Structural basis for tRNA decoding and aminoacylation sensing by T-box riboregulators

Robert A. Battaglia, Jason C. Grigg and Ailong Ke

T-box riboregulators are a class of cis-regulatory RNAs that govern the bacterial response to amino acid starvation by binding, decoding and reading the aminoacylation status of specific transfer RNAs. Here we provide a high-resolution crystal structure of a full-length T-box from Mycobacterium tuberculosis that explains tRNA decoding and aminoacylation sensing by this riboregulator. Overall, the T-box consists of decoding and aminoacylation sensing modules bridged by a rigid pseudoknot structure formed by the mid-region domains. Stem-I and the Stem-II S-turn assemble a claw-like decoding module, while the antiterminator, Stem-III, and the adjacent linker form a tightly intertwined aminoacylation sensing module. The uncharged tRNA is selectively recognized by an unexpected set of favorable contacts from the linker region in the aminoacylation sensing module. A complex structure with a charged tRNA mimic shows that the extra moiety dislodges the linker, which is indicative of the possible chain of events that lead to alternative base-pairing and altered expression output.

A substantial portion of bacterial genes are regulated at the messenger RNA level by cis-acting riboswitches in their 5′ untranslated regions. A typical riboswitch encodes an aptamer domain and an expression platform. Ligand binding to the aptamer domain triggers alternative base-pairing in the expression platform that alters the transcription or translation levels of the associated mRNA. The T-box riboregulator controls gene expression using the same conformational switching mechanism, however, instead of responding to small molecule ligands it recognizes a macromolecule, tRNA. A typical T-box contains five structural domains: Stem-I, Stem-II, Stem-IIA/B, Stem-III and antiterminator (AntiT) or antisequestrator (AntiS) (Fig. 1a and inset). Two variant classes exist, glycine and atypical, which contain truncations outside the AntiT/S domain. Previous work has attributed distinct roles to structural domains: (1) a specifier motif in Stem-I decodes tRNA via codon-anticodon base-pairing; (2) a conserved double T-loop structure in the distal half of Stem-I stacks against the D/T loops of the tRNA elbow and measures anticodon arm length in conjunction with the codon-anticodon pairing; (3) the T-box sequence in AntiT/S specifies an uncharged tRNA-3′-NCCA tail through Watson–Crick (WC) pairing. Whereas the Stem-I-tRNA contacts have been defined by high-resolution structures, the mechanism for aminoacylation sensing and the roles of the mid-region remain unclear.

In this study, we present crystal structures of the full-length isoleucine T-box riboregulator from Mycobacterium tuberculosis (Mtb-ileS) in complex with its cognate uncharged tRNA or a charged tRNA mimic that contains a 2′-3′-cyclic phosphate at its terminus. Together they provide the structural basis for tRNA decoding and aminoacylation sensing by the T-box. We can now ascribe functions to the ‘variable’ mid-region domains (Stem-II, Stem-IIA/B and the linker), which play important roles in organizing the overall T-box structure. At the 5′-end, Stem-II assembles with the previously characterized Stem-I to form a decoding module. At the opposite end, Stem-IIA/B forms a rigid pseudoknot structure with the linker region. The mid-region domains combine to assemble an elongated structure that projects the aminoacylation sensing module near the tRNA-3′-NCCA tail. An unexpected finding of our work is that aminoacylation sensing not only requires the T-box antitermination domain, as previously reported, but further involves Stem-III and the adjacent linker region. Together they form a tightly wove aminoacylation sensing module that coaxially stacks against the tRNA acceptor arm, guides the 3′-NCCA tail into a pocket through WC base-pairing and cradles the uncharged tRNA terminus with favorable contacts. We further show that the aminoacylation sensing module is elastic enough to accommodate a tRNA with a 2′-3′-cyclic phosphate group at its 3′-end. However, the resulting steric hindrance tears the bottom of the binding pocket and increases overall conformational entropy in the T-box. This depicts a plausible scenario for how a charged tRNA could bind and actively initiate the transition to the alternative base-pairing scheme that switches the expression output. The T-box is the only other structured RNA capable of tRNA decoding and aminoacylation sensing besides the ribosome. A high-resolution understanding of this system opens discussions on its possible function in the primordial RNA world.

Results

Overall structure and domain organization in the T-box–tRNA complex. Molecular dissection of the T-box–tRNA interaction has been difficult due to the lack of robust in vitro reconstitution. Speculating that this is because most T-boxes are transcriptional riboregulators whose folding is influenced by transcription speed and pausing elements, we turned our attention to a class of translational T-boxes from actinobacteria, reasoning that their folding is governed by thermodynamics. In vitro reconstitution was indeed very robust (Extended Data Fig. 1a and b). We subsequently solved the co-crystal structure of a full-length isoleucine T-box from M. tuberculosis bound to its uncharged cognate tRNA at 2.95 Å resolution (Fig. 1b, Table 1). The Mtb-ileS T-box includes all conserved structural features except the distal stacking module. It instead uses a degenerate Stem-I to decode tRNA (Fig. 1a). Overall, the structure reveals that the Mtb-ileS T-box consists of two functional modules connected in the middle by a continuous structural unit assembled from the mid-region domains. On one side, the tail-to-tail stacked
Stem-II and Stem-IA/B further assemble with Stem-I to form the decoding module that latches onto the tRNA anticodon loop (Fig. 1c). On the opposite side, Stem-III and the nearby linker region weave around the AntiS domain to form the aminoacylation sensing module that stacks to the end of the tRNA acceptor arm to interrogate the aminoacylation status of tRNA (Fig. 1d).

