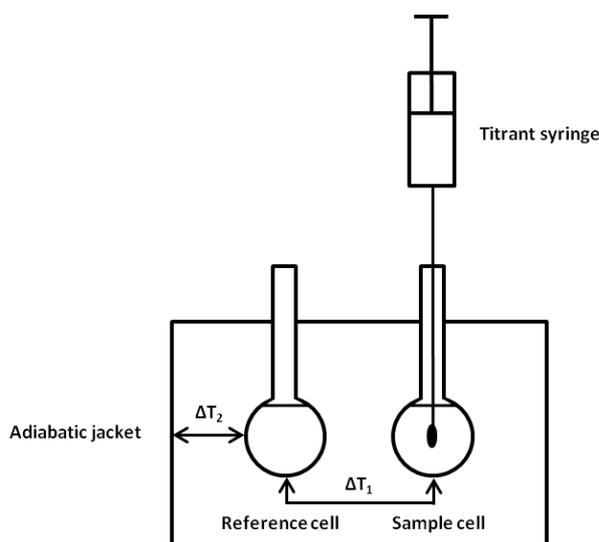


Figure 56 - Diagram showing the apparatus of a typical isothermal titration calorimeter.

A typical experimental set up comprises sequential small volume injections of a solution of one binding partner – typically the small-molecule – into the sample cell containing the other binding partner. The change in enthalpy resulting from the interaction between the two species is then measured by the instrument. Since the majority of ITC-studied molecular interactions are exothermic, generally the change in heat is measured as a change in the power required to heat the reference cell to compensate for the rise in the temperature of the sample cell. It is then possible to take this raw data of differential power consumption versus time, and integrate it to determine the total heat evolved as a result of each injection. The change in heat recorded with each injection is directly related to the relative population of binding complexes formed between the two interacting molecules. Thus, plotting the measured enthalpy per mole of injectant against the molar ratio of the two binding partners generates a binding curve from which the desired thermodynamic parameters can be extracted. The curve is typically sigmoidal in shape, with the early injections representing the maximal enthalpy of binding (when the amount of free analyte in the cell is highest and therefore all titrant is bound) and the final injections representing the background enthalpy of mixing as a result of saturation of the analyte. This background enthalpy of mixing is typically subtracted from the whole dataset. The binding curve can be used to directly determine ΔH , the binding affinity K_A (from the gradient at the inflection point on the curve) and stoichiometry of the binding interaction (the molar ratio at the inflection point). From these parameters, the dissociation constant K_D , the change in Gibbs free energy ΔG , and the change in entropy of the interaction ΔS can all be calculated using the equations presented in **Figure 57**.

$$\Delta G = -RT \ln K_A$$

$$\Delta G = \Delta H - T\Delta S$$

$$K_D = \frac{1}{K_A}$$

Where:

ΔG = Change in Gibbs Energy (Kcal mol⁻¹)

ΔH = Change in enthalpy (Kcal mol⁻¹)

ΔS = Change in entropy (Kcal mol⁻¹)

T = Temperature (K)

R = 1.987 x 10⁻³ Kcal mol⁻¹

K_D = Dissociation constant

K_A = Binding constant

Figure 57 – Equations for determining the binding parameters of a molecular interaction. The change in enthalpy (ΔH) and binding constant (K_A) are directly determined from the ITC experiment and the remaining parameters are calculated using these equations.

Over the last 30 years, ITC has principally been utilised to monitor biochemical interactions of proteins, with particular application in probing small-molecule-protein interactions in drug discovery¹⁷⁵. Over the last 15 years, improvements in the sensitivity of titration calorimeters, as well as advances in the synthesis and purification of *in vitro* derived RNAs, has facilitated the study of RNA binding interactions using ITC. In the case of riboswitches, ITC is primarily used to explore the ligand binding specificity of the aptamer domain, as a means of ratifying *in vivo* observations.

ITC has been used to explore RNA-ligand interactions for a variety of riboswitch classes including purine^{79, 176}, PreQ⁸⁶, 2'-deoxyguanosine¹⁷⁷ and SAM¹⁷⁸ aptamers. For example, Batey *et al.* have used ITC to probe the ligand binding specificity of the THF riboswitch aptamer¹¹⁰, identifying a number of commercially available purines, pyrimidines and folate derivatives which were able to bind. In addition to detailed structural information obtainable from x-ray crystallography, ITC can be a useful complimentary technique in identifying binding site residues and key ligand functional groups required for binding interactions. For instance, Batey *et al.* have shown that ITC can be used to probe the ability of purine riboswitches to accept a series of non-natural purine and pyrimidine derivatives¹⁷⁶. The group was able to use the relative binding affinities of the aforementioned ligands to identify the functional groups necessary for the RNA-ligand binding interaction. In particular, they identified a Watson-Crick interaction between a key binding site residue and the accepted ligands.

The concentration of divalent cations such as Mg²⁺ is known to have a significant impact on RNA tertiary structure and in the case of riboswitches, this effect has been studied using ITC. Ferre-D'Amare *et al.*¹⁷⁹ reported an ITC study of the thiamine pyrophosphate (TPP) riboswitch, whereby the ligand binding parameters were obtained at a variety of Mg²⁺ concentrations. The group observed binding affinities more than an order of magnitude greater at higher magnesium concentrations, as well as a reduction in the unfavourable entropy of binding as the Mg²⁺ concentration increased. Taken together, these observations indicated that divalent cations are able to pre-organise the TPP riboswitch aptamer domain into a conformation more amenable to ligand binding.

ITC has also been used in combination with other analytical techniques such as NMR spectroscopy to probe the structure, function and regulatory mechanisms of riboswitches. Wilson *et al.* investigated the functional and thermodynamic implications of the alternate conformations of the S-adenosylmethionine (SAM)-responsive S_{mk} box riboswitch in the ligand bound and unbound states¹⁷⁸. The group used ITC in combination with NMR to measure the ligand binding affinities of two S_{mk} box RNA

constructs, one of which was the wild-type and the other a shorter sequence designed to adopt a conformation similar to the ligand-bound state. From these studies they hypothesised a switching mechanism that occurs via ligand sampling of an equilibrium of distinct RNA conformations¹⁷⁸.

4.1.2 *In vitro* Transcription Assays

For riboswitches which operate via transcriptional mechanisms, it is possible to explore the ligand specificity using an *in vitro* transcription assay. This has a number of advantages over using ITC. In particular, this method allows for the identification of ligands which both bind to the riboswitch aptamer *and* elicit a regulatory response. Assays are typically performed by incubating the DNA template, and an RNA polymerase with increasing concentrations of the ligand of interest and then analysing the transcripts produced by urea PAGE. The ratio of truncated to full-length mRNA transcripts can be used to quantify the ability of the ligand of interest to activate the riboswitch. In the case of a transcriptional-‘ON’ switch, increasing concentrations of the effector ligand will favour the formation of full length transcripts, whilst increasing concentrations of ligands which activate ‘OFF’ switches will lead to more truncated mRNAs. Assays of this type have been previously employed to study a variety of riboswitch classes^{66, 180, 181, 182}. For example, Breaker *et al.* utilised a single-round transcription assay to demonstrate that the glycine riboswitch responded selectively to glycine whilst rejecting serine and related glycine analogues³³. Furthermore, by plotting the truncated transcript levels versus ligand concentration and fitting these data points to a Hill equation, the group were able to demonstrate cooperative ligand binding by the tandem glycine riboswitch system.

Transcription termination assays have also proven to be a useful tool to explore the relationship between riboswitch structure and regulatory ability. La Fontaine *et al.* were able to utilise such an assay to probe the transcriptional-‘OFF’ SAM-I riboswitch. The group were able to show that strengthening the aptamer domain P1 stem by the insertion of additional base pairs resulted in an increase in transcriptional termination in the absence of the effector ligand. Conversely, removing base pairs to destabilise the P1 stem resulted in a riboswitch which lacked the ability to terminate transcription in the presence of the cognate ligand. From this, it was possible to determine that the ligand binding event stabilised the formation of the aptamer P1 stem, ultimately leading to the conformational changes in the expression platform required for the formation of a

transcriptional terminator. Furthermore, the group utilised *in vitro* transcription assays to demonstrate that the presence of the aptamer domain P4 stem gave a 3-fold reduction in the T_{50} (the concentration of ligand required to effect a 50% increase in truncated transcript formation), thereby highlighting the function of this seemingly redundant structural feature¹¹⁹.

In the development of chimeric riboswitches as gene expression tools, whereby aptamer domains and expression platforms from distinct riboswitch classes and organisms are combined, *in vitro* transcription assays have proven a valuable means of demonstrating and optimising their functionality. Indeed, assays of this type have formed the basis of a number of studies, notably in work by Batey *et al.*, who created an extensive library of chimeric riboswitches, comprised of natural and non-natural aptamer domains fused to a selection of transcriptional-‘OFF’ expression platforms. The assays were again utilised to both optimise the regulatory output of the resulting chimeras and to demonstrate that the switching mechanism proceeds through P1 stem formation/disruption upon ligand binding¹⁰⁰.

4.1.3 In-line probing analysis

In-line probing is an *in vitro* analytical technique which exploits the inherent instability of RNA to elucidate secondary and tertiary structural features, and has been utilised extensively in the study of riboswitch-ligand interactions^{38, 63, 74, 183}. The phosphodiester bonds comprising the backbone of RNA molecules are vulnerable to intramolecular hydrolysis by the nucleophilic attack of the adjacent 2'-OH moiety, and the rate of this reaction is dependent on the relative orientations of 2'-OH, the phosphorus centre and the 5' oxygen of the following ribonucleotide residue⁶. The reaction is most rapid when the three atoms are in an ‘in-line’ conformation (when the 5'-O – P – 2'-OH angle is 180°), and thus hydrolysis occurs preferentially at non-base paired regions whereby the adjacent nucleotides have a greater degree of conformational flexibility allowing them to freely sample a range of geometries. Conversely, highly structured regions of RNA are less likely to adopt the required conformation for hydrolysis to occur and hence the extent of cleavage at these positions is reduced. In the analysis of riboswitches, typical experimental protocols involve the incubation of the RNA of interest with and without the cognate ligand for a period of up to a week and then analysing the resulting RNA fragments using urea PAGE. As well as revealing key nucleotides involved in ligand recognition by the riboswitch, performing multiple experiments at a range of ligand

concentrations allows for an estimation of the ligand binding affinity. This is achieved by plotting the fraction of the total sample of RNA which has been cleaved at a specified position (determined by urea PAGE) as a function of ligand concentration.

4.1.4 *In vitro* characterisation of the THF riboswitch

The ligand binding specificity of the THF riboswitch has been extensively explored *in vitro*, with more than 20 compounds identified which bind the aptamer domain^{57, 104, 110}. A summary of this data is presented in **Table 4**. ITC and in-line probing analysis by Breaker *et al.* and Batey *et al.* has shown that THF riboswitches selectively bind reduced forms of folate such as THF and the partially reduced dihydrofolate (DHF) whilst rejecting folic acid⁵⁷. X-ray crystallography studies have revealed the importance of a reduced C7-N8 bond in allowing hydrogen bond donation from N8 to the adjacent U7 and U25 residues of the two ligand binding sites, and this is thought to be the mechanism by which the THF aptamer distinguishes the reduced forms of folate. For a more in-depth analysis of the x-ray crystallography data, see Chapter 1. THF riboswitches have been shown to tolerate ligand functionalization at both the N5 and N10 positions with binding having been observed for folinic acid (5-formyl THF), 5-methyl THF, 10-formyl THF, 5,10-methenyl THF and 5,10-methylene THF. These folate derivatives are known cellular metabolites involved in one-carbon transfer reactions and their binding suggests that the THF riboswitch may be able to sample a large pool of related metabolites in the determination of a regulatory response.

The *S. mutans* THF riboswitch was investigated by Batey *et al.* utilising an *in vitro* transcription termination assay. The THF aptamer domain was fused to the expression platform of a *B. subtilis* SAM-I riboswitch, thus creating a chimeric system which functions by transcriptional termination upon ligand binding¹¹⁰. An overview of the proposed switching mechanism is found in **Figure 31**. Briefly, in the absence of an effector ligand, the expression platform forms an anti-terminator stem and the RNAP transcribes the full length mRNA. Ligand binding stabilises the P1 stem and induces the formation of a terminator stem (and poly-uridylate tract) leading to premature abortion of transcription. Interestingly, this analysis revealed a discrepancy between ligands which were able to bind the THF riboswitch aptamer domain, versus ligands which were able to both bind and elicit a regulatory response. A variety of natural and non-natural folate analogues have been shown by ITC to bind the THF aptamer domain including: tetrahydrobiopterin; 6,7-dimethyl-tetrahydrobiopterin; 7-

deazaguanine; and pemetrexed (see **Figure 17** for structures and **Table 2** for binding parameters).

However, whilst all of these compounds exhibited binding affinities within the same order of magnitude as THF/folinic acid, they had a significantly diminished ability to elicit the riboswitch regulatory response *in vitro*. The working hypothesis of Batey *et al.* is that a specific conformation of the THF *p*ABA side chain, which allows it to interact with the backbone RNA, is a required for the riboswitch regulatory response¹¹⁰. This is supported by the fact that the aforementioned ligands all either lack this structural feature, or (in the case of pemetrexed) possess a planar pterin moiety which positions the side-chain in an alternative orientation. Moreover, whilst the binding affinities of the 6*S*- and 6*R*- diastereoisomers of THF are almost identical, the 6*R*- isomer is more than 3-fold less effective in eliciting the riboswitch regulatory response¹¹⁰. This again suggests that the correct orientation of the *p*ABA side chain within the THF binding site is an important requirement for regulating gene expression. This phenomenon is further exemplified by the observation that adenine and 2,6-diaminopurine are able to bind the THF PK site via a 'retro binding' motif but have little regulatory ability considering their strong binding affinities.

4.1.5 Aims and Objectives

The advantage of *in vitro* investigations of riboswitch-ligand interactions is that the myriad potential participating species constituting the cellular milieu can be reduced to the minimal necessary components in order to ratify that the observed *in vivo* effects are actually a direct result of aptamer-ligand binding. ITC was the favoured analytical technique for a number of reasons. Firstly, riboswitch aptamer domains have been extensively characterised using ITC^{67, 79, 81, 176}, and so there are numerous benchmarks in the literature for the evaluation of binding parameters. Secondly, the *S. mutans* wild-type THF aptamer domain has also been well characterised by ITC^{104, 110}, and thus maintaining the same experimental technique allows for more meaningful comparisons to be drawn between the datasets. Finally, ITC is a particularly attractive *in vitro* analytical technique as given the optimum conditions, a wide range of thermodynamic parameters can be obtained from a single experiment.

The interaction between the *S. mutans* wild-type aptamer and folinic acid has already been studied¹⁰⁴, and this pairing was selected for ITC analysis to provide a benchmark for comparing the two datasets. Despite the lack of an *in vivo* effect (see **Section 3.2.1.5**), it was anticipated that the *S. mutans* wild-type aptamer would bind **T0A**, based on the structural similarity to folinic acid/THF (**Figures 22A & 22B**). Additionally, whilst compounds **T1A** and **T2B** had no *in vivo* effect on the **M1** and **M2** riboswitches respectively (see **Sections 3.2.1.5** and **3.2.1.6**), it was anticipated from the x-ray crystallography models that the **M1** (U25C/C53U) and **M2** (U35C/U42C) mutations would provide complementary hydrogen bond directionality to allow binding to occur (**Figure 22**). Consequently, the **WT-T0A**, **M1-T1A** and **M2-T2B** interactions were selected for ITC analysis.

In addition to the *S. mutans* THF aptamer, the analogous RNA-ligand interactions for the *L. casei* aptamer were also investigated. The *L. casei* THF aptamer domain has not been characterised previously by ITC or by any other biophysical technique. The *in vivo* data generated as part of this project had shown that the basal expression levels for the *L. casei* riboswitch were significantly lower than those for the *S. mutans* native and chimeric riboswitches (see **Sections 3.2.1.3** and **3.2.2.4**). It was anticipated that this may be due to the *L. casei* aptamer domain having greater binding affinity for folinic acid/THF than the *S. mutans* aptamer. Assuming the same concentration of intracellular folate in both cases, a higher binding affinity would lead to a larger population of ligand-bound ('OFF') riboswitch structures in the cell, and hence greater background repression of gene expression. Comparison of the folinic acid binding affinities of the *S. mutans* and *L. casei* aptamer domains by ITC would allow for the evaluation of this hypothesis. Additionally, since gene regulatory effects had been observed *in vivo* with the *L. casei* **WT-T0A** and the **M1-T1A** riboswitch-ligand pairings (see **Sections 3.2.2.7** and **3.2.2.11**), it was anticipated that binding would be observed *in vitro* between these ligands and the corresponding aptamer domains.

Finally, in order to confirm that the chimeric *Smu folT/Bsu metE* THF riboswitch employed in the *in vivo* assay was functioning by transcriptional termination, the same riboswitch construct was deployed in an *in vitro* transcription assay as has been previously outlined by Batey *et al.*¹⁰⁴.

4.2 Results and Discussion

4.2.1 Synthesis and purification of RNA for ITC analysis

A series of THF aptamer domains were designed and synthesised by *in vitro* transcription. *In vitro* transcription for the isolation of preparative-scale quantities of RNA requires the first two nucleotides of the transcribed sequence to be guanosine in order to maximise yields when using the T7-RNA polymerase enzyme¹⁸⁴.

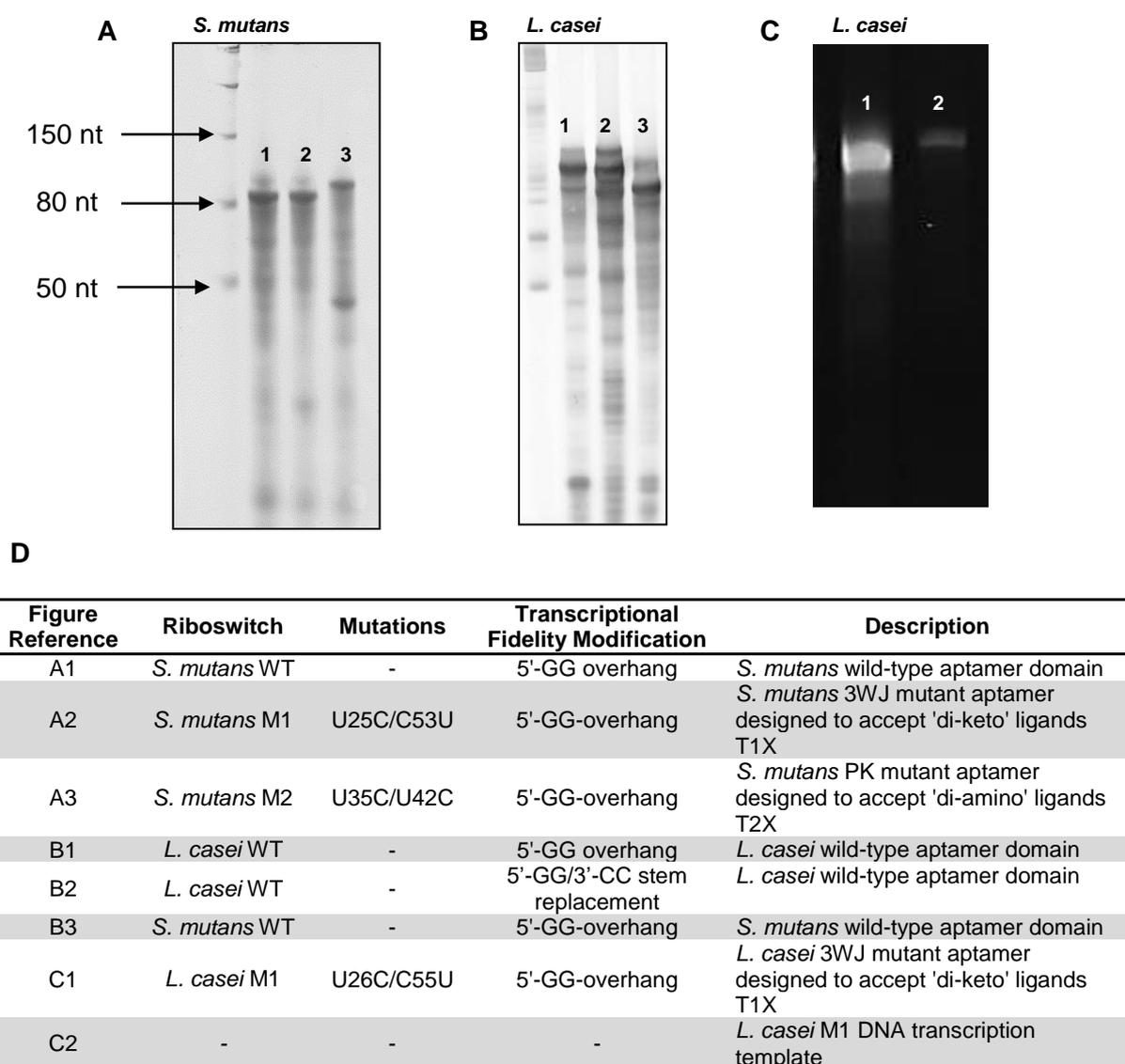


Figure 58 – Urea PAGE analysis of *in vitro* transcription reactions for THF aptamer domains. A) *S. mutans*. B) and C) *L. casei*. See table D) for the identity of the gel lanes and descriptions of the aptamer domain sequences.

In the case of riboswitch aptamer domains, this typically involves the addition of GG to the 5' end of the aptamer domain leaving an overhang at the base of the P1 stem, or alternatively substituting the first two nucleotides of the P1 stem for GG and introducing the complementary C or U residues at the 3' end. In most cases the former is preferable to the latter since the introduction of two strong GC pairs at the base of the P1 stem can impact the relative stabilities of the ligand-bound and unbound conformations.

4.2.1.1 *In vitro* transcription of *S. mutans* THF aptamer domains

The *S. mutans* THF aptamer domain has already been characterised using ITC^{104, 110} and this sequence was used here as the basis for the design of appropriate DNA transcription templates. Template DNA sequences were designed for three *S. mutans* THF aptamer domains: wild-type (**WT**), **M1** (U25C/C53U) and **M2** (U35C/U42C). The sequences consisted of the T7 promoter fused to the 5'-end of the respective *S. mutans* THF aptamer templates (see **Section 6.8**). The *in vitro* transcription protocol was modified from a procedure by Gurevich¹⁸⁴, whereby ssDNA templates (Sigma Aldrich) are utilised in combination with a complimentary primer such that the T7 promoter region of the template sequence is double stranded, whilst the transcribed region remains single stranded. Briefly, the DNA templates are incubated in a Mg²⁺-containing transcription buffer with NTPs and T7-RNA polymerase, and the transcription products are analysed by urea PAGE. *In vitro* transcription reactions were initially carried out on a pilot scale of 50 µl to optimise conditions before being scaled up to 10 ml. A urea PAGE gel image of the *S. mutans* aptamer *in vitro* transcription reactions is shown in **Figure 58A**. Both the **WT** and **M1** templates produced one major transcription product corresponding to the desired 86 nt THF aptamer domain. However, the **M2** template yielded two transcription products, one apparently longer than the expected 86 nt product and the other shorter. Following this observation the *in vitro* transcription of the **M2** aptamer domain was passed on to colleague Dr. Chris Robinson for troubleshooting. The desired **M2** transcription product was successfully synthesised from a PCR-amplified dsDNA template of the same sequence as used in the ssDNA-templated *in vitro* transcription reaction described in this section.

4.2.1.2 FPLC purification of *S. mutans* RNA transcription products

Following DNase I digestion of the DNA transcription templates, the **WT**, and **M1** transcription products were purified under native conditions using size exclusion FPLC, equipped with UV-vis detection. The principle of size exclusion chromatography (also known as gel filtration in the case of aqueous systems) is the separation of molecules by hydrodynamic volume based on their relative ability to permeate a column matrix, and is usually applied to large biomolecules such as proteins and RNA. The column matrix contains pores with a distribution of sizes. Larger molecules are unable to infiltrate the entire matrix and are eluted earlier, whilst lower molecular weight species are able to permeate a larger volume of the matrix and hence experience a longer flow path and are eluted at later time points. The *S. mutans* **WT** and **M1** transcription products were applied to a Superdex 75 prep-grade column and the elution of the desired aptamer domains was detected by monitoring absorbance at 260 and 280 nm. The UV absorbance elution profile for the purification of the *S. mutans* wild-type *in vitro* transcription reaction is shown as an example in **Figure 59A**. The chromatogram shows two sets of peaks, the second of which was detected at an elution volume in the range of the maximum volume of the column (right hand peaks on chromatogram). This peak is characteristic of small-molecule species and is attributed to the presence of unreacted NTPs and other components of the transcription buffer which absorb at these wavelengths. The peaks which eluted first (left hand peaks on chromatogram) were deemed to be the RNA products of the transcription reaction, and so aliquots from these fractions were analysed by urea PAGE to confirm this (**Figure 59B**) and identify suitable fractions for pooling. It can be seen from the gel image that separation of the desired 86 nt RNA transcript from lower molecular weight bi-products was successful, with the chromatogram absorbance intensity correlating with relative RNA concentration as determined visually by urea PAGE. The appropriate fractions were combined, concentrated by filter centrifugation, and sample homogeneity was confirmed by further urea PAGE analysis. The so described purification procedure was repeated for the *S. mutans* **M1** THF aptamer domain, which similarly yielded a monodisperse transcription product. The purification of the **M2** aptamer domain was performed by Dr. Chris Robinson and followed the same procedure as described for **WT** and **M1**. All three samples were dialysed into ITC buffer ready for analysis.

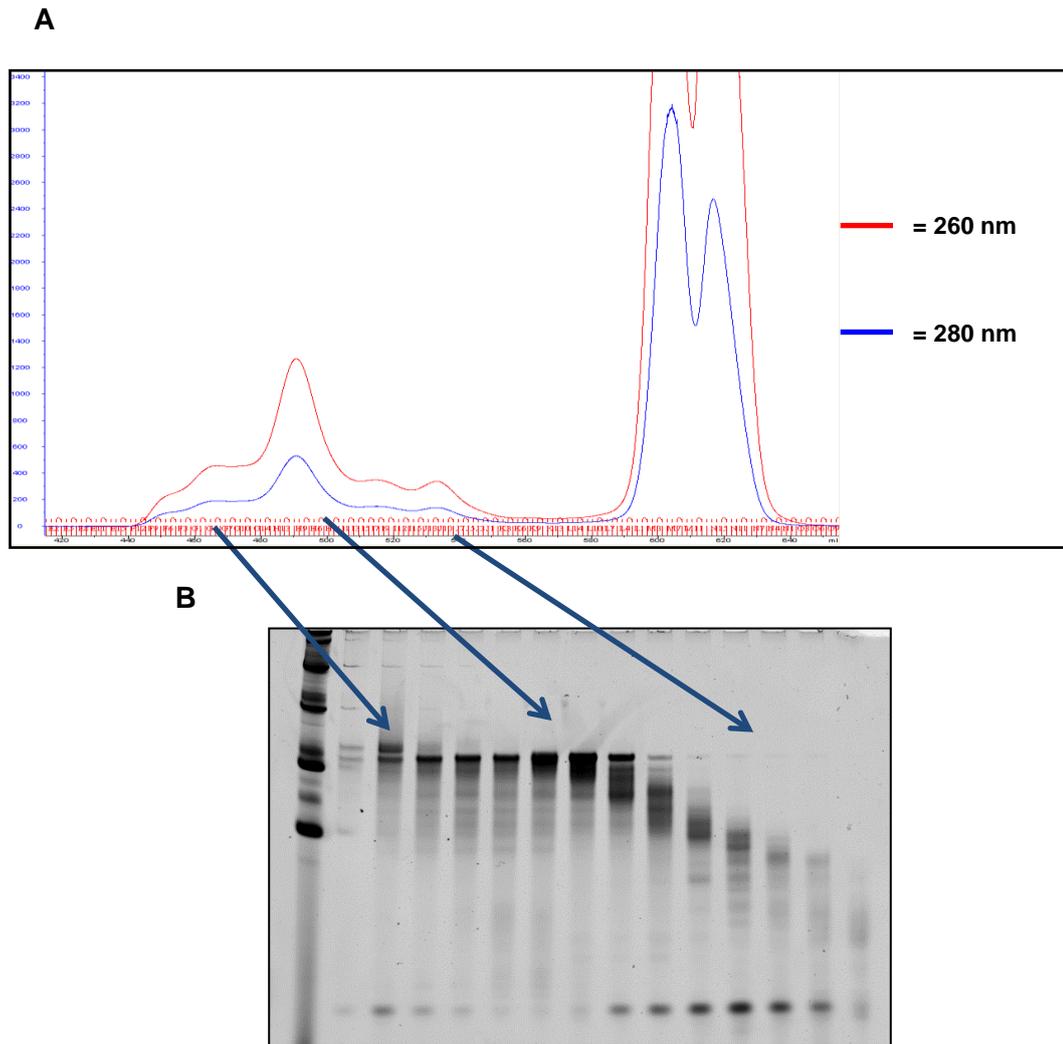


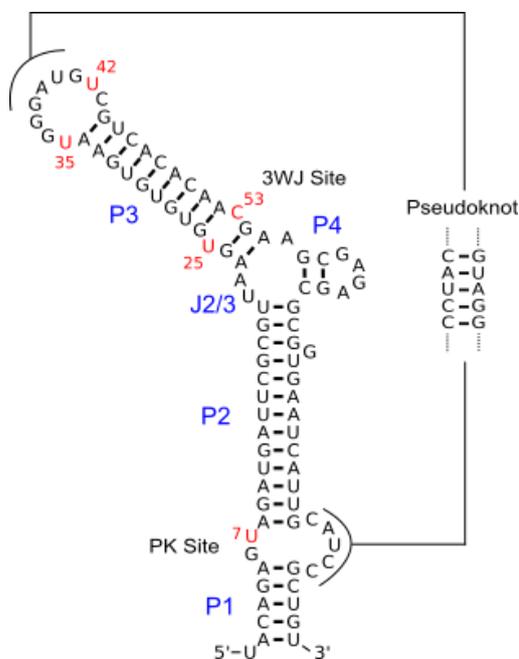
Figure 59 – A) UV absorbance elution profile for the FPLC size exclusion purification of the *S. mutans* wild-type THF aptamer domain showing absorbance at 260 (red) and 280 (blue) nm. The separation protocol utilised a GE healthcare Superdex 75 Hiload 26/600 pg gel filtration column run isocratically in FPLC buffer at 2 ml/min on an AKTA purifier FPLC system. B) Urea PAGE analysis of the indicated FPLC fractions.

