TITLE: Role of riboswitches in gene regulation and their potential for algal biotechnology

RUNNING TITLE: Algal riboswitches

AUTHORS: Ginnie T.D.T. Nguyen, Mark A. Scaife, Katherine E. Helliwell, Alison G. Smith

AUTHOR ADDRESSES: Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA, UK.

***CORRESPONDING AUTHOR:**

Alison G. Smith, Tel: +44 1223 333952, email: as25@cam.ac.uk

KEY WORDS: riboswitches; microalgae; metabolism; biotechnology

TOTAL WORD COUNT (including references):

Abstract

Riboswitches are regulatory elements in messenger RNA to which specific ligands can bind directly in the absence of proteins. Ligand binding alters the mRNA secondary structure, thereby affecting expression of the encoded protein. Riboswitches are widespread in prokaryotes, with over 20 different effector ligands known, including amino acids, cofactors and Mg^{2+} ions, and gene expression is generally regulated by affecting translation or termination of transcription. In plants, fungi and microalgae, riboswitches have been found, but only those bind thiamine pyrophosphate. These eukaryotic riboswitches operate through alternative splicing of the transcript, a highly conserved process. Here we review the current status of riboswitch research with specific emphasis on microalgae. We discuss new riboswitch discoveries and insights into the underlying mechanism of action, and how next generation sequencing technology provides the motivation and opportunity to improve our understanding of these rare but important regulatory elements. We also highlight the potential of microalgal riboswitches as a tool for synthetic biology and industrial biotechnology.

Introduction

Riboswitches are regulatory elements in messenger RNA that function through binding of a ligand, typically an endogenous metabolite, in the absence of protein factors. The binding induces alteration in the mRNA secondary structure, which then affects gene expression (Mironov et al. 2002; Winkler 2005). Riboswitches generally regulate genes involved in the metabolism of the ligand, and respond to levels of the metabolite in the environment, but the term 'riboswitch' has now been expanded to include RNA-based regulators that sense temperature change, as well as those that bind tRNAs or small metal ions (Serganov & Patel 2007; Furukawa et al. 2015; Dambach et al. 2015). Riboswitches consist of an aptamer domain to which the metabolite binds, and an expression platform, which affects the translation of the mRNA (Figure 1). Since these are downstream in the same mRNA, they act in *cis*. So-called marooned riboswitches, located >200 nucleotides from a coding region, have also been found, which act *in trans* through small RNAs and sequestration of proteins (Mellin et al. 2014; DebRoy et al. 2014).

Riboswitches are abundant in viruses (Ooms et al. 2004) and prokaryotes (Winkler 2005). Ligands to riboswitches are generally compounds of primary metabolism including amino acids (glycine, lysine), nucleotides (adenine, guanine, or deoxyguanosine), and cofactors such as S-adenosylmethionine, S-adenosylhomocysteine, adenosylcobalamin (coenzyme B_{12}), flavin mononucleotide, and thiamine pyrophosphate (TPP) (Barrick & Breaker 2007), and the genes that are regulated encode proteins involved in the biosynthesis, transport or utilisation of the ligand. Riboswitches have not been found in animals, but they have been demonstrated in fungi, plants and algae, although in all cases only those that respond to TPP (Cheah et al. 2007; Wachter et al. 2007; Bocobza et al. 2007; Croft et al. 2007; Moulin et al. 2013; McRose et al. 2014). The first riboswitches in algae were found in *Chlamydomonas reinhardtii* and *Volvox carteri*, where they regulate expression of two thiamine biosynthesis

genes (Croft et al. 2007). Analysis of algal genome sequence data indicates that TPP riboswitches are present in all algal lineages, and their activity may extend beyond genes involved in thiamine metabolism (Worden et al. 2010; McRose et al. 2014). In this review we explain the mechanism of action of TPP riboswitches, discuss their distribution in algae and other organisms, and describe how they can be exploited for biotechnology.

Thiamine pyrophosphate riboswitches

Thiamine, vitamin B_1, is the precursor of TPP, an important cofactor for many enzymes of central metabolism. Most bacteria, algae and higher plants are able to produce thiamine *de novo,* whereas animals, including humans, cannot and instead they obtain it through their diet. In organisms that synthesise thiamine *de novo,* an external source down-regulates the biosynthetic pathway, and in many cases this is via one or more TPP riboswitches. In prokaryotes TPP riboswitches were first identified by searching for the presence of *thi*-box elements (essentially the TPP binding aptamer, and referred to hereafter as such) with conserved secondary structures (Miranda-Ríos et al. 2001). Pioneering research used site-directed mutagenesis to investigate the 5'-untranslated region (UTR) of the *Escherichia coli thiCOGE* and *thiMD* operons. The identified aptamer sequence was shown to be able to regulate the reporter, *LacZ,* in response to exogenous thiamine (Miranda-Ríos et al. 2001). Winkler *et al.* (2002) analysed secondary structures of the *E. coli thiC* and *thiM* 5' UTR and were able to demonstrate that there was considerable change in secondary structure upon binding TPP, such that a new stem formed, sequestering a portion of the ribosome-binding site or start codon, illustrated schematically in Figure 1. Bacterial TPP riboswitches have since been found upstream of thiamine-related operons containing biosynthetic (*thiC)*, salvage (*thiMD),* transport (*thiBPQ)*, and degradation (*tenA)* genes (Rodionov et al. 2002; Sudarsan et al. 2005). In gram-negative bacteria like *E. coli,* binding of TPP by the aptamer

usually impacts translation, whereas in gram-positive bacteria, e.g. *Bacillus subtilis*, premature transcription termination is commonly observed (Mironov et al. 2002).

TPP riboswitches in fungi such as *Neurospora crassa* and *Aspergillus oryzae*, have been found that regulate both thiamine biosynthetic (*THIA, NMT1*) (Kubodera et al. 2003; Sudarsan et al. 2003; Cheah et al. 2007) and transporter (*NCU01977*) genes (Li & Breaker 2013). Growth of fungal cells in the presence of exogenous thiamine results in alternatively spliced mRNAs, and inspection of the sequence of the introns containing riboswitches revealed that there are multiple GU splice sites that can combine with the constitutive AG splice site (Figure 2). For *NMT1* and *THIA* the TPP riboswitch is located in an intron of the 5' UTR (Cheah et al. 2007). Alternative-splicing of the *NMT1* mRNA in results in partial retention of the intron, which contains an upstream open reading frame (uORF). This is preferentially translated instead of the downstream NMT1 (Cheah et al. 2007). Similarly, in *A. oryzae* the riboswitch in *THIA* encoding the first enzyme of the thiazole branch of thiamine biosynthesis, is located in the second intron of the 5' UTR (Kubodera et al. 2003), and intron retention disrupts translation initiation. In contrast, the riboswitch-harbouring intron is within the coding region of the *NCU01977* pre-mRNA (Li & Breaker 2013), and retention of a partial intron introduces a premature stop codon, preventing the production of a full-length NCU01977 polypeptide (Li & Breaker 2013).