The long single-stranded linker between the two epicenters is largely absorbed through tertiary structure formation, leaving little flexibility therein (Fig. 1e). The 5′-half of the linker emerges from the stacking interface between Stem-II and IA/B and travels along the minor groove of Stem-IIA/B, forming an extensive pseudoknot (Fig. 2a). Three WC pairs and a ladder of stacked tertiary contacts to Stem-IIA/B are evident (Fig. 2a). The sharp bend required to redirect the linker from the stacking junction to the top of the pseudoknot is stabilized by a type-I A-minor interaction from A81 to the G68-C77 pair in Stem-IIA/B (Fig. 2a). The essentiality of this tertiary interaction explains the perplexing ‘F-box’ sequence reduction (Fig. 1b). The key to this specificity is tRNA decoding via codon-anticodon pairing—a theme once thought to be unique to the ribosome. Previous structures of the glycine (glyQS) Stem-I–tRNA complex revealed an S-turn mediated codon-anticodon decoding mechanism and the further requirement of a second stacking contact to the tRNA D/T-loop elbow. The Mtb-ileS Stem-I lacks the stacking module and its decoding S-turn is replaced by a codon-embedded terminal specifier loop. It was puzzling how such T-boxes could use a simple stem-loop to efficiently decode tRNA. Here the structure reveals that decoding by Stem-I is supported by Stem-II; together they form a ‘claw’ module that clamps onto the tRNA anticodon from both the major and minor groove sides (Fig. 1c and Fig. 2c).

Stem-II consists of a stem-loop with a conserved S-turn motif in the middle (Fig. 1b). It is one of the mid-region structural domains ubiquitous among typical and atypical T-boxes yet absent from glycine T-boxes. Thus, the pseudoknot contributes to T-box function not by making direct tRNA contact but by introducing the necessary geometric constraints to place the aminoacylation sensing module close to the tRNA acceptor.

**tRNA identity is specified by the Stem-I/II decoding module.** With rare exceptions, each T-box riboregulator selectively binds to one tRNA species to respond to the corresponding amino acid level. The key to this specificity is tRNA decoding via codon-anticodon pairing—a theme once thought to be unique to the ribosome. Previous structures of the glycine (glyQS) Stem-I–tRNA complex revealed an S-turn mediated codon-anticodon decoding mechanism and the further requirement of a second stacking contact to the tRNA D/T-loop elbow. The Mtb-ileS Stem-I lacks the stacking module and its decoding S-turn is replaced by a codon-embedded terminal specifier loop. It was puzzling how such T-boxes could use a simple stem-loop to efficiently decode tRNA. Here the structure reveals that decoding by Stem-I is supported by Stem-II; together they form a ‘claw’ module that clamps onto the tRNA anticodon from both the major and minor groove sides (Fig. 1c and Fig. 2c).

Stem-II consists of a stem-loop with a conserved S-turn motif in the middle (Fig. 1b). It is one of the mid-region structural domains ubiquitous among typical and atypical T-boxes yet absent from glycine T-boxes. Thus, the tRNA contacts mediated by Stem-II that specify the aminoacylation status of the tRNA are determined by the Stem-I/II decoding module. This mechanism allows for a single T-box to sense multiple tRNA species via decoding interactions.

**tRNA identity is specified by the Stem-I/II decoding module.** With rare exceptions, each T-box riboregulator selectively binds to one tRNA species to respond to the corresponding amino acid level. The key to this specificity is tRNA decoding via codon-anticodon pairing—a theme once thought to be unique to the ribosome. Previous structures of the glycine (glyQS) Stem-I–tRNA complex revealed an S-turn mediated codon-anticodon decoding mechanism and the further requirement of a second stacking contact to the tRNA D/T-loop elbow. The Mtb-ileS Stem-I lacks the stacking module and its decoding S-turn is replaced by a codon-embedded terminal specifier loop. It was puzzling how such T-boxes could use a simple stem-loop to efficiently decode tRNA. Here the structure reveals that decoding by Stem-I is supported by Stem-II; together they form a ‘claw’ module that clamps onto the tRNA anticodon from both the major and minor groove sides (Fig. 1c and Fig. 2c).

Stem-II consists of a stem-loop with a conserved S-turn motif in the middle (Fig. 1b). It is one of the mid-region structural domains ubiquitous among typical and atypical T-boxes yet absent from glycine T-boxes. Thus, the tRNA contacts mediated by Stem-II that specify the aminoacylation status of the tRNA are determined by the Stem-I/II decoding module. This mechanism allows for a single T-box to sense multiple tRNA species via decoding interactions.
of the Stem-I specifier loop and the Stem-II S-turn is important for T-box function. Increasing the Stem-II length by two basepairs (Stem-II_2bp_shift) projects the S-turn beyond the decoding ground level (Fig. 2b, dark blue). At the interface of Stem-I and II, 

certain modifications may require the Stem-II S-turn to interact with the ribosome, the first two layers of codon-anticodon pairing are supported by type-I and type-II A-minor contacts, respectively18–20. Conversely, the A-minor contacts made by the T-box are focused on the second and wobble position pairs and are more extensive (Fig. 2d, middle and bottom). Whereas the ribosome tolerates mismatches in the wobble position, the T-box is expected to have a much more stringent preference for tRNA isoaacpetors2. In fact, a bias for a cytosine residue in the wobble position of codons found in T-box specifier loops has been noted21. The specific contacts in our structure between S-turn A38 and the sugar edge of tRNA G34 could explain the wobble cytosine preference from the complementary side (Fig. 2d, bottom).