4.2.1.3 *In vitro* transcription of *L. casei* THF aptamer domains

The sequence and ligand-bound secondary structure of the *S. mutans* THF aptamer domain is modelled from published x-ray crystallography studies¹⁰⁴ and is shown in **Figure 62A**. However, to date there has been no published characterisation of the *L. casei* aptamer domain. These two riboswitch sequences were aligned in order to ascertain the minimal *L. casei* aptamer domain for use in *in vitro* studies. This analysis identified a 93 nt predicted aptamer sequence from the *L. casei* riboswitch, which was

then subjected to conformational energy minimisation calculations using the Mfold web server¹⁶⁴.

A) *S. mutans*



B) *L. casei*

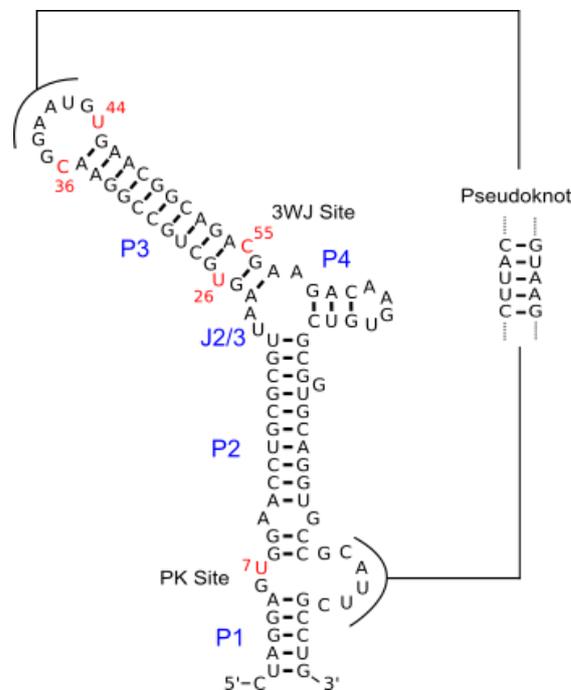


Figure 60 – Comparison of THF aptamer domain secondary structures. A) Ligand-bound *S. mutans* secondary structure as determined by x-ray crystallography¹⁰⁴. B) Predicted ligand bound *L. casei* secondary structure based on *S. mutans* sequence alignment and minimal energy conformation determined using the Mfold web server.

The resulting energy minimised structure (in the ligand un-bound state) was utilised along with the knowledge of the *S. mutans* ligand bound structure to postulate the ligand bound conformation of the *L. casei* THF aptamer domain (**Figure 60B**). The aptamer domains are highly similar in structure, with key binding site residues conserved across both sequences. Both aptamers feature a short P1 stem and a longer P2 stem between which is a series of residues which constitute the pseudoknot and PK binding site. The two aptamer domains also contain a highly conserved three way junction region, which joins the P2, P3 and P4 stems. The overall sequence identity is 59%, with the composition of the stems being particularly variable between the two sequences. This is to be expected, as these regions do not directly contribute to ligand recognition or the switching mechanism. There are also a number of other differences: The *L. casei* P2 stem contains two unpaired A and G residues at its base;

the P3 stem is one base pair longer than in the *S. mutans* sequence; the P3 loop in the *L. casei* aptamer is shorter; and the P4 stem of the *L. casei* aptamer is one base pair longer than *S. mutans*.

Based on this analysis the identified 93 nt *L. casei* aptamer domain sequence was taken forward for further investigation *in vitro*. Templates were designed in a similar manner to the *S. mutans* equivalents, utilising an ssDNA template with the T7 promoter fused immediately upstream of the transcribed region, and two template designs were explored. For the first template, two guanosine residues were added to the 5' end of the sequence to create a 'GG-overhang' at the base of the P1 stem (giving a 95 nt transcript). The second template was designed such that the first two base pairs of the P1 stem were replaced by GC, thus maintaining the transcript size but inevitably modifying P1 stem strength (referred to as 'GC-stem replacement'). As with the *S. mutans in vitro* transcription reactions, the ssDNA templates were combined with primers such that the T7 promoter region would be double stranded, whilst the transcribed sequence remained single stranded. Reactions were initially performed at pilot scale (50 µl) and the resulting urea PAGE analysis is shown in **Figure 58B**. Both the 'GG-overhang' and 'GC-stem replacement' templates (lanes 1 & 2 respectively) yielded a major transcription product of approximately the desired length as determined by comparison with an ssRNA ladder and the 86 nt *S. mutans* aptamer domain (lane 3). Whilst the yield under standard transcription conditions appeared higher for the 'GC-stem replacement' template, the number of additional shorter transcription products was also more significant. Therefore the 'GG-overhang' sequence was selected for scale up and purification, and was also used as the basis for template design during the synthesis of the *L. casei* **M1** (U26C/C55U) aptamer domain. The **M1** aptamer was prepared in a similar fashion to the wild-type sequence and the results of the pilot scale reaction are shown in **Figure 58C**. Both the wild-type and **M1** transcription products were then taken forward for purification by gel filtration FPLC.

4.2.1.4 FPLC purification of *L. casei* RNA transcription products

The *L. casei* wild-type aptamer transcription products were purified using gel filtration FPLC, utilising the same procedure as for the *S. mutans* aptamers. The resulting chromatogram and corresponding urea PAGE analysis is shown in **Figure 61**. Separation of the desired 95 nt transcription product from the numerous lower molecular weight transcripts was only partially successful. The UV absorbance elution

profile (**Figure 61A**) shows two sets of peaks, with the eluted earliest (left hand side of figure) corresponding to the transcription products, and the later eluting peaks (right hand side of figure) corresponding to unreacted NTPs and other low molecular weight components of the transcription buffer.

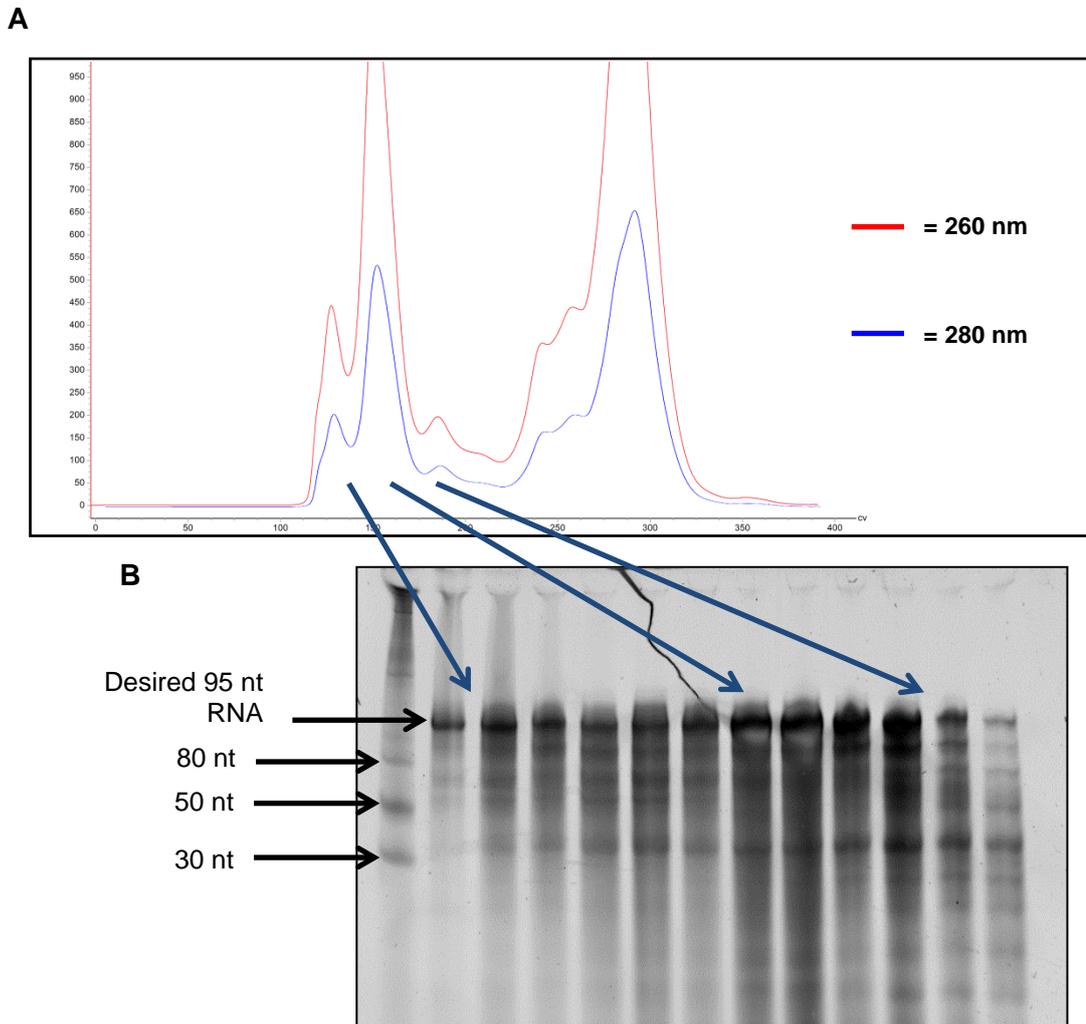


Figure 61 – A) UV absorbance elution profile for the FPLC size exclusion purification of the *L. casei* wild-type THF aptamer domain showing absorbance at 260 (red) and 280 (blue) nm. The separation protocol utilised a GE Healthcare Superdex HiLoad 26/600 pg gel filtration column run isocratically in FPLC buffer at 2 ml/min on an AKTA purifier FPLC system. **B)** Urea PAGE analysis of the indicated FPLC fractions.

It can be seen from the corresponding urea PAGE analysis (**Figure 61B**) that in this case, size exclusion FPLC was only partially successful in separating the mixture of transcription products. The desired 95 nt RNA was the major product across the three

peaks eluted between 110 – 200 ml, and with the exception of the very earliest fractions, there were also additional low molecular weight transcription products present. The fact that the 95 nt *L. casei* aptamer appears across three separate peaks on the chromatogram may be due to the adoption of multiple conformations with differing hydrodynamic volumes under native conditions. In the interest of obtaining samples of greater homogeneity, a narrow range of fractions eluted between 110 – 130 ml were pooled and taken forward for dialysis in preparation for ITC analysis.

4.2.1.5 Preparative-scale denaturing polyacrylamide gel electrophoresis (PAGE) purification of *L. casei* RNA transcription products

Since gel filtration FPLC had proven to be only moderately successful in purifying the 95 nt *L. casei* aptamer domain, large-scale urea PAGE was explored as a means of obtaining a more homogeneous RNA product. This procedure follows the same principle as the analytical-scale urea PAGE described in the previous sections, whereby a mixture of RNA transcripts is denatured and resolved according to transcript length. The advantage of denaturing PAGE purification is that the desired transcript can be selectively excised from the gel and the RNA extracted, allowing the creation of highly homogeneous samples at the expense of overall yield. A further consideration when choosing FPLC (native) versus denaturing PAGE purification methods is that RNA purified under denaturing conditions may not refold correctly. Briefly, the crude *in vitro* transcription reaction mixture is denatured in 8 M urea at 95 °C before being loaded onto a large denaturing PAGE gel and the transcription products resolved by running the gel for ~ 18h at a relatively low voltage (~100 V). The RNA is visualised using UV shadowing and the desired transcription product is excised from the gel then repeatedly extracted into (0.5x) TBE buffer until recovered RNA yields are sufficient. RNA extracts were pooled, concentrated using filter centrifugation, and dialysed into ITC buffer ready for analysis. A detailed experimental procedure can be found in Chapter 5.

4.2.2 *In vitro* analysis of THF aptamer domains using isothermal titration calorimetry (ITC)

4.2.2.1 ITC analysis of the *Streptococcus mutans* wild-type THF aptamer domain

ITC analysis of the *S. mutans* **WT** THF aptamer domain was carried out using folic acid, the folic acid analogue **T0A** and 'diketo-faced' analogue **T1A**, with the results are shown in **Figure 62**. As expected (having been previously demonstrated by Batey *et al.*¹¹⁰) a binding interaction between the wild-type riboswitch and folic acid was observed, with an apparent equilibrium dissociation constant of 11.1 μM . This value is comparable to the published 13 μM affinity, obtained using ITC under similar temperature and buffer conditions¹¹⁰. It was therefore taken with some confidence that the procedures utilised in this study to synthesise and purify the *S. mutans* THF aptamer domain were effective, and that the RNA was adopting the required tertiary conformation to allow for specific ligand binding under these conditions. Binding was also observed with ligand **T0A**, with an apparent K_D of 21.8 μM - roughly twice as high as that of folic acid (ie. weaker binding is observed). In both cases, the binding interaction was driven by favourable entropic and enthalpic contributions. The binding stoichiometry of both interactions was approximately 2:1 ligand/RNA, consistent with binding at each of the two distinct ligand binding sites within the THF aptamer domain. Indeed, compound **T0A** binds to the THF aptamer with a roughly comparable affinity to folic acid (a natural effector compound) and identical stoichiometry, but does not stimulate a gene regulatory effect *in vivo* (see **Section 3.2.1.5**). These observations are consistent with the working hypothesis of Batey *et al.* that some ligands are able to bind the *S. mutans* THF aptamer without eliciting a regulatory response, and that the presence of a *pABA* (or related) side chain at the C6-position is a requirement for a compound to induce the gene regulatory effect of the THF riboswitch¹¹⁰.

Compound **T1A** was designed to bind to the mutant **M1** (U25C/C52U) riboswitch at the modified 3WJ site. Encouragingly, no binding was observed between the *S. mutans* wild-type aptamer and compound **T1A**, which is the first step towards the development of an orthogonal riboswitch system.

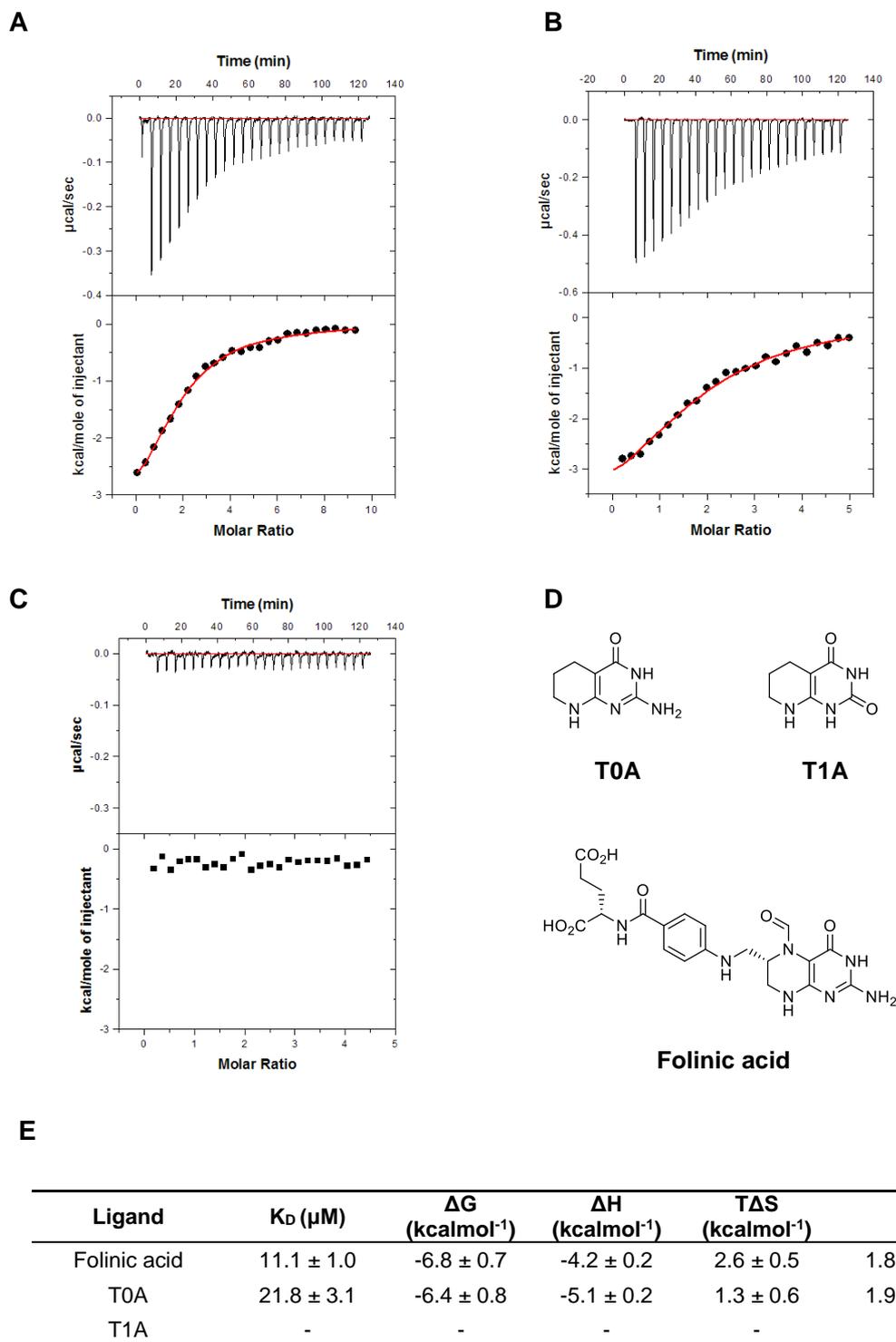


Figure 62 – ITC analysis of ligand binding by the *S. mutans* wild-type THF aptamer domain. A) Binding isotherm for the titration of folinic acid. B) Binding isotherm for the titration of T0A. C) Binding isotherm for the titration of T1A. D) Structures of folinic acid, T0A and T1A E) Thermodynamic parameters calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0.

4.2.2.2 ITC analysis of the *Streptococcus mutans* M1 THF aptamer domain

ITC binding experiments were performed with the *S. mutans* M1 (U25C/C53U) THF aptamer domain utilising the compounds T1A, which had been designed to bind the mutated 3WJ site, and a structural analogue, 7-deazaxanthine (structures shown in Figure 63D). Experiments were also performed using folinic acid, which was anticipated to bind the unmodified PK site of the M1 aptamer. The compound 7-deazaxanthine was chosen for analysis since the related compound 7-deazaguanine (structures shown in Figure 63D) had been shown to bind to the wild-type aptamer domain *in vitro*¹⁰⁴, and so it was considered that reversal of hydrogen bonding directionality provided by the U25C/C53U mutations may similarly allow the binding of 7-deazaxanthine to the M1 aptamer. The resulting ITC binding curves are shown in Figure 63. Compound T1A was found to bind the M1 aptamer with an apparent K_D of 84.7 μ M. The stoichiometry of the binding interaction was approximately 1:1 ligand/RNA, indicating that T1A only binds to one of the two ligand binding sites, presumably the modified 3WJ site. The binding affinity is lower in comparison with that of folinic acid for the wild-type aptamer (11.1 μ M), which may be a reflection of the fact that only one binding site is being occupied, or alternatively may reflect a less favourable interaction between T1A and M1 compared with the wild-type system. The latter of these two possibilities is considered more likely; it has been noted by the Micklefield group for other riboswitch systems that mutant aptamer-ligand pairs discovered using a rational design approach generally have binding affinities which are lower in comparison to the wild-type aptamer and its cognate ligand^{79, 81, 86}. Combining these observations with the *in vivo* data discussed Chapter 3, it appears that compound T1A is able to bind the M1 aptamer domain *in vitro*, but is not able to trigger the riboswitch regulatory response *in vivo*. Analogous to the case of T0A and the wild-type riboswitch, the lack of an *in vivo* regulatory effect is attributed to the absence of a *p*AABA or related moiety at the C6-position of the synthetic ligand.

No binding interaction was observed for the titration of 7-deazaxanthine (Figure 63B), which may have been expected to bind given its structural similarity to T1A and the fact that 7-deazaguanine was found to bind the wild-type aptamer¹⁰⁴. It is possible that a puckered ring conformation is a requirement for binding to the M1 3WJ site, which is possible for T1A but not 7-deazaxanthine. Finally, folinic acid was shown to bind to the M1 aptamer (Figure 63C), presumably via interaction at the unmodified PK site.

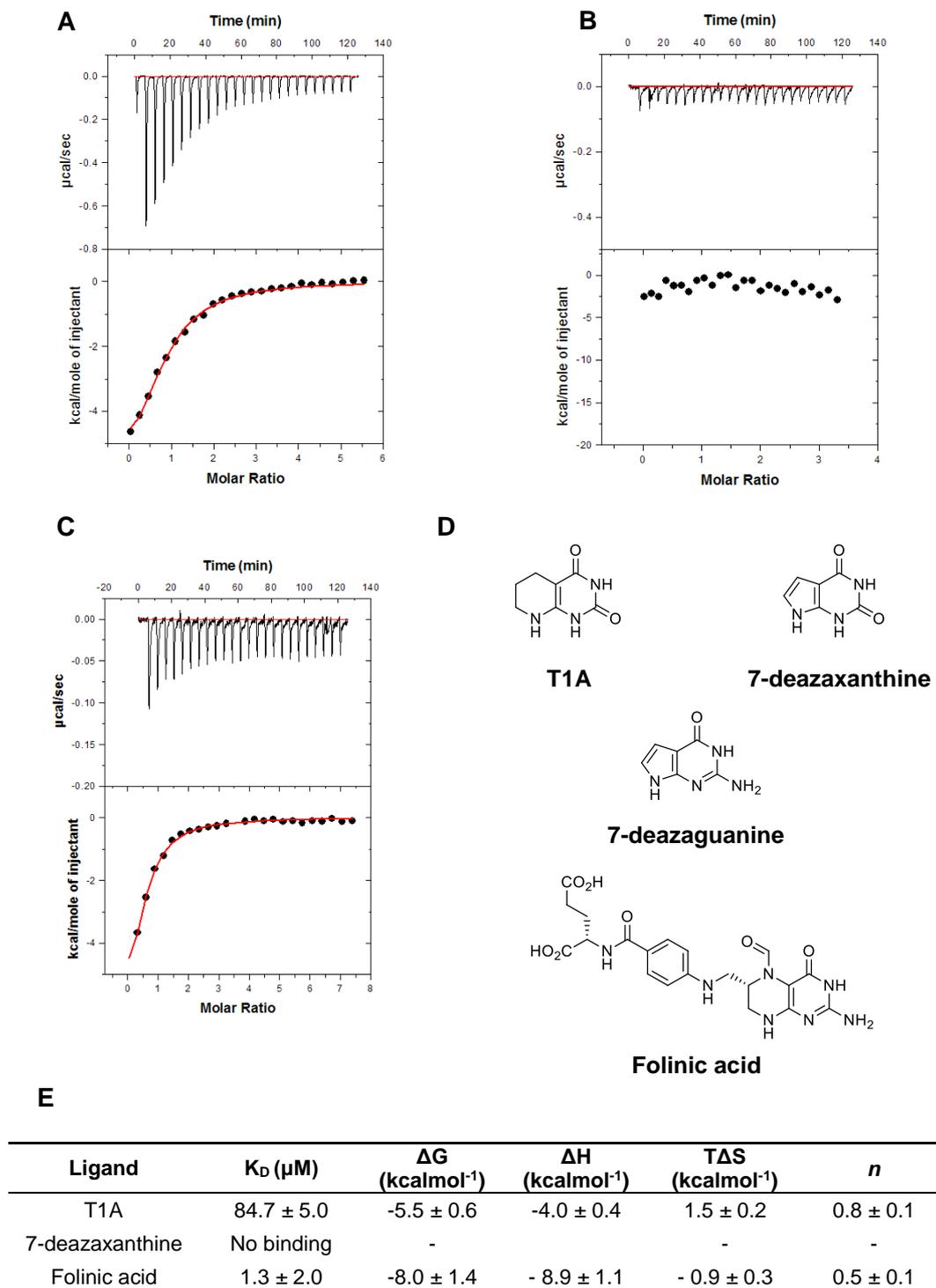
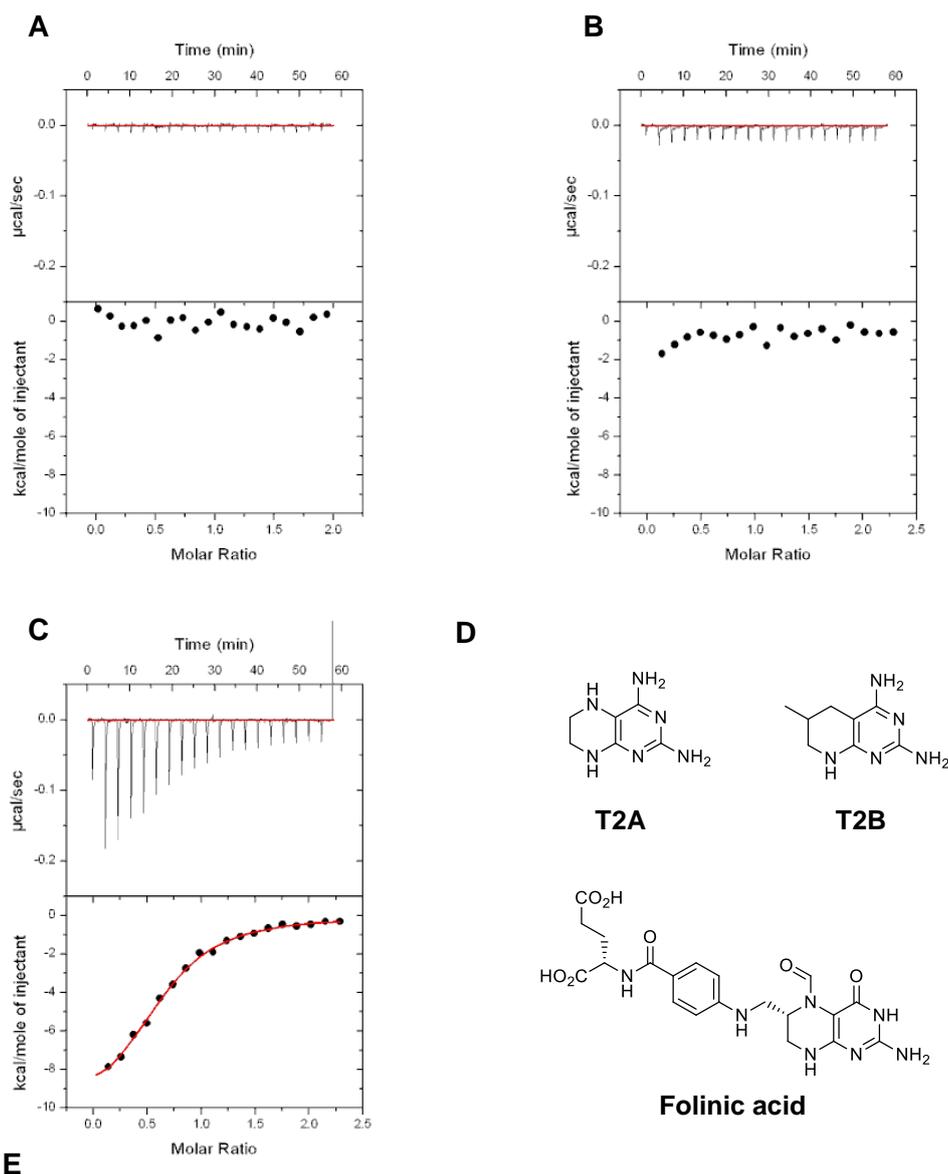


Figure 63 - ITC analysis of ligand binding by the *S. mutans* M1 THF aptamer domain. A) Binding isotherm for the titration of T1A. B) Binding isotherm for the titration of 7-deazaxanthine. C) Binding isotherm for the titration of folinic acid. D) Structures of T1A, 7-deazaxanthine, 7-deazaguanine and folinic acid. E) Thermodynamic parameters calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0.