In the higher plant *Arabidopsis thaliana*, a single riboswitch has been identified, which is located in the second intron of the 3'UTR of the *THIC* gene, encoding the first enzyme of the pyrimidine branch of thiamine biosynthesis (Wachter et al. 2007; Bocobza et al. 2007) (Figure 2). Growth of seedlings in the presence of TPP results in splicing of this intron, removing the polyadenylation motif from the mature mRNA. This causes polyadenylation to occur at several sites in the mRNA, creating unstable mRNA and reducing protein expression. In the absence of TPP the second intron of the 3' UTR is not spliced, and stable

mRNA is produced. In further investigations into the TPP riboswitch in the plant kingdom, homologs of the *A. thaliana THIC* TPP riboswitch were found in the 3' UTR of *THIC* and/or *THI1* (equivalent to fungal *THIA*) in species ranging from bryophytes (basal/ancestral land plants) through to angiosperms (higher land plants) (Bocobza et al 2007; Watcher et al 2007). Although the expression platform of these riboswitches vary, the aptamer region is highly conserved, as evidenced by the very similar X-ray crystal structures of the *E. coli thiM* and *A. thaliana THIC* aptamers (Thore et al. 2006; Kulshina et al. 2010; Serganov et al. 2006; Noeske et al. 2006; Lang et al. 2007; Warner et al. 2014; Edwards & Ferré-D'Amaré 2006).

The first microalgal TPP riboswitches were identified as regions with high levels of sequence conservation between *C. reinhardtii* and *Volvox carteri* in non-coding regions of thiamine biosynthesis genes. Two TPP riboswitches were identified, one located in an intron within the 5' UTR of *THI4*, equivalent to fungal *THIA*, and the second in intron 6 of *THIC* (Croft et al. 2007). As in *THIA,* the *THI4* riboswitch contains multiple alternative GU and AG splice sites that may interact with constitutive GU and AG splice sites, located at the boundary of the riboswitch containing intron, to trigger alternative splicing (Figure 3). These longer transcripts retain an 81 bp uORF that interferes with translation of THI4. Alternative splicing of the *C. reinhardtii THIC* riboswitch causes the intron to be retained, a process that introduces a premature stop codon and causes the translation of a truncated protein (Croft et al. 2007; Moulin et al. 2013). This is equivalent to the mechanism of the TPP riboswitch in *NCU01977* (Figure 2).

In addition to TPP itself, the *C. reinhardtii* riboswitches also bind intermediates of the thiamine biosynthesis pathway (Moulin et al. 2013). The *THI4* 5' UTR containing the TPP riboswitch was shown to regulate reporter gene expression *in vivo* in response to both thiamine or the thiazole precursor, hydroxyethylthiazole (HET), when added to the growth medium. Similarly, the *THIC* riboswitch aptamer was observed to bind the pyrimidine

precursor hydroxymethylpyrimidine-pyrophosphate (HMP-PP) *in vitro*, and to cause alternative splicing in *in vivo* in cells grown with HMP (Moulin et al. 2013). This is the first evidence that heterocyclic thiamine precursors directly regulate endogenous gene expression via the TPP riboswitch, and suggests exquisite fine-tuning of the levels of this cofactor and its intermediates within the cell. Subsequently, studies of environmental samples have shown that there are sufficient levels of HMP to support the growth of the thiamine-dependent marine algae *E. huxleyi* (McRose et al. 2014), *Pavlova calceolate* and *P. lutheri* (Paerl et al. 2015).

Conservation of TPP riboswitches in algae

Microalgae have an extraordinarily complex evolutionary history, with different taxonomic groups being derived from distinct endosymbiotic events, followed by random reorganisation to eliminate or diversify duplicated genes and pathways (Dorrell & Smith 2011). In addition, horizontal gene transfer has shaped algal genomes and resultant physiology extensively (Qiu et al. 2013). Microalgae as a group are therefore extremely diverse, a fact that is clearly demonstrated by the phylogenetic distance between species of microalgae (Dorrell & Smith 2011), and through analysis of the number of unique and shared genes in those with sequenced genomes. For example, comparison of the green alga, *C. reinhardtii* (Merchant et al. 2007), the red alga, *Cyanidioschyzon merolae* (Matsuzaki et al. 2004), the diatom, *Phaeodactylum tricornutum* (Bowler et al. 2008), and the eustigmatophyte, *Nannochloropsis gaditana* (Radakovits et al. 2012) genomes revealed that although a subset of core genes are conserved, the vast majority are lineage specific. However, around 30% of identified genes are unique to individual species (Radakovits et al. 2012). In spite of this divergence, TPP riboswitches have been identified in all lineages of algae. Including the photosynthetic marine picoeukaryote *Micromonas*. In *Micromonas*, two lineages, which share 90% of their predicted genes, have distinct TPP riboswitch

arrangements (Worden et al. 2009). *Micromonas* RCC299 has three putative TPP riboswitches in the 3' UTR of *NMT1*, *FOLR-like* and *EFG-DC*, while *Micromonas* CCMP1545 contains putative TPP riboswitches present in the 5' and 3' UTR of *SSSF* and *SSSP*, but not in the *NMT1*, *FOLR-like* or *EFG-DC* genes (Worden et al. 2009). Subsequently, using publically available whole genome and transcriptome sequence data, as well as novel data sets, forty-three putative riboswitches were found in twenty five different algal species (Supplementary Table 1; McRose et al. 2014). The TPP riboswitches identified were associated with genes involved in thiamine biosynthesis (*THI4* or *THIC*), sodium: solute symporters (*SSS*) to transport metabolites (*SSSF, SSSP,* or *SSSQ)*, as well as genes that have not been previously linked to thiamine metabolism (*ATS1* and *UNK1*) (McRose et al. 2014). Analysis of the riboswitch aptamers demonstrated conservation of all the nucleotides previously identified as responsible for the interaction with the pyrimidine and pyrophosphate groups in bacterial and plant TPP aptamers. Moreover, of fifty-five structurally important nucleotides, forty-seven show nucleotide conservation >50% across the different species (Figure 4 and Supplementary Table 1). Although this is less than in higher plants, which are 80% similar, algal riboswitches are more conserved than those known in fungi (<40%) (Wachter et al. 2007), which is remarkable given the diversity of algal lineages. The exception to conserved regions is in the P3 stem, which is known as the variable loop and which encodes the uORF in the *THI4* and *NMT1* genes.

The expression of several genes containing TPP riboswitches was shown to be regulated by exogenous thiamine (McRose et al. 2014), but direct involvement of the riboswitches in this response remains to be confirmed. Based upon their gene context, many of the newly identified riboswitches are unlikely to affect splicing. This is said because those identified in *Micromonas, Ostreococcus, Aureococcus,* and *Fragilariopsis* species are located in genes containing a single intron (or none) and are thus unlikely to undergo alternative splicing.