AntiS alone cannot distinguish the tRNA aminoacylation state. Whereas the tRNA decoding mechanism varies to some extent among the three T-box variants, the aminoacylation sensing mechanism is essentially identical. The AntiT or AntiS domains have long been established as the main contributors to aminoacylation sensing due to their characteristic ‘T-box’ sequence in an asymmetric bulge that complements the tRNA NCCA tail2. In our structure, the AntiS helix is bent at the central asymmetric bulge to an L-shape (Fig. 3a). The T-box sequence inside the asymmetric bulge mediates four consecutive WC base-pairs with the tRNA NCCA tail (Fig. 3b). The resulting four-base-pair helix is coaxially stacked on either side by the tRNA acceptor arm and the AntiS bottom helix, forming a continuous superhelix (Fig. 3a and Extended Data Fig. 2b). This arrangement ensures that only mature tRNA, but not precursors, are examined by T-boxes. An important finding from our structure is that the AntiS domain alone lacks specific contacts with the tRNA that interrogate its aminoacylation status. A residue (C161) from its bottom helix partially stacks with the tRNA terminal adenosine and a wobble pair (G132•U160) from the base of the AntiS top helix partially blocks the tRNA terminus (Extended Data Fig. 2a). However, these blockages are insufficient to fully distinguish a charged tRNA. For example, changing the wobble pair to a C-G pair (U160C) in the tyrS T-box should alleviate the partial blockage, but the mutation only decreases T-box induction by ~2-fold (Fig. 3f, green). Overall, the AntiS forms a surprisingly incomplete binding pocket where the tRNA terminal adenosine is exposed in a large opening that could easily accommodate an aminoacylated tRNA terminus (Extended Data Fig. 2b).

**Aminoacylation sensing requires the nearby linker region.** The formation of a functional aminoacylation sensing module necessarily involves Stem-III and the nearby linker region (Fig. 3a). This came as a surprise because neither the sequence nor the length of Stem-III is conserved23 and the linker sequence conservation was

### Table 1 | Data collection and refinement statistics

|                      | MtbsdileS_tRNA-OH_IrHex | MtbsdileS_tRNA-OH_native (PDB 6UFG) | MtbsdileS_tRNA-cP_native (PDB 6UFG) |
|----------------------|-------------------------|-------------------------------------|-------------------------------------|
| **Data collection**   |                         |                                     |                                     |
| Space group          | P41                     | P41                                 | P12,1                               |
| Cell dimensions      |                         |                                     |                                     |
| a, b, c (Å)          | 76.769, 76.769, 170.857 | 77.209, 77.209, 172.754             | 76.276, 61.43, 122.52               |
| α, β, γ (°)          | 90, 90, 90              | 90, 90, 90                          | 90, 90, 90                          |
| **Peak**             |                         |                                     |                                     |
| Wavelength           | 1.105                   |                                     |                                     |
| Resolution (Å)*      | 76.77–3.12 (3.232–3.12) | 77.21–2.929 (3.034–2.929)           | 62.68–3.104 (3.215–3.104)           |
| R_merge              | 0.1096 (1.143)          | 0.08105 (1.111)                     | 0.08573 (1.761)                     |
| I/σ (I)              | 10.54 (1.37)            | 12.77 (1.23)                        | 7.19 (0.83)                         |
| Cc1/2                | 0.994 (0.813)           | 0.994 (0.751)                       | 0.999 (0.626)                       |
| Completeness (%)     | 99.84 (99.66)           | 99.33 (96.16)                       | 96.34 (81.92)                       |
| Redundancy           | 6.9 (7.0)               | 6.8 (5.7)                           | 3.5 (3.3)                           |
| **Refinement**       |                         |                                     |                                     |
| Resolution (Å)       | 77.21–2.929 (3.034–2.929) | 62.68–3.104 (3.215–3.104)           |                                     |
| No. reflections      | 21,611                  | 20,047                              |                                     |
| R_work / R_free      | 19.7 / 23.5             | 22.0 / 26.9                         |                                     |
| No. atoms            | RNA: 5,082              | MG: 11                               | K: 4                                |
|                      |                         | 6                                    | –                                    |
| B factors            |                         |                                     |                                     |
| T-box                | 112.65                  | 171.16                               |                                     |
| tRNA                 | 126.34                  | 147.07                               |                                     |
| MG                   | 91.5                    | 97.5                                 |                                     |
| K                    | 125.2                   | –                                    |                                     |
| R.m.s. deviations    | Bond lengths (Å)        | 0.003                                | 0.003                               |
|                      | Bond angles (°)         | 0.69                                 | 0.76                                |

*Data in each column are from one crystal. Values in parentheses are for highest-resolution shell.
noted but not understood. We can now ascribe three functions to the linker and Stem-III: they contribute key tertiary contacts to stabilize AntiS into the aminoacylation sensing conformation, are an integral component of the tRNA binding pocket and play pivotal roles in aminoacylation sensing. Beginning with a pair of type-II A-minor interactions from A90 and G91 (Fig. 3c), the linker mandates the minor groove of the AntiS bottom helix to the opening where the tRNA terminus is pointed. Here, a GAG sequence (G94, A95 and G96) encompassing the previously noted AG motif specifies the aminoacylation status of the tRNA (Fig. 3d). Based on our structure and further inspection of T-box alignments, we propose that a RAG sequence in the linker region adjacent to Stem-III is conserved across T-boxes and maintains this essential function.