4.2.2.3 ITC analysis of the *Streptococcus mutans* M2 aptamer domain

The *S. mutans* **M2** (U35C/U42C) THF aptamer domain was analysed using ITC by colleague Dr. Chris Robinson, with titrations performed with the compounds **T2A**, **T2B** and folic acid. It was anticipated that the mutations present in **M2** would enable binding of **T2A** and **T2B** to the PK site, whilst folic acid would bind the unmodified 3WJ site. The results of the ITC experiments are shown in **Figure 64**.

No binding was observed for either **T2A** or **T2B** to the THF **M2** aptamer domain (**Figures 64A & 64B**), indicating that the U35C/U42C mutations do not allow the PK binding site to accommodate 'diamino-faced' ligands of this type. However, binding was observed for folic acid (**Figure 64C**), presumably at the unmodified 3WJ site. This is evidence that the U35C/U42C mutations do not drastically alter the overall conformation of the aptamer domain, and therefore the **M2** riboswitch might still be a target for future studies in the development of an orthogonal PK site mutant.



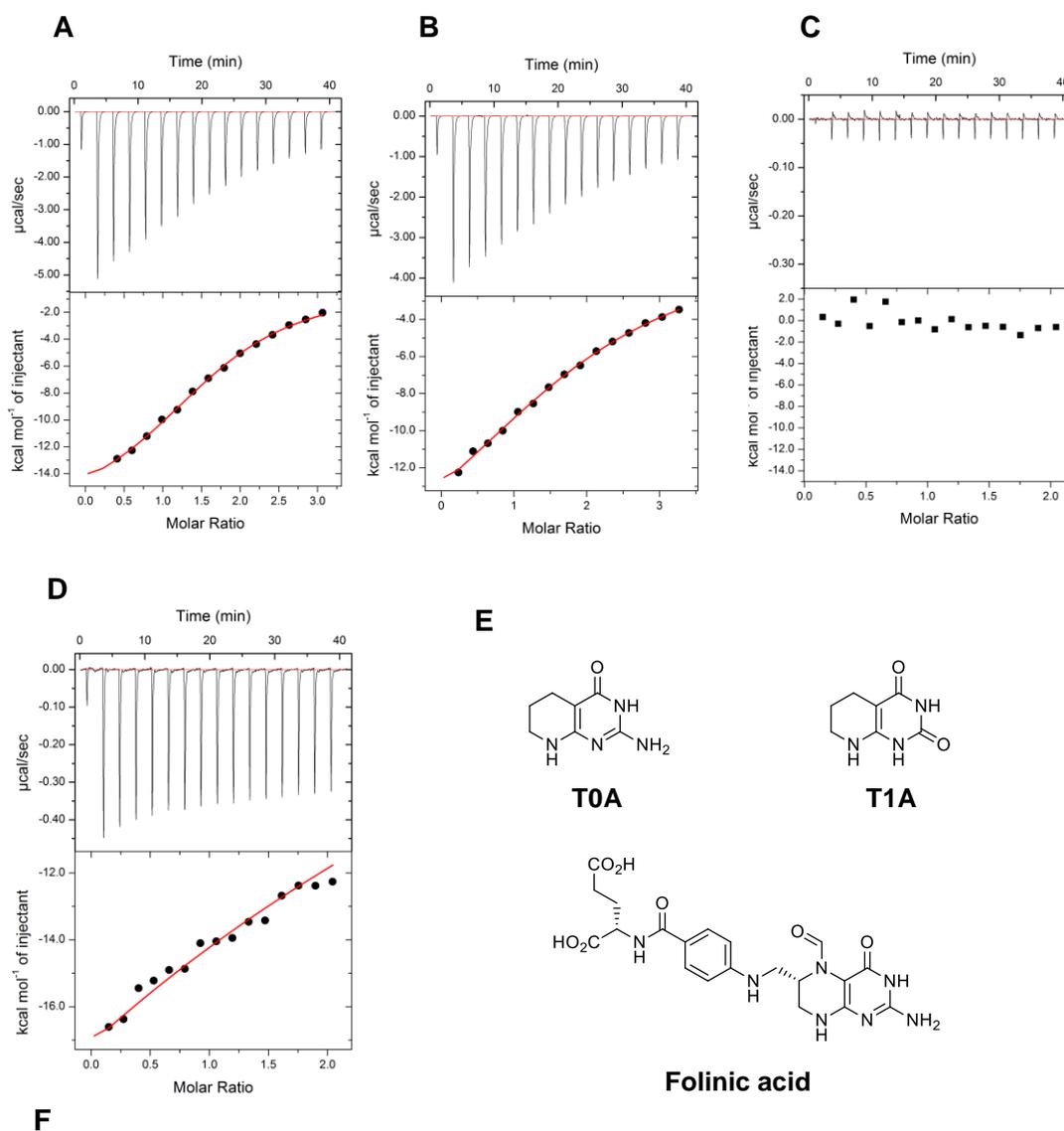
Ligand	K_D (μM)	ΔG (kcalmol^{-1})	ΔH (kcalmol^{-1})	$T\Delta S$ (kcalmol^{-1})	n
T2A	No binding	-	-	-	-
T2B	No binding	-	-	-	-
Folinic acid	2.0 ± 1.2	-7.8 ± 0.6	-10.5 ± 0.3	2.2 ± 0.3	0.6 ± 0.1

Figure 64 - ITC analysis of ligand binding by the *S. mutans* M2 THF aptamer domain (data provided by Dr. Chris Robinson). A) Binding isotherm for the titration of T2A. B) Binding isotherm for the titration of T2C. C) Binding isotherm for the titration of folinic acid. D) Structures of T2A, T2B and folinic acid. E) Thermodynamic parameters calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0.

4.2.2.4 ITC analysis of the *Lactobacillus casei* wild-type aptamer domain

ITC analysis was performed on the *L. casei* wild-type THF aptamer domain using the compounds **T0A** and folic acid. Titrations were carried out using RNA which had been purified by either size exclusion FPLC (native conditions), or by preparative-scale urea PAGE (denaturing conditions). See **Sections 4.2.3.5** and **4.2.3.6** for details. The results of these ITC experiments are shown in **Figure 65**. Folic acid was found to bind the *L. casei* aptamer domain with an apparent K_D of 59.9 μM (**Figure 65A**), giving a binding affinity approximately 5-fold lower than for the wild-type *S. mutans* aptamer. Compound **T0A** was also shown to bind, however the K_D was 109.0 μM (**Figure 65B**), approximately 80% higher than folic acid (ie. weaker binding was observed). Both interactions showed binding stoichiometries of approximately 2:1 ligand/RNA, indicating that both the 3WJ and PK sites are able to accept folic acid and **T0A** (the slightly low n values are attributed to RNA of sub-optimal purity leading to an overestimation of the concentration of the *L. casei* aptamer domain in solution based on absorbance). The difference in binding affinities between these two ligands is a possible explanation for some of *in vivo* effects described in Chapter 3. Folic acid and **T0A** had been successfully combined to create a two-ligand expression system whereby the de-repressive effect of **T0A** was exploited to raise the basal expression levels and thereby increase the dynamic range of accessible expression outputs. It was noted that adding whilst folic acid could be used to repress expression following the addition of 2 mM **T0A** (which raised the basal expression levels), the reverse was not possible (see **Section 3.2.2.9**). When 0.5, 1.0 or 2.0 mM folic acid was added to the cell culture, it was not possible to subsequently de-repress expression even at saturating **T0A** concentrations. The fact that **T0A** has an apparent K_D for the wild-type aptamer domain which is approximately 80% higher than that of folic acid may be a potential explanation for this.

As with the *S. mutans* aptamer domain, no binding was observed between the compound **T1A** and the *L. casei* wild-type aptamer – an important requirement for the development of an orthogonal gene expression system (**Figure 65C**). The RNA purified by denaturing PAGE failed to yield reliable binding curves (**Figure 65D**), indicating that THF aptamers purified under denaturing conditions may not refold correctly and thus this method of aptamer purification was not pursued further.



Ligand	K_D (μM)	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n
Folinic acid	59.9 ± 3.0	-5.8 ± 1.0	-21.0 ± 0.5	-15.2 ± 0.5	1.7 ± 0.1
T0A	109.0 ± 2.5	-5.4 ± 0.9	-21.3 ± 0.3	-15.9 ± 0.6	1.8 ± 0.2

Figure 65 - ITC analysis of ligand binding by the *L. casei* wild-type THF aptamer domain. A) Binding isotherm for the titration of folinic acid with the aptamer domain purified under native conditions. B) Binding isotherms for the titration of T0A with the aptamer domain purified under native conditions. C) Binding isotherms for the titration of T1A with the aptamer domain purified under native conditions. D) Binding isotherms for the titration of folinic acid with the aptamer domain purified under denaturing conditions. D) Structures of T0A, T1A and folinic acid. E) Thermodynamic parameters for the interactions observed using the aptamer domain purified under native conditions. Parameters were calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0.

4.2.2.5 ITC analysis of the *Lactobacillus casei* M1 aptamer domain

ITC analysis was performed on the *L. casei* M1 (U26C/C55U) aptamer domain with the ligand T1A. The result of this ITC experiment is shown in Figure 66. T1A was found to bind the *L. casei* aptamer domain with an apparent K_D of 67.6 μM . It was expected that T1A would only bind the modified 3WJ binding site of the M1 aptamer domain, whilst not interacting with the unmodified PK site. The stoichiometry of the interaction (n) was found to be 0.89, which indicates that T1A only binds one of the two M1 binding sites and it is assumed that this is the mutated 3WJ site.

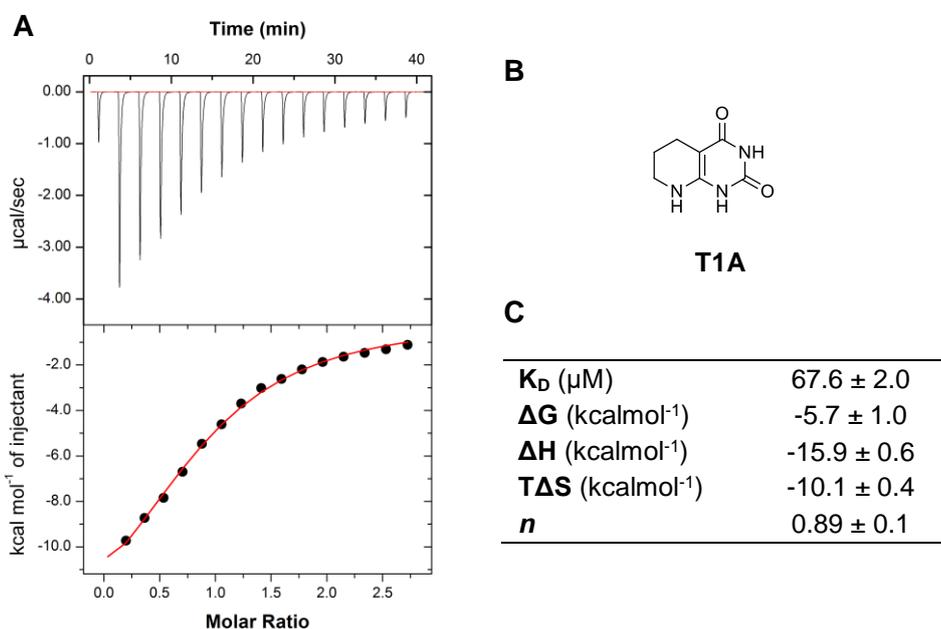


Figure 66 – ITC analysis of ligand binding by the *L. casei* M1 (U26C/C55U) aptamer domain. A) Binding isotherm for the titration of T1A with the aptamer domain. B) Structure of T1A. C) Thermodynamic parameters for the interactions observed using the aptamer domain purified under native conditions. Parameters were calculated from the binding data using the MicroCal™ plug-in for Origin® 7.0.

4.2.3 *in vitro* transcription termination assays using the *Smu folT/Bsu metE* chimeric riboswitch

In order to confirm that the *Smu folT/Bsu metE* chimeric riboswitch employed in the *in vivo* assays (see Section 3.2.1) functions correctly as a transcriptional terminator, a transcription termination assay previously reported by Batey *et al.*¹⁰⁴ was performed.

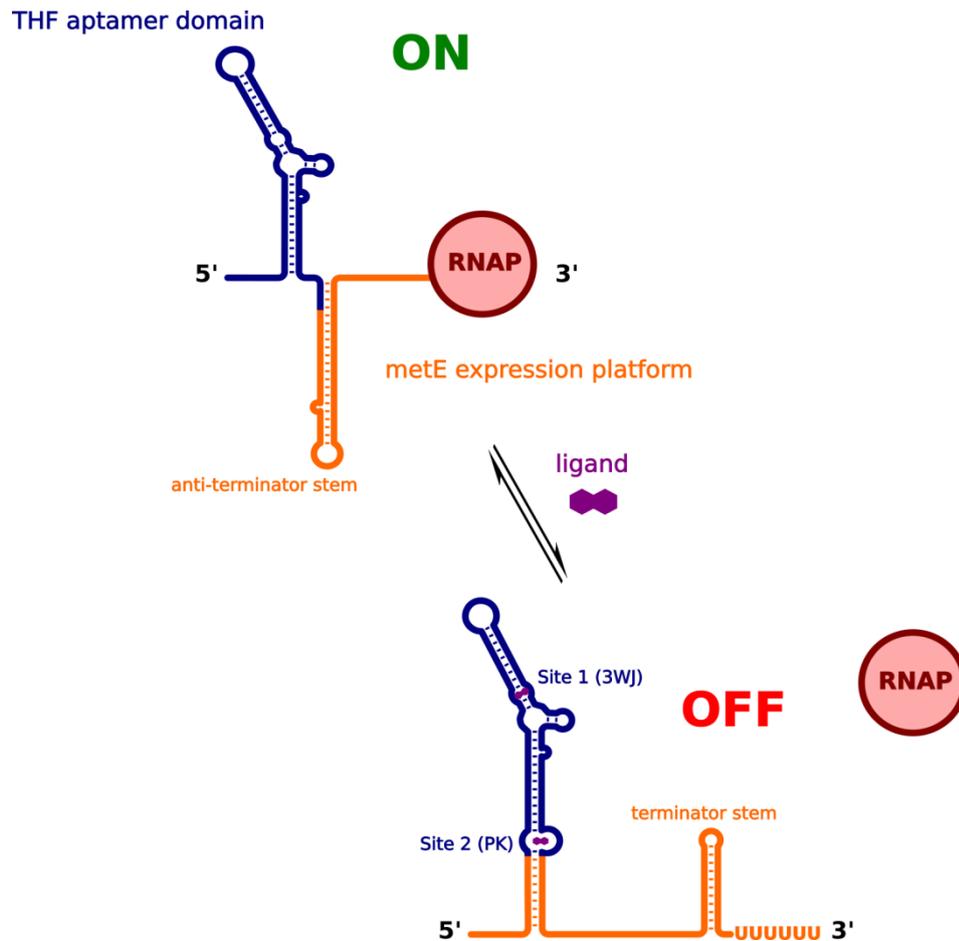


Figure 67 – Proposed secondary structure, ligand binding and transcriptional termination model for the *foIT/metE* chimeric riboswitch. The *foIT* aptamer domain is shown in blue and the *metE* expression platform is shown in orange. Ligand binding induces the formation of a downstream intrinsic transcriptional terminator stem in the expression platform causing the RNA polymerase to abort transcription.

The proposed secondary structure and switching mechanism of the chimeric riboswitch is shown in **Figure 67**. Ligand binding induces P1 stem formation, which in turn induces the formation of a downstream terminator stem which aborts transcription of the gene before the coding sequence is reached. The ratio of full-length to truncated transcripts can be analysed by urea PAGE and the ability of the ligand of interest to elicit transcriptional termination by the riboswitch can be assessed. The assay can be performed in a non-limiting manner (multiple turnovers) or as a single-turnover assay, whereby each RNA polymerase enzyme is limited to just one round of transcription. The concept of the single-turnover *in vitro* transcription assay is shown in **Figure 68**. A

DNA template is first incubated with the RNA polymerase and a ^{32}P labelled nucleotide (**Figure 68A**) to allow the initiation to occur.

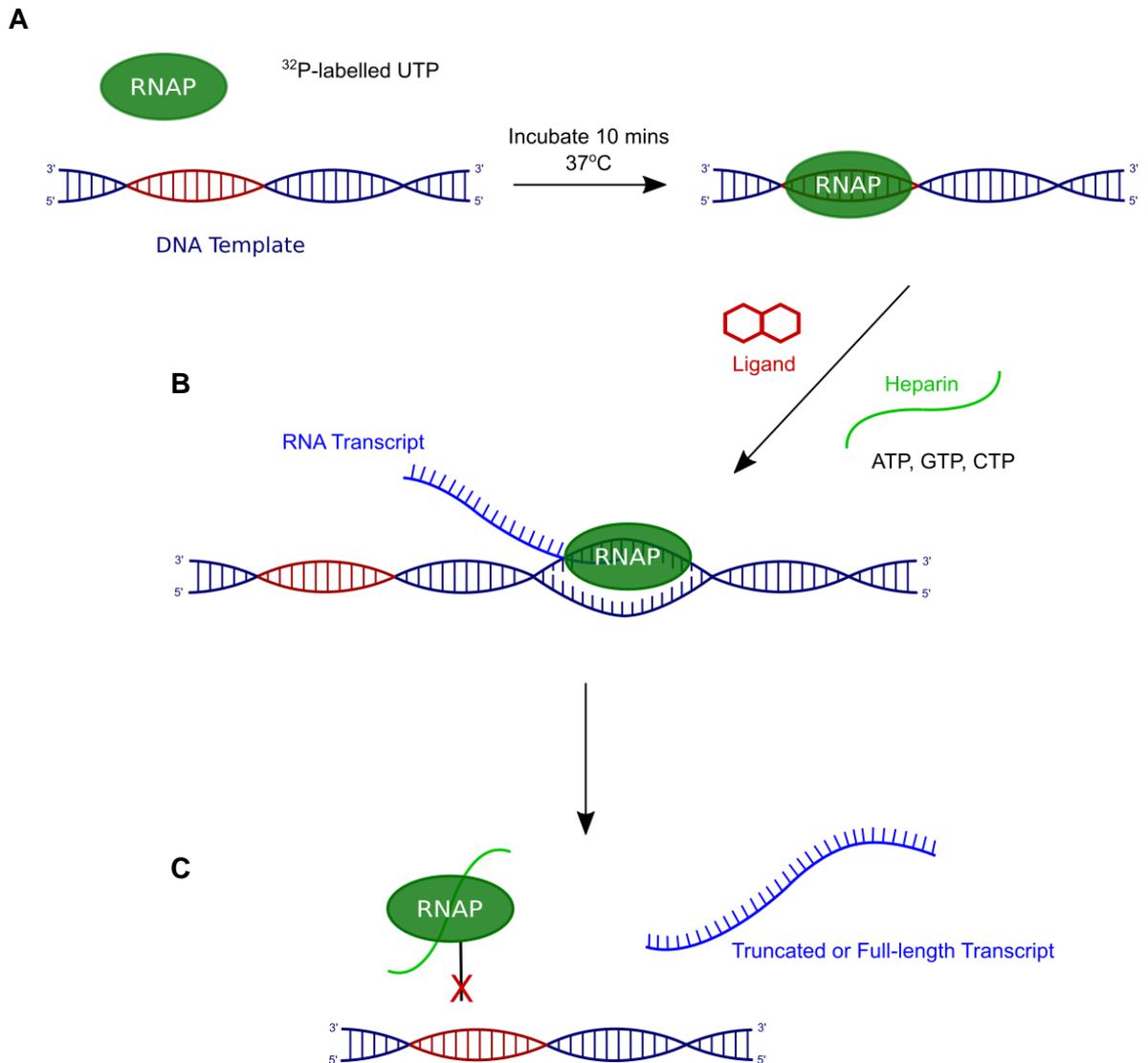


Figure 68 – Overview of the single-turnover *in vitro* transcription assay. A) The DNA template containing the T7 promoter is pre-incubated with T7-RNAP and radiolabelled UTP; **B)** Transcription is initiated by the addition of a mixture of the remaining NTPs, the ligand of interest and an excess of the polyanionic saccharide heparin; **C)** Upon transcription termination the T7-RNAP binds to heparin, preventing further rounds of transcription. The relative levels of read-through and truncated transcripts can be assessed by urea PAGE.

The template is comprised of a promoter, a transcriptional riboswitch and a short sequence which usually constitutes the first ~100 nucleotides of the native gene controlled by the riboswitch. Transcriptional elongation is induced by the addition of the remaining NTPs, as well the ligand of interest and heparin. Heparin is a highly

sulphated glycosaminoglycan and a potent inhibitor of RNA polymerase. Following termination of transcription, either prematurely due to the transcriptional riboswitch, or after the transcription of the full-length template, heparin binds the released RNA polymerase enzyme, thus preventing further rounds of transcription from initiating. A single-turnover assay is a more faithful representation of cellular conditions as it maintains a stable NTP concentration, which is known to affect transcription rates and hence the effectiveness of transcriptional termination²⁰. However, single-turnover assays require the use of ³²P-radiolabelling to detect the low concentrations of transcribed RNA, which is less convenient, and hence it was decided that conditions would be optimised using the multiple-turnover assay before switching to the single-turnover protocol.

4.2.3.1 Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis

The DNA template for the transcription termination assays was synthesised using a PCR-based gene synthesis method developed by Harris *et al.*¹⁸⁵ The procedure uses successive pairs of overlapping primers which have been designed such that their annealing temperatures are almost identical. This enables the one-pot PCR synthesis of short DNA sequences up to around 360 base pairs.

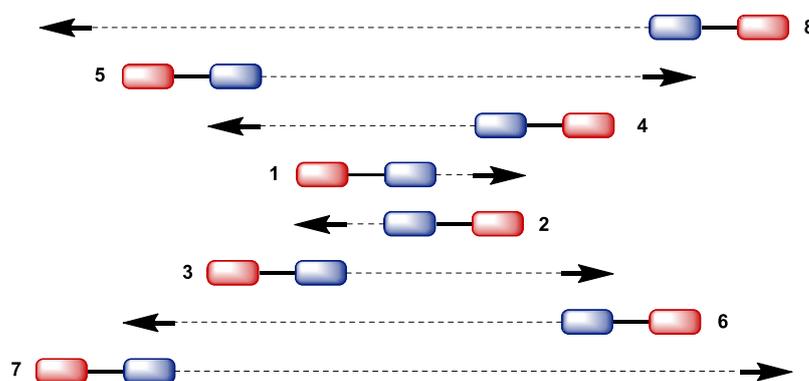


Figure 69 – Overview of TBIO PCR-based gene synthesis. See body text for details.

The concept is demonstrated in **Figure 69**. Successive pairs of overlapping primers are elongated to produce sequentially longer sequences. In the first PCR round on the thermocycler, primers 1 and 2 bind and subsequent elongation produces a DNA sequence of length indicated by the arrows. In the second round, primers 3 and 4 bind

to the product of the first round of PCR and elongation produces a longer sequence. The process is repeated for primers 5 and 6 and finally primers 7 and 8. The 5'-red- and 3'-blue- coloured sequences correspond to the red- and blue-coloured regions of primer overlap. Following 20 elongation cycles, the crude PCR product is re-amplified using primers 7 and 8 to produce the full length PCR amplified DNA sequence in good yield (**Figure 70**). Primers were designed using the DNA Works server¹⁸⁶; each primer was approximately 60 nt in length and the overlapping regions were around 25-30 nt (see **Section 6.8**) The total synthesised DNA sequence encompassing the T7 promoter, the chimeric *folT/metE* riboswitch and a 5' fragment of the *metE* gene was 256 nt. The PCR product was purified using a Qiagen mini-elut gel extraction kit.

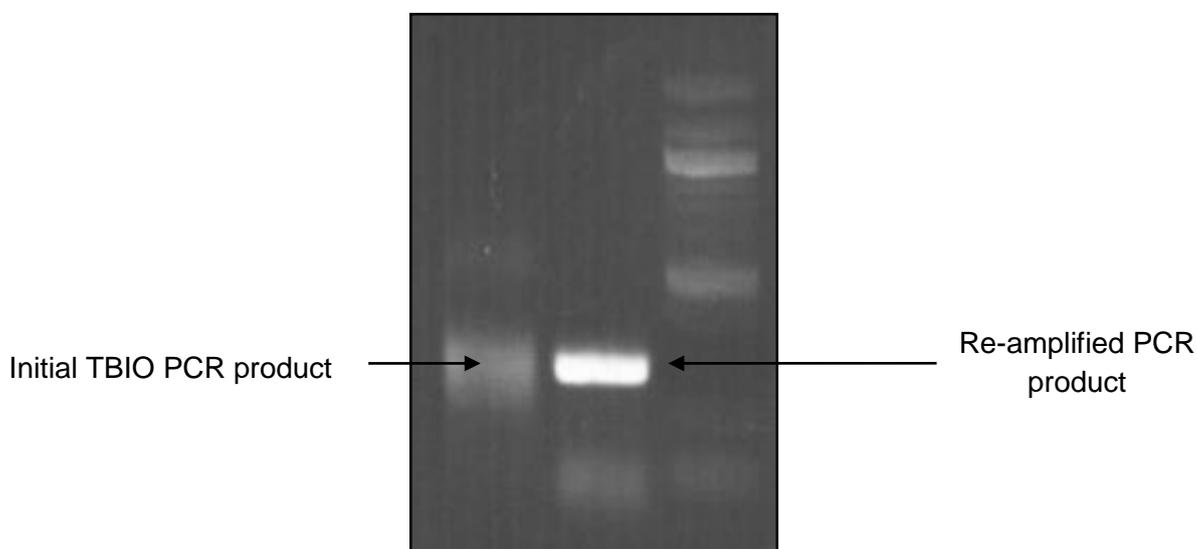


Figure 70 - Agarose gel of PCR products; lanes (L->R) Initial TBIO PCR product; re-amplified PCR product; ladder.

4.2.3.2 *In vitro* transcription assays utilising the *Smu folT/Bsu metE* chimeric riboswitch

The *Smu folT/Bsu metE* riboswitch was first investigated using a multi-round assay. The DNA template prepared as described above was mixed with all four NTPs in a buffer solution containing $MgCl_2$.