Similarly, the proposed *SSSQ* riboswitch in *Guillardia theta* is split in one of two splice forms of the pre-mRNA, suggesting the riboswitch is not involved in the splicing process (McRose et al. 2014). If these riboswitches are indeed functional *in vivo*, they are perhaps more likely to regulate transcriptional or translational processes via mechanisms similar to those found in prokaryotes, such as physical inhibition of translation initiation (Worden et al. 2009; McRose et al. 2014).

Potential of riboswitches for algal biotechnology

Several species of microalgae demonstrate robust, well-characterised growth in defined culture conditions, and produce metabolites of commercial value. These properties mean that microalgae have considerable potential for industrial biotechnology, not just for the production of existing compounds, but also for metabolic engineering and as platforms for expression of novel proteins (Rasala & Mayfield 2011; Gangl et al. 2015; Scaife et al. 2015). For this potential to be realised it will be essential to develop robust molecular tools for the manipulation of algal genomes, and to regulate the expression of transgenes. Because algal biotechnology is in its infancy, now is an ideal opportunity to consider novel approaches to these questions, and to take advantage of the increasing number of genomic resources that are being developed for microalgae, both in terms of genome sequence information (Merchant et al. 2007; Bowler et al. 2008; Radakovits et al. 2012), and omics technologies (Reijnders et al. 2014; Zhang et al. 2014).

One aspect that will be key for successful metabolic engineering is to be able to express a transgene in a predictable fashion so as to modify metabolism reliably. At the same time, it is important to avoid pleiotropic and unexpected effects on endogenous metabolism. There is thus interest in identifying regulatory elements that can be used to modulate expression of the transgene, for example to induce it only once the culture is established, avoiding toxic effects, or undesirable allocation of metabolic substrates or reducing power when the cells are

growing rapidly. One approach to develop regulatory sequences is to recruit endogenous systems within the host. In microalgae such as *C. reinhardtii*, decades of research have elucidated several fundamental processes that maybe built upon to facilitate transgene expression (Scaife et al. 2015). Characterised promoters used for regulated transgene expression include those regulated by light (*PSAD;* Fischer & Rochaix 2001), nitrate (*NIT1;* Ohresser et al. 1997), copper (*CYC6*; Quinn & Merchant 1995), and vitamin B_{12}, (*METE;* Helliwell et al. 2014). The *PSAD* promoter has developed into one of the most commonly used and best characterised parts for the expression of transgenes in *C. reinhardtii*. In addition, the PSAD promoter has been shown to give the highest level nuclear transgene expression when compared directly with promoters of β-tubulin (*B2-TUB*; Davies et al. 1992), ferrireductase (*FEA1*; Allen et al. 2007), actin (Tang et al. 1995) and *HSP70A/RBCS2* (Schroda et al. 2000) (Kumar et al. 2013). Although much less advanced than work in *C. reinhardtii*, parallel studies in other microalgae are starting to be reported, which take advantage of experience and know-how from the former. For example, in *Nannochloropsis* sp. endogenous promoters such as violaxanthin/chlorophyll-binding proteins (*VCP1* and *VCP2*) (Kilian et al. 2011), β-tubulin (β-*TUB*), heat shock protein 70 (*HSP*) and the ubiquitin extension protein (*UEP*) (Radakovits et al. 2012) have been used, as well as α-tubulin (α-*TUB*), lipid droplet size protein (*LDSP*) and the heterologous 35S promoter (Vieler et al. 2012). However, for regulated gene expression promoters have several limitations, including the fact that the level of inducer/repressor (such as nitrate, or copper ions) can be difficult to control, especially in commercial scale processes where cheap potable water sources are likely to be employed. Also, the use of regulatory promoters, like *CYC6, NIT1, CA1,* and *METE*, constrains transgene expression to the level and/or regulatory profile of the endogenous gene, and unknown regulatory processes may impact this, such as circadian regulation or nutrient stress. Finally, the use of endogenous promoters largely limits the

utility of the regulatory sequence to one, or a small number of species, making parallel development a necessity. In contrast the riboswitch is unique. It is mRNA encoded, functions in response to a ligand which is universal (not a species specific protein), can impose regulation via splicing which is a highly conserved eukaryotic process, and when present in an intron a riboswitch may be integrated into an existing genetic circuit to introduce a novel regulatory function. The riboswitch therefore lends itself to synthetic biology. Moreover, from a pragmatic standpoint, the TPP riboswitch can be regulated by nano-molar concentrations of thiamine, a benign and in expensive vitamin. The *C. reinhardtii THI4* riboswitch has already been used to regulate transgene expression in combination with strong constitutive promoters such as *PSAD* and *RBCS2* (Croft et al. 2007; Moulin et al. 2013). This system has been extended in a biological context to regulate the expression of plastid genes. Ramundo et al (2013) built on the knowledge that nuclear encoded proteins can regulate the expression of chloroplast genes to develop a novel synthetic regulatory circuit (Figure 5). The nuclear encoded protein, *NAC2,* is required to stabilise the 5' leader region of chloroplast-encoded *psbD* mRNA for translation (Boudreau et al. 2000). A construct, in which the *C. reinhardtii THI4* riboswitch was included in the wild-type *NAC2* gene, was introduced into the *nac2* mutant, so that addition of thiamine repressed *NAC2* production (Ramundo et al. 2013. The endogenous *psbD* 5' leader was replaced with the *psaA* leader, enabling functional photosynthesis even when *NAC2* was repressed. Then fusion of the *psbD* 5' leader to essential chloroplast-encoded genes caused them to be repressed by addition of thiamine, allowing their roles to be investigated (Ramundo et al. 2014).

Towards new riboswitches in algae

As we have discussed riboswitches are an ancient regulatory system, likely retained from the RNA world (Breaker 2012). They are frequently involved in regulation of fundamental

metabolic processes in prokaryotes such as the biosynthesis and uptake of small metabolites and metal ions, as well as associated proteins that use these as cofactors. The discovery of new natural riboswitches in microalgae and other eukaryotes would offer the potential to employ these in a combinatorial manner or to develop synthetic riboswitches, providing a mechanism to carefully balance transgene expression within a given network (Groher & Suess 2014; Berens & Suess 2015). Yet, in spite of focused research on *C. reinhardtii*, and an exponential increase in genomic and transcriptomic data for this and other eukaryotes, the only riboswitches identified to date are those that bind TPP. We attempted to identify novel riboswitches in *C. reinhardtii* by screening for highly conserved residues present in riboswitch aptamer sequences retrieved from the RFAM database (Griffiths-Jones et al. 2003), using these data to search non-coding regions of genes known to be involved in purine, glycine, lysine, methionine and folate metabolism (Mandal & Breaker 2004; Mandal et al. 2004; Grundy et al. 2003). In addition secondary structures of riboswitch aptamers in the non-coding regions were screened using the RibEx prediction tool (Abreu-Goodger & Merino 2005). TPP riboswitches were found in algae initially because of unusual sequence conservation in non-coding regions (Croft et al. 2007; McRose et al. 2014), we therefore manually inspected orthologues of these genes using the JGI browser. However, no sequences found that might constitute putative riboswitches.

