

Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions

Shuang Li^{1,9}, Zhaoming Su^{2,9}, Jean Lehmann³, Vassiliki Stamatopoulou⁴, Nikoleta Giarimoglou⁴, Frances E. Henderson¹, Lixin Fan⁵, Grigore D. Pintilie², Kaiming Zhang^{6,2}, Muyuan Chen⁶, Steven J. Ludtke⁶, Yun-Xing Wang^{5,7}, Constantinos Stathopoulos⁴, Wah Chiu^{2,8*} and Jinwei Zhang^{6,1*}

Amino acid availability in Gram-positive bacteria is monitored by T-box riboswitches. T-boxes directly bind tRNAs, assess their aminoacylation state, and regulate the transcription or translation of downstream genes to maintain nutritional homeostasis. Here, we report cocrystal and cryo-EM structures of *Geobacillus kaustophilus* and *Bacillus subtilis* T-box-tRNA complexes, detailing their multivalent, exquisitely selective interactions. The T-box forms a U-shaped molecular vise that clamps the tRNA, captures its 3' end using an elaborate 'discriminator' structure, and interrogates its aminoacylation state using a steric filter fashioned from a wobble base pair. In the absence of aminoacylation, T-boxes clutch tRNAs and form a continuously stacked central spine, permitting transcriptional readthrough or translation initiation. A modeled aminoacyl disrupts tRNA-T-box stacking, severing the central spine and blocking gene expression. Our data establish a universal mechanism of amino acid sensing on tRNAs and gene regulation by T-box riboswitches and exemplify how higher-order RNA-RNA interactions achieve multivalency and specificity.

Amino acids are among the most ubiquitous and essential nutrients for all known cellular life. Diverse mechanisms of amino acid sensing and regulation have evolved to appropriately supply the translation machinery and other cellular processes such as cell wall synthesis¹. Free amino acids can be sensed by direct binding to protein sensors, including Sestrins², CASTORs³, and TRAP⁴, or RNA sensors, as exemplified by the glycine, lysine and glutamine riboswitches^{5,6}. Alternatively, amino acids can also be sensed in conjugated forms, especially in the form of aminoacyl tRNAs, which are the immediate substrates of protein synthesis. Sensing amino acids on the tRNAs can take advantage of the well-established coupling specificities between tRNAs and their aminoacyl-tRNA synthetases⁷⁻⁹. A number of cellular proteinaceous systems sense aminoacylation on tRNAs to detect nutrient limitation, such as translating ribosomes¹⁰ and ribosome-associated ppGpp synthase RelA in bacteria^{11,12} and eIF2 α kinase GCN2 in eukaryotes¹³. Interestingly, mRNA elements have also been proposed to sense aminoacylation on tRNAs, such as the T-box riboswitches that are widespread in Gram-positive bacteria. However, direct sensing of tRNA aminoacylation by another RNA without any proteins, including the transcribing RNA polymerase (RNAP), has not been demonstrated¹⁴.

More than one thousand T-box riboswitches have been identified, including in most key human pathogens, such as *Bacillus*, *Clostridium*, and *Staphylococci*¹⁵⁻¹⁸. They control essential genes in amino acid biosynthesis, transport and tRNA aminoacylation^{15,16,19}. T-boxes are attractive antimicrobial targets that exhibit drastically

reduced antibiotic resistance²⁰, presumably because their universal core structure allows the same compounds to simultaneously target multiple T-boxes, and thus, manifold metabolic pathways. T-boxes share a phylogenetically conserved architecture of two or three domains (Fig. 1a). An obligate 5' stem I domain initially selects a cognate tRNA through base-pairing interactions to the tRNA anticodon and platform-stacking interactions to the tRNA elbow²¹⁻²³ (Fig. 1a). An intervening stem II domain of unknown structure and function plays important yet uncharacterized roles in most T-boxes^{15,16}. Curiously, the stem II domain is absent from the most-studied glycine-specific *glyQ* T-boxes. Next, a 3' antiterminator or antisequester domain supposedly probes the docked tRNA 3' end to sense aminoacylation and directs RNA conformational switching that governs transcription attenuation^{24,25} or translation initiation^{14,26,27}. However, the nature of the T-box mRNA-tRNA interactions, especially those involving the tRNA 3' end, is poorly understood, due to the lack of high-resolution structural information. Two recent low-resolution studies employing small-angle X-ray scattering (SAXS) led to divergent structural models and do not provide insights into the mechanism of amino acid sensing^{28,29}. The sensory and regulatory functions of the T-boxes are anchored by a presumed ability to sense aminoacylation on a tRNA directly. This complex function is generally accomplished on the ribosome by large, specialized protein sensors, such as RelA^{11,12} and Gcn2 (ref. ¹³). Recently, a new class of T-box riboswitches were identified in Actinobacteria that regulate the initiation of translation instead of transcription termination²⁷. These translational T-boxes harbor a