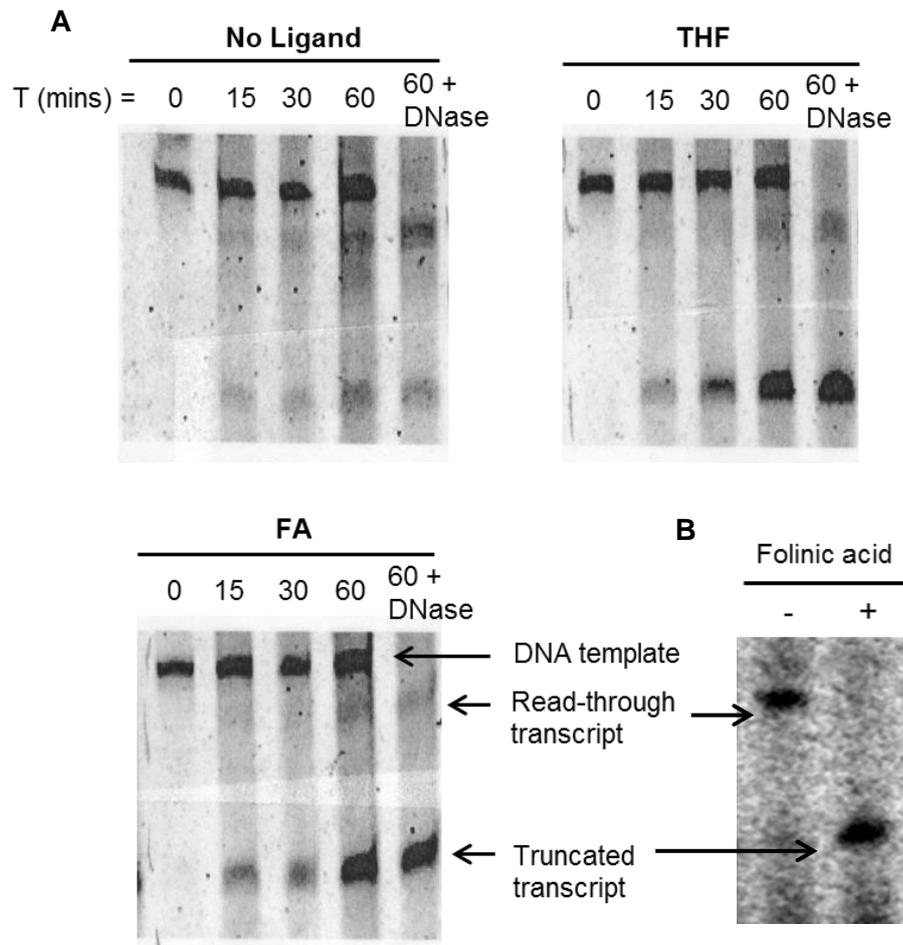


Figure 71 - Urea PAGE analysis of *in vitro* transcription assays. A) Multi-round assays, conducted in the absence of ligand or in the presence of THF or folinic acid; B) Single-turnover assay conducted in the presence or absence of folinic acid. Multi-round assays were stained using SYBR gold (Life Technologies), whilst the single-turnover assay was imaged by autoradiography through incorporation of ^{32}P in the RNA transcript from ^{32}P -UTP.

Three samples were prepared, containing THF, folinic acid and a no-ligand control to which only water was added. Transcription was initiated by the addition of *E. coli* RNA polymerase and the samples were incubated at 37°C for 60 minutes, with aliquots taken every 15 minutes for analysis. The sample was then DNase I treated and re-incubated for a further 30 minutes, before EDTA was added to stabilise the RNA products and the samples were then analysed by urea PAGE (**Figure 71A**). The DNA template can be easily identified on the gel as it retains a steady concentration over time, and is absent in the control lanes treated with DNase I. The levels of terminated transcript could be seen to increase steadily over time in the presence of both THF and folinic acid, but remained low and constant in the absence of ligand. The full-length

transcripts are less clear in the multi-round assay, but can be seen below the DNA template.

The *Smu folT/Bsu metE* was then investigated using a single-round transcription assay. The single-round *in vitro* transcription assays were carried out using the principles described in **Section 4.3.3**, using a modification of the protocol outlined by Trausch¹⁰⁴. The DNA template prepared as described in the previous section (**4.3.3.1**) was incubated with *E. coli* RNA polymerase and α -³²P-UTP to allow transcription initiation to occur. Elongation was induced by the addition of the remaining unlabelled NTPs, along with heparin and 2 mM folinic acid. The resulting transcription products were then analysed using urea PAGE and visualised with a phosphor screen, the results of which can be seen in **Figure 71B**. The single-round assay clearly demonstrates the formation of full-length mRNA transcripts in the absence of folinic acid and the formation of truncated transcripts when folinic acid is present, confirming that the chimeric *Smu folT/Bsu metE* riboswitch functions efficiently as a transcriptional terminator.

4.3 Chapter conclusions

The assay results outlined in Chapter 3 demonstrated that the *S. mutans* and *L. casei* THF riboswitches differed in their response to a number of ligands in an *in vivo* setting. Notably, the ligand **T0A** had little *in vivo* effect on the *S. mutans* riboswitch, but resulted in a significant de-repression of riboswitch-controlled gene expression in the *L. casei* system (see **Section 3.2.2.7**). Additionally, the ligand **T1A** was able to repress gene expression with the mutant *L. casei* **M1** (U26C/C55U) system, but had no effect at all on the equivalent *S. mutans* **M1** riboswitch (see **Sections 3.2.1.5** and **3.2.2.11**). However, the analysis delineated in this chapter revealed that the differences observed in terms of *in vivo* activity do not correspond to differences in *in vitro* binding behaviour.

This study constitutes the first *in vitro* characterisation of the *L. casei* THF riboswitch aptamer domain. Both the *S. mutans* and *L. casei* wild-type aptamer domains were shown to bind folinic acid, with the *L. casei* aptamer exhibiting ~5-fold reduced binding affinity compared with *S. mutans*. Additionally, both wild-type riboswitches were shown to bind the synthetic compound **T0A**, in both cases with binding affinities ~2-fold lower than for folinic acid (see **Sections 4.2.2.1** and **4.2.2.4**).

Both the *L. casei* and *S. mutans* mutant **M1** riboswitches were shown to bind the synthetic compound **T1A**, although a regulatory effect was only observed with the *L. casei* system in an *in vivo* setting. Moreover, in both cases, **T1A** had no apparent affinity for the wild-type riboswitch. This represents a significant step towards in the re-engineering of an orthogonally selective THF riboswitch. The **M1** riboswitch was rationally designed to accept the compound **T1A** whilst rejecting folinic acid at the mutated 3WJ site. This has been demonstrated using ITC. Moreover, compound **T1A** was designed not only to bind the modified 3WJ site of **M1** but also to be rejected by the wild-type riboswitch. This has also been confirmed through the presented ITC analysis (see **Sections 4.2.2.2** and **4.2.2.5**).

However, similar attempts to re-engineer the THF riboswitch PK site were not met with success. The rationally designed *S. mutans* **M2** riboswitch and their putative ligands **T2A** and **T2B** showed no binding interaction *in vitro* (see **Section 4.2.2.3**). Since ligand binding to the PK site has been shown to be more important in mediating the riboswitch regulatory effect, re-engineering of this site should be a priority in any future studies.

Chapter 5

Project Conclusions and Future Work

5. Project conclusions and future work

The goals of this project were split into two main strands. Firstly, the characterisation of the wild-type THF riboswitch *in vivo*, using a series of synthesised THF analogues to probe the ligand binding/regulatory specificity, and exploring the regulatory function of the two ligand binding sites through a series of mutations known to abrogate binding at each site independently¹⁰⁴. The second goal of this project was the re-engineering of the THF riboswitch to create an orthogonal mutant riboswitch gene expression system, which no longer responds to THF/folinic acid but is instead activated by two distinct synthetic ligands.

5.1 Conclusions: Characterisation of the wild-type THF riboswitch *in vivo*

Through the establishment of an *in vivo* reporter gene assay in *E. coli*, it has been demonstrated the THF riboswitch is able to dose-responsively control the expression of GFP in a non-native host. This was demonstrated using a well characterised *folT/metE* chimeric THF riboswitch (see **Section 3.2.1**) and further exemplified using the previously uncharacterised native THF riboswitch from *Lactobacillus casei* (see **Section 3.2.2**). These two riboswitch systems were used to elucidate the role of the two THF ligand binding sites *in vivo*, which confirmed the *in vitro* observation by Batey *et al.*¹¹⁰ that the ligand binding at the PK site is the primary mechanism responsible for the riboswitch regulatory response. The regulatory ability of the chimeric *folT/metE* riboswitch was almost completely negated on abrogation of PK site ligand binding, however the *L. casei* THF riboswitch retained significant regulatory function when the PK site was disrupted, thereby highlighting the importance of the 3WJ binding site in this system.

It had been noted by Batey *et al.* that there was a discrepancy between ligands which were able to bind the THF riboswitch and ligands which were able to repress gene expression in an *in vitro* model¹¹⁰. The group hypothesised that a structural requirement for ligands which have a regulatory ability is a side-chain at the C6-position which is capable of adopting a conformation such that it interacts with the backbone RNA. However, it was not clear if this phenomenon was merely an artefact of the non-native expression platform within the chimeric riboswitch used in the *in vitro* model. In this project, it has been shown that the disconnect between binding and regulating ligands

is also observed *in vivo* with the chimeric *folT/metE* riboswitch and with the previously uncharacterised native THF riboswitch from *L. casei*. For the chimeric *folT/metE* riboswitch, the compound **T0A** had no *in vivo* effect on GFP expression (see **Section 3.2.1.5**). However, **T0A** was found to bind the aptamer domain *in vitro* by ITC (see **Section 4.2.2.1**), with comparable binding affinity to that of folinic acid. Interestingly, when **T0A** was tested *in vivo* with the *L. casei* THF riboswitch, rather than eliciting a repressive effect, it was found to increase gene expression by up to four-fold (see **Section 3.2.2.7**). ITC analysis found that **T0A** also binds the *L. casei* aptamer domain, with a slightly lower, but comparable affinity to that of folinic acid (see **Section 4.2.2.4**). Therefore, it is hypothesised that **T0A** is a competitive inhibitor of the THF riboswitch, and that the observed increase in gene expression seen with the *L. casei* system is actually a de-repressive effect. In other words, **T0A** binds the THF riboswitch without inducing the regulatory response, thus it is able to displace cellular THF and other reduced folates which are acting to repress the basal expression levels; therefore, an increase in GFP expression is observed.

It is not clear why the de-repressive effect is only observed with the *L. casei* THF riboswitch and not for the chimeric *folT/metE* or native *S. mutans* systems. The lower basal level observed with the *L. casei* THF riboswitch could be indicative of a higher binding affinity for its cognate ligand(s) already present in the cell. This would therefore correspond to a greater level of background repression by endogenous reduced folates and hence provide a more significant opportunity for de-repression with **T0A**. However, the binding affinity of the *L. casei* aptamer domain for folinic acid was found to be around five-fold lower than that of the *S. mutans* aptamer, therefore not supporting this hypothesis. As delineated in **Section 1.7.3**, there are many forms of reduced folate present in the cell, a number of which have been found to bind the THF riboswitch⁵⁷. It may be that the observed background repression of the *L. casei* THF riboswitch is mediated through interaction with other intracellular reduced folates for which the riboswitch has a particularly high binding affinity. Clearly, further ITC studies are required to assess this hypothesis.

5.2 Conclusions: Re-engineering of the THF riboswitch

With respect to the second goal of this project, the re-engineering of the THF riboswitch to create an orthogonal gene expression system, the work presented here represents a significant step towards that objective. A mutant THF riboswitch, **M1**, containing a

modified 3WJ binding site was designed and shown to bind the synthetic ligand **T1A** by ITC (see **Sections 4.2.2.2** and **4.2.2.5**). It was also demonstrated using ITC that these mutations prevented the binding of folinic acid at the 3WJ site. Moreover, for the *L. casei* THF riboswitch, the **M1** mutant was found to repress GFP expression *in vivo* by up to 40% upon addition of exogenous **T1A** (see **Section 3.2.2.11**). This represents a significant step forward in the engineering of a dual-input orthogonal THF riboswitch. However, it should be noted that this repressive effect was not observed when the PK site was knocked out; indicating that the presence of endogenous reduced folate at an unmodified PK site is required for the **M1-T1A** expression response.

Whilst the riboswitch 3WJ binding site was successfully re-engineered, attempts to modify the PK site in a similar fashion were not met with success. The ‘diamino-faced’ ligands designed to bind the PK site mutant **M2** were found to neither repress gene expression *in vivo* (see **Section 3.2.1.6**) nor bind the **M2** aptamer *in vitro* (see **Section 4.3.4.3**). Given that it has now been established *in vivo* that binding to the THF PK site is more important for the riboswitch regulatory function than binding at the 3WJ site, the re-engineering of the PK site should take priority in any future work.

The de-repressive effect of compound **T0A** discussed in **Section 5.1** was exploited to improve the performance of the wild-type *L. casei* THF riboswitch as a gene expression tool (see **Section 3.2.2.9**). When the basal levels of expression were de-repressed with **T0A**, it was possible to achieve up to 95% repression of expression with folinic acid in the *L. casei* system. Hence, this two-ligand solution was able to overcome the intrinsically low basal levels, which is one of the major drawbacks of utilising wild-type ‘OFF’ riboswitches as gene expression devices.

5.3 Future work

The work presented in this thesis represents a step forward in the understanding of the *in vivo* function of the THF riboswitch. However a number of questions remain unanswered. The mechanism by which some ligands are able to bind the THF riboswitch but not elicit a regulatory effect is still not clear and further structure-activity relationship studies with THF analogues would be useful. These studies would combine *in vitro* measurements of ligand binding affinity by ITC with *in vivo* reporter gene assays to assess the regulatory effect of the ligands in question. X-ray crystallography structures of THF analogues which induce a regulatory effect

compared with those that bind but do not regulate may provide an insight into the mechanism by which this discrepancy occurs. An overview of compounds desirable for further structure-activity relationship studies with the THF riboswitch is shown in **Figure 74**. One of the interesting observations of this study was that folinic acid was able to elicit the regulatory effect of the THF riboswitch *in vivo*, whereas the close structural analogue lometrexol was not (structures shown in **Figure 74**). One possible reason for this is that the presence of the N10 functional group in THF/folinic – which is absent in lometrexol – provides the restriction of rotation along the N10-C11 bond necessary for the adoption of the active ligand conformation within the aptamer domain binding sites. Compound **T0D** was shown to have no riboswitch-mediated effect on gene expression *in vivo*. However, the presence of the N10-formyl group in **T0D** reduces the availability of the N10 lone pair electrons for delocalisation into the adjacent phenyl ring and hence increases the rotational freedom of the N10-C11 bond. If this hypothesis is correct, compound **T0E** (see **Figure 74**), which does not possess an N10-formyl moiety, would be anticipated to elicit the regulatory effect of the THF riboswitch. This could be further expounded by the synthesis and *in vivo* testing of the N10-deaza compound **T0F**; a synthesis target of this project, but which was not successfully obtained due to difficulties with the selectivity of the hydrogenation step (see **Section 2.2.7**).

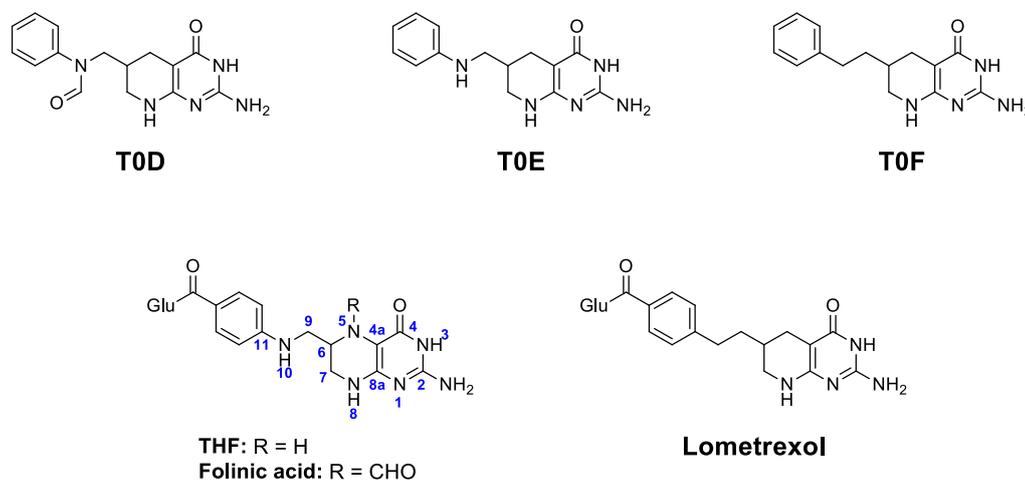


Figure 72 – Structures of compounds required for structure-activity relationship studies of the THF riboswitch

Additionally, this line of study may identify a THF analogue which is able to penetrate *E. coli* cell membranes without the need for a heterologously expressed folate transport protein. This would be highly desirable if the wild-type THF riboswitch were considered for use as a gene expression device.

Finally, the development of the **M1-T1A** riboswitch-ligand system demonstrated modest applicability as a gene expression tool. However, this riboswitch-ligand combination may now present an opportunity for studying the cooperativity of the two THF ligand binding sites which has not previously been possible. It has been demonstrated in this project that it is possible to selectively populate the THF **M1** riboswitch 3WJ binding site with the synthetic ligand **T1A**, whilst leaving the PK site unoccupied. Similarly, the **M1** PK site can be populated with folinic acid without ligand binding at the 3WJ site (see **Section 4.2.2.2**). An interesting experiment could be carried out *in vitro* using ITC to assess how ligand binding at each of the two binding sites affects the binding affinity at the other binding site. An experimental protocol would involve two ITC experiments. In one experiment, folinic acid is titrated into a solution the **M1** riboswitch (3WJ site unbound) and the binding affinity is measured. In a second experiment, folinic acid is titrated into a solution containing the **M1** riboswitch with **T1A** bound at the 3WJ site. A comparison of the two binding affinities would indicate the extent to which binding at the 3WJ site pre-organises the PK site for folinic acid binding and *vice versa*.

Chapter 6

Experimental

6.1 Overview of experimental apparatus and techniques

Nuclear magnetic spectra were recorded for solutions in DMSO- d_6 , D₂O (with minimum KOH dissolved for sufficient solubility) or TFA- d on Bruker 400MHz and 800MHz instruments (as specified). Chemical structures were assigned using ¹H-, ¹³C-, COSY, DEPT, HMBC and HMQC NMR data where appropriate. Proton and carbon chemical shifts (δ_C and δ_H) are quoted in parts per million (ppm) to the nearest 0.01 ppm, and include integration and multiplicity (s, singlet; d, doublet; m, multiplet; b, broad). Integration values and coupling constants (J_{HH}) were calculated using the MestReNova and ACDLabs software packages.

Mass spectrometry data was obtained from a Fissions VG Trio 2000 quadrupole mass spectrometer for Electrospray (ES +/-) conditions. High resolution MS data was recorded using a Thermo Finnigan MAT 95XP mass spectrometer. Molecular ions, sodium adducts and other major peaks are reported as mass/charge (m/z) ratios.

Analytical thin layer chromatography (TLC) was performed on Merck silica gel 60 Å plates (40-63 μ m), with visualisation under UV light, or the following staining agents: ninhydrin, vanillin or potassium permanganate.

Melting points were recorded using a Sanyo Gallenkamp MPD350 heater.

Solvents which were required to be anhydrous were dried using 3Å molecular sieves and reactions with anhydrous conditions were run under an atmosphere of N₂. Removal of solvents *in vacuo* was performed using a Büchi RE111 Rotavapor and a Büchi 461 water bath.

Centrifugation was carried out utilising: an Eppendorf 5804R Centrifuge; and a Sigma 4K15 centrifuge (speeds specified).

Flash Column Chromatography was carried out using Merck silica gel 60 Å, 2020-440 mesh particle size (35-75 μ m), utilising dry loading technique and preparing the sample by dissolving in the appropriate solvent and adding two mass equivalents of silica, drying *in vacuo*, and loaded onto the column.

Plate Reader measurements were obtained using a Synergy HT Microplate Reader (using 540 nm and 620 nm absorbance filters).

Fast Protein Liquid Chromatography (FPLC) was carried out using an Amersham Biosciences ÄKTA prime plus FPLC system. Separations were achieved utilising a, a Superdex[™] 75 HiLoad[™] 26/60 size exclusion column.

High Performance Liquid Chromatography (HPLC) was carried out on a Varian ProStar 210 system using a Varian ProStar 320 detector. Separations were achieved on a reverse-phase Phenomenex 5 µm C18 Column.

Isothermal titration Calorimetry was performed using a MicroCal VP-ITC microcalorimeter and MicroCal Auto-ITC200 instrument.

Sonication was performed using a Bandelin Sonopuls.

Gel imaging was performed using a GE Life Sciences Typhoon Trio variable mode imager.

6.2 Methods – *in vivo* reporter gene assays

6.2.1 Plasmid extractions, sequencing and glycerol cell stocks

Plasmids were extracted from *E. coli* cells using a Qiagen Qiaprep spin miniprep kit, following a standard protocol:

1. 5 ml overnight cultures in LB and the appropriate antibiotics were pelleted by centrifugation, re-suspended in 250 μ l of P1 buffer and transferred to a microfuge tube.
2. 250 μ l of P2 Buffer was added, the tube was inverted 10 times to thoroughly mix the solution, and the solution was allowed to equilibrate for 3 mins.
3. 350 μ l of P3 Buffer was added and the tube inverted 10 times to thoroughly mix the solution.
4. The sample was centrifuged for 10 min at 13,000 rpm.
5. The resulting supernatant was transferred to a QIAprep spin column, centrifuged for 1 minute, and the flow through discarded.
6. The spin column was washed with 500 μ l of Buffer PB, centrifuged for 60 seconds, and the flow through discarded.
7. The spin column was washed a second time with 750 μ l of Buffer PB, centrifuged for 1 minute, and the flow through discarded.
8. The spin column was centrifuged dry for a further 30 secs at 13,000 rpm to remove any excess wash buffer.
9. The spin column was placed in a 1.5 ml microfuge tube, 50 μ l of dH₂O was added and the column was allowed to stand for 1 minute before being centrifuged for 1 minute at 13,000 rpm to elute the DNA.

Glycerol cell stocks were prepared by taking cell culture from an overnight starter culture in LB and made to a 20% final glycerol concentration and stored at -80°C. Sequencing of all the plasmids that were used over the course of this investigation was carried out by GATC Biotech.

6.2.2 Site-directed mutagenesis

Mutagenesis was performed on plasmids so extracted by the above procedure following a standard protocol.

Primer design:

Mutagenic oligonucleotide primers were designed to contain the desired mutation and to anneal to the same sequence on opposite strands of the template plasmid. Primers were between 20 and 30 bases in length and were designed such that they began and ended with C or G, had a T_m of 65 – 75°C, and had the desired mutation approximately 7-10 bases from the 5' –end.

Protocol for site-directed mutagenesis:

1. Mutagenesis reactions were prepared in 200 µl thin-walled PCR tubes, containing the following:
 - 10 µl 5x New England Biolabs (NEB) Q5 buffer
 - 5 µl 2 mM dNTPs
 - 50 ng plasmid DNA
 - 125 ng of each primer as described above
 - 0.5 µl NEB Q5 DNA polymerase (2000 U/ml)
 - dH₂O to 50 µl total reaction volume

2. The mutagenesis reactions were then placed in a thermocycler using the following conditions:

Step	Temperature (°C)	Time	Number of Cycles
Initial Denaturation	98	30 s	1
Denaturation	98	10 s	
Annealing	$T_m - 5$	30 s	12
Extension	72	144 s	
Final Extension	72	10 mins	1
Storage	4	∞	

3. To each mutagenesis reaction was then added 0.5 µl restriction endonuclease DpnI (NEB, 20,000U/ml) to digest the parent plasmid and reactions were placed in the incubator at 37°C for 2 hours.
4. The so-prepared mutagenic plasmids were then transformed into chemically competent *E. coli* DH5α cells as described below.

6.2.3 Preparation of chemically competent *E. coli* DH5 α cells

Chemically competent *E. coli* DH5 α cells were prepared by a modification to the method reported by Inoue¹⁸⁷. Briefly, a 5 ml overnight LB culture of commercial DH5 α cells (NEB) was diluted 100-fold into LB media and grown until an OD₆₀₀ of 0.55 was reached. The culture was placed on ice for 10 minutes and then centrifuged at 3000 rpm for 10 minutes to pellet the cells. Cells were re-suspended in ice-cold Inoue buffer (see **Section 6.7**) and harvested by centrifugation at 3000 rpm for 10 minutes. The cells were finally re-suspended in Inoue buffer containing 6% DMSO, incubated on ice for 10 minutes, split into 100 μ l aliquots and flash frozen with liquid N₂ to store at -80°C.

6.2.4 Heat-shock transformation of *E. coli* DH5 α cells with mutagenic plasmids

E. coli DH5 α cells were transformed using a standard heat-shock procedure. To 80 μ l of chemically competent cells (prepared as previously described) was added 10 μ l of plasmid DNA from the PCR mutagenesis reaction. The mixture was incubated on ice for 30 minutes and then heated at 42°C in a heating block for 45 seconds, before being placed on ice for a further 5 minutes. 300 μ l of LB media was added and the culture was incubated at 37°C for 2 hours. From this, two aliquots of 50 μ l and 200 μ l were taken and plated onto two pre-prepared LB agar plates (containing 37 μ g/ml chloramphenicol) using a sterile rod. The plates were grown overnight at 37°C with shaking and three colonies were picked, grown up in an overnight starter culture (LB, 37 μ g/ml chloramphenicol) and then the plasmids harvested the following day using the previously described mini-prep plasmid extraction method. Plasmids were sent for sequencing to confirm the success of the mutagenesis.

6.2.5 Preparation of electro-competent *E. coli* DH5 α pLOI707HE/slr0642 cells

E. coli DH5 α cells containing the plasmid pLOI707HE/slr0642 were prepared by colleague Dr. Helen Vincent using the heat-shock transformation procedure outlined above. From these glycerol stocks, a 5 ml starter culture (containing 10 μ g/ml tetracycline) was prepared and grown overnight at 37°C with shaking. This was then diluted 500-fold into 100 ml LB media (containing 10 μ g/ml tetracycline) and incubated at 37°C with shaking until a OD₆₀₀ of 0.35 – 0.4 was reached. Cells were placed on ice

for 20 minutes and harvested by centrifugation at 2500 RCF for 10 minutes at 4°C. Cells were then re-suspended in 50 ml ice-cold filter-sterilised water by gentle swirling and re-harvested by centrifugation at 2500 RCF for 10 minutes at 4°C. The previous re-suspension and harvesting steps were repeated a further three times, before the cells were finally re-suspended in 1 ml of ice-cold filter-sterilised water, glycerol added to 10% concentration and the cells divided into 100 µl aliquots for flash freezing and storing at -80°C.

6.2.6 Transformation of *E. coli* DH5α pLOI707HE/slr0642 with mutagenic plasmids by electroporation

To 100 µl of *E. coli* DH5α pLOI707HE/slr0642 competent cells (prepared as described above) was added 2 ng of plasmid DNA (in no more than 10 µl H₂O) and the mixture was incubated on ice for 5 minutes. The cells were then transferred to an ice-cold electroporation cuvette (0.1 cm) and electroporated using a BioRad MicroPulser in Ec1 mode. LB media (900 µl) was immediately added and the cells were incubated at 37°C with shaking for 1.5 hours. Aliquots of 50 µl and 200 µl were then taken and plated onto pre-prepared LB agar plates (containing chloramphenicol 37 µg/ml; and tetracycline 10 µg/ml) using a sterile rod and then incubated at 37°C overnight. The following day, three colonies were picked, grown up in overnight cultures, the plasmids extracted by mini-prep as previously described and sent for sequencing.

6.2.7 *E. coli* GFP-*lv* reporter gene expression assay

Strains were plated onto pre-incubated LB agar plates (containing chloramphenicol, 37 µg/ml; and tetracycline 10 µg/ml) straight from -80°C glycerol stocks using a sterile pipette tip and plates were then transferred to a 37°C incubator overnight. The following day, three colonies from each plate were subsequently picked and grown in 5 ml starter cultures (containing chloramphenicol, 37 µg/ml; and tetracycline 10 µg/ml) overnight. The following day, the starter cultures were diluted 20-fold into freshly prepared M9-cas media (containing chloramphenicol, 37 µg/ml; and tetracycline 10 µg/ml) and grown for 6 hours at 37 °C. These cultures were then diluted 20-fold into M9-cas media a second time immediately prior to the preparation of the assay. IPTG (12.5 µl, 10 mM in H₂O) and ligand solutions (25 µl, 10x final desired concentration) were added to a 96 deep-well plate along with 212.5 µl of the appropriate cell culture prepared as previously described, giving a total volume of 250 µl. Control wells

consisted of 1) fresh M9-cas media and each ligand (without any cells) and; 2) the appropriate cell culture and water to replace the ligand. The 96 deep-well plate was then incubated with shaking at 500 rpm at 37°C for 16 hours. Following incubation, 50 μ l aliquots were then taken from the samples, diluted with 150 μ l of 1x PBS and added to a black-sided clear-bottom 96-well plate. This was then transferred to a Synergy HT Microplate reader to record *GFP-IV* expression and cell density by monitoring fluorescence intensity units using 485_{ex}/535_{em} filters and measuring UV absorbance at 600 nm respectively.