A number of studies in bacteria have already demonstrated the plasticity and versatility of riboswitch-mediated control of transgene expression by coupling different aptamers and expression platforms in artificial systems (Ceres, Trausch, et al. 2013; Ceres, Garst, et al. 2013; Rudolph et al. 2013). For example, neomycin, theophylline or tetracycline-sensing aptamers have been used in bacterial synthetic riboswitches in a highly specific and dose-dependent manner to regulate both recombinant and endogenous gene expression (reviewed

in Berens & Suess 2014). Similar approaches may be adapted to develop aptamers of interest in microalgae with a minimal effort of gene design.

The study of eukaryotic riboswitches furthers our understanding of RNA based regulatory mechanisms that are ancient in origin and have likely evolved as eukaryotic cell biology became more complex. The presence of TPP riboswitches across the complex and diverse algal lineages suggests that these elements offer considerable selective advantage, and imply that regulation of thiamine metabolism is more important than previously recognised. Moreover, as the interest in use of microalgae for biotechnological purposes continues to increase, the use of riboswitches has the potential to develop novel and robust regulatory tools for metabolic engineering and synthetic biology. However, a lack of knowledge regarding eukaryotic riboswitches, other than those which bind TPP or its precursor compounds, means that in the short term development efforts should focus on synthetic alternatives.

Acknowledgements

The authors declare no conflict of interest. They would like to thank Mr Vaibhav Bhardwaj and Dr Payam Mehrshahi for thoughtful discussions related to this paper. G.T.D.T.N was funded in part by Murray Edwards College and the Cambridge Philosophical Society. M.A.S was funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/I00680X/1 and the European Commission 7th Framework Programme (FP7) project SPLASH (Sustainable PoLymers from Algae Sugars and Hydrocarbons), grant agreement number 311956. K.E.H was funded by BBSRC grant BB/ I013164/1.

References

Abreu-Goodger, C. & Merino, E. 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res.* 33:W690–2.

Allen, M.D., del Campo, J.A., Kropat, J. & Merchant, S.S. 2007. FEA1, FEA2, and FRE1, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed

coordinately with FOX1 and FTR1 in iron-deficient *Chlamydomonas reinhardtii*. *Eukaryot. Cell*. 6:1841–52.

Barrick, J.E. & Breaker, R.R. 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* 8:R239.

Berens, C. & Suess, B. 2015. Riboswitch engineering - making the all-important second and third steps. *Curr. Opin. Biotechnol.* 31:10–5.

Bocobza, S., Adato, A., Mandel, T., Shapira, M., Nudler, E. & Aharoni, A. 2007. Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev.* 21:2874–9.

Boudreau, E., Nickelsen, J., Lemaire, S.D., Ossenbühl, F. & Rochaix, J.D. 2000. The *Nac2* gene of *Chlamydomonas* encodes a chloroplast TPR-like protein involved in *psbD* mRNA stability. *EMBO J.* 19:3366–76.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U. et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature*. 456:239–44.

Breaker, R.R. 2012. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* 4.

Ceres, P., Garst, A.D., Marcano-Velázquez, J.G. & Batey, R.T. 2013a. Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. *ACS Synth. Biol.* 2:463–72.

Ceres, P., Trausch, J.J. & Batey, R.T. 2013b. Engineering modular “ON” RNA switches using biological components. *Nucleic Acids Res.* 41:10449–61.

Cheah, M.T., Wachter, A., Sudarsan, N. & Breaker, R.R. 2007. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*. 447:497–500.

Croft, M.T., Moulin, M., Webb, M.E. & Smith, A.G. 2007. Thiamine biosynthesis in algae is regulated by riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* 104:20770–5.

Dambach, M., Sandoval, M., Updegrove, T.B., Anantharaman, V., Aravind, L., Waters, L.S. & Storz, G. 2015. The ubiquitous *yybP-ykoY* riboswitch is a manganese-responsive regulatory element. *Mol. Cell*. 57:1099–109.

Davies, J.P., Weeks, D.P. & Grossman, A.R. 1992. Expression of the arylsulfatase gene from the beta-tubulin promoter in *Chlamydomonas reinhardtii*. 20:2959–65.

DebRoy, S., Gebbie, M., Ramesh, A., Goodson, J.R., Cruz, M.R., van Hoof, A., Winkler, W.C. et al. 2014. A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. *Science.* 345:937–40.

Dorrell, R.G. & Smith, A.G. 2011. Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell*. 10:856–68.

Edwards, T.E. & Ferré-D’Amaré, A.R. 2006. Crystal structures of the *thi*-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure*. 14:1459–68.

Fischer, N. & Rochaix, J.D. 2001. The flanking regions of *PsaD* drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. *Mol. Genet. Genomics*. 265:888–94.

Furukawa, K., Ramesh, A., Zhou, Z., Weinberg, Z., Vallery, T., Winkler, W.C. & Breaker, R.R. 2015. Bacterial riboswitches cooperatively bind Ni(2+) or Co(2+) ions and control expression of heavy metal transporters. *Mol. Cell*. 57:1088–98.

Gangl, D., Zedler, J.A.Z., Rajakumar, P.D., Martinez, E.M.R., Riseley, A., Włodarczyk, A., Purton, S. et al. 2015. Biotechnological exploitation of microalgae. *J. Exp. Bot.* erv426 – .

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., Eddy, S.R. 2003. Rfam: an RNA family database. *Nucleic Acids Res*. 31:439-41.

Groher, F. & Suess, B. 2014. Synthetic riboswitches - A tool comes of age. *Biochim. Biophys. Acta*. 1839:964–73.

Grundy, F.J., Lehman, S.C. & Henkin, T.M. 2003. The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc. Natl. Acad. Sci. U. S. A.* 100:12057–62.

Helliwell, K.E., Scaife, M.A., Sasso, S., Araujo, A.P.U., Purton, S. & Smith, A.G. 2014. Unraveling vitamin B12-responsive gene regulation in algae. *Plant Physiol.* 165:388–97.

Kilian, O., Benemann, C.S.E., Niyogi, K.K. & Vick, B. 2011. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc. Natl. Acad. Sci. U. S. A.* 108:21265–9.

Kubodera, T., Watanabe, M., Yoshiuchi, K., Yamashita, N., Nishimura, A., Nakai, S., Gomi, K. et al. 2003. Thiamine-regulated gene expression of *Aspergillus oryzae thiA* requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS Lett.* 555:516–20.