In Mtb-ileS, this RAG sequence begins with G94, which makes a sequence-specific contact to the tRNA-3'-adenosine (A77) by contributing its 2'OH to a type-II A-minor interaction with the tRNA and its N7 to an H-bond with the AntiS ribose (U125) across the base pair (Fig. 3d). A transversion mutation (R94U) at this position in tyrS reduces induction to around background levels, illustrating the importance of the purine specific contacts we observe (Fig. 3f, yellow). Following G94, A95 plays a pivotal role by mediating continuous stacking between AntiS A129 and C131 residues while H-bonding with the riboses of tRNA C76 and AntiS C130 using its Hoogsteen edge (Fig. 3e). The function of the tyrS T-box is especially sensitive to nucleotide identity at this position; an A95G mutation reduces T-box induction to near background (Fig. 3f, light blue). The AntiS ACC (129–131) sequence that interacts with linker A95 (Fig. 3e) is almost universally found downstream of the NCCA-complementary region across T-boxes. Our structure clearly shows that it serves as a platform to position the linker for aminoacylation sensing (Fig. 3e). Finally, G96 forms a long-range base pair with AntiS C131 (Fig. 3d) and positions its Hoogsteen edge to make an H-bond to the terminal 3'OH of the tRNA (Fig. 3d). This is the crux interaction for sensing uncharged tRNA because the 3’OH is strictly specified as a H-bond donor. As expected, a mutation to G96 (G96C) in the tyrS T-box precludes any induction (Fig. 3f pink). Importantly, T-box function could not be rescued with a purine-pyrimidine switch (bp_switch, G96C+C131G) that restores the linker-AntiS base pair between G96 and C131 (Fig. 3f, teal). Together, the linker region interactions with the tRNA leave no room for a side chemical moiety and are highly specific for an uncharged tRNA with a terminal adenosine containing intact 2’ and 3’ hydroxyls. Previous in vitro studies using cognate tRNAs without terminal hydroxyls or with a terminal 2-aminopurine decreased the T-box response. We can now rationalize these results as being due to the modifications disrupting the observed linker contacts.
The bottom of Stem-III inserts into the AntiS minor groove with the rest of the stem-loop extending away (Fig. 3a). Its configuration in the aminoacylation sensing module suggests it probably functions as a twist-tie that maintains tension between the tRNA binding pocket and the AntiS bottom helix. All of Stem-III’s contacts with the AntiS are mediated by residues near the base of the stem, explaining the lack of sequence or length conservation in the body of the stem-loop. The noncanonical G98•A117 pair expands the base of Stem-III, allowing it to span the minor groove of the AntiS top helix and position G98 to make cross-groove ribose contacts (Extended Data Fig. 2c,d). The following A118 residue connects Stem-III to the beginning of the AntiS bottom helix while anchoring to the third base pair of the AntiS top helix through a type-I A-minor interaction (Extended Data Fig. 2c,e). The close proximity of Stem-III and the linker allows a continuous belt of tertiary contacts, which act like a Velcro strip to patch up the tRNA binding pocket and stabilize the AntiS conformation.

Fig. 3 | Details of the aminoacylation sensing module. a, Tertiary structure of the Mtb-ileS aminoacylation sensing module bound to tRNA^{ile} (pale purple) with the linker (yellow), Stem-III (pink) and AntiS (cyan) domains. b, Enlarged tRNA^{ile} NCCA base-pairing with AntiS. Hydrogen bonds are represented by black dashes. Asterisks denote residues mutated from wild type sequence. c, Minor groove interactions between the linker and AntiS. d, Aminoacylation sensing by the linker rAG (G94, A95 and G96) sequence. e, Aminoacylation sensing pocket showing the interaction between the linker and conserved AntiS ACC (A129, C130 and C131) sequence. f, Induction of tyrS T-box after tyrosine starvation. WT (purple), WT+ (orange), G96C (pink), U160C (green), R94U (yellow), A95G (light blue) and bp_switch (teal) are shown. G96C and bp_switch overlap at the bottom. Mutants are represented with dashed lines. Error bars represent s.e.m. with n = 3 biologically independent bacterial cultures. WT and WT+ data are the same as plotted in Fig. 2b. Source data for f are available in Supplementary Table 3.

Structure of T-box interacting with a charged tRNA mimic. There are two mechanistic possibilities for how T-boxes distinguish charged and uncharged tRNAs and thereby mediate switching between sequestator and AntiS structures: (1) The aminoacylation sensing pocket may be rigid and sterically rejects the charged tRNA tail from binding or (2) the pocket may be elastic enough to accommodate the extra amino acid on the charged tRNA, but the deformation nucleates a cascade of conformational changes that facilitate switching to the sequestator base-pairing scheme. To differentiate these two models, we determined the structure of another Mtb-ileS construct bound to tRNA with a terminal 2′,3′-cyclic phosphate (tRNA^{ile-cP}) to mimic a charged tRNA. We hypothesized that the presence of a relatively small moiety at the tRNA terminus would not completely disrupt the T-box–tRNA complex but instead would allow us to capture changes indicative of the actual mechanism. The resulting 3.1 Å-resolution co-crystal structure (Fig. 4a, Extended Data Fig. 3a and Table 1) is consistent with the second mechanistic model—the
gain from continuous WC pairing with the tRNA tail offsets the penalty from steric hindrance-induced disruption of tertiary contacts. Indeed, complex formation between $\text{Mtb-ileS}$ and tRNA$\text{Ile}$-cP is comparable to uncharged tRNA (Extended Data Fig. 3c,d).