To account for the absorbance/fluorescence of the media and ligand, the control well readings (containing M9-cas media and ligand but no cells, as described above) were subtracted from the readings for the wells containing cells. Fluorescence values were normalised for cell density by dividing by the absorbance at 600_{nm} (OD). Relative gene expression was calculated by dividing normalised fluorescence levels for each strain-ligand combination by the normalised fluorescence levels for each strain with water only. This procedure was replicated for all three colonies picked for each strain and the mean and standard deviation was calculated using Microsoft Excel. In the case of the dose-response experiments, this raw data was then imported into SigmaPlot (v12.0) and fitted to a three-parameter logistic function using iterative analysis. The resulting sigmoidal curve was then used to calculate the expression parameters: T_{50} , IC_{50} and 90/10 dynamic range.

6.3 Methods - Enzyme over-expression and purification

6.3.1 Expression of recombinant T7 RNA Polymerase (T7-RNAP)

Over-expression of recombinant T7 RNAP was performed using an *E. coli* library strain containing the T7-RNAP expression vector, under the control of the lac promoter. A 5 ml starter culture (containing 50 µg/ml ampicillin) was prepared from -80 °C glycerol stocks, and grown at 37 °C with shaking overnight. The following day, the culture was diluted 100-fold into 400 ml fresh LB media (containing 50 µg/ml ampicillin) and incubated at 37 °C with shaking until an OD₆₀₀ of 0.6 was reached. Cultures were then induced with IPTG (1M, 400 µl) and incubated at 37 °C for a further 3 h, before being pelleted by centrifugation to harvest the cells. Pelleted cells were re-suspended in 30 ml of cell lysis buffer, lysozyme (Sigma Aldrich) and protease-inhibitor cocktail (Roche) were added and the cells were incubated on ice for 15 minutes. Cells were lysed by sonication (Bandelin Sonoplus), using 10 x 30 second bursts at 30 second intervals. The cellular debris was pelleted by centrifugation (10 krpm, 10 min, 4 °C) the supernatant collected, and the volume reduced to 2 ml using a Centricon spin column (3000 M_w cut-off) and stored at -80°C.

6.3.2 Purification of T7 RNA Polymerase (T7-RNAP)

Crude T7-RNAP (cell lysate) was purified using Ni²⁺ affinity gravity flow chromatography using a modification of a standard protocol from Life Technologies. The His-tagged T7-RNAP was bound to the Ni-NTA resin, washed with low concentration imidazole and then eluted at high imidazole concentration. The protocol utilised was as follows:

1. 2.5 ml of Ni-NTA resin (Qiagen) was added to a 10 ml purification column and allowed to settle under gravity for 5 – 10 minutes.
2. The supernatant was removed and the resin was suspended in 10 ml dH₂O.
3. The resin was again allowed to settle under gravity, the supernatant was removed and 10 ml of Native Purification Buffer (see buffer recipe page) was added to re-suspend the resin.
4. Step 3 was then repeated a further three times to equilibrate the resin in the purification buffer.

5. The supernatant was removed a final time and the previously prepared cell lysate was loaded onto the resin by eluting slowly under gravity and then equilibrating for 15 minutes.
6. The resin was then washed with 10 ml Native Wash Buffer (50 mM imidazole, see buffer recipe page) and pressure was applied to the column to elute the wash buffer more rapidly.
7. Step 6 was then repeated an additional three times.
8. The T7-RNAP was eluted by adding 12 ml of Native Elution Buffer (250 mM imidazole, see buffer recipe page) and applying pressure to the column to elute.
9. Aliquots of the cell lysate, the wash flow-through and the final eluent were retained and analysed by SDS-PAGE.
10. The so-purified T7-RNAP in 250 mM imidazole elution buffer was concentrated to ~ 2 ml and immediately dialysed into T7-RNAP buffer (see buffer recipe page).
11. From this, 50% glycerol stocks were prepared and stored in 0.5 ml aliquots at -80 °C.

6.3.3 SDS-Polyacrylamide Gel Electrophoresis (PAGE) analysis

SDS-PAGE analysis was performed using pre-cast 10% SDS gels (Amersham Biosciences) in an Amersham ECL Gel Electrophoresis System. Samples were prepared by mixing 5 µl of the analyte with an equal volume of loading dye and denaturing at 60 °C for 10 minutes. The samples were loaded onto the gel with the appropriate controls and ladder (Fermentas PAGERuler), and the gel was run at 200 V in SDS buffer (see buffer recipes) for 45 min. The gel was stained with instant blue (50 ml) for 30 minutes, and then scanned using a standard optical scanner for record keeping.

6.4 Methods - RNA synthesis and purification

6.4.1 *In vitro* transcription reactions

In vitro transcription was carried out under strict RNase-free conditions, utilising a modification to the procedure outlined by Gurevich¹⁸⁴. Reactions were performed using ssDNA anti-sense templates (Sigma Aldrich) containing the T7 promoter at the 5'-end and a short, reverse complement primer for the T7 promoter region only. Reaction conditions were first investigated at the pilot scale (total volume: 50 µl) before being scaled up for preparation of large quantities of RNA for purification and *in vitro* analysis. The large scale reactions were conducted at a total 10 ml volume, using the following components:

Reaction components (total volume: 10 ml):

10 x IVT Buffer	1000 µl
MgCl ₂ (1 M)	280 µl
DTT (1 M)	100 µl
ssDNA anti-sense template (100 µM)	200 µl
T7 promoter region primer (100 µM)	400 µl
DEPC H ₂ O	5120 µl

Then:

NTPs (25 mM)	2400 µl
T7-RNA polymerase	500 µl

The above components, excluding the NTPs and T7-RNA polymerase, were added to a 15 ml falcon tube and denatured in a water bath at 95 °C for 10 minutes and then allowed to cool gradually to room temperature. The NTPs and T7-RNAP were then added and the reaction mixture was then incubated at 37 °C for 3 hours, during which time a white precipitate of magnesium pyrophosphate was formed. DNase I (10U, New England Biolabs) was then added and the reaction was incubated for a further 1 hour at 37°C, before the reaction was halted by the addition of 100 µl EDTA (0.5 M, pH 8.0). Excess precipitate was removed by centrifugation and the supernatant was analysed by urea PAGE.

6.4.2 Urea PAGE analysis

The gel was prepared by mixing 5 g urea, 2 ml of 5x TBE, 3 ml of acrylamide (40%, 19:1) and 1 ml H₂O and then heating the mixture in a microwave (750W, 5 second bursts) until the urea had completely dissolved. The mixture was then allowed to cool to room temperature, set by the addition of 100 µl APS (10%) and 10 µl TEMED, and immediately added to a 1 mm gel cassette. The gel was then left for 2 hours to set thoroughly, before addition of the samples. RNA samples were prepared by mixing 5 µl of the analyte with an equal volume of 2 x loading dye (see buffer recipe page), and then denaturing the mixture at 95°C for 10 min in a heating block. After cooling to room temperature, the samples were loaded onto the gel and ran at 110 V for 10 minutes and then at 220 V for a further 50 minutes. Staining was performed using solution of 5 µl SYBR gold in 50 ml of 1x TBE, and the bands were visualised by fluorescence imaging (Typhoon™ Trio).

6.4.3 RNA purification using size exclusion FPLC

Following *in vitro* transcription, samples were reduced in volume to 2 ml using a Centricon spin column (3000 M_w cut-off) prior to purification. The FPLC instrument was fitted with a Superdex 75 HiLoad 26/600 size exclusion column, housed in a chiller cabinet (4°C) and strict RNase-free procedures were followed. The FPLC system was equilibrated with de-gassed RNA FPLC Buffer (10 mM phosphate, 100 mM NaCl, see buffer recipe page) for 6 hours at 2 ml/min flow rate. RNA samples were loaded onto the system and purified at a flow rate of 2 ml/min, with dual-wavelength UV monitoring at 260 and 280 nm. Fractions (1 ml) were collected in 96 well plates and samples were taken for urea PAGE analysis. The desired fractions were pooled according to integrity before being concentrated to ~2 ml using a Centricon spin column (3000 M_w cut-off), prior to ITC analysis.

6.4.4 Urea PAGE preparative scale purification

All of the steps in this section were carried out using RNase-free conditions. The gel cassette was assembled using glassware treated with RNase-zap and washed with DEPC-treated H₂O. The gel contained a mixture of urea (96 g), acrylamide (60 ml) and 5x TBE (40 ml) heated at 50 °C until the urea had completely dissolved and the solution was then made up to 200 ml with DEPC-treated water. The gel was set by adding 500 µl of 10 % APS and 50 µl TEMED and poured immediately into the gel cassette. Gel setting typically required around 3-4 hours. The crude *in vitro* transcription products were concentrated with a Centricon filter unit (3000 M_w cut-off) to 0.5 ml volume and then mixed with an equal volume of a denaturing loading dye. The so-prepared sample was then denatured at 95 °C for 10 min before being allowed to cool to room temperature. The sample was then loaded on to the gel and allowed to run in a buffer of 1x TBE at 150V for 18 hours.

The gel was removed from the cassette apparatus and covered in Saran Wrap to protect from RNase contamination. RNA bands were visualised using a UV and desired product band was excised from the gel with a scalpel blade. Gel fragments excised in this manner were crushed by forcing through a 50 ml syringe, 25 ml of (0.5x) TBE was added to the and the mixture was mixed on a roller at 4 °C for 3 h. Samples were then centrifuged for 15 min at 5000 rpm to pellet the crushed gel fragments, 20 ml of the solution recovered, and the volume restored to 25 ml with (0.5x) TBE. This process was repeated a further 3-6 times until RNA extraction into the buffer solution ceased, as determined by urea PAGE. Samples were then pooled and concentrated using a Centricon filter unit (3000 M_w cut-off), ready for dialysis prior to ITC analysis or x-ray crystallography trials.

6.5 Methods - Isothermal Titration Calorimetry (ITC)

RNA samples prepared as described in **Sections 6.4.1 – 6.4.4** were dialysed overnight at 4 °C into a thoroughly degassed buffer of 70 mM Na HEPES pH 8.0, 70 mM NaCl, 2.5 mM MgCl₂ and 1 mM DTT (buffer conditions outlined by Trausch¹¹⁰). Ligand solutions were then made in the same dialysis buffer, degassed in a sonicator bath and incubated at 50 °C for 20 minutes to ensure complete dissolution.

ITC experiments were performed using a MicroCal VP-ITC microcalorimeter, or a MicroCal Auto-ITC200 automated system. ITC experiments were performed using standard parameters shown in **Tables 5 and 6**. RNA concentrations in the cell were between 50-200 µM, with the concentration of the ligand in the injection syringe between 10-20x greater than the RNA concentration.

Table 5: Standard parameters used for performing ITC experiments with the MicroCal VP-ITC microcalorimeter

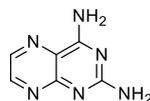
MicroCal VP-ITC Parameters		
Injections	25	
Cell Temperature	25 °C	
Reference Power	5 µCal/s	
Initial Delay	60 s	
Injection parameters	(1st)	Others
Volume (µl)	2	12
Duration (s)	4.8	14.4
Spacing (s)	300	300
Filter Period (s)	2	2

Table 6: Standard parameters used for performing ITC experiments with the MicroCal PEAQ-ITC automated microcalorimeter

MicroCal Auto-ITC200 Parameters		
Injections	16	
Cell Temperature	25 °C	
Reference Power	10 μ Cal/s	
Initial Delay	60 s	
Injection parameters	(1st)	Others
Volume (μ l)	0.5	2.5
Duration (s)	1.0	5.0
Spacing (s)	150	150
Filter Period (s)	5	5

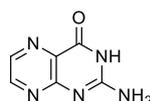
6.6 Chemical Synthesis

Synthesis of pteridine-2,4-diamine (28)¹⁵²



2,4,5,6-tetraaminopyrimidine sulphate salt (2.00 g, 8.40 mmol) was added slowly to a stirred solution of NaHCO₃ (1.40 g, 16.8 mmol) in H₂O (50 ml, HPLC grade). The reaction mixture was heated to 70°C and glyoxal (40% soln in H₂O, 1.20 ml, 8.4 mmol) was added dropwise. The solution was refluxed at 70 °C for 3 h and then immediately cooled to 10°C in a water bath. The resulting light-brown precipitate was collected by filtration and washed with EtOAc (3 x 100 ml), CHCl₃ (3 x 100 ml) and EtOH (3 x 100 ml) yielding a light-brown solid (0.95 g, 5.86 mmol, 69.8%). m.p. >250°C (decomp.). ¹H-NMR (400MHz, DMSO-d₆): δ = 6.70 ppm (bs, 2H, NH₂); δ = 7.70 (bs, 2H, NH₂); δ = 8.30 (s, 1H, Ar-H); δ = 8.70 (s, 1H, Ar-H). ¹³C-NMR (100 MHz, DMSO-d₆): δ = 163.07 (C2), 162.99 (C4), 156.14 (C8a), 150.29 (C7), 137.13 (C6) 122.96 (C4a) ppm. IR: 3322, 3201, 1638, 1587, 1542, 1449 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 163.1, found = 163.0.

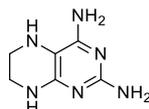
Synthesis of 2-aminopteridin-4(3H)-one (29)¹⁵²



4-hydroxy-2,5,6-triaminopyrimidine sulphate salt (1.0 g, 4.4 mmol) and glyoxal (0.7 ml, 40% in H₂O, 4.8 mmol) were added to H₂O (20 ml, HPLC grade). KOH (1.0 g, 17.9 mmol, 4 eq.) was then added portion-wise, which solubilised the suspension, and the reaction mixture was stirred at room temperature for 24 hours. The solution was neutralised with 5 M HCl, and the resulting precipitate was collected by filtration and washed with H₂O (3 x 50 ml), MeOH (3 x 50 ml), CHCl₃ (3 x 50 ml) and acetone (1 x 50 ml). The resulting orange/brown solid was then dissolved in ~20 ml 5 M KOH and stirred over active carbon for 15 h. The carbon was removed by filtering through celite, and the resulting solution was neutralised with 5 M HCl and filtered to yield an orange/brown solid (604 mg, 3.7 mmol, 84.2%). m.p: > 350°C (decomp.). ¹H-NMR (400MHz, DMSO-d₆): δ = 8.44 ppm (d, *J* = 2.2 Hz, 1H, Ar-H); δ = 8.15 (d, *J* = 2.2 Hz,

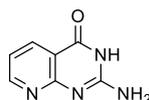
1H, Ar-H); δ = 6.88 (bs, 2H, NH₂). ¹³C-NMR (100 MHz, D₂O/NaOH): δ = 173.19 (C4), 163.89 (C2), 156.50 (C8a), 148.29 (C7), 138.62 (C6), 129.29 (C4a). IR: 3329, 3058, 1686, 1618, 1573, 1539, 1514 cm⁻¹. ESI-MS (negative ion mode): calculated for [M – H]⁻ = 162.0, found = 162.0.

Synthesis of 5,6,7,8-tetrahydropteridine-2,4-diamine (31)¹⁵¹



H₂O (30 ml, HPLC grade) was de-oxygenated by purging with N₂ for 30 minutes before PtO₂ (90 mg, 390 μ mol, 0.2 eq.) was added and the mixture was exposed to H₂ for 1 hour to activate the catalyst. Compound **28** (300 mg, 1.86 mmol) was added and the reaction mixture was hydrogenated for 72 h (1 atm) and monitored by TLC (CHCl₃/MeOH, 5:1). When all the starting material was consumed, the reaction mixture was filtered through celite and the water removed in vacuo to yield a yellow, sticky solid. This residue was dissolved in a minimum of MeOH/HCl (~5 ml) and stored at -20 °C. After 24 hours, the resulting cream precipitate was collected by filtration and washed with MeOH (3 x 20 ml). Yield: (92 mg, 455 μ mol, 24.5%). m.p. 252-254 °C. ¹H-NMR (400MHz, DMSO-d₆): δ = 8.51 ppm (bs, 2H, NH); δ = 7.93 (bs, 2H, NH₂); δ = 7.60 (bs, 2H, NH₂); δ = 3.52 (bt, 2H, CH₂); δ = 3.34 (bt, 2H, CH₂). ¹H-NMR (400MHz, D₂O): δ = 3.47 ppm (t, *J* = 4.5 Hz, 2H, CH₂); δ = 3.40 ppm (t, *J* = 4.5 Hz, 2H, CH₂). ¹³C-NMR (100 MHz, DMSO-d₆): δ = 162.59 (C2), 154.11 (C4), 150.03 (C8a), 133.05 (C4a), 44.19, 40.22 (C6 + C7) ppm. IR: 3221, 3061, 1614, 1573, 1339 cm⁻¹. ESI-MS (positive ion mode): found = 167.1, calculated for [M + H]⁺ = 167.1. HRMS (ES, m/z): calculated for [M + H]⁺ = 167.1045, found = 167.1041.

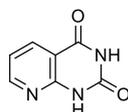
Synthesis of 2-aminopyrido[2,3-d]pyrimidin-4(3H)-one (36)¹⁵⁶



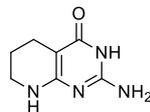
2,4-diamino-6-hydroxypyrimidine (1.0g, 7.93 mmol) was dissolved in hot H₂O (50 ml). AcOH (1.2 ml) and 1,1,3,3-tetraethoxypropane (700 mg, 3.18 mmol) were added, and the reaction mixture was refluxed for 3.5 h. The resulting white precipitate was

collected by filtration, was with H₂O (3 x 20 ml) and EtOH (3 x 20 ml) and dried in the oven at 90 °C for 4 h and analysed by ¹H-NMR. After drying, the white powder was refluxed in AcOH (50 ml) for 10 hours and the resulting white precipitate was collected by filtration and washed with H₂O (3 x 20 ml) and EtOH (3 x 20 ml). Yield: (473 mg, 2.92 mmol, 92%). m.p. > 300 °C (decomp.); ¹H NMR (400 MHz, DMSO): δ = 11.41 ppm (bs, 1H, CONH), δ = 8.68 (m, 1H, H7), δ = 8.28 (d, *J* = 7.7Hz, 1H, H5), δ = 7.20 (dd, *J*₁ = 7.8 Hz, *J*₂ = 4.8 Hz, 1H, H6), δ = 6.87 (bs, 2H, NH₂). ¹³C-NMR (100 MHz, D₂O/NaOD): δ = 174.69 (C8a), 164.36 (C2), 160.56 (C4), 153.47 (C7), 135.91 (C5), 117.36 (C6), 111.72 (C4a) ppm. IR: 3239, 2982, 2677, 1665, 1594, 1568. ESI-MS (positive ion mode): found = 163.0, calculated for [M + H]⁺ = 163.1.

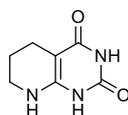
Synthesis of pyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione (37)¹⁵⁷



2-aminonicotinic acid (500 mg, 3.62 mmol) and urea (4.34 g, 72.4 mmol, 20 eq.) were mixed together in the absence of solvent and heated at 150 °C for 12 h. The mixture was then cooled to 90 °C, water (~10 ml) was added to dissolve the excess urea and the reaction was allowed to cool to room temperature. The unsolubilised white solid was collected by filtration, dissolved in 20 ml 1 M KOH and stirred for 1 h before being neutralised with acetic acid, resulting in re-precipitation of a white solid which was collected by filtration. Yield: (538 mg, 3.30 mmol, 91.2%). m.p. > 310 °C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 11.59 ppm (bs, 2H, NH); δ = 8.59 (dd, *J*₁ = 4.7 Hz, *J*₂ = 1.7 Hz, 1H, H7); δ = 8.25 (dd, *J*₁ = 7.7 Hz, *J*₂ = 1.6 Hz, 1H, H5); δ = 7.23 (dd, *J*₁ = 7.7 Hz, *J*₂ = 4.8 Hz, 1H, H6). ¹³C-NMR (100 MHz, TFA-d/D₂O): δ = 154.66 (C4), 144.08 (C2), 142.52 (C7), 142.05 (C5), 139.90 (C8a), 115.52 (C6), 108.76 (C4a) ppm. IR: 3275, 3153, 1636, 1539, 1403 cm⁻¹. ESI-MS (positive ion mode): found = 164.0, calculated for [M + H]⁺ = 164.0.

Synthesis of 2-amino-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-4(3H)-one (38, T0A)¹⁵⁸

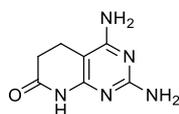
AcOH (60% in H₂O, 20ml) was de-oxygenated by purging with N₂ for 30 minutes, before PtO₂ (28 mg, 123 μmol, 0.2 eq.) was added and the mixture was hydrogenated under H₂ (1 atm) for 1 hour to activate the catalyst. Compound **36** (100 mg, 612 μmol) was added and the reaction mixture was hydrogenated (1 atm) for 48 hours and monitored by TLC (CHCl₃/MeOH 5:1). Upon consumption of **36** the reaction was filtered through celite and the solvent was removed *in vacuo* to yield a colourless oil. Crystallisation was achieved by the repeated addition and evaporation (*in vacuo*) of Et₂O and the resulting off-white precipitate was collected by centrifugation, washed with Et₂O (3 x 25 ml) and MeOH (3 x 5 ml) and dried under a stream of N₂. Yield: (65 mg, 391 μmol, 63.9%). m.p. > 300°C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 11.32 ppm (bs, 1H, CONH); δ = 7.55 (bs, 2H, NH₂); δ = 7.09 (bs, 1H, NH); δ = 3.21 (m, 2H, H7); δ = 2.24 (m, 2H, H5); δ = 1.69 (m, 2H, H6). ¹³C-NMR (100 MHz, D₂O/NaOD): δ = 173.67 (C8a), 161.21 (C2), 160.79 (C4), 88.11 (C4a), 40.58 (C7), 21.31 (C6), 19.60 (C5) ppm. IR: 3247, 2971, 1709, 1586, 1538 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 167.1 found = 167.1. HRMS (ES, m/z): calculated for C₇H₁₀N₄O [M + H]⁺ = 167.0933, found = 167.0936.

Synthesis of 5,6,7,8-tetrahydropyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione (39, T1A)

AcOH (60% in H₂O, 20 ml) was de-oxygenated by purging with N₂ for 30 minutes, before PtO₂ (28 mg, 123 μmol, 0.2 eq.) was added and the mixture was hydrogenated under H₂ (1 atm) for 1 h to activate the catalyst. Compound **37** (100 mg, 613 μmol) was added and the reaction mixture was hydrogenated for 20 h (1 atm) and monitored by TLC (CHCl₃/MeOH 2:1). The reaction was filtered through celite and the solvent was removed *in vacuo* to yield a straw coloured oil. This was dissolved in minimum MeOH and allowed to stand at room temperature for 5 minutes, affording a fine cream precipitate, which was collected by centrifugation and dried under stream of N₂. The

remaining solution was concentrated *in vacuo* and the process was repeated three times, yielding a cream solid (59 mg, 353 μmol , 57.6%). m.p. > 300°C (decomp.); ^1H NMR (400 MHz, DMSO): δ = 10.09 ppm (bs, 2H, CONH); δ = 6.10 (bs, 1H, NH); δ = 3.17 (m, 2H, H7); δ = 2.16 (t, J = 6.1 Hz, 2H, H5); δ = 1.67 (m, 2H, H6). ^{13}C NMR (100 MHz, DMSO): δ = 163.09 (C8a), 150.50, 150.03 (C2 + C4), 79.88 (C4a), 40.22 (C7) 20.54 (C6), 18.29 (C5). IR: 3244, 2944, 1728, 1628, 1556 cm^{-1} . ESI-MS (positive ion mode): calculated for $[\text{M} + \text{H}]^+$ = 168.1 found = 168.0. HRMS (ES, m/z): calculated for $\text{C}_7\text{H}_9\text{N}_3\text{O}_2$ $[\text{M} + \text{H}]^+$ = 168.0772, found = 168.0772.

Synthesis of 2,4-diamino-5,6-dihydropyrido[2,3-d]pyrimidin-7(8H)-one (40)¹⁵⁹



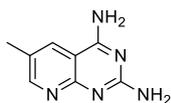
Sodium methoxide (825 mg, 15.25 mmol) was added to MeOH (15 ml) and refluxed at 60°C for 15 minutes to solubilise. Guanidine carbonate (900 mg, 15 mmol) was added, the solution was refluxed for a further 15 minutes and the resulting Na_2CO_3 precipitate was collected by filtration. To the so-prepared solution of guanidine in methanol was placed in a microwave process vial and malononitrile (395 mg, 6 mmol), and methyl acrylate (500 mg, 5 mmol) were added. The vial was sealed and subjected to microwave radiation for 10 minutes at 140 °C. After cooling, the resulting off-white precipitate was collected by filtration and washed with MeOH (3 x 10 ml). Yield (530 mg, 2.96 mmol, 90%). m.p. > 300°C (decomp.). ^1H NMR (400 MHz, DMSO): δ = 10.45 ppm (s, 1H, CONH); δ = 6.23 (s, 2H, NH_2); δ = 5.97 (s, 2H, NH_2); δ = 2.47 (m, 4H, CH_2). IR: 3410, 3327, 3128, 1619, 1565, 1438 cm^{-1} . ESI-MS (positive ion mode): calculated for $[\text{M} + \text{H}]^+$ = 180.1, found = 180.0.

Procedure A: Example procedure for the condensations of 2,4-diamino-6-hydroxypyrimidine and 2,4,6-triaminopyrimidine with α -substituted acrylaldehydes

The synthesis of 6-substituted-2,4-diaminopyrido[2,3-*d*]pyrimidines and 6-substituted-2-amino-4-hydroxypyrido[2,3-*d*]pyrimidines was achieved using a modification of the procedure outlined by Troschutz¹⁴⁸. The appropriate pyrimidine (1.0 eq.) was completely dissolved in minimum refluxing AcOH and then the appropriate α -substituted acrylaldehyde (1.1 equivs) was added rapidly and the reaction mixture stirred under reflux at 120°C for 12 hours. During this time, a precipitate formed which was then collected by centrifugation and washed with water (3 x 50 ml), MeOH (3 x 20 ml), EtOAc (3 x 50 ml), DCM (3 x 10 ml), Me₂CO (3 x 50 ml) and Et₂O to yield a powder, which was then characterised by NMR and mass spec.

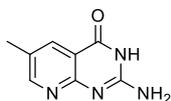
Procedure B: Example procedure for the condensation of 4-amino-2,6-dihydroxypyrimidine with α -substituted acrylaldehydes

The synthesis of 6-substituted-2,4-dihydroxypyrido[2,3-*d*]pyrimidines was achieved using a modification of the procedure outlined by Troschutz¹⁴⁸. 4-amino-2,6-dihydroxypyrimidine (1.0 eq.) was dissolved in minimum refluxing TFA, the appropriate α -substituted acrylaldehyde (1.1 eq.) was added and the reaction mixture stirred under reflux at 75°C and monitored by TLC (5:1 CHCl₃, MeOH) for the consumption of 4-amino-2,6-dihydroxypyrimidine. On completion, the reaction was allowed to cool to room temperature, and Et₂O was added dropwise to effect precipitation. The resulting precipitate was collected by centrifugation and washed with water (3 x 50 ml), MeOH (3 x 20 ml), EtOAc (3 x 50 ml), DCM (3 x 10 ml), Me₂CO (3 x 50 ml) and Et₂O to yield a powder, which was then characterised by NMR and mass spec.

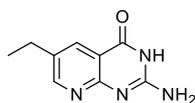
Synthesis of 6-methylpyrido[2,3-d]pyrimidine-2,4-diamine (43)¹⁴⁸

Compound **43** was prepared according to Procedure A, with 2,4,6-triaminopyrimidine (1.0 g, 8.0 mmol) and methacrolein (720 μ l, 610 mg, 8.8 mmol) and then purified by C18 reverse phase chromatography (10:1 H₂O/MeOH) to give product **43**^{*} (201 mg, 13.0%) as a yellow solid. ¹H NMR (400 MHz, DMSO): δ = 8.51 (s, 1H, H7); δ = 8.20 (s, 1H, H5); δ = 7.49 (bs, 2H, NH₂); δ = 6.32 (bs, 2H, NH₂); δ = 2.31 (s, 3H, Me). IR: 3300, 3115, 1632, 1616, 1664, 1432 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 176.1 found = 176.1; calculated for [M + Na]⁺ = 198.2 found = 198.2.

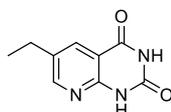
*Note – product was approximately 70% pure following chromatography and carried through to next stage.