Kulshina, N., Edwards, T.E. & Ferré-D'Amaré, A.R. 2010. Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. *RNA*. 16:186–96.

Kumar, A., Falcao, V.R. & Sayre, R.T. 2013. Evaluating nuclear transgene expression systems in *Chlamydomonas reinhardtii*. *Algal Res.* 2:321–32.

Lang, K., Rieder, R. & Micura, R. 2007. Ligand-induced folding of the *thiM* TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res.* 35:5370–8.

Li, S. & Breaker, R.R. 2013. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Res.* 41:3022–31.

Mandal, M. & Breaker, R.R. 2004. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* 11:29–35.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L. & Breaker, R.R. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science*. 306:275–9.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.-Y., Mori, T. et al. 2004. Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature*. 428:653–7.

McRose, D., Guo, J., Monier, A., Sudek, S., Wilken, S., Yan, S., Mock, T. et al. 2014. Alternatives to vitamin B1 uptake revealed with discovery of riboswitches in multiple marine eukaryotic lineages. *ISME J.* 8:2517–29.

Mellin, J.R., Koutero, M., Dar, D., Nahori, M.-A., Sorek, R. & Cossart, P. 2014. Riboswitches. Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. *Science*. 345:940–3.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A. et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science.* 318:245–50.

Miranda-Ríos, J., Navarro, M. & Soberón, M. 2001. A conserved RNA structure (*thi* box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 98:9736–41.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A. et al. 2002. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*. 111:747–56.

Moulin, M., Nguyen, G.T.D.T., Scaife, M.A., Smith, A.G. & Fitzpatrick, T.B. 2013. Analysis of *Chlamydomonas* thiamin metabolism in vivo reveals riboswitch plasticity. *Proc. Natl. Acad. Sci. U. S. A.* 110:14622–7.

Noeske, J., Richter, C., Stirnal, E., Schwalbe, H. & Wöhnert, J. 2006. Phosphate-group recognition by the aptamer domain of the thiamine pyrophosphate sensing riboswitch. *Chembiochem*. 7:1451–6.

Ohresser, M., Matagne, R.F. & Loppes, R. 1997. Expression of the arylsulphatase reporter gene under the control of the *nit1* promoter in *Chlamydomonas reinhardtii*. *Curr. Genet.* 31:264–71.

Ooms, M., Huthoff, H., Russell, R., Liang, C. & Berkhout, B. 2004. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J. Virol.* 78:10814–9.

Paerl, R.W., Bertrand, E.M., Allen, A.E., Palenik, B. & Azam, F. 2015. Vitamin B1 ecophysiology of marine picoeukaryotic algae: Strain-specific differences and a new role for bacteria in vitamin cycling. *Limnol. Oceanogr.* 60:215–28.

Qiu, H., Yoon, H.S. & Bhattacharya, D. 2013. Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Front. Plant Sci.* 4:366.

Quinn, J.M. & Merchant, S. 1995. Two copper-responsive elements associated with the *Chlamydomonas Cyc6* gene function as targets for transcriptional activators. *Plant Cell Online*. 7:623–8.

Radakovits, R., Jinkerson, R.E., Fuerstenberg, S.I., Tae, H., Settlage, R.E., Boore, J.L. & Posewitz, M.C. 2012. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*. *Nat. Commun.* 3:686.

Ramundo, S., Rahire, M., Schaad, O. & Rochaix, J.-D. 2013. Repression of essential chloroplast genes reveals new signaling pathways and regulatory feedback loops in *Chlamydomonas*. *Plant Cell*. 25:167–86.

Ramundo S, Casero D, Mühlhaus T, Hemme D, Sommer F, Crèvecoeur M, Rahire M, Schroda M, Rusch J, Goodenough U, Pellegrini M, Perez-Perez ME, Crespo JL, Schaad O, Civic N, **Rochaix** JD. (2014) Conditional depletion of the *Chlamydomonas* chloroplast clpP protease activates nuclear genes involved in autophagy and plastid protein quality control. Plant Cell. 2014 May 30;26(5):2201-2222

Rasala, B.A. & Mayfield, S.P. 2011. The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioeng Bugs*. 2:50–4.

Reijnders, M.J.M.F., van Heck, R.G.A., Lam, C.M.C., Scaife, M.A., Santos, V.A.P.M. dos, Smith, A.G. & Schaap, P.J. 2014. Green genes: bioinformatics and systems-biology innovations drive algal biotechnology. *Trends Biotechnol.*

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. & Gelfand, M.S. 2002. Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J. Biol. Chem.* 277:48949–59.

Rudolph, M.M., Vockenhuber, M.-P. & Suess, B. 2013. Synthetic riboswitches for the conditional control of gene expression in *Streptomyces coelicolor*. *Microbiology*. 159:1416–22.

Scaife, M.A., Nguyen, G.T., Rico, J., Lambert, D., Helliwell, K.E. & Smith, A.G. 2015. Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *Plant J.* 82:532–46.

Schroda, M., Blocker, D. & Beck, C. 2000. The *HSP70A* promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant J.* 21:121–31.

Serganov, A. & Patel, D.J. 2007. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* 8:776–90.

Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R. & Patel, D.J. 2006. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*. 441:1167–71.

Sudarsan, N., Barrick, J.E. & Breaker, R.R. 2003. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA*. 9:644–7.

Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G.M. & Breaker, R.R. 2005. Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem. Biol.* 12:1325–35.

Tang, D.K.H., Qiao, S.Y. & Wu, M. 1995. Insertion mutagenesis of *Chlamydomonas reinhardtii* by electroporation and heterologous DNA. *Biochem. Mol. Biol. Int.* 36:1025–35.

Thore, S., Leibundgut, M. & Ban, N. 2006. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science*. 312:1208–11.

Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., Cornish, A.J., Harvey, C., Reca, I.-B. et al. 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *Nannochloropsis oceanica* CCMP1779. *PLoS Genet.* 8:e1003064.

Wachter, A., Tunc-Ozdemir, M., Grove, B.C., Green, P.J., Shintani, D.K. & Breaker, R.R. 2007. Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell*. 19:3437–50.

Warner, K.D., Homan, P., Weeks, K.M., Smith, A.G., Abell, C. & Ferré-D'Amaré, A.R. 2014. Validating fragment-based drug discovery for biological RNAs: lead fragments bind and remodel the TPP riboswitch specifically. *Chem. Biol.* 21:591–5.

Winkler, W., Nahvi, A. & Breaker, R.R. 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*. 419:952–6.

Winkler, W.C. 2005. Metabolic monitoring by bacterial mRNAs. *Arch. Microbiol.* 183:151–9.

Winkler, W.C. & Breaker, R.R. 2003. Genetic control by metabolite-binding riboswitches. *Chembiochem*. 4:1024–32.

Worden, A.Z., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M.P., Aerts, A.L., Allen, A.E. et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*. 324:268–72.