In the structure, the entire tRNA NCCA tail is accommodated with continuous WC pairing in the AntiS binding pocket (Extended Data Fig. 3b). Steric hindrance from the extra terminal moiety is absorbed locally by the linker region, leading to the disruption of its tertiary interaction with the sensing core and readjustments thereafter. As a result, the important long-range G96–C131 pairing is broken; G96 undergoes a $\text{syn}$-to-$\text{anti}$ flip and makes alternative H-bonds (Fig. 4c,d). In fact, the entire linker appears to be peeling away: G94 is shifted out of its type-II A-minor interaction with the tRNA terminus (Fig. 4c,d), and B-factors for the linker at the start of its interaction with the AntiS are much higher in this structure (Fig. 4a,b), indicating that it is no longer stably docked in its preferred conformation. Additionally, the weakened linker-AntiS interaction increases the conformational entropy of the nearby Stem-IIA/B pseudoknot (Fig. 4a,b). Electron density corresponding to parts of the linker region and Stem-IIA/B diminishes, suggesting these residues are disordered on binding to the cyclic phosphate-containing tRNA (Fig. 4e,f). It is worth noting that even the smallest amino acid is bigger than a cyclic phosphate; a charged tRNA would introduce more severe steric clashing, particularly against G96, and would likely push the entire GAG motif out of the pocket. The complete ejection of the GAG sequence from the pocket could be the first domino to fall in a series of events that nucleate alternative base-pairing. Overall, our data suggest that small changes to the linker in the aminoacylation sensing pocket can nucleate large global rearrangements.
The uncharged tRNA was shown to induce even longer-lived binary complexes, which are probably stabilized by the linker contacts revealed in our structure. Due to the limitations of the labeling scheme, the existing studies have not resolved the conformational differences and interconversion rates among different states. This limits our ability to draw more definitive conclusions. The improved spatial resolution from our structural study could stimulate future single-molecule based mechanistic studies to probe deeper into the tRNA sensing and gene regulatory mechanisms of the T-box.

By synthesizing previous structural information with the current study, we can generate a family portrait depicting all three classes of T-boxes: glycine, typical and atypical (Fig. 5a–c). These snapshots allow us to begin rationalizing the differences among the three classes. Glycine T-boxes probably shed most of the mid-region because the GC-rich codon-anticodon interaction does not require further stabilization. Likewise, transcriptional T-boxes may rely on the distal stacking module in Stem-I to improve the kinetics of tRNA recruitment, whereas this is not as critical a consideration for the thermodynamically driven translational T-boxes. As stated above, the T-box and the ribosome are the only RNA molecules capable of decoding and aminoacylation sensing. An intriguing hypothesis is that the ancestors of T-boxes may have played essential roles in translation in the primordial RNA world, possibly serving as a primitive tRNA synthetase. The high-resolution information from this study could open the door to a more detailed investigation of these hypotheses.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, and statements of code and data availability are available at https://doi.org/10.1038/s41594-019-0327-6.

Received: 8 July 2019; Accepted: 7 October 2019; Published online: 18 November 2019

Fig. 5 | T-box family portrait. T-box classes are distinguished based on regulatory mechanism and characteristics of tRNA interaction. a, Glycine T-box class. Top: secondary structure representation. Bottom: tertiary structure model combining glyQS Stem-I-tRNA structure (PDB 4LCK) with aminoacylation sensing module from Mtb-ileS. Linker represented with yellow dashes. b, Typical T-box class. Top: secondary structure representation. Bottom: tertiary structure model combining glyQS Stem-I-tRNA (PDB 4MGN) and Mtb-ileS. c, Atypical T-box class. Top: secondary structure representation. Bottom: overall structure of Mtb-ileS.

Discussion
The discovery that T-box riboregulators are able to simultaneously decode and sense the aminoacylation status of tRNA has fascinated researchers for the past three decades. Such sophisticated tasks were thought to be reserved for the megadalton ribosome complex. Our T-box–tRNA structures finally provide a complete high-resolution explanation of both processes. The structures best support a stepwise model for tRNA sensing. In the first step, Stem-I and Stem-II capture the tRNA and specify the anticodon irrespective of its aminoacylation status (Extended Data Fig. 4a). The next step is aminoacylation sensing, which our structures suggest proceeds through an intermediate where the AntiS interacts with the NCNA sequence of the tRNA while the linker samples the aminoacylation status (Extended Data Fig. 4b). An important concept from our work is that the T-box may be capable of accommodating both versions of its ligand at this step, although they ultimately progress to different fates (Extended Data Fig. 4c). Interaction with the uncharged tRNA leads to stable complex formation, which traps the T-box in the AntiS conformation (Extended Data Fig. 4d). Alternatively, a charged tRNA is guided into the aminoacylation sensing pocket through WC pairing to the tRNA-3′-tail, however, the extra moeity at the tail dislodges the linker region, which probably nucleates subsequent conformational changes (Extended Data Fig. 4e). We therefore conclude that charged tRNA binding leads to a short-lived intermediate and the aminoacylated tRNA terminus may kinetically facilitate the AntiS-to-sequestator transition.