¹Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, USA. ²Department of Bioengineering and Department of Microbiology and Immunology, James H. Clark Center, Stanford University, Stanford, CA, USA. ³Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Campus Paris-Saclay, Gif-sur-Yvette, France. ⁴Department of Biochemistry, School of Medicine, University of Patras, Patras, Greece. ⁵Small-Angle X-ray Scattering Core Facility, Center for Cancer Research of the National Cancer Institute, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc, Frederick, MD, USA. ⁶Verna Marrs and McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA. ⁷Structural Biophysics Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD, USA. ⁸Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo Park, Stanford University, Stanford, CA, USA. ⁹These authors contributed equally: Shuang Li, Zhaoming Su. *e-mail: wahc@stanford.edu; jinwei.zhang@nih.gov

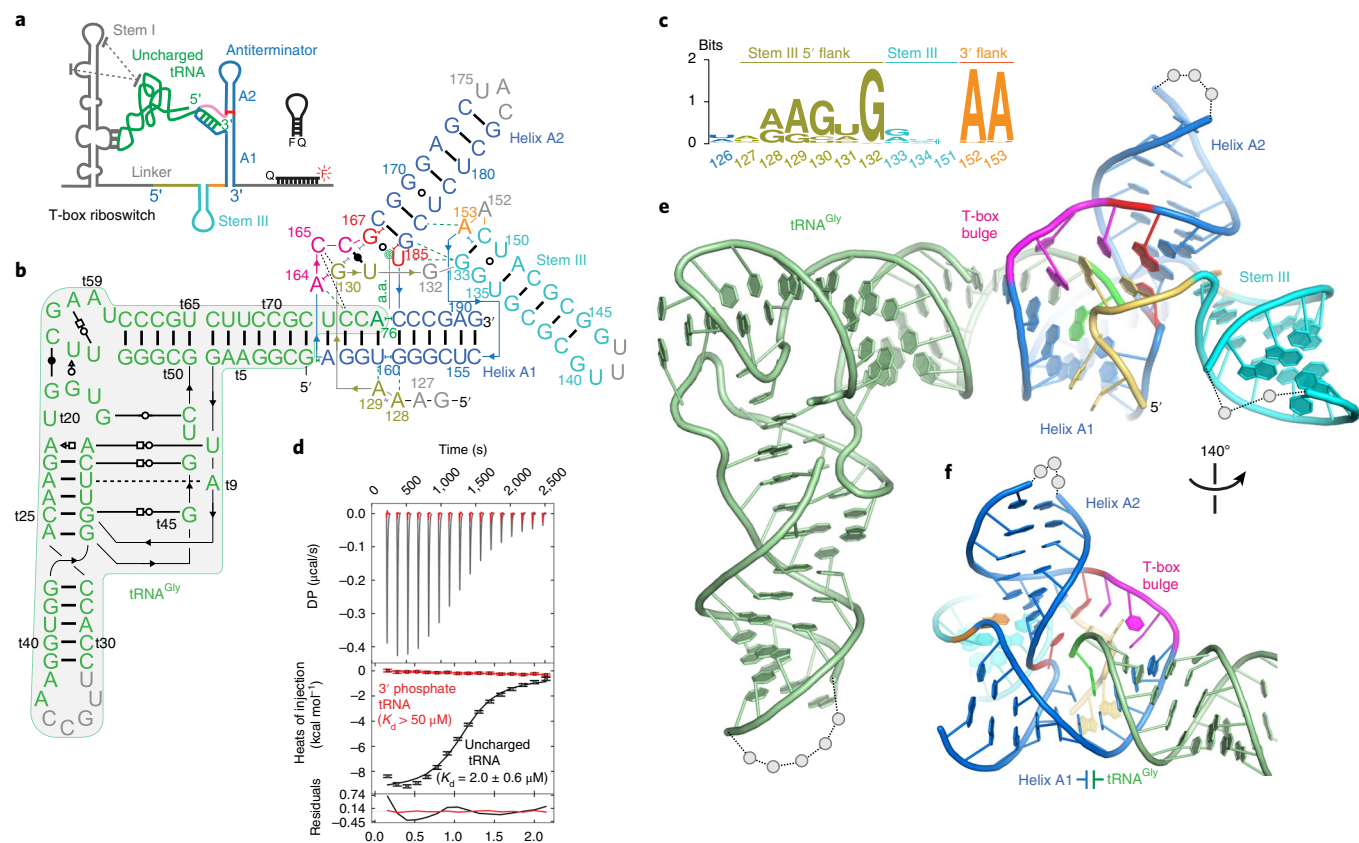


Fig. 1 | Overall structure of the T-box discriminator–tRNA complex. **a**, *glyT* T-box riboswitch–tRNA^{Gly} complex. Dotted lines denote stacking interactions between the tRNA elbow and the distal T loops of stem I. The molecular beacon harboring a 5'-Cy5 fluorophore (F) and 3'-dabcyl quencher (Q) to measure transcription readthrough is rendered in black. **b**, Sequence and secondary structure of the cocrystallized *G. kaustophilus glyT* T-box discriminator and tRNA^{Gly}. Noncanonical base pairs are denoted by Leontis–Westhof⁶⁰ symbols. Lines with embedded arrowheads denote chain connectivity. Green and gray dotted lines denote the purine minor groove and other hydrogen bonding interactions, respectively. The tRNA (shaded) is conventionally numbered, where a 't' precedes tRNA residue numbering. **c**, Sequence logo of stem III and its flanking regions. **d**, Representative ITC analysis of discriminator binding by uncharged (black) and 3'-phosphate (red) tRNAs. The error bars are statistical estimates of the uncertainties of individual injection heats calculated using NITPIC (Methods). The K_d value is the mean and s.d., $n=3$ biologically independent samples. **e, f**, Front and rear views of the overall cocystal structure. tA76 is highlighted in bright green. Parallel lines connected to orthogonal lines denote stacking. Residues that are not modeled are indicated as gray spheres.