Zhang, R., Patena, W., Armbruster, U., Gang, S.S., Blum, S.R. & Jonikas, M.C. 2014. High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell*. 26:1398–409.

Figure Legends

Figure 1. Schematic diagram of the action of a representative riboswitch. The aptamer region (black dashed box) upon binding of its ligand (**gray** oval) changes conformation to affect the expression platform (gray dashed box). In this example binding of the ligand prevents access of the ribosome to the initiation codon (AUG) of the mRNA, inhibiting translation. In other examples ligand binding and conformational changes may prevent access to the ribosome binding site, splice sites or other important RNA motifs.

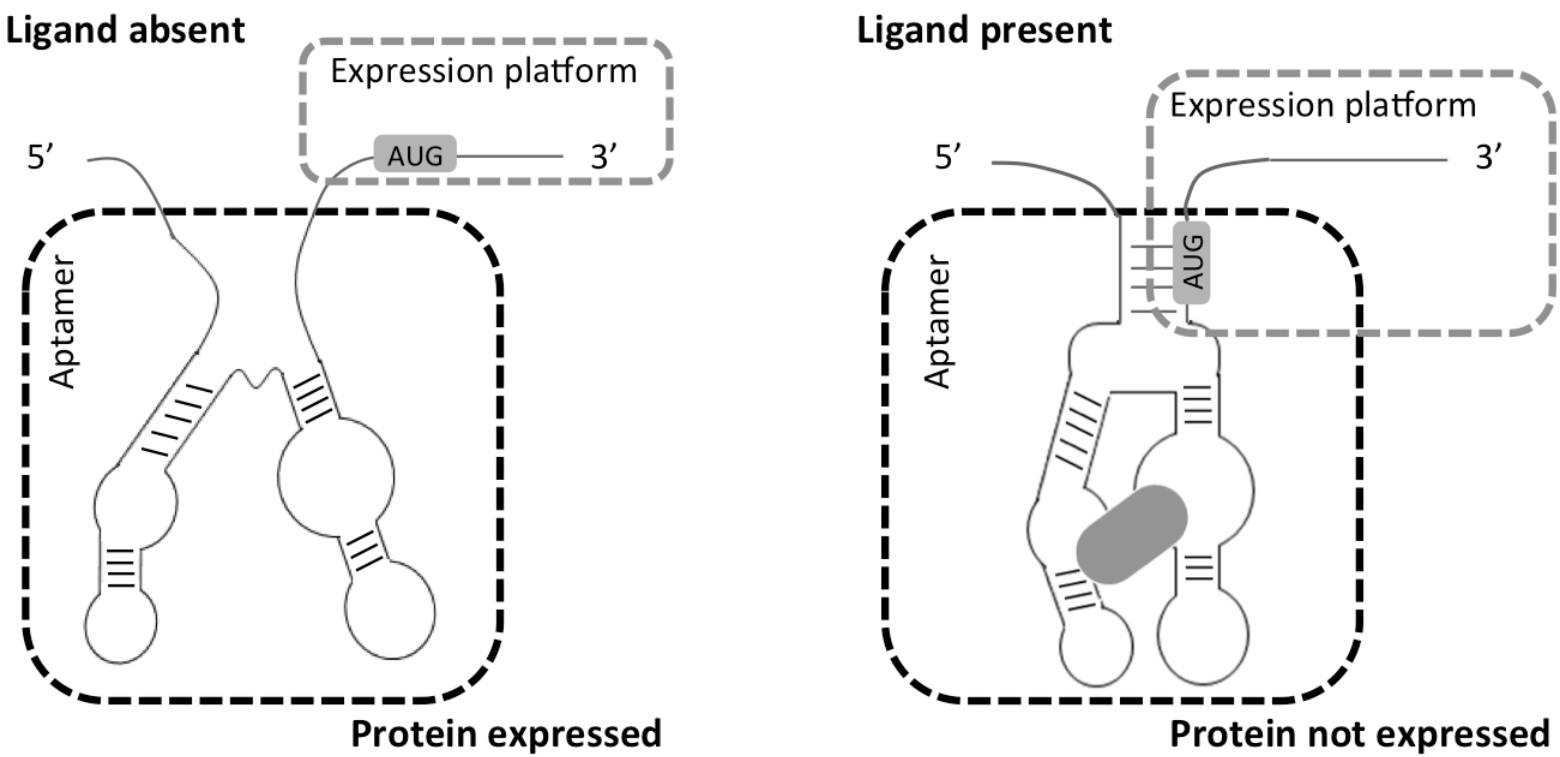
Figure 2. Mechanism of action of different eukaryotic TPP riboswitches. The changes in secondary structure on binding of TPP alter the accessibility of the consensus splice donor (GU) and acceptor (AG) sites, leading to alternative splice site variants to that of the transcript encoding the functional protein. In *NMT1* and *THI4* riboswitches alternative splicing reveals an upstream open reading frame (uORF), which is translated instead of the coding region (green block). For *THIC* in *C. reinhardtii* and *NCU01977* in *N. crassa* alternative splicing introduces a premature in-frame stop codon, resulting in the translation of a truncated protein. In *A. thaliana* alternative splicing results in the removal of the polyadenylation motif from the 3'UTR. Black blocks depict UTRs, both 5' and 3', black lines describe introns, exons are represented by green blocks, upper case letter depict splice sites used in the absence of thiamine, also highlighted by the solid arrow. Lower case letters show the alternative splice sites employed in the presence of thiamine, additionally highlighted by dashed arrows. The aptamer region is shown by a small RNA structured cartoon. Where present the uORF is depicted by a red block, and the polyadenylation motif by a black circle with upper case letter A in it. The question mark indicates the absence of splicing in the presence of thiamine, however the actual mechanism is not currently known.

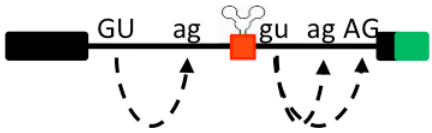
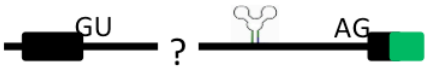
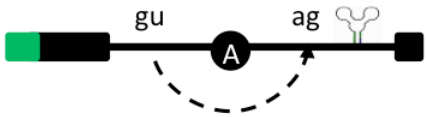
Figure 3. Detailed representation of the *THI4* riboswitch of *C. reinhardtii* showing the different transcripts and proteins produced in the absence and presence of the ligand, thiamine. The black arrow describes the *THI4* promoter (P_{THI4}), black blocks the 5' UTR and the blackline the intron. The first exon of *THI4* is represented by the green block. For splicing upper case letters depict splice sites used in the absence of thiamine and lower case letters show the alternative splice sites employed in the presence of thiamine. The aptamer region is highlighted by the black box and the uORF depicted by the red block. In the absence thiamine the intron is spliced out and a short transcript produced which allows translation of THI4 (depicted by a green tear drop). In the presence of thiamine two larger transcripts are produced, which include the uORF. From these the uORF is translated instead of THI4 (represented by a red tear drop).