Synthesis of 2-amino-6-methylpyrido[2,3-d]pyrimidin-4(3H)-one (44)

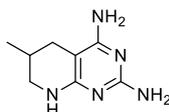
Compound **44** was prepared according to Procedure A, with 2,4-diamino-6-hydroxypyrimidine (1.0 g, 8.0 mmol) and methacrolein (720 μ l, 610 mg, 8.8 mmol) to give product **44** (806 mg, 4.58 mmol 57.2%) as a white solid. m.p. > 300°C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 11.36 ppm (bs, 1H, CONH); δ = 8.47 (s, 1H, H7); δ = 8.02 (s, 1H, H5); δ = 6.72 (bs, 2H, NH₂); δ = 2.31 (s, 3H, Me). ¹³C-NMR (100 MHz, D₂O/NaOD): δ = 174.43 (C8a), 163.71 (C2), 158.46 (C4), 153.97 (C7), 135.10 (C5), 127.12 (C6), 111.02 (C4a), 16.89 (CH₃) ppm. IR: 3252, 2999, 2770, 1655, 1600. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 177.1 found = 177.1. HRMS (ES, m/z): calculated for C₈H₈N₄O [M + Na]⁺ = 199.0596 found = 199.0604.

Synthesis of 2-amino-6-ethylpyrido[2,3-d]pyrimidin-4(3H)-one (45)

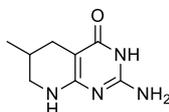
Compound **45** was prepared according to Procedure A, with 2,4-diamino-6-hydroxypyrimidine (1.0 g, 8.0 mmol) and 2-ethylacrolein (860 μ l, 740 mg, 8.8 mmol) to give product **45** (880 mg, 4.63 mmol 57.9%) as a white solid. m.p. > 300°C (decomp.) ^1H NMR (400 MHz, DMSO): δ = 8.26 (d, J = 2.5 Hz, 1H, H7); δ = 7.80 (d, J = 2.5 Hz, 1H, H5); δ = 6.72 (bs, 2H, NH_2); δ = 2.42 (q, J = 7.5 Hz, 2H, CH_2); δ = 0.98 ppm (t, J = 7.5 Hz, 3H, CH_3);. ^{13}C -NMR (100 MHz, $\text{D}_2\text{O}/\text{NaOD}$): δ = 174.63 (C8a), 163.89 (C2), 158.81 (C4), 153.52 (C7), 134.02 (C6), 133.49 (C5), 111.29 (4a), 24.70 (C9), 14.37 (C10). IR: 3232, 3004, 2826, 1672, 1598, 1561. ESI-MS (positive ion mode): calculated for $[\text{M} + \text{H}]^+$ = 191.1 found = 191.2. HRMS (ES, m/z): calculated for $\text{C}_9\text{H}_{10}\text{N}_4\text{O}$ $[\text{M} + \text{H}]^+$ = 191.0933, found = 191.0932.

Synthesis of 6-ethylpyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione (46)

Compound **46** was prepared according to Procedure B, with 4-amino-2,6-dihydroxypyrimidine (1.0 g, 8.0 mmol) and 2-ethylacrolein (860 μ l, 740 mg, 8.8 mmol) to give product **46** (696 mg, 3.64 mmol, 45.5%) as an off-white powder. m.p. > 300°C (decomp.). ^1H NMR (400 MHz, DMSO): δ = 11.60 ppm (s, 1H, CONH); δ = 11.43 (s, 1H, CONH); δ = 8.50 (d, J = 2.4 Hz, 1H, H7); δ = 8.10 (d, J = 2.5 Hz, 1H, H5); 2.67 (q, J = 7.5 Hz, 2H, H9); δ = 1.20 (t, J = 7.6 Hz, 3H, H10). ^{13}C -NMR (100 MHz, $\text{D}_2\text{O}/\text{NaOD}$): δ = 173.32 (C8a), 164.50 (C2), 156.82 (C4), 153.40 (C7), 134.59 (C5), 133.23 (C6), 110.34 (C4a), 24.72 (C9), 14.53 (C10). IR: 3048, 2870, 2823, 1729, 1699, 1601 cm^{-1} . ESI-MS (positive ion mode): calculated for $[\text{M} + \text{H}]^+$ = 192.1 found = 192.1. HRMS (ES, m/z): calculated for $\text{C}_9\text{H}_9\text{N}_3\text{O}_2$ $[\text{M} + \text{H}]^+$ = 192.0773, found = 192.0768.

Synthesis of 6-methyl-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidine-2,4-diamine (47, T2B)

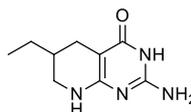
TFA (20 ml) was de-oxygenated by purging with N₂ for 30 minutes, before PtO₂ (51 mg, 228 μmol, 0.2 eq.) was added and the mixture was hydrogenated under H₂ for 1 hour to activate the catalyst. Compound **43** (200 mg, 1.14 mmol) was added and the reaction mixture was hydrogenated (1 atm) for 36 hours and monitored by TLC (CHCl₃/MeOH 5:1). Upon consumption of **43** the reaction was filtered through celite and the solvent was removed *in vacuo* to yield a pale yellow oil. Crystallisation was achieved by the repeated addition and evaporation (*in vacuo*) of Et₂O and the resulting off-white precipitate was collected by centrifugation, washed with Et₂O (3 x 25 ml) and MeOH (3 x 5 ml) and dried under a stream of N₂. The solid was re-crystallised from MeOH to give a white powder (104 mg, 0.58 mmol, 51.0%). m.p. > 300°C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 5.94 ppm (bs, 1H, NH); δ = 5.39 (bs, 2H, NH₂); δ = 5.10 (bs, 2H, NH₂); δ = 3.10 (m, 1H, H7a); δ = 2.72 (m, 1H, H7b); δ = 2.35 (m, 1H, H5a); δ = 1.83 (m, 2H, H5b + H6); δ = 0.98 (d, *J* = 6.1 Hz, 3H, Me). ¹³C-NMR (100 MHz, DMSO-d₆): δ = 161.09, 160.87, 159.70 (C2 + C4 + C8a), 80.55 (C4a), 46.81 (C7), 28.24 (C5), 26.80 (C6), 18.89 (CH₃) ppm. IR: 3382, 3330, 3164, 2948, 1597. HRMS (ES, *m/z*): calculated for C₈H₁₃N₅ [M + H]⁺ = 180.1249, found = 180.1246.

Synthesis of 2-amino-6-methyl-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-4(3H)-one (48, T0B)

Compound **48** was prepared as described for the synthesis of compound **47**, with compound **44** (200 mg, 1.14 mmol), TFA (20 ml) and PtO₂ (51 mg, 228 μmol, 0.2 eq.) to give a white powder (158 mg, 0.88 mmol, 77.0%). m.p. > 300°C (decomp.) ¹H NMR (400 MHz, DMSO): δ = 11.42 ppm (bs, 1H, CONH); δ = 7.57 (bs, 2H, NH₂); δ = 7.14 (bs, 1H, NH); δ = 3.27 (d, *J* = 12.1 Hz, 1H, H7a); δ = 2.79 (m, 1H, H7b); δ = 2.42 (d, *J* = 10.9 Hz, 1H, H5a); δ = 1.82 (m, 2H, H5b + H6); δ = 0.92 (d, *J* = 6.6 Hz, 3H, Me). ¹³C-NMR (100 MHz, D₂O/NaOD): δ = 173.76 (C8a), 161.22 (C2), 160.43 (C4), 87.37 (C4a),

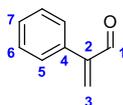
47.00 (C7), 27.88 (C5), 26.43 (C6), 18.02 (CH₃) ppm. IR: 3301, 2967, 2938, 1710, 1623 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 181.1 found = 181.1. HRMS (ES, m/z): calculated for C₈H₁₂N₄O [M + H]⁺ = 181.1089, found = 181.1094.

Synthesis of 2-amino-6-ethyl-5,6,7,8-tetrahydropyrido[2,3-d]pyrimidin-4(3H)-one (49, T0C)



Compound **49** was prepared as described for the synthesis of compound **47**, with compound **45** (200 mg, 1.05 mmol), TFA (20 ml) and PtO₂ (47 mg, 210 μmol, 0.2 eq.) to give a white powder (133 mg, 0.69 mmol, 65.3%). m.p. > 300°C (decomp.) ¹H NMR (400 MHz, DMSO): δ = 7.71 ppm (bs, 2H, NH₂); δ = 7.13 (bs, 1H, NH); δ = 3.27 (d, *J* = 11.4 Hz, 1H, H7a); δ = 2.80 (m, 1H, H7b); δ = 2.42 (m, 1H, H5a); δ = 1.79 (dd, *J*₁ = 15.2 Hz, *J*₂ = 9.5 Hz, 1H, H5b); δ = 1.53 (bm, 1H, H6); δ = 1.23-1.33 (m, 2H, H9); δ = 0.91 (t, *J* = 7.3 Hz, 1H, H10). ¹³C-NMR (100 MHz, DMSO): δ = 160.43 (C_q), 152.98 (C_q), 151.92 (C_q), 82.30 (C4a), 44.78 (C7), 32.06 (C6), 25.44 (C9), 24.32 (C5), 11.32 (C10). IR: 3246, 2959, 2822, 2784, 1663, 1601 cm⁻¹. (ESI-MS (positive ion mode): calculated for [M + H]⁺ = 195.1 found = 195.1. HRMS (ES, m/z): calculated for C₉H₁₄N₄O [M + H]⁺ = 195.1246, found = 195.1244.

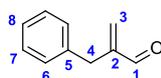
Synthesis of 2-phenylacrylaldehyde (51) (modified from Erkillä¹⁶⁰)



4-(dimethylamino)benzoic acid (274 mg, 1.66 mmol, 0.2 eq.) was added to stirred DCM (10 ml) and pyrrolidine (68 μl, 59 mg, 0.83 mmol, 0.1 eq.) was added dropwise to solubilise. Formaldehyde (37% in H₂O, 673 μl, 8.3 mmol, 1.0 eq.) was added dropwise before phenylacetaldehyde (1.0 g, 8.3 mmol) was added in a fast manner and the reaction mixture was immediately refluxed at 40 °C for 1 h. The reaction was worked up by adding saturated NaHCO₃ (20 ml) and extracting with DCM (3 x 20 ml). The combined organic extracts were dried over MgSO₄ and the solvent removed *in vacuo* to yield a yellow oil (1.02 g, 93.1%). Due to the instability of the product this was then

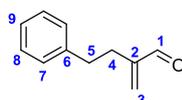
used immediately in subsequent reactions without further purification. ^1H NMR (400 MHz, CDCl_3): δ = 9.75 ppm (s, 1H, H1); δ = 7.14-7.20 (m, 5H, Ar-H); δ = 6.57 (s, 1H, H3a); δ = 6.13 (s, 1H, H3b). ^{13}C -NMR (100 MHz, CDCl_3): δ = 193.07 (C1), 148.51, 136.29 (C2 + C4), 135.57 (C3), 128.87, 128.54, 128.20 (C5, C6 + C7). IR: 3101, 2916, 1721, 1651, 1495, 1454 cm^{-1} . ESI-MS (positive ion mode): calculated for $[\text{M} + \text{H}]^+$ = 133.1, found = 133.1.

Synthesis of 2-benzylacrylaldehyde (52)



Compound **52** was prepared as described for the synthesis of compound **51** with 4-(dimethylamino)benzoic acid (825 mg, 5 mmol, 0.2 eq.), pyrrolidine (212 μl , 184 mg, 2.5 mmol, 0.1 eq.), formaldehyde (37% in H_2O , 1.0 ml, 25 mmol, 1.0 eq), hydrocinnamaldehyde (3.35 ml, 3.41 g, 25 mmol), and DCM (25 ml). The product was purified by flash chromatography (20:1 hexane/EtOAc) to yield a colourless oil (2.52 g, 17.3 mmol, 69.0%). ^1H NMR (400 MHz, CDCl_3): δ = 9.67 ppm (s, 1H, H1); δ = 7.21-7.36 (m, 5H, Ar-H); δ = 6.14 (s, 1H, H3a); δ = 6.09 (s, 1H, H3b); δ = 3.60 (s, 2H, H4). ^{13}C -NMR (100 MHz, CDCl_3): δ = 193.99 (C1), 149.76 (C2), 138.15 (C5), 135.23 (C3), 129.16, 128.57, 126.46 (C6, C7, C8), 34.16 (C4) ppm. IR: 3051, 2910, 1720, 1584, 1496, 1453 cm^{-1} . ESI-MS (positive ion mode): calculated for $[\text{M} + \text{H}]^+$ = 146.1, found = 146.1. HRMS (ES, m/z): calculated for $\text{C}_{10}\text{H}_{10}\text{O}$ $[\text{M} + \text{H}]^+$ = 146.0732, found = 146.0729.

Synthesis of 2-methylene-4-phenylbutanal (53)

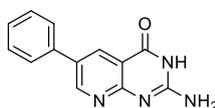


Step 1: Swern oxidation. To a stirred solution of $(\text{COCl})_2$ (6.8 ml, 10.0 g, 79.4 mmol, 1.2 eq.) in dry DCM (150 ml) at -78°C under an atmosphere of N_2 was added DMSO (5.63 ml, 6.20 g, 79.4 mmol, 1.2 eq.) in a fast manner to prevent freezing. The reaction mixture was then stirred for 15 mins before 4-phenyl-1-butanol (10.0 g, 66.7 mmol) was added dropwise over a period of 45 minutes. The reaction was then stirred at -78°C for 3 hours before the addition of Et_3N (25 ml) and stirred for a further 30 mins whilst being

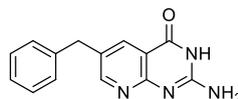
allowed to warm to room temperature. The reaction was diluted with Et₂O (100 ml) and washed with saturated NH₄Cl solution (200 ml). The organic layer was isolated and the aqueous was washed with Et₂O (3 x 100 ml) before combining the organic extracts, drying over MgSO₄ and concentrating to give a dark yellow oil (9.20 g, 62.2 mmol, 93.2%) which was confirmed as containing 4-phenylbutanal by proton NMR and mass spec and used in the next step without further purification.

Step 2: α -methenylation: The α -methenylation reaction was carried out as described for the synthesis of compound **51** with 4-(dimethylamino)benzoic acid (2.08 g, 12.6 mmol, 0.2 eq.), pyrrolidine (450 μ l, 6.1 mmol, 0.1 eq.), formaldehyde (37% in H₂O, 4.86 ml, 60.8 mmol, 1.0 eq), the crude product of Step 1 (9.0 g, 60.8 mmol) and DCM (80 ml). The crude product was purified in batches by flash chromatography (20:1 hexane/EtOAc) to yield a colourless oil (4.21 g, 39.4% over two steps). ¹H NMR (400 MHz, CDCl₃): δ = 9.58 ppm (s, 1H, H1); δ = 7.21-7.35 (m, 5H, Ar-H); δ = 6.23 (s, 1H, H3a); δ = 6.03 (s, 1H, H3b), δ = 2.81 (m, 2H, H5), δ = 2.61 (m, 2H, H4). IR: 2969, 2911, 1684, 1596, 1498 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + Na]⁺ = 183.1 found = 183.1.

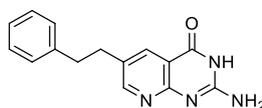
Synthesis of 2-amino-6-phenylpyrido[2,3-d]pyrimidin-4(3H)-one (**54**)



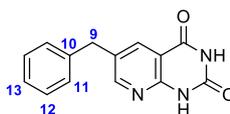
Compound **54** was prepared according to Procedure A with 2,4-diamino-6-hydroxypyrimidine (500 mg, 4.0 mmol), **51** (580 mg, 4.4 mmol) and AcOH (5 ml) to give product **54** (518 mg, 2.17 mmol 54.3%) as a pale yellow solid. m.p. > 300 °C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 12.24 ppm (bs, 1H, CONH); δ = 8.95 (d, J = 2.4 Hz, 1H, H7); δ = 8.40 (d, J = 2.4 Hz, 1H, H5); δ = 7.76 (d, J = 7.5 Hz, 2H, H9); δ = 7.50 (m, 2H, H10); δ = 7.16 (m, 1H, H11). ¹³C-NMR (100 MHz, D₂O/NaOD): δ = 174.53 (C8a), 164.17 (C2), 159.70 (C4), 151.33 (C7), 135.97 (C6), 132.85 (C5), 129.07 (C9), 128.78, 127.20, 125.52 (C10, C11, C12), 111.52 (C4a) ppm. IR: 3233, 3022, 1694, 1634, 1557, 1472 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 239.1 found = 239.2; calculated for [M + Na]⁺ = 261.2 found = 261.3. HRMS (ES, m/z): calculated for C₁₃H₁₀N₄O [M + H]⁺ = 239.0933, found = 239.0925.

Synthesis of 2-amino-6-benzylpyrido[2,3-d]pyrimidin-4(3H)-one (55)

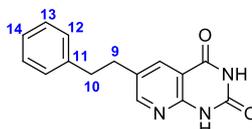
Compound **55** was prepared according to Procedure A with 2,4-diamino-6-hydroxypyrimidine (500 mg, 4.0 mmol), **52** (640 mg, 4.4 mmol) and AcOH (5 ml) to give product **54** (695 mg, 2.75 mmol, 68.8%) as a yellow solid. m.p. > 300°C (decomp.) ¹H NMR (400 MHz, DMSO): δ = 11.19 ppm (bs, 1H, CONH); δ = 8.59 (s, 1H, H7); δ = 8.00 (s, 1H, H5); δ = 7.10-7.35 (m, 5H, Ar-H); δ = 6.67 (bs, 2H, NH₂); δ = 4.00 (s, 2H, H9). ¹³C-NMR (100 MHz, TFA): 160.63 (C8a), 155.72 (C2), 152.58 (C4), 147.50 (C7), 144.43 (C5), 138.15, 135.02 (C6 + C10), 129.53, 129.26, 127.73 (C11, C12 + C13), 114.99 (C4a), 37.84 (C9). IR: 3205, 1672, 1585, 1529, 1485 cm⁻¹. ESI-MS (negative ion mode): calculated for [M - H]⁻ = 251.1 found = 251.2. HRMS (ES, m/z): calculated for C₁₄H₁₂N₄O [M + H]⁺ = 253.1089, found = 253.1079.

Synthesis of 2-amino-6-phenethylpyrido[2,3-d]pyrimidin-4(3H)-one (56)

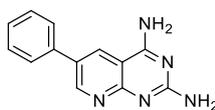
Compound **56** was prepared according to Procedure A with 2,4-diamino-6-hydroxypyrimidine (500 mg, 4.0 mmol), **53** (704 mg, 4.4 mmol) and AcOH (5 ml) to give product **56** (196 mg, 0.73 mmol, 18%) as a pale yellow solid. m.p. > 300°C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 8.46 (d, *J* = 2.0 Hz, 1H, H7); δ = 8.05 (d, *J* = 2.1 Hz, 1H, H5); δ = 7.22-7.29 (m, 5H, Ar-H); δ = 2.88-2.95 (m, 4H, H9 + H10). IR: 3020, 2923, 2823, 1733, 1670, 1619 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 267.1 found = 267.0. HRMS (ES, m/z): calculated for C₁₅H₁₄N₄O [M + H]⁺ = 267.1246, found = 267.1253.

Synthesis of 6-benzylpyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione (57)

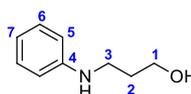
Compound **57** was prepared according to Procedure B with 4-amino-2,6-dihydroxypyrimidine (500 mg, 3.9 mmol), **52** (625 mg, 4.3 mmol) and TFA (5 ml) to give product **57** (259 mg, 1.02 mmol, 26.2%) as an off-white solid. m.p. > 300°C (decomp.) ¹H NMR (400 MHz, DMSO): δ = 11.63 ppm (s, 1H, CONH); δ = 11.43 (s, 1H, CONH); δ = 8.56 (d, *J* = 2.4 Hz, 1H, H7); δ = 8.06 (d, *J* = 2.5 Hz, 1H, H5); δ = 7.28–7.34 (m, 4H, H11 + H12); δ = 7.22 (m, 1H, H13); δ = 4.03 (s, 2H, H9). ¹³C-NMR (100 MHz, TFA): δ = 160.73 (C4), 150.12 (C8a), 147.34 (C2), 146.44 (C7), 145.59 (C5), 137.44, 136.62 (C6 + C10), 129.65, 129.27 (C11 + C12), 127.96 (C13), 114.07 (C4a), 37.80 (C9). IR: 3061, 1705, 1618, 1603, 1386 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 254.1 found = 254.1. HRMS (ES, *m/z*): calculated for C₁₄H₁₃N₃O₂ [M + H]⁺ = 254.0924, found = 254.0917

Synthesis of 6-phenethylpyrido[2,3-d]pyrimidine-2,4(1H,3H)-dione (58)

Compound **58** was prepared according to Procedure B with 4-amino-2,6-dihydroxypyrimidine (500 mg, 3.9 mmol), **53** (690 mg, 4.3 mmol) and TFA (5 ml) to give product **58** (310 mg, 1.16 mmol, 29.8%) as an off-white solid. m.p. > 300°C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 11.60 ppm (bs, 1H, CONH); δ = 11.43 (bs, 1H, CONH); δ = 8.43 (d, *J* = 2.4 Hz, 1H, H7); δ = 8.12 (d, *J* = 2.2 Hz, 1H, H5); δ = 7.16–7.30 (m, 5H, Ar-H); δ = 2.89–2.96 (m, 4H, H9 + H10). IR: 3237, 3024, 2917, 2822, 1672, 1599, 1564 cm⁻¹. HRMS (ES, *m/z*): calculated for C₁₅H₁₃N₃O₂ [M + H]⁺ = 268.1086, found = 268.1087.

Synthesis of 6-phenylpyrido[2,3-d]pyrimidine-2,4-diamine (59)

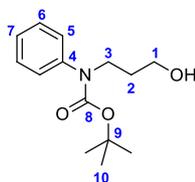
2,4,6-triaminopyrimidine (1750 mg, 14 mmol) was dissolved in AcOH (20 ml) and stirred under reflux at 120°C for 6 hours. The reaction was then cooled to room temperature and the solvent removed *in vacuo*. The resulting residue was dissolved in EtOH (20 ml) and the insoluble material was collected by centrifugation, washed with EtOH (3 x 20 ml) and Et₂O (3 x 20 ml) to yield a yellow solid (1.45 g, 6.1 mmol, 44%). ¹H NMR (400 MHz, DMSO): δ = 9.04 ppm (d, *J* = 2.4 Hz, 1H, H7); δ = 8.78 (d, *J* = 2.3 Hz, 1H, H5); δ = 7.88 (bs, 2H, NH₂); δ = 7.79 (d, *J* = 7.2 Hz, 2H, H10); δ = 7.52 (m, 2H, H11); δ = 7.40 (m, 1H, H12). ¹³C-NMR (100 MHz, TFA-*d*): δ = 162.75 (C2), 155.96 (C8a), 151.54 (C4), 149.31 (C7), 136.84 (C9/10), 136.18 (C5), 134.44 (C9/10), 130.17, 130.02, 127.32 (C10, C11 + C12), 107.19 (C4a) ppm. IR: 3025, 2937, 1673, 1600, 1563, 1486 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 238.1 found = 238.3. HRMS (ES, *m/z*): calculated for C₁₅H₁₁N₅ [M + H]⁺ = 238.1093, found = 238.1082.

Synthesis of 3-(phenylamino)propan-1-ol (62)¹⁶²

Iodobenzene (20.4 g, 0.1 mol), KOH (11.2 g, 0.2 mol) and CuCl (1 g, 10 mmol) were added to a round bottom flask and 3-aminopropanol (22.5 g, 0.3 mol) was added portion-wise at room temperature. The reaction mixture was stirred at room temperature for 12 hours and monitored by TLC (5:1 hex/EtOAc) for consumption of iodobenzene. On completion, the reaction mixture was diluted with water (20 ml) and extracted with ethyl acetate (3 x 50 ml), dried over MgSO₄ and the solvent removed *in vacuo*. The crude product was purified flash chromatography (5:1 hex/EtOAc) to yield a colourless oil (13.44 g, 89 mmol, 89%). (¹H NMR (400 MHz, CDCl₃): δ = 7.21 (t, *J* = 7.5 Hz, 2H, H6); δ = 6.75 (t, *J* = 7.5 Hz, 1H, H7); δ = 6.68 (d, *J* = 7.5 Hz, 2H, H5); δ = 3.82 (t, *J* = 5.9 Hz, 2H, H1); δ = 3.29 (t, *J* = 6.5 Hz, 2H, H1); δ = 1.86-1.92 (m, 2H, H2). ¹³C-NMR (100 MHz, CDCl₃): δ = 148.41 (C4), 129.32 (C6), 117.65 (C7), 113.19 (C5),

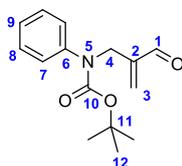
61.29 (C1), 41.83 (C3), 31.88 (C2) ppm. ESI-MS (positive ion mode): calculated for $[M + H]^+ = 152.1$ found = 152.1.

Synthesis of tert-butyl (3-hydroxypropyl)(phenyl)carbamate (**63**)¹⁸⁸



Boc₂O (15.7 g, 72 mmol) was added portion-wise to a stirred solution of **62** (10 g, 66 mmol) in EtOH (80 ml) and the reaction mixture was stirred at room temperature for 18 hours. Removal of the solvent yielded an oil, which was diluted with water (50 ml), extracted with EtOAc (3 x 50 ml), dried over MgSO₄ and concentrated *in vacuo* to yield a clear oil (15.08 g, 60.1 mmol, 91%). The so extracted compound was of acceptable purity and was carried through to the next step without any additional purification. ¹H NMR (400 MHz, CDCl₃): δ = 7.22-7.26 ppm (m, 2H, H₆); δ = 7.07-7.13 (m, 3H, H₅ + H₆); δ = 3.78 (bs, 1H, OH); δ = 3.68 (t, *J* = 6.5 Hz, 2H, H₃); δ = 3.54 (t, *J* = 6.0 Hz, 2H, H₁); δ = 1.62 (m, 2H, H₂); δ = 1.33 (s, 9H, H₁₀). ¹³C-NMR (100 MHz, CDCl₃): δ = 155.54 (C₈), 142.06 (C₄), 128.76 (C₆), 127.07, 126.22 (C₅, C₇), 80.34 (C₉), 58.84 (C₁), 46.44 (C₃), 30.94 (C₂), 28.16 (C₁₀) ppm. IR: 3432, 2978, 2936, 2871, 1673 cm⁻¹. ESI-MS (positive ion mode): calculated for $[M + H]^+ = 152.1$ found = 151.9.

Synthesis of tert-butyl (2-methylenebut-3-en-1-yl)(phenyl)carbamate (**67**)

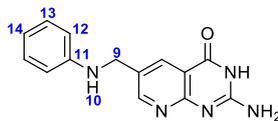


Step 1: Swern oxidation. To a stirred solution of (COCl)₂ (2.0 ml, 3.0 g, 24 mmol, 1.2 eq.) in dry DCM (30 ml) at -78°C under an atmosphere of N₂ was added DMSO (1.7 ml, 1.8 g, 24 mmol, 1.2 eq.) in a fast manner to prevent freezing. The reaction mixture was then stirred for 15 mins before **63** (5.0 g, 20 mmol) was added dropwise over a period of 15 minutes. The reaction was then stirred at -78 °C for 2 hours before the addition of Et₃N (10 ml) and stirred for a further 20 mins whilst being allowed to warm to room

temperature. The reaction was diluted with Et₂O and saturated NH₄Cl. The organic layer was isolated and the aqueous was washed with Et₂O (3 x 30 ml) before combining the organic extracts, drying over MgSO₄ and concentrating to give a yellow oil (4.3 g, 17.3 mmol, 86%) which was confirmed as tert-butyl (3-oxopropyl)(phenyl)carbamate by proton NMR and mass spec and used in the next step without further purification.

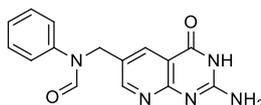
Step 2: α -methenylation: To a stirred suspension of 4-(dimethylamino)benzoic acid (532 mg, 4 mmol, 0.2 eq.) in DCM (20 ml) at room temperature was added dropwise pyrrolidine (132 μ l, 1.6 mmol, 0.1 eq.) until the suspension was solubilised. Formaldehyde (37% in H₂O, 1.28 ml, 16 mmol, 1.0 eq.) was added dropwise and the solution was stirred at room temperature for 15 mins. The previously prepared aldehyde (4.0 g, 16 mmol) was then added portion-wise as a solution in DCM (5 ml) and the reaction mixture was then stirred at room temperature for 1 h. A saturated NaHCO₃ solution (25 ml) was added and the mixture was extracted with DCM (3 x 25 ml). The combined organic extracts were washed with brine, dried over MgSO₄ and concentrated in vacuo to give a yellow oil, which was confirmed as the desired product by proton NMR and mass spec. The crude product was purified by flash chromatography (5:1, hex/EtOAc) to yield a colourless oil (3.03 g, 11.6 mmol, 58% over two steps). ¹H NMR (400 MHz, CDCl₃): δ = 9.54 ppm (s, 1H, CHO); δ = 7.35-7.60 (m, 2H, Ar-H); δ = 7.10-7.30 (m, 3H, Ar-H); δ = 6.33 (s, 1H, H3a); δ = 6.10 (s, 1H, H3b); δ = 4.43 (s, 2H, H4); δ = 1.72 (s, 9H, H12). ¹³C-NMR (100 MHz, CDCl₃): δ = 193.39 (C1), 154.41 (C10), 146.23 (C2), 142.57 (C6), 133.65 (C3), 128.71, 125.84, 125.75 (C7, C8, C9), 80.89 (C11), 48.13 (C4), 28.22 (C12). IR: 2968, 1685, 1596, 1499, 1476 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 262.1 found = 262.1, calculated for [M + Na]⁺ = 284.1 found = 284.1. HRMS (ES, m/z): calculated for C₁₅H₁₉NO₃ [M + Na]⁺ = 284.1263, found = 284.1270.