This model is in line with the recent single-molecule studies that propose hierarchical tRNA recognition consisting of two distinct binding states. One of these studies also showed the emergence of a significantly longer-lived binding state for both charged and uncharged tRNA when a full-length T-box rather than the decoding-only fragment was used. This is suggestive of the intermediate step we propose, where the tRNA terminus may be bound and examined by the aminoacylation sensing module.
13. Seliverstov, A. V., Putzer, H., Gelfand, M. S. & Lyubetsky, V. A. Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol.* 5, 54 (2005).

14. Sherwood, A. V., Grundy, F. J. & Henkin, T. M. T box riboswitches in Actinobacteria: translational regulation via novel tRNA interactions. *Proc. Natl Acad. Sci. USA* 112, 1113–1118 (2015).

15. Rollins, S. M., Grundy, F. J. & Henkin, T. M. Analysis of cis-acting sequence and structural elements required for antitermination of the *Bacillus subtilis tyrS* gene. *Mol. Microbiol.* 25, 411–421 (1997).

16. Saad, N. Y. et al. Two-codon T-box riboswitch binding two tRNAs. *Proc. Natl Acad. Sci. USA* 110, 12756–12761 (2013).

17. Shepherd, J. & Ibba, M. Bacterial transfer RNAs. *FEMS Microbiol. Rev* 39, 280–300 (2015).

18. Agris, P. F., Vendeix, F. A. P. & Graham, W. D. tRNA’s wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* 366, 1–13 (2007).

19. Putzer, H., Condon, C., Brechemier-Baey, D., Brito, R. & Grunberg-Manago, M. Transfer RNA-mediated antitermination in vitro. *Nucleic Acids Res.* 30, 3026–3033 (2002).

20. Ogle, J. M. et al. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292, 897–902 (2001).

21. Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* 111, 721–732 (2002).

22. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* 484, 256–259 (2012).

23. Vitreschak, A. G., Mironov, A. A., Lyubetsky, V. A. & Gelfand, M. S. Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA* 14, 717–735 (2008).

24. Kalvari, I. et al. Non-coding RNA analysis using the Rfam database. *Curr. Protoc. Bioinformatics.* 62, e51 (2018).

25. Henkin, T. M., Glass, B. L. & Grundy, F. J. Analysis of the *Bacillus subtilis tyrS* gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. *J. Bacteriol.* 174, 1299–1306 (1992).

26. Grundy, F. J., Yousef, M. R. & Henkin, T. M. Monitoring uncharged tRNA during transcription of the *Bacillus subtilis glyQS* gene. *J. Mol. Biol.* 346, 73–81 (2005).

27. Murakami, H., Ohita, A., Goto, Y., Sako, Y. & Suga, H. Flexizyme as a versatile tRNA acylation catalyst and the application for translation. *Nucleic Acids Symp. Ser. (Oxf.)* 50, 35–36 (2006).

28. Golden, B. L., Chen, J. & Luptak, A. Ribozyme with tRNA synthetase activity and methods of manufacturing and using the same. US patent 28,160,348,108 (2016).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

Constructs, plasmids and strains. Strains and plasmids are described in Supplementary Table 1. The sequences of the constructs used in the biochemical and crystallization experiments are documented in Supplementary Table 2.

RNA preparation. RNA constructs were cloned and produced as described previously\(^3\). T-box and tRNA sequences were cloned into the pUC19 plasmid downstream of a 17 RNA polymerase promoter and upstream of the hepatitis delta virus ribozyme (HDV) to produce homogeneous ends. The hammerhead ribozyme sequence was added upstream of tRNA\(^{3\prime}\) sequences. Guanosine residues were added at the beginning of each construct to increase expression and nonconserved loops were mutated to GAAA tetraloops to facilitate crystallization as noted in Supplementary Table 1. Plasmid templates for transcription of T-box constructs were prepared with Quagen MegaPrep kits and linearized by restriction digestion after the HDV sequence. The transcription template for tRNA\(^{3\prime}\)-3′-hydroxyl was produced by PCR with a 2′-O-methylated reverse primer. The plasmid transcription template for tRNA\(^{3\prime}\)-3′-cyclic phosphate was produced as described for T-box constructs. Transcription reactions (10 µl) in vitro were performed as described previously\(^4\). RNA was gel purified by urea denaturing polyacrylamide gel electrophoresis (PAGE). The bands were excised from the gel, crushed and heating at 95 °C for 2 min, cooling to 65 °C, adding 10 mM MgCl\(_2\) and equal volume was produced by PCR with a 2′-O-methylated reverse primer. The plasmid sequence was added upstream of tRNAIle sequences. Guanosine residues were prepared with Qiagen MegaPrep kits and linearized by restriction digestion sites were: 20 µM Mtb-ileS\(^{3\prime}\) tRNA-3′-cyclic phosphate were: 20 µM Mtb-ileS\(^{3\prime}\) tRNA-3′-hydroxyl were: 20 µM RNA in a mother liquid of 10% PEG 4,000, 50 mM HEPES pH 7.0, 150 mM KCl, 40 mM MgCl\(_2\), at 16 °C with a 1:1 (RNA:mother liquid) drop ratio. Optimized conditions for Mtb-ileS tRNA-3′-cyclic phosphate were: 20 µM RNA in a mother liquid of 10% PEG 4,000, 50 mM sodium caccodylate pH 6.5, 200 mM KCl, 15 mM MgCl\(_2\), at 16 °C with 1:1 (RNA:mother liquid) drop ratio. For phasing, 10 mM iridium hexamine and 20% ethylene glycol were added to the crystals overnight at 4 °C before flash freezing in liquid N\(_2\).