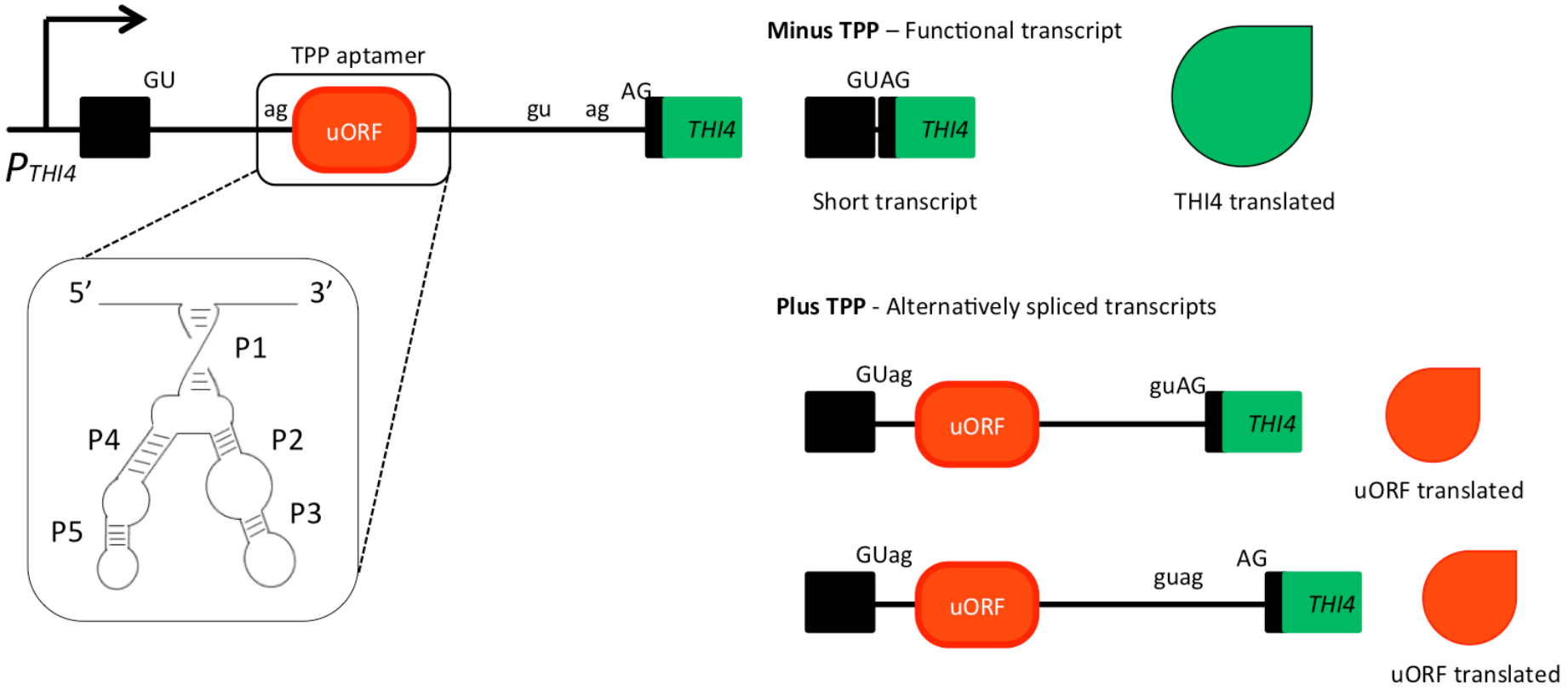
Figure 4. Nucleotide conservation in microalgal TPP aptamers. The primary sequence of 43 TPP riboswitch aptamers were aligned at a single nucleotide level within the conserved secondary structure. Stems (P) and loops (L) are indicated. Locations of pyrimidine binding (blue triangles), α- & β- pyrophosphate binding (red & orange triangles, respectively), and joining loop between L5 and J2/3 (grey triangles) are shown. Nucleotides conserved in $\geq$ 50% of analysed sequences are described by their letter, and highlighted to demonstrate the degree of conservation (increasing from yellow to black). Non-conserved nucleotides are represented by a solid line or black dot. Details of these aptamers and their associated regulated genes are listed in Table S1.

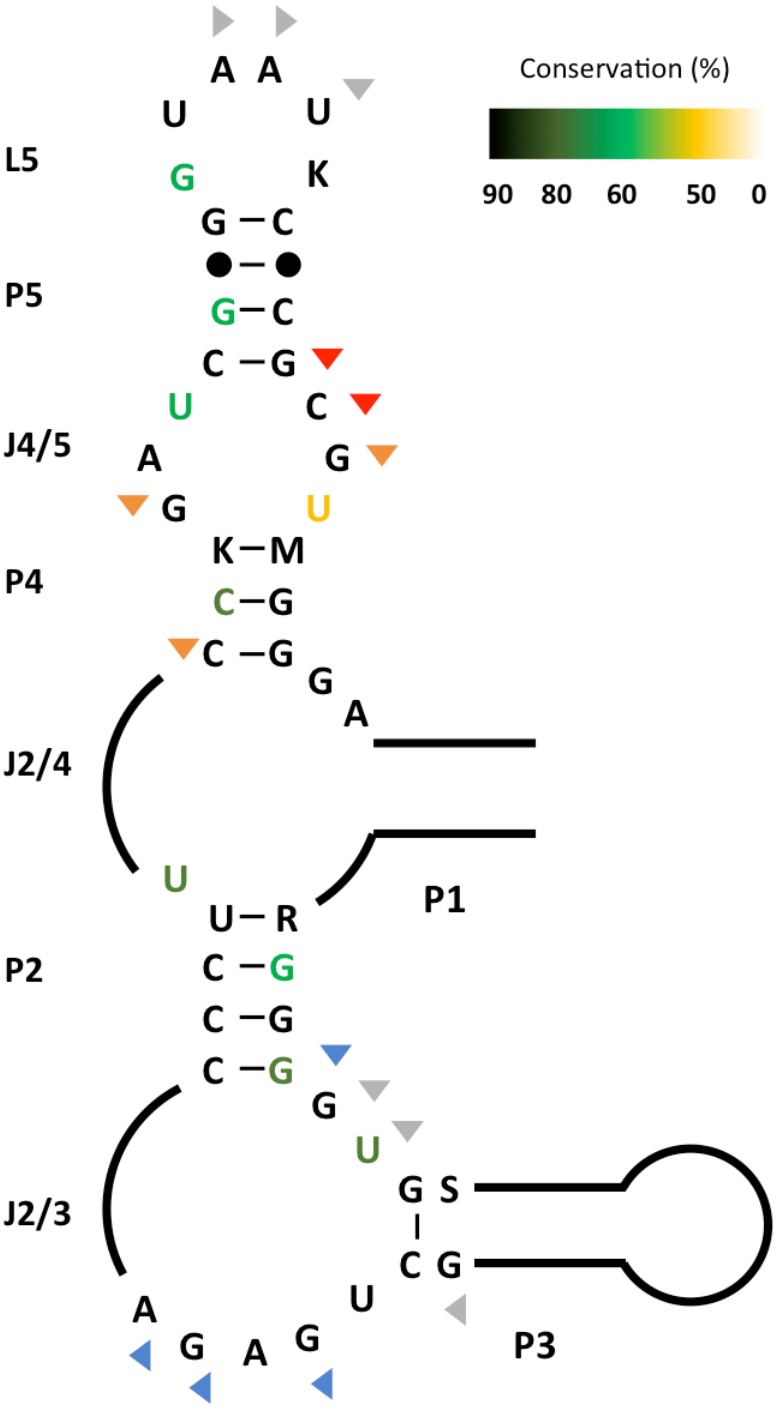
Figure 5. Schematic representation of a thiamine responsive synthetic circuit to control expression of genes in the *C. reinhardtii* chloroplast. This system, developed by Ramundo et al (2013) is based on the role of NAC2, a nucleus-encoded protein required for stabilization of chloroplast encoded *psbD* gene. The *nac2* mutant, in which *psbD* is expressed