Synthesis of 2-amino-6-((phenylamino)methyl)pyrido[2,3-d]pyrimidin-4(3H)-one (68)



Compound **68** was prepared according to Procedure A with 2,4-diamino-6-hydroxypyrimidine (250 mg, 2.0 mmol), **67** (575 mg, 2.2 mmol) and AcOH (2.5 ml) to give product **68** (154 mg, 0.58 mmol 29.0%) as a yellow solid. m.p. > 300°C (decomp.). ¹H NMR (400 MHz, DMSO): δ = 11.25 ppm (bs, 1H, CONH); δ = 8.63 (s, 1H, H7); δ = 8.18 (s, 1H, H5); δ = 7.04 (t, *J* = 7.8 Hz, 2H, H13); δ = 6.70 (bs, 2H, NH₂); δ = 6.59 (d, *J* = 7.7 Hz, 2H, H12); δ = 6.52 (m, 1H, H14); δ = 6.31 (m, 1H, NH); δ = 4.30 (d, *J* = 6.2 Hz, 2H, H9). ¹³C-NMR (200 MHz, TFA): δ = 164.07 (C8a), 154.95 (C2), 149.88 (C4), 148.12 (C7), 140.71 (C5), 129.03, 127.79 (C6 + C11), 127.07 (C14), 126.60 (C13), 118.11 (C4a), 117.82 (C12), 59.45 (C9) ppm. IR: 2967, 1712, 1669, 1626, 1406, 1356 cm⁻¹. ESI-MS (positive ion mode): calculated for [M + H]⁺ = 268.1 found = 268.1. HRMS (ES, m/z): calculated for C₁₄H₁₃N₅O [M + H]⁺ = 268.1198, found = 268.1189.

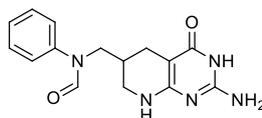
Synthesis of N-((2-amino-4-oxo-3,4-dihydropyrido[2,3-d]pyrimidin-6-yl)methyl)-N-phenylformamide (69)



68 (110 mg, 420 μmol) was dissolved in HCOOH (98%, 5 ml) and to this stirred solution was added Ac₂O (2.5 ml) and the resulting solution was stirred at room temperature and the reaction monitored by TLC for consumption of **68**. After 4 hours the reaction mixture was then poured onto 50 ml of cold Et₂O to effect precipitation of a pale yellow solid which was collected by centrifugation. The solid was then washed with water (3 x 10 ml), Me₂CO (3 x 10 ml), DCM (3 x 5 ml) and Et₂O (3 x 20 ml) and dried under high-vacuum overnight. The resulting crude compound was then recrystallised from DMF, further washed with water, MeOH, DCM and Et₂O and dried overnight under high-vacuum to yield a pale yellow solid (68 mg, 230 μmol, 54.9%). ¹H NMR (400 MHz, DMSO): δ = 11.28 ppm (bs, 1H, CONH); δ = 8.64 (s, 1H, H7); δ = 8.48 (s, 1H, CHO); δ = 8.02 (s, 1H, H5); δ = 7.22-7.41 (m, 5H, Ar-H); δ = 6.77 (bs, 2H, NH₂); δ = 5.09 (s, 2H, H9). ESI-MS (positive ion mode): calculated for [M + H]⁺ = 296.1 found

= 296.0. ^{13}C -NMR (200 MHz, TFA): δ = 164.40 (CHO), 162.04 (C8a) 155.71 (C2), 150.76 (C4), 148.25 (C7), 140.75 (C5), 134.15, 129.44 (C6 + C11), 126.39, 125.81, 125.46 (C12, C13 + C14), 120.55 (C4a), 42.56 (C9). IR: 3236, 3071, 2772, 1669, 1594, 1563. HRMS (ES, m/z): calculated for $\text{C}_{15}\text{H}_{13}\text{N}_5\text{O}_2$ $[\text{M} + \text{H}]^+ = 296.1147$, found = 296.1136.

Synthesis of N-((2-amino-4-oxo-3,4,5,6,7,8-hexahydropyrido[2,3-d]pyrimidin-6-yl)methyl)-N-phenylformamide (70, T0D)



TFA (10 ml) was de-oxygenated by purging with N_2 for 30 minutes, before PtO_2 (4 mg, 17 μmol , 0.1 eq.) was added and the mixture was hydrogenated under H_2 for 1 hour to activate the catalyst. Compound **69** (50 mg, 170 μmol) was added and the reaction mixture was hydrogenated for 24 hours (1 atm) and monitored by TLC ($\text{CHCl}_3/\text{MeOH}$ 5:1). Upon consumption of **69** the reaction was filtered through celite and the solvent was removed *in vacuo* to yield a pale yellow oil. Crystallisation was achieved by the repeated addition and evaporation (*in vacuo*) of Et_2O and the resulting off-white precipitate was collected by centrifugation, washed with Et_2O (3 x 25 ml) and MeOH (3 x 5 ml) and dried under a stream of N_2 . The crude reaction product was then purified by C18 reverse phase chromatography (10:1 $\text{H}_2\text{O}/\text{MeOH}$) to give an off-white solid (5 mg, 17 μmol , 10%). m.p. > 300°C (decomp.). ^1H NMR (800 MHz, DMSO): δ = 8.48 ppm (s, 1H, CHO); δ = 7.52 (bs, 2H, NH_2); δ = 7.40-7.47 (m, 4H, H12 + H13); δ = 7.30 (t, J = 7.0 Hz, 1H, H14); δ = 7.08 (bs, 1H, NH); δ = 3.85 (d, J = 6.6 Hz, 2H, H9); δ = 3.22 (d, J = 11.1 Hz, 1H, H7a); δ = 2.88 (dd, J_1 = 11.9 Hz, J_2 = 9.0 Hz, 1H, H7b); δ = 2.36 (d, J = 11.0 Hz, 1H, H5a); δ = 1.81-1.92 (m, 2H, H5b + H6). ^{13}C -NMR* (200 MHz, DMSO- d_6): δ = 162.62 (CHO); 140.61 (C_q), 129.60 (CH), 126.22 (C14), 123.16 (CH), 81.83 (C4a), 45.04 (C9), 42.73 (C7), 29.41 (C6), 22.17 (C5) ppm. IR: 3201, 2844, 2779, 1650, 1601, 1561, 1482. HRMS (ES, m/z): calculated for $\text{C}_{15}\text{H}_{17}\text{N}_5\text{O}_2$ $[\text{M} + \text{H}]^+ = 300.1461$, found = 300.1462. *Note 3x quaternary carbons missing from ^{13}C -NMR data. It was not possible to improve this with the amount of material available.

6.7 Methods - Buffers and Media Recipes

DEPC H₂O (2L)

1.0 ml DEPC

Add dH₂O to 2L and autoclave

5 x M9 Salts (1L)

Na₂HPO₄ 85.7 g

KH₂PO₄ 15.0 g

NH₄Cl 5.0 g

NaCl 2.5 g

Add dH₂O to 1 litre then autoclave

M9-cas Media (250 ml)

5 x M9 salts 50 ml

100 x trace metal solution 2.5 ml

Thiamine 10 mg/ml 2.5 ml

20% w/v glucose 5 ml 0.4% w/v

20% casamino acids 5 ml 0.4% w/v

1 M MgSO 250 µl 1 mM

1 M CaCl₂ 75 µl 300 µM

Add dH₂O to 250 ml then filter sterilise

1 M Tris Buffer pH 8.0 (1L)

Tris base 36.73 g

Tris-HCl 109.78 g

Adjust pH if necessary

Add dH₂O to 1L

Inoue transformation buffer (1L)

MnCl₂.4H₂O 10.88 g 55 mM

CaCl₂.2H₂O 2.20 g 15 mM

KCl 18.65 g 250 mM

PIPES (0.5 M, pH 6.7) 20 ml 10 mM

Add dH₂O to 1L

Cell lysis buffer pH 7.4 (100 ml)

Na ₂ HPO ₄ ·12H ₂ O	501 mg	
NaH ₂ PO ₄ ·2H ₂ O	78 mg	
NaCl	2.92 g	0.5 M
Sucrose	10 g	10% w/v

3 x protease inhibitor tablets

Add 100 mg lysozyme immediately prior to use

5 x Native Purification Buffer (200 ml)

NaH ₂ PO ₄	7.0 g	250 mM
NaCl	29.2 g	2.5 M

Adjust to pH 8.0 with NaOH

Bring final volume to 200 ml with dH₂O

3 M Imidazole pH 6.0 (100 ml)

Imidazole	20.6 g	3 M
NaH ₂ PO ₄	242 mg	
Na ₂ HPO ₄	35 mg	(20 mM total phosphate)
NaCl	2.93 g	500 mM

Adjust to pH 6.0 with dH₂O

5 x Native Wash Buffer (100 ml)

3 M Imidazole pH 6.0	1.67 ml	50 mM
5 x Native Purification Buffer	98.33 ml	

5 x Native Elution Buffer (100 ml)

3 M Imidazole pH 6.0	8.35 ml	250 mM
5 x Native Purification Buffer	91.65 ml	

5 x SDS Buffer (1L)

Tris base	15.1 g
Glycine	94 g
SDS	5 g

Add dH₂O to 1L

T7 RNAP Buffer pH 7.9 (2L)

Tris-HCl	22.54 g
----------	---------

Trizma base	6.89 g	100 mM
NaCl	23.38 g	200 mM
EDTA	1.49 g	2 mM
β -mercaptoethanol	5.59 ml	40 mM
Add DEPC H ₂ O to 2L		

NTPs 25 mM (10 ml)

250 μ mol of each NTP
 Add 50 mM Tris buffer to 8 ml
 Adjust pH with 3 M HCl
 Add 50 mM Tris Buffer to 10 ml
 Divide into 1.2 ml aliquots

10 x IVT Buffer (50 ml)

1 M Tris buffer pH 8.0	15 ml	300 mM
Triton X-100	50 μ l	0.1% v/v
Spermidine	267 μ l	34 mM
1 M DTT	5 ml	100 mM
Add DEPC H ₂ O to 50 ml		

0.5 M EDTA pH 8.0 (100 ml)

EDTA disodium salt	18.60 g	0.5 M
Adjust to pH 8.0 using NaOH (5M)		
Add dH ₂ O to 100 ml		

RNA FPLC Buffer (2L)

NaH ₂ PO ₄	1.64 g	
Na ₂ HPO ₄	1.69 g	(10 mM total phosphate)
NaCl	11.7 g	(100 mM)
Add dH ₂ O to 2L		

2 x Denaturing PAGE loading dye (10 ml)

90% Formamide	9 ml	
0.5 M EDTA	269 μ l	0.1%
Xylene cyanol	10 mg	0.1%
Bromphenol blue	10 mg	0.1%
DEPC H ₂ O to 10 ml		

ITC Buffer pH 7.5 (2L)

K HEPES	6.72 g	
HEPES	6.10 g	50 mM total
MgCl ₂ .6H ₂ O	4.07 g	20 mM
KCl	7.45 g	100 mM

6.8 Sequences of primers and *in vivo* expression constructs

6.8.1 *In vivo* expression assay construct sequences

Riboswitch-controlled expression constructs were cloned into a pACYCDuet-1 expression vector (Invitrogen) at the *EcoI* site using standard molecular biology techniques. The constructs contain the T7A1 promoter (black) the THF riboswitch (blue) and in the case of the chimeric riboswitches, the *metE* expression platform (red). The GFP coding region is shown in green.

6.8.1.1 *S. mutans* THF/*metE* chimera GFP expression construct

```
TTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCCAAAAAATTAAT
AACATTTTCTCTTAAGTAGATGATTTCGCGTTAAGTGTGTGTGAATGGGATGTCGTCACACAA
CGAAGCGAGAGCGCGGTGAATCATTGCATCCGTGAGAGAGGCAGTGTTTTACGTAGAAAA
GCCTCTTTCTCTCATGGGAAAGAGGCTTTTTGTTGTGAGAAAACCTCTTAGCAGCCTGTATC
CGCGGGTGAAAGAGAGTGTTTTACATATAAAGGAGGAGAAACAATGATCATGAGTAAAGGA
GAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGTA
CAAATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACCTACCCTTAAA
TTTATTTGCACTACTGGAAAACCTACCTGTTCCATGGCCAACACTTGTCACTACTTTGACTTAT
GGTGTTC AATGCTTTTCAAGATACCCAGATCACATGAAACAGCATGACTTTTTCAAGAGTGC
CATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGGAACTACAAG
ACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTAT
TGATTTAAAGAAGATGGAACATTCTTGACACAAATTGGAATACA ACTATAACTCACACA
ATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTAACTTCAA AATTAGACAC
AACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCG
ATGGCCCTGTCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTTT CGAAAGAT
CCCAACGAAnAGAGAGACCACATGGTCCTTCTTGAGTTTGTAAACAGCT
```

6.8.1.2 *S. mutans* THF wild-type GFP expression construct

```
TTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCAGTAGATGATT
CGCGTTAAGTGTGTGTGAATGGGATGTCGTCACACAACGAAGCGAGAGCGCGGTGAATCA
TTGCATCCGTGAGAGAGGCAGTGTTTTACGTAGAAAAGCCTCTTTCTCTCATGGGAAAGAG
GCTTTTTGTTGTGAGAAAACCTCTTAGCAGCCTGTATCCGCGGGTGAAAGAGAGTGTTTTA
CATATAAAGGAGGAGAACATCATGATCATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTG
TCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGTACAAATTTTCTGTGAGTGGAGAG
GGTGAAGGTGATGCAACATACGGAAAACCTACCCTTAAATTTATTTGCACTACTGGAAAACCT
ACCTGTTCCATGGCCAACACTTGTCACTACTTTGACTTATGGTGTTC AATGCTTTTCAAGAT
ACCCAGATCACATGAAACAGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACA
GGAAAGA ACTATATTTTTCAAAGATGACGGGAACTACAAGACACGTGCTGAAGTCAAGTTT
GAAGGTGATACCCTTGTTAATAGAATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAA
CATTCTTGACACAAATTGGAATACA ACTATAACTCACACAATGTATACATCATGGCAGACA
AACAAAAGAATGGAATCAAAGTAACTTCAA AATTAGACACAACATTGAAGATGGAAGCGTT
CAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCTTTTACCAGA
CAACCATTACCTGTCCACACAATCTGCCCTTTT CGAAAGATCCCAACGAAnAGAGAGACCAC
ATGGTCCTTCTTGAGTTTGTAAACAGCT
```

6.8.1.3 *L. casei* THF wild-type GFP expression construct

TTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCGGGGAGTAGGA
 ACCTGCGCGTTAAGTGCTGCCGGAACGGAATGTGAACGGCAGACGAAGACAAGTGTCGC
 GGTGCAGGTGCCGCATTGCCCCAGAATCATTGCATCCGTGAGAGAGGCAGTGTTTTACG
 TAGAAAAGCCTCTTTCTCTCATGGGAAAGAGGCTTTTTGTTGTGAGAAAACCTCTTAGCAGC
 CTGTATCCGCGGGTGAAAGAGAGTGTTTTACATATAAAGGAGGAGAACATCATGATCATGA
 GTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTT
 AATGGGTACAAATTTTCTGTCTAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACCTTA
 CCCTTAAATTTATTTGACTACTGGAAAACCTACCTGTTCCATGGCCAACACTTGTCACTACT
 TTGACTTATGGTGTTCATGCTTTTCAAGATACCCAGATCACATGAAACAGCATGACTTTTT
 CAAGAGTGCCATGCCCGAAGGTTATGTACAGGAAAGAACTATATTTTTCAAAGATGACGGG
 AACTACAAGACACGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATAGAATCGAGTT
 AAAAGGTATTGATTTTAAAGAAGATGGAACATTCTTGGACACAAATTGGAATACA ACTATA
 ACTCACACAATGTATACATCATGGCAGACAAACAAAAGAATGGAATCAAAGTTAACTTCAA
 ATTAGACACAACATTGAAGATGGAAGCGTTCAACTAGCAGACCATTATCAACAAAATACTCC
 AATTGGCGATGGCCCTGTCTTTTACCAGACAACCATTACCTGTCCACACAATCTGCCCTTT
 CGAAAGATCCCAACGAAnAGAGAGACCACATGGTCTTCTTGAGTTTGTAAACAGCT

Table 5 - DNA templates used in *in vitro* transcription reactions. T7 promoter region is underlined for clarity.

Sequence	Description	Sequence (5'-3')
pRSW143	<i>S. mutans</i> wild-type aptamer	TGGAGCGGATGCAATGATTCACCGCGCTCTCGCTTCGTTGTGT GACGACATCCCATTACACACACTTAACGCGAATCATCTACTC <u>TCCTATAGTGAGTCGTATTA</u>
pRSW149	<i>S. mutans</i> M1 aptamer (U25C/C53U)	TGGAGCGGATGCAATGATTCACCGCGCTCTCGCTTCATTGTGT GACGACATCCCATTACACACGCTTAACGCGAATCATCTACTC <u>TCCTATAGTGAGTCGTATTA</u>
pRSW151	<i>S. mutans</i> M2 aptamer (U35C/U42C)	TGGAGCGGATGCAATGATTCACCGCGCTCTCGCTTCGTTGTGT GACGGCATCCCGTTACACACACTTAACGCGAATCATCTACTC <u>TCCTATAGTGAGTCGTATTA</u>
pJL023	<i>L. casei</i> wild-type aptamer (GG-overhang)	CAGGCGAATGCGGCACCTGCACCGCGACACTTGTCTTCGTCT GCCGTTACATTCCGTTCCGGCAGCACTTAACGCGCAGGTTT CTACTCCTACCTATAGTGAGTCGTATTA
pJL024	<i>L. casei</i> wild-type aptamer (GC-stem replacement)	TGGGGCGAATGCGGCACCTGCACCGCGACACTTGTCTTCGTCT TGCCGTTACATTCCGTTCCGGCAGCACTTAACGCGCAGGTTT CTACTCCCTATAGTGAGTCGTATTA
pJL030	<i>L. casei</i> M1 aptamer (U26C/C55U; GG-overhang)	CAGGCGAATGCGGCACCTGCACCGCGACACTTGTCTTCATCT GCCGTTACATTCCGTTCCGGCAGCGCTTAACGCGCAGGTTT CTACTCCTACCTATAGTGAGTCGTATTA
pCJR032	Universal primer	TAATACGACTCACTATA

Table 6 - Sequence of construct used in *in vitro* transcription assay and primers used for its creation. The full construct sequence contains the T7A1 promoter (green), the *foI* THF aptamer domain (blue) and the *metE* expression platform (red).

Primer	Sequence
Full construct	GGTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCAAAAAAT TAATAACATTTTCTCTTAAGTAGATGATTCGCGTTAAGTGTGTGTGAATGGGATGTCGTCA CACAAACGAAGCGAGAGCGCGGTGAATCATTGCATCCGTGAGAGAGGCAGTGTTTTACGT AGAAAAGCCTCTTTCTCTCATGGGAAAGAGGCTTTTTGTGTGAGAAAACCTCTTAGCAG CCTGTATCCGCGGGTGAAAGAGAGTGTTTTACATATAAAGGAGGAGAAACAATG
1	TTAAGTGTGTGTGAATGGGATGTCGTCACACAACGAAGCGAGAGCGCGGTGAATCATTG
2	GAGGCTTTTCTACGTAAAACACTGCCTCTCTCACGGATGCAATGATTCACCGCGCTCTC
3	ATTAATAACATTTTCTCTTAAGTAGATGATTCGCGTTAAGTGTGTGTGAATGGGATGTC
4	TCTCACAAACAAAAAGCCTCTTTCCCATGAGAGAAAGAGGCCTTTTCTACGTAAAACACTG
5	CCTATAGGATACTTACAGCCAAAAAATTAATAACATTTTCTCTTAAGTAGATGATTCCG
6	TCTTTACCCGCGGATACAGGCTGCTAAGAGTTTTTCTCACAAACAAAAAGCCTCTTTCC
7	GGTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAGCCAAAA
8	CATTGTTTCTCCTCTTTATATGTAAAACACTCTCTTTACCCGCGGATACAG

7. References

1. Mortimer, S. A.; Kidwell, M. A.; Doudna, J. A., Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **2014**, *15* (7), 469-479.
2. Leontis, N. B.; Westhof, E., Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7* (4), 499-512.
3. Sundaralingam, M., Non-watson—crick base pairs in ribonucleic acids. *Int. J. Quant. Chem.* **1977**, *12* (S4), 11-23.
4. Svoboda, P.; Cara, A. D., Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.* **2006**, *63* (7-8), 901-908.
5. Spitale, R. C.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y., RNA structural analysis by evolving SHAPE chemistry. *RNA* **2014**, *5* (6), 867-881.
6. Regulski, E.; Breaker, R., In-Line Probing Analysis of Riboswitches. *Post-Transcriptional Gene Regulation*, Wilusz, J., Ed. Humana Press **2008**, 53-67.
7. Westhof, E., Twenty years of RNA crystallography. *RNA* **2015**, *21* (4), 486-487.
8. Clemons, W. M.; May, J. L. C.; Wimberly, B. T.; McCutcheon, J. P.; Capel, M. S.; Ramakrishnan, V., Structure of a bacterial 30S ribosomal subunit at 5.5Å resolution. *Nature* **1999**, *400* (6747), 833-840.
9. Spitale, R. C.; Crisalli, P.; Flynn, R. A.; Torre, E. A.; Kool, E. T.; Chang, H. Y., RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **2013**, *9* (1), 18-20.
10. Burgess, D. J., RNA: Detailed probing of RNA structure in vivo. *Nat. Rev. Genet.* **2015**, *16* (5), 255-255.
11. von Hippel, P. H., An Integrated Model of the Transcription Complex in Elongation, Termination, and Editing. *Science* **1998**, *281* (5377), 660-665.
12. Maston, G. A.; Evans, S. K.; Green, M. R., Transcriptional Regulatory Elements in the Human Genome. *Ann. Rev. Genom. Human Genet.* **2006**, *7* (1), 29-59.
13. Gralla, J. D., S. L. McKnight and K. R. Yamamoto. *Transcriptional Regulation*. Cold Spring Harbor Laboratory Press **1992**, 26-29.
14. Johnson, D. H. a. F. B., Synergism in transcriptional activation: a kinetic view. *Genes Dev.* **1993**, *7* (173), 88-95.
15. von Hippel, P. H.; Rees, W. A.; Rippe, K.; Wilson, K. S., Specificity mechanisms in the control of transcription. *Biophys. Chem.* **1996**, *59* (3), 231-246.
16. Hippel, T.; Wachsmuth, M., *E. coli and S. typhimurium: Cellular and Molecular Biology*. American Society for Microbiology, Washington DC: **1987**, 103-140
17. Schneider, T. D.; Stormo, G. D.; Gold, L.; Ehrenfeucht, A., Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **1986**, *188* (3), 415-431.
18. Berg, O. G.; von Hippel, P. H., Selection of DNA binding sites by regulatory proteins: Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* **1987**, *193* (4), 723-743.
19. von Hippel, P.H., *Biological Regulation and Development*. Plenum Press, New York, **1979**, 301-349.
20. von Hippel, P.H., *Transcription Regulation*. Cold Spring Harbor Laboratories **1992**, 179-201.
21. Landick, R., The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* **2006**, *34* (6), 1062-1066.
22. Martin, F. H.; Tinoco, I., DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res.* **1980**, *8* (10), 2295-2300.
23. Kaczanowska, M.; Rydén-Aulin, M., Ribosome Biogenesis and the Translation Process in Escherichia coli. *Microbiol. Mol. Biol. Rev.* **2007**, *71* (3), 477-494.

24. Kozak, M., Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **2005**, *361* (0), 13-37.
25. Pape, T.; Wintermeyer, W.; Rodnina, M. V., Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. *EMBO* **1998**, *17* (24), 7490-7497.
26. Garst, A. D.; Batey, R. T., A switch in time: Detailing the life of a riboswitch. *Biochim. Biophys. Acta* **2009**, *1789* (9-10), 584-591.
27. Winkler, W. C.; Breaker, R. R., Regulation of bacterial gene expression by riboswitches. In *Ann. Rev. Microbiol.* **2005**, *59*, 487-517.
28. Breaker, R. R., RNA Switches Out in the Cold. *Mol. Cell* **2010**, *37* (1), 1-2.
29. Breaker, R. R., Riboswitches and the RNA World. *Perspect. Biol.*, **2012**, *4* (2).
30. Roth, A.; Breaker, R. R., The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Ann. Rev. Biochem.*, **2009**; *78*, 305-334.
31. Borovok, I.; Gorovitz, B.; Schreiber, R.; Aharonowitz, Y.; Cohen, G., Coenzyme B12 controls transcription of the *Streptomyces* class Ia ribonucleotide reductase *nrdABS* operon via a riboswitch mechanism. *J. Bacteriol.* **2006**, *188* (7), 2512-2520.
32. Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J., Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. & Biol.* **2004**, *11* (12), 1729-1741.
33. Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R., A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **2004**, *306* (5694), 275-279.
34. Ames, T. D.; Breaker, R. R., Bacterial aptamers that selectively bind glutamine. *RNA Biol.* **2011**, *8* (1), 82-89.
35. Ferre-D'Amare, A. R., The *glmS* ribozyme: use of a small molecule coenzyme by a gene-regulatory RNA. *Quart. Rev. of Biophys.* **2010**, *43* (4), 423-447.
36. Cromie, M. J.; Shi, Y. X.; Latifi, T.; Groisman, E. A., An RNA sensor for intracellular Mg²⁺. *Cell* **2006**, *125* (1), 71-84.
37. Watson, P. Y.; Fedor, M. J., The *ydaO* motif is an ATP-sensing riboswitch in *Bacillus subtilis*. *Nat. Chem. Biol.* **2012**, *8* (12), 963-965.
38. Baker, J. L.; Sudarsan, N.; Weinberg, Z.; Roth, A.; Stockbridge, R. B.; Breaker, R. R., Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science* **2012**, *335* (6065), 233-235.
39. Mandal, M.; Breaker, R. R., Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 29-35.
40. Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R., Riboswitches Control Fundamental Biochemical Pathways in *Bacillus subtilis* and Other Bacteria. *Cell* **2003**, *113* (5), 577-586.
41. Miranda-Ríos, J.; Navarro, M.; Soberón, M., A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *PNAS* **2001**, *98* (17), 9736-9741.
42. Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R., An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.* **2003**, *10* (9), 701-707.
43. Corbino, K. A.; Barrick, J. E.; Lim, J.; Welz, R.; Tucker, B. J.; Puskarz, I.; Mandal, M.; Rudnick, N. D.; Breaker, R. R., Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. *Genome Biol.* **2005**, *6* (8).
44. Fuchs, R. T.; Grundy, F. J.; Henkin, T. M., The S-MK box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol.* **2006**, *13* (3), 226-233.

45. Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R., An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **2003**, *17* (21), 2688-2697.
46. Barrick, J. E.; Corbino, K. A.; Winkler, W. C.; Nahvi, A.; Mandal, M.; Collins, J.; Lee, M.; Roth, A.; Sudarsan, N.; Jona, I.; Wickiser, J. K.; Breaker, R. R., New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *PNAS* **2004**, *101* (17), 6421-6426.
47. Serganov, A.; Nudler, E., A Decade of Riboswitches. *Cell* **2012** *152* (1), 17-24.
48. Mandal, M.; Breaker, R. R., Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 29-35.
49. Kim, J. N.; Roth, A.; Breaker, R. R., Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *PNAS* **2007**, *104* (41), 16092-7.
50. Miranda-Rios, J.; Navarro, M.; Soberon, M., A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *PNAS* **2001**, *98* (17), 9736-41.
51. Winkler, W.; Nahvi, A.; Breaker, R. R., Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **2002**, *419* (6910), 952-956.
52. Nahvi, A.; Sudarsan, N.; Ebert, M. S.; Zou, X.; Brown, K. L.; Breaker, R. R., Genetic Control by a Metabolite Binding mRNA. *Chem. & Biol.* **2002**, *9* (9), 1043-1049.
53. Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R., An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.* **2003**, *10* (9), 701-7.
54. Fuchs, R. T.; Grundy, F. J.; Henkin, T. M., The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol.* **2006**, *13* (3), 226-33.
55. Weinberg, Z.; Regulski, E. E.; Hammond, M. C.; Barrick, J. E.; Yao, Z.; Ruzzo, W. L.; Breaker, R. R., The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA* **2008**, *14* (5), 822-828.
56. Poiata, E.; Meyer, M. M.; Ames, T. D.; Breaker, R. R., A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA* **2009**, *15* (11), 2046-2056.
57. Ames, T. D.; Rodionov, D. A.; Weinberg, Z.; Breaker, R. R., A Eubacterial Riboswitch Class That Senses the Coenzyme Tetrahydrofolate. *Chem. & Biol.* **2010**, *17* (7), 681-685.
58. Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R., Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **2003**, *113* (5), 577-86.
59. Rodionov, D. A.; Vitreschak, A. G.; Mironov, A. A.; Gelfand, M. S., Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res.* **2003**, *31* (23), 6748-57.
60. Sudarsan, N.; Wickiser, J. K.; Nakamura, S.; Ebert, M. S.; Breaker, R. R., An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **2003**, *17* (21), 2688-97.
61. Barrick, J. E.; Corbino, K. A.; Winkler, W. C.; Nahvi, A.; Mandal, M.; Collins, J.; Lee, M.; Roth, A.; Sudarsan, N.; Jona, I.; Wickiser, J. K.; Breaker, R. R., New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *PNAS* **2004**, *101* (17), 6421-6.
62. Lee, E. R.; Baker, J. L.; Weinberg, Z.; Sudarsan, N.; Breaker, R. R., An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. *Science* **2010**, *329* (5993), 845-848.
63. Nelson, J. W.; Sudarsan, N.; Furukawa, K.; Weinberg, Z.; Wang, J. X.; Breaker, R. R., Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat. Chem. Biol.* **2013**, *9* (12), 834-839.
64. Montange, R. K.; Batey, R. T., Riboswitches: Emerging themes in RNA structure and function. *Ann. Rev. Biophys.* **2008**, *37*, 117-133.

65. Klein, D. J.; Edwards, T. E.; Ferre-D'Amare, A. R., Cocrystal structure of a class I preQ(1) riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* **2009**, *16* (3), 343-344.
66. McDaniel, B. A. M.; Grundy, F. J.; Artsimovitch, I.; Henkin, T. M., Transcription termination control of the S box system: Direct measurement of S-adenosylmethionine by the leader RNA. *PNAS* **2003**, *100* (6), 3083-3088.
67. Wilson-Mitchell, S. N.; Grundy, F. J.; Henkin, T. M., Analysis of lysine recognition and specificity of the *Bacillus subtilis* L box riboswitch. *Nucleic Acids Res.* **2012**.
68. Haller, A.; Rieder, U.; Aigner, M.; Blanchard, S. C.; Micura, R., Conformational capture of the SAM-II riboswitch. *Nat. Chem. Biol.* **2011**, *7* (6), 393-400.
69. Neupane, K.; Yu, H.; Foster, D. A. N.; Wang, F.; Woodside, M. T., Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res.* **2011**, *39* (17), 7677-7687.
70. Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M., The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochem.* **2005**, *44* (40), 13404-13414.
71. Wunderlich, C. H.; Spitzer, R.; Santner, T.; Fauster, K.; Tollinger, M.; Kreutz, C., Synthesis of (6-C-13)Pyrimidine Nucleotides as Spin-Labels for RNA Dynamics. *JACS* **2012**, *134* (17), 7558-7569.
72. Greenleaf, W. J.; Frieda, K. L.; Foster, D. A.; Woodside, M. T.; Block, S. M., Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **2008**, *319* (5863), 630-633.
73. Ryals, J.; Little, R.; Bremer, H., Temperature dependence of RNA synthesis parameters in *Escherichia coli*. *J. Bacteriol.* **1982**, *151* (2), 879-887.
74. Sudarsan, N.; Lee, E. R.; Weinberg, Z.; Moy, R. H.; Kim, J. N.; Link, K. H.; Breaker, R. R., Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **2008**, *321* (5887), 411-413.
75. Lee, E. R.; Blount, K. F.; Breaker, R. R., Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol.* **2009**, *6* (2), 187-194.
76. Welz, R.; Breaker, R. R., Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA* **2007**, *13* (4), 573-582.
77. Kim, J. N.; Roth, A.; Breaker, R. R., Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *PNAS* **2007**, *104* (41), 16092-16097.
78. Lemay, J.-F.; Lafontaine, D. A., Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA* **2007**, *13* (3), 339-350.
79. Dixon, N.; Duncan, J. N.; Geerlings, T.; Dunstan, M. S.; McCarthy, J. E.; Leys, D.; Micklefield, J., Reengineering orthogonally selective riboswitches. *PNAS* **2010**, *107* (7), 2830-2835.
80. Dixon, N.; Robinson, C. J.; Geerlings, T.; Duncan, J. N.; Drummond, S. P.; Micklefield, J., Orthogonal Riboswitches for Tuneable Coexpression in Bacteria. *Angew. Chem.* **2012**, *51* (15), 3620-3624.
81. Robinson, C. J.; Vincent, H. A.; Wu, M.-C.; Lowe, P. T.; Dunstan, M. S.; Leys, D.; Micklefield, J., Modular Riboswitch Toolsets for Synthetic Genetic Control in Diverse Bacterial Species. *JACS* **2014**, *136* (30), 10615-10624.
82. Iwata-Reuyl, D., Biosynthesis of the 7-deazaguanosine hypermodified nucleosides of transfer RNA. *Bioorg. Chem.* **2003**, *31* (1), 24-43.
83. Reader, J. S.; Metzgar, D.; Schimmel, P.; de Crécy-Lagard, V., Identification of Four Genes Necessary for Biosynthesis of the Modified Nucleoside Queuosine. *J. Biol. Chem.* **2004**, *279* (8), 6280-6285.
84. Rieder, U.; Lang, K.; Kreutz, C.; Polacek, N.; Micura, R., Evidence for Pseudoknot Formation of Class I preQ1 Riboswitch Aptamers. *ChemBioChem* **2009**, *10* (7), 1141-1144.

85. Roth, A.; Winkler, W. C.; Regulski, E. E.; Lee, B. W. K.; Lim, J.; Jona, I.; Barrick, J. E.; Ritwik, A.; Kim, J. N.; Welz, R.; Iwata-Reuyl, D.; Breaker, R. R., A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **2007**, *14* (4), 308-317.
86. Wu, M.-C.; Lowe, P. T.; Robinson, C. J.; Vincent, H. A.; Dixon, N.; Leigh, J.; Micklefield, J., Rational Re-engineering of a Transcriptional Silencing PreQ1 Riboswitch. *JACS* **2015**, *137* (28), 9015-9021.
87. Sudarsan, N.; Hammond, M. C.; Block, K. F.; Welz, R.; Barrick, J. E.; Roth, A.; Breaker, R. R., Tandem riboswitch architectures exhibit complex gene control functions. *Science* **2006**, *314* (5797), 300-304.
88. Pejchal, R.; Ludwig, M. L., Cobalamin-independent methionine synthase (MetE): A face-to-face double barrel that evolved by gene duplication. *Plos Biology* **2005**, *3* (2), 254-265.
89. Keasling, J. D., Synthetic Biology for Synthetic Chemistry. *ACS Chem. Biol.* **2008**, *3* (1), 64-76.
90. Terpe, K., Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **2006**, *72* (2), 211-222.
91. Lutz, R.; Bujard, H., Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I-1-I-2 regulatory elements. *Nucleic Acids Res.* **1997**, *25* (6), 1203-1210.
92. Donovan, R. S.; Robinson, C. W.; Glick, B. R., Review: Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *J. Ind. Microbiol.* **1996**, *16* (3), 145-154.
93. Kærn, M.; Blake, W. J.; Collins, J. J., The Engineering of Gene Regulatory Networks. *Ann. Rev. Biomed. Eng.* **2003**, *5* (1), 179-206.
94. Adhya, S., *Regulation of Gene Expression in Escherichia coli*, Springer US **1996**, 181-200.
95. Roderick, S. L., The lac operon galactoside acetyltransferase. *C. R. Biol.* **2005**, *328* (6), 568-575.
96. Woolston, B. M.; Edgar, S.; Stephanopoulos, G., Metabolic Engineering: Past and Future. *Ann. Rev. Chem. Biomol. Eng.* **2013**, *4* (1), 259-288.
97. Wittmann, A.; Suess, B., Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* **2012**, *586* (15), 2076-2083.
98. Ellington, A. D.; Szostak, J. W., In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346* (6287), 818-822.
99. Tuerk, C.; Gold, L., Systematic evolution of ligands by exponential enrichment - RNA ligands to bacteriophage-T4 DNA-polymerase. *Science* **1990**, *249* (4968), 505-510.
100. Ceres, P.; Garst, A. D.; Marcano-Velázquez, J. G.; Batey, R. T., Modularity of Select Riboswitch Expression Platforms Enables Facile Engineering of Novel Genetic Regulatory Devices. *ACS Syn. Biol.* **2013**.
101. Lynch, S. A.; Gallivan, J. P., A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res.* **2009**, *37* (1), 184-192.
102. Stoddard, C. D.; Widmann, J.; Trausch, J. J.; Marcano-Velázquez, J. G.; Knight, R.; Batey, R. T., Nucleotides Adjacent to the Ligand-Binding Pocket are Linked to Activity Tuning in the Purine Riboswitch. *J. Mol. Biol.* **2013**, *425* (10), 1596-1611.
103. Stoddard, C. D.; Batey, R. T., Mix-and-match riboswitches. *ACS Chem. Biol.* **2006**, *1* (12), 751-754.
104. Trausch, J.; Ceres, P.; Reyes, Francis E.; Batey, R. T., The Structure of a Tetrahydrofolate-Sensing Riboswitch Reveals Two Ligand Binding Sites in a Single Aptamer. *Structure* **2011**, *19*(10), 1413-1423 .
105. Darmostuk, M.; Rimpelová, S.; Gbelcová, H.; Ruml, T., Current approaches in SELEX: An update to aptamer selection technology. *Biotechnol. Adv.* **2015**, *33*, 1141-1161

106. Kang, K.-N.; Lee, Y.-S., RNA Aptamers: A Review of Recent Trends and Applications. *Fut. Trend. Biotechnol.* **2013**, 131, 153-169.
107. Suess, B.; Fink, B.; Berens, C.; Stentz, R.; Hillen, W., A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.* **2004**, 32 (4), 1610-1614.
108. Jang, S.; Yang, J.; Seo, S. W.; Jung, G. Y., Riboselector: Riboswitch-Based Synthetic Selection Device to Expedite Evolution of Metabolite-Producing Microorganisms. *Meth. Enzymol.* **2015**; 550, 341-362.
109. Davidson, M. E.; Harbaugh, S. V.; Chushak, Y. G.; Stone, M. O.; Kelley-Loughnane, N., Development of a 2,4-Dinitrotoluene-Responsive Synthetic Riboswitch in E. coli Cells. *ACS Chem.Biol.* **2013**, 8 (1), 234-241.
110. Trausch, Jeremiah J.; Batey, Robert T., A Disconnect between High-Affinity Binding and Efficient Regulation by Antifolates and Purines in the Tetrahydrofolate Riboswitch. *Chem. & Biol.* **2014**, 21 (2), 205-216.
111. Xayaphoummine, A.; Viasnoff, V.; Harlepp, S.; Isambert, H., Encoding folding paths of RNA switches. *Nucleic Acids Res.* **2007**, 35 (2), 614-622.
112. Fontecave, M.; Atta, M.; Mulliez, E., S-adenosylmethionine: nothing goes to waste. *Trend. Biochem. Sci.* **2004**, 29 (5), 243-249.
113. Sekowska, A.; Kung, H. F.; Danchin, A., Sulfur metabolism in Escherichia coli and related bacteria: Facts and fiction. *J. Mol. Microbiol. Biotechnol.* **2000**, 2 (2), 145-177.
114. Barrick, J. E.; Breaker, R. R., The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **2007**, 8 (11).
115. Epshtein, V.; Mironov, A. S.; Nudler, E., The riboswitch-mediated control of sulfur metabolism in bacteria. *PNAS* **2003**, 100 (9), 5052-5056.
116. Batey, R. T., Recognition of S-adenosylmethionine by riboswitches. *RNA* **2011**, 2 (2), 299-311.
117. Gilbert, S. D.; Rambo, R. P.; Van Tyne, D.; Batey, R. T., Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* **2008**, 15 (2), 177-182.
118. Andre, G.; Even, S.; Putzer, H.; Burguiere, P.; Croux, C.; Danchin, A.; Martin-Verstraete, I.; Soutourina, O., S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of Clostridium acetobutylicum. *Nucleic Acids Res.* **2008**, 36 (18), 5955-5969.
119. Heppell, B.; Blouin, S.; Dussault, A. M.; Mulhbach, J.; Ennifar, E.; Penedo, J. C.; Lafontaine, D. A., Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat. Chem. Biol.* **2011**, 7 (6), 384-392.
120. Wachsmuth, M.; Findeiss, S.; Weissheimer, N.; Stadler, P. F.; Morl, M., De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res.* **2013**, 41 (4), 2541-2551.
121. Sharma, V.; Nomura, Y.; Yokobayashi, Y., Engineering Complex Riboswitch Regulation by Dual Genetic Selection. *JACS* **2008**, 130 (48), 16310-16315.
122. Muranaka, N.; Yokobayashi, Y., A synthetic riboswitch with chemical band-pass response. *Chem. Comm.* **2010**, 46 (36), 6825-6827.
123. Rieder, R.; Lang, K.; Graber, D.; Micura, R., Ligand-Induced Folding of the Adenosine Deaminase A-Riboswitch and Implications on Riboswitch Translational Control. *Chem. Bio. Chem.* **2007**, 8 (8), 896-902.
124. de Crecy-Lagard, V.; El Yacoubi, B.; de la Garza, R. D.; Noiriél, A.; Hanson, A. D., Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC Genom.* **2007**, 8.
125. Schirch, V.; Strong, W. B., Interaction of folylpolglutamates with enzymes in one-carbon metabolism. *Arch. Biochem. Biophys.* **1989**, 269 (2), 371-380.
126. Herrmann, K. M., The shikimate pathway as an entry to aromatic secondary metabolism. *Plant Physiol.* **1995**, 107 (1), 7-12.

127. Cossins, E. A.; Chen, L. F., Foliates and one-carbon metabolism in plants and fungi. *Phytochem.* **1997**, *45* (3), 437-452.
128. Lucock, M., Folic acid: Nutritional biochemistry, molecular biology, and role in disease processes. *Molec. Genet. Metabol.* **2000**, *71* (1-2), 121-138.
129. Fox, J. T.; Stover, P. J., Folate-Mediated One-Carbon Metabolism. *Vit. & Hormone.*, **2008**; *79*, 1-44.
130. Hartman, S. C.; Buchanan, J. M., Nucleic Acids, Purines, Pyrimidines (Nucleotide Synthesis). *Ann. Rev. Biochem.* **1959**, *28* (1), 365-410.
131. Umbarger, H. E., Amino Acid Biosynthesis and its Regulation. *Ann. Rev. Biochem.* **1978**, *47* (1), 533-606.
132. Reed, L. S.; Archer, M. C., Oxidation of tetrahydrofolic acid by air. *J. Agri. Food Chem.* **1980**, *28* (4), 801-805.
133. Hill, A., The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *J. Physiol.* **1910**, *40*, 1-7.
134. Buckstein, M. H.; He, J.; Rubin, H., Characterization of Nucleotide Pools as a Function of Physiological State in *Escherichia coli*. *J. Bacteriol.* **2008**, *190* (2), 718-726.
135. Bochner, B. R.; Ames, B. N., Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography. *J. Biol. Chem.* **1982**, *257* (16), 9759-9769.
136. Gangjee, A. J., Hiteshkumar; Kurup, Sonal, Recent Advances in Classical and Non-Classical Antifolates as Antitumor and Antiopportunistic Infection Agents: Part I. *Anti-Cancer Agents Med. Chem.* **2007**, *7* (5), 524-542.
137. Schweitzer, B. I.; Dicker, A. P.; Bertino, J. R., Dihydrofolate reductase as a therapeutic target. *FASEB* **1990**, *4* (8), 2441-52.
138. Gangjee, A.; Shi, J.; Queener, S. F.; Barrows, L. R.; Kisliuk, R. L., Synthesis of 5-methyl-5-deaza nonclassical antifolates as inhibitors of dihydrofolate reductases and as potential antipneumocystis, antitoxoplasma, and antitumor agents. *J. Med. Chem.* **1993**, *36* (22), 3437-3443.
139. DeGraw, J. I.; Christie, P. H.; Kisliuk, R. L.; Gaumont, Y.; Sirotnak, F. M., Synthesis and antifolate properties of 10-alkyl-5,10-dideaza analogs of methotrexate and tetrahydrofolic acid. *J. Med Chem.* **1990**, *33* (2), 673-677.
140. Gangjee, A.; Adair, O. O.; Queener, S. F., Synthesis and Biological Evaluation of 2,4-Diamino-6-(arylaminoethyl)pyrido[2,3-d]pyrimidines as Inhibitors of *Pneumocystis carinii* and *Toxoplasma gondii* Dihydrofolate Reductase and as Antiopportunistic Infection and Antitumor Agents. *J. Med. Chem.* **2003**, *46* (23), 5074-5082.
141. Temple, C.; Elliott, R. D.; Montgomery, J. A., Pyrido[2,3-d]pyrimidines. The synthesis of the 5-deaza analogs of aminopterin, methotrexate, folic acid, and N10-methylfolic acid. *The J. Org. Chem.* **1982**, *47* (5), 761-764.
142. Piper, J. R.; McCaleb, G. S.; Montgomery, J. A.; Kisliuk, R. L.; Gaumont, Y.; Sirotnak, F. M., Syntheses and antifolate activity of 5-methyl-5-deaza analogs of aminopterin, methotrexate, folic acid, and N10-methylfolic acid. *J. Med. Chem.* **1986**, *29* (6), 1080-1087.
143. Yang, Y.; Coward, J. K., Synthesis of p-Aminophenyl Aryl H-Phosphinic Acids and Esters via Cross-Coupling Reactions: Elaboration to Phosphinic Acid Pseudopeptide Analogues of Pteroyl Glutamic Acid and Related Antifolates. *J. Org. Chem.* **2007**, *72* (15), 5748-5758.
144. Rosowsky, A.; Forsch, R. A.; Null, A.; Moran, R. G., 5-Deazafolate Analogues with a Rotationally Restricted Glutamate or Ornithine Side Chain: Synthesis and Binding Interaction with Folylpolylglutamate Synthetase. *J. Med. Chem.* **1999**, *42* (18), 3510-3519.
145. Graffner-Nordberg, M.; Kolmodin, K.; Åqvist, J.; Queener, S. F.; Hallberg, A., Design, Synthesis, Computational Prediction, and Biological Evaluation of Ester Soft Drugs as Inhibitors of Dihydrofolate Reductase from *Pneumocystis carinii*. *J. Med Chem.* **2001**, *44* (15), 2391-2402.
146. Kuwada T, H. K., Nobuhiro J, A New Synthesis of 6-Substituted Pyrido[2,3-d]pyrimidines. *Heterocycles* **2002**, *57* (11), 2081-2090.

147. Stark, E.; Breitmaier, E., 5-Desazapteridine, synthese und NMR-spektroskopie. *Tetrahedron* **1973**, *29* (14), 2209-2217.
148. Troschütz, R.; Dennstedt, T., Verbesserte Synthese von methyl- und phenylsubstituierten Pyrido[2,3-d]pyrimidin-2,4-diaminen. *Archiv. Pharm.* **1994**, *327* (4), 221-224.
149. Teixido, J.; Borrell, J. I.; Colominas, C.; Deupi, X.; Matallana, J. L.; Falco, J. L.; Martinez-Teipel, B., Selective hydrolysis of 2,4-diaminopyrimidine systems: A theoretical and experimental insight into an old rule. *J. Org. Chem.* **2001**, *66* (1), 192-199.
150. Palanki, M. S. S.; Dneprovskaja, E.; Doukas, J.; Fine, R. M.; Hood, J.; Kang, X.; Lohse, D.; Martin, M.; Noronha, G.; Soll, R. M.; Wrasidlo, W.; Yee, S.; Zhu, H., Discovery of 3,3-(2,4-Diaminopteridine-6,7-diyl)diphenol as an isozyme-selective inhibitor of PI3K for the treatment of ischemia reperfusion injury associated with myocardial infarction. *J. Org. Chem.* **2007**, *50* (18), 4279-4294.
151. Konrad, G.; Pfeleider, W., Synthesis and properties of 2,4-diaminopteridines and their 5,6,7,8-tetrahydroderivatives. *Chem. Ber.* **1970**, *103* (3), 722-728.
152. Kenyon, C. O., LC; Van Der Westhuyzen, CW Modulation of phosphoryl transferase activity of glutamine synthetase. **2007**, WO/2007/105023.
153. Albert, A.; Matsuura, S., 415. Pteridine studies. Part XVIII. The reduction of hydroxypteridines. *J. Chem. Soc.* **1962**, (0), 2162-2171.
154. Pfeleiderer, W., Pteridine, XLIII. Synthese und Eigenschaften von 5.6.7.8-Tetrahydrolumazinen und ihren 5-Acetyl-Derivaten. *Just. Lieb. Ann. Chem.* **1971**, *747* (1), 111-122.
155. Vankayalapati, X L., Pyrimidine-2,4-diamine derivatives and their use as jak2 kinase inhibitors. **2008**, WO/2008/106645A1
156. Pfeleiderer, M.; Pfeleiderer, W., Synthesis and properties of pyrido[2,3-d]pyrimidine-2,4-diones (5-deazalumazines). *Heterocycles* **1992**, *33* (2), 905-929.
157. Li, P. F.; Zhan, C. L.; Zhang, S. L.; Ding, X. L.; Guo, F. Q.; He, S. G.; Yao, J. N., Alkali metal cations control over nucleophilic substitutions on aromatic fused pyrimidine-2,4-1H,3H-diones: towards new PNA monomers. *Tetrahedron* **2012**, *68* (43), 8908-8915.
158. Kubota, K. Novel bi-cyclic pyrimidine derivatives having antagonistic activity on H4 receptor. **2009**, WO/2009/107767A1
159. Mont, N.; Teixido, J.; Borrell, J. I.; Kappe, C. O., A three-component synthesis of pyrido 2,3-d pyrimidines. *Tet. Lett.* **2003**, *44* (29), 5385-5387.
160. Erkkilä, A.; Pihko, P. M., Rapid Organocatalytic Aldehyde-Aldehyde Condensation Reactions. *Eur. J. Org. Chem.* **2007**, *2007* (25), 4205-4216.
161. Davies, H. M. L.; Dai, X., Lewis Acid-Catalyzed Tandem Diels–Alder Reaction/Retro-Claisen Rearrangement as an Equivalent of the Inverse Electron Demand Hetero Diels–Alder Reaction. *J. Org. Chem.* **2005**, *70* (17), 6680-6684.
162. Yin, H.; Jin, M.; Chen, W.; Chen, C.; Zheng, L.; Wei, P.; Han, S., Solvent-free copper-catalyzed N-arylation of amino alcohols and diamines with aryl halides. *Tet. Lett.* **2012**, *53* (10), 1265-1270.
163. Singh, S. K.; Singer, S. C.; Ferone, R.; Waters, K. A.; Mullin, R. J.; Hynes, J. B., Synthesis and biological evaluation of N.alpha.-(5-deaza-5,6,7,8-tetrahydropteroyl)-L-ornithine. *J. Med. Chem.* **1992**, *35* (11), 2002-2006.
164. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31* (13), 3406-3415.
165. Klaus, S. M. J.; Kunji, E. R. S.; Bozzo, G. G.; Noiriél, A.; de la Garza, R. D.; Basset, G. J. C.; Ravel, S.; Rébeillé, F.; Gregory, J. F.; Hanson, A. D., Higher Plant Plastids and Cyanobacteria Have Folate Carriers Related to Those of Trypanosomatids. *J. Biol. Chem.* **2005**, *280* (46), 38457-38463.

166. Andersen, J. B.; Sternberg, C.; Poulsen, L. K.; Bjørn, S. P.; Givskov, M.; Molin, S., New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. *App. Environ. Microbiol.* **1998**, *64* (6), 2240-2246.
167. McDowell, J. C.; Roberts, J. W.; Jin, D. J.; Gross, C., Determination of intrinsic transcription termination efficiency by RNA polymerase elongation rate. *Science* **1994**, *266* (5186), 822-825.
168. Cho, B.-K.; Federowicz, S. A.; Embree, M.; Park, Y.-S.; Kim, D.; Palsson, B. Ø., The PurR regulon in Escherichia coli K-12 MG1655. *Nucleic Acids Res.* **2011**, *39* (15), 6456-6464.
169. Lorenz, R.; Bernhart, S. H.; Höner zu Siederdisen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; Hofacker, I. L., ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26-26.
170. Shine, J.; Dalgarno, L., Determinant of cistron specificity in bacterial ribosomes. *Nature* **1975**, *254* (5495), 34-38.
171. Bennett, B. D.; Kimball, E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D., Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat. Chem. Biol.* **2009**, *5* (8), 593-599.
172. Schirch, V.; Hopkins, S.; Villar, E.; Angelaccio, S., Serine hydroxymethyltransferase from Escherichia coli: purification and properties. *J. Bacteriol.* **1985**, *163* (1), 1-7.
173. Huang, L.; Ishibe-Murakami, S.; Patel, D. J.; Serganov, A., Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch. *PNAS* **2011**, *108* (36), 14801-6.
174. Salim, N. N.; Feig, A. L., Isothermal titration calorimetry of RNA. *Methods* **2009**, *47* (3), 198-205.
175. Freyer, M. W.; Lewis, E. A., Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. *Methods Cell Biol.* **2008**, *84*, 79-113.
176. Gilbert, S. D.; Mediatore, S. J.; Batey, R. T., Modified Pyrimidines Specifically Bind the Purine Riboswitch. *JACS* **2006**, *128* (44), 14214-14215.
177. Pikovskaya, O.; Polonskaia, A.; Patel, D. J.; Serganov, A., Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat. Chem. Biol.* **2011**, *7* (10), 748-755.
178. Wilson, R. C.; Smith, A. M.; Fuchs, R. T.; Kleckner, I. R.; Henkin, T. M.; Foster, M. P., Tuning Riboswitch Regulation through Conformational Selection. *J. Mol. Biol.* **2011**, *405* (4), 926-938.
179. Kulshina, N.; Edwards, T. E.; Ferré-D'Amaré, A. R., Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. *RNA* **2010**, *16* (1), 186-196.
180. Artsimovitch, I.; Henkin, T. M., In vitro approaches to analysis of transcription termination. *Methods* **2009**, *47* (1), 37-43.
181. Blouin, S.; Lafontaine, D. A., A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA* **2007**, *13* (8), 1256-1267.
182. Blouin, S.; Chinnappan, R.; Lafontaine, D. A., Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation. *Nucleic Acids Res.* **2011**, *39* (8), 3373-3387.
183. Sudarsan, N.; Cohen-Chalamish, S.; Nakamura, S.; Emilsson, G. M.; Breaker, R. R., Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem. & Biol.* **2005**, *12* (12), 1325-1335.
184. Gurevich, V. V., Use of bacteriophage RNA polymerase in RNA synthesis. *Vir. Poly. Rel. Prot.* **1996**, *275*, 382-397.

185. Gao, X. X.; Yo, P.; Keith, A.; Ragan, T. J.; Harris, T. K., Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res.* **2003**, *31* (22), 901-904
186. Hoover, D. M.; Lubkowski, J., DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **2002**, *30* (10), 43-46.
187. Inoue, H.; Nojima, H.; Okayama, H., High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **1990**, *96* (1), 23-28.
188. Vilaivan, T., A rate enhancement of tert-butoxycarbonylation of aromatic amines with Boc2O in alcoholic solvents. *Tet. Lett.* **2006**, *47* (38), 6739-6742.