Data were collected at the Advanced Photon Source (APS) 24-ID-C beamline, Northeastern Collaborative Access Team (NE-CAT). Datasets were processed using HKL-2000 (ref. 19) or XDS\(^2\) as part of NE-CAT’s RAPID pipeline. The Mtb-ileS tRNA-3′-hydroxy structure was phased by the single-wavelength anomalous dispersion (SAD) method. The iridium hexamine sites were located by direct methods using SHELX\(^2\). These sites were fed to PHENIX AutoSol and further refined therein\(^2\). A model was built into the experimental map using the program COOT\(^4\). The partial model was then used in MR-SAD to improve experimental map quality. Iterative rounds of such efforts eventually led to the completion of the T-box–tRNA model. Structural refinement was done in phenix.refine\(^3\). Multiple rounds of simulated annealing refinement were carried out to remove possible model bias. The final tRNA-3′-cyclic phosphate model was built by molecular replacement of the SAD model into a higher resolution native dataset using PHENIX Phaser-MR followed by additional refinement. Simulated annealing omit maps were generated to ensure the correct interpretation of weaker electron densities. The tRNA-3′-cyclic phosphate structure was solved by molecular replacement using Phaser-MR and the model was built with subsequent rounds of rigidbody, positional and temperature factor refinement.

T-box induction assay. The induction assay was adapted from Grundy and Henkin\(^1\). Briefly, the B. subtilis tyrS leader (native promoter and T-box) was cloned upstream of lacZ in the pDG1661 integration vector\(^1\). The reporter construct was transformed into a tyrosine auxotroph B. subtilis strain (BKE22610) from the Bacillus Genetic Stock Center and integrated into the amyE locus by homologous recombination. Mutant T-box sequences were generated using site-directed mutagenesis of the wild type (WT) tyrS reporter plasmid and transformed into the BKE22610 strain.

For T-box induction experiments, cells were grown overnight in Spizizen minimal medium\(^1\) with 50 µg ml\(^{-1}\) tyrosine and the appropriate antibiotics at 37 °C. Cells from the overnight cultures were subcultured into the same media and grown to mid-log phase. The cells were pelleted by centrifugation and resuspended in Spizizen minimal medium with or without tyrosine supplementation. Cells (1 ml) were harvested at 0, 1, and 3 h timepoints after resuspension. Beta-galactosidase activity was measured as described previously\(^1\). All measurements were done in triplicate.

Reporting Summary. Further information on the research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates and structure factors for Mtb-ileS\(^{3\prime}\) tRNA-OH native and Mtb-ileS\(^{3\prime}\) tRNA-cP native have been deposited in the Protein Data Bank with the accession codes 6JPF and 6UZH, respectively. Source data for Figs. 2b and 3f are available in Supplementary Table 3. Other data are available upon reasonable request.

References

29. Ke, A. & Doudna, J. A. Crystallization of RNA and RNA–protein complexes. Methods 34, 408–414 (2004).

30. Otwinoski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).

31. Kabsch, W. XDS. Acta Crystallogr. D 66, 125–132 (2010).

32. Pape, T. et al. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. J. Appl. Crystallogr. 37, 843–844 (2004).

33. Sheldrick, G. M. et al. Experimental phasing with SHELXD/E: combining chain tracing with density modification. Acta Crystallogr. D 66, 479–485 (2010).

34. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D 66, 213–221 (2010).

35. Emney, P. et al. Features and development of Coot. Acta Crystallogr. D 66, 486–491 (2010).

36. Guérout-Fleury, A.-M., Frandsen, N. & Stragier, P. Plasmids for ectopic integration in Bacillus subtilis. Gene 180, 57–61 (1996).

37. Anagnostopoulos, C. & Spizizen, J. Requirements for transformation in Bacillus subtilis. J. Bacteriol. 81, 741–746 (1961).

38. Chen, L., James, L. P. & Helmann, J. D. Metalloregulation in Bacillus subtilis: isolation and characterization of two genes differentially repressed by metal ions. J. Bacteriol. 175, 5428–5437 (1993).

39. Miller, J. H. Experiments in Molecular Genetics (Cold Spring Harbor Laboratory, 1972).

Acknowledgements

The authors thank A.J. Sachla and J.D. Helmann for advice on the β-galactosidase assay, the Bacillus Genetic Stock Center, the Nicholson Lab for the use of their spectrophotometer and the Fromme Lab for the use of their microscope. This work is supported by the National Institutes of Health (NIH) (grant nos. GM118174 and GM116632 to A.K.). This work is based on research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (grant no. P30 GM124165). The Pilatus 6M detector on the 24-ID-C beamline is funded by a NIH-ORIP HEI grant (no. S10 RR029205). This research used the resources of the Advanced Photon Source, a U.S. Department of Energy Office of Science User Facility operated for the US Department of Energy Office of Science by Argonne National Laboratory (contract no. DE-AC02-06CH11357).