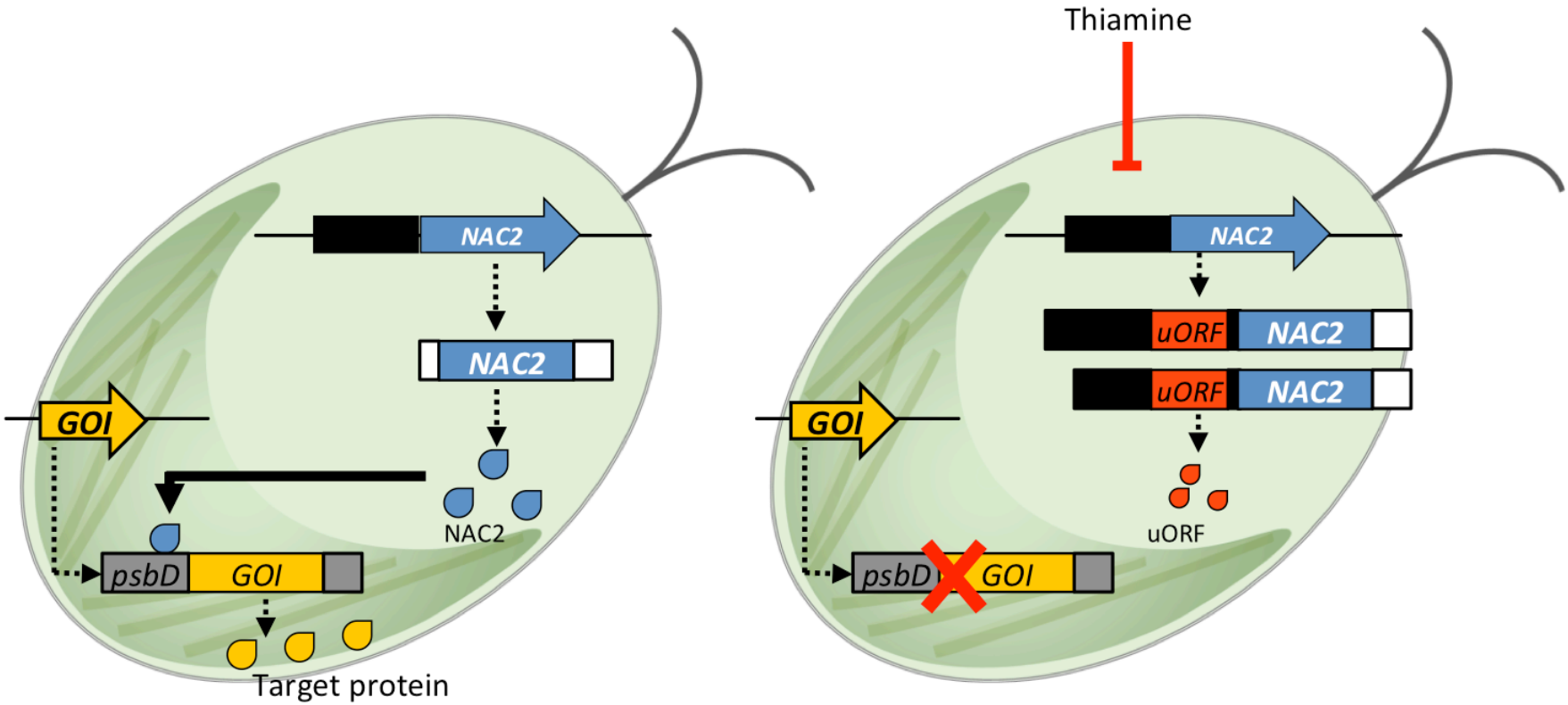
constitutively, is complemented by the wild-type *NAC2* gene under the control of the THI4 riboswitch (black box). The *psbD* 5'UTR (gray block) is used in a transgene expression construct for a specific Gene Of Interest (GOI, yellow arrow (DNA) and block (mRNA)). In the absence of thiamine (left) *NAC2* is expressed and allows translation of the GOI (yellow tear drop). In the presence of thiamine (right) *NAC2* is absent, so that although the target GOI is transcribed the mRNA is degraded (red cross) and translation is down regulated.


Ligand absent
Expression platform
5’
AUG
3’
Aptamer
Protein expressed
Ligand present
Expression platform
5’
3’
AUG
Aptamer
Protein not expressed

Location	Mechanism in presence of TPP	Examples
5’ UTR		*N. crassa* (*NMT1, NCU01977*)
		*C. reinhardtii* (*THI4*)
		*A. oryzae* (*THIA*)
Coding region		*C. reinhardtii* (*THIC*)
3’ UTR		*A. thaliana* (*THIC*)


TPP aptamer
GU
ag
uORF
gu
ag
AG
THI4
P_{THI4}
5'
3'
P1
P4
P2
P3
P5
Minus TPP – Functional transcript
GUAG
THI4
Short transcript
THI4 translated
Plus TPP - Alternatively spliced transcripts
GUag
uORF
guAG
THI4
uORF translated
GUag
uORF
guag
AG
THI4
uORF translated


Conservation (%)
90 80 60 50 0
L5
P5
J4/5
P4
J2/4
P1
P2
J2/3
P3


Thiamine
NAC2
NAC2
GOI
NAC2
psbD
GOI
Target protein
NAC2
uORF
NAC2
uORF
NAC2
GOI
uORF
psbD
GOI