Author contributions

R.A.B., J.C.G. and A.K. designed the research. R.A.B. was the main contributor to the structure-function analysis. J.C.G. contributed to experimental design and structural refinement. R.A.B. and A.K. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0327-6.

Correspondence and requests for materials should be addressed to A.K.

Peer review information Anke Sparmann was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
Extended Data 1 | Size exclusion chromatographic analysis and mutagenesis guide. a, Size exclusion chromatogram of Mtb-ileS folded with (blue) and without (orange) tRNA\textsuperscript{+}. b, Urea denaturing PAGE gel of peak fractions from T-box + tRNA peak. c, Secondary structure model of tyrS T-box from \textit{Bacillus subtilis} (Bsub-tyrS). Boxes highlight location and identity of each mutation. d, Conversion table for residue notation in Mtb-ileS and Bsub-tyrS. Fold difference from WT is the difference between the average of three induction measurements at 1 hour for WT versus mutation of indicated residue.
Extended Data 2 | Details of AntiS-tRNA and AntiS-Stem-III interactions. a, Detail of G132-U160 wobble pair in the aminoacylation sensing pocket showing AntiS (cyan), tRNA (pale blue) and linker (yellow). Hydrogen bonds represented by black dashes. Magnesium ion represented by green sphere. b, View of AntiS bound to tRNA showing tRNA 3′-end exposed in a large opening. c, Stem-III (pink) and AntiS minor groove interactions. d, Details of G98 interactions with AntiS. e, Details of A118 type-I A-minor interaction with AntiS.
Extended Data 3 | Structural details and size exclusion chromatography of Mtb-ileS with tRNA\textsuperscript{\textit{\textalpha}}-cP. a, Tertiary structure model of Mtb-ileS in complex with tRNA\textsuperscript{\textit{\textalpha}}-cP. b, Detail of AntiS interaction with tRNA\textsuperscript{\textit{\textalpha}}-cP NCCA sequence. Hydrogen bonds represented by black dashes. Asterisk indicates residues mutated from wildtype sequence. c, Size exclusion chromatograms of Mtb-ileS folded with tRNA\textsuperscript{\textit{\textalpha}}-cP (blue) and with tRNA\textsuperscript{\textit{\textalpha}}-OH (orange). d, Urea denaturing PAGE gel of peak fractions from T-box + tRNA\textsuperscript{\textit{\textalpha}}-cP peak.
Extended Data 4 | Mechanistic diagram of atypical T-box translational regulation. a, tRNA recruitment and decoding. b, Pseudoknot formation positions AntiS to interact with tRNA NCCA sequence. c, Transient intermediate where aminoacylation sensing module interacts with uncharged (top) and charged (bottom) tRNA. d, Favorable interactions between uncharged tRNA and linker locks T-box into ON conformation. Exposed Shine–Dalgarno (SDS) allows translation initiation. e, Steric clashing between charged tRNA and linker unravels the aminoacylation sensing module leading to alternative base pairing. Sequestrator formation prevents ribosome access to SDS and prevents translation.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referrees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed
---|---
☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒ | The statistical test(s) used AND whether they are one- or two-sided
    Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☒ | A description of all covariates tested
☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒ | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
    Give P values as exact values whenever suitable.
☒ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒ | Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

| Data collection | OD 420 and 550 measurements for beta-gal assay were collected using a Thermo Scientific Nanodrop 2000c spectrophotometer and corresponding software. |
|----------------|----------------------------------------------------------------------------------------------------------------------------------|
| Data analysis  | HKL-2000 and XDS were used to process collected diffraction data. SHELEX and PHENIX (version 1.14-3260-000) Autosol were used for SAD phasing. COOT and PHENIX Refine were used for model building and refinement. PHENIX Phaser-MR was used for molecular replacement. Microsoft Excel was used to compute and graph Miller units for beta-gal assay. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Structure data will be deposited to the Protein Data Bank at rcsb.org with accession codes 6UFG for Mtβ-lleS_SRNA-OH and 6UFH for Mtβ-lleS_SRNA-cP.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑️ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size  Sample size for beta-gal assay experiments was determined according to established convention. Diffraction data was collected in a manner to ensure completeness near 100% and high multiplicity.

Data exclusions  Refinement statistics from the dataset used for initial SAD phasing are not included in the data table. This model was only used for molecular replacement into a higher resolution native dataset where refinement was completed.

Replication  Three biological replicates were done for each beta-gal assay experiment.

Randomization  Randomization was not relevant to the beta-gal assay because replicates can account for differences between samples outside of the established variables. During refinement, a random set of the diffraction data is set aside for comparison to the model in order to generate the R-free statistic.

Blinding  Blinding was not relevant to the study because results could not be influenced by subjective bias. Bias introduced during model building is accounted for in the established statistical analysis of crystal structures. This data is included in Table 1.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems  n/a  Involved in the study

☑️ Antibodies
☑️ Eukaryotic cell lines
☑️ Palaeontology
☑️ Animals and other organisms
☑️ Human research participants
☑️ Clinical data

Methods  n/a  Involved in the study

☑️ ChiP-seq
☑️ Flow cytometry
☑️ MRI-based neuroimaging