more divergent stem I domain and a similar 3' amino acid-responsive domain that either sequesters or exposes the Shine–Dalgarno sequence to control translation initiation^{26,27}. The conserved secondary structure of the 'antisequester' domain is nearly identical to that of the transcriptional antiterminator, suggesting a unified mechanism of amino acid sensing and conformational switching²⁷.

To investigate how the T-box riboswitches sense and respond to tRNA aminoacylation, we first define a minimal T-box region that selectively binds an uncharged tRNA, which we term the 'T-box discriminator'. We then elucidate how the T-box discriminator, merely 66 nucleotides (nts) in size, directly senses tRNA aminoacylation by solving its 2.66-Å resolution cocystal structure bound to an uncharged tRNA. Furthermore, we also determine a 4.9-Å resolution cryo-EM structure of a full-length T-box riboswitch–tRNA complex, which reveals how the U-shaped T-box mRNA acts as a molecular vise to clamp its cognate uncharged tRNA to construct a stable antitermination complex. Together, our structural and functional analyses define a universal mechanism of amino acid sensing and gene regulation that operates in all known T-box riboswitches.

Results

A compact T-box discriminator domain directly senses tRNA aminoacylation. The T-box stem I initially docks a cognate tRNA through contacts to its anticodon and elbow and sets the stage for

the T-box 3' region to interact with the tRNA 3' end²². Located 3' to stem I is a single-stranded linker, the highly variable stem III of unknown function, and an antiterminator domain consisting of two helices (A1 and A2) flanking a 7-nucleotide (nt) bulge^{15,16} (Fig. 1a and Extended Data Fig. 1). Surprisingly, the antiterminator alone, long assumed to be sufficient for tRNA 3'-end binding and aminoacylation sensing, did not appreciably bind tRNAs, despite exhaustive trials (Extended Data Fig. 2). This finding suggested that additional T-box sequences required for tRNA binding may be missing. Located immediately upstream of the antiterminator is a stem III hairpin that is ubiquitously present in all known T-boxes, but the length and sequences of its stem and loop are highly variable (Extended Data Fig. 1). Stem III was proposed to be a part of the flexible linker, to provide a RNAP pause site³⁰ or to stack underneath helix A1 to stabilize the antiterminator³¹. When we folded the variable stem III and removed it from the sequence alignment, we discovered that the sequences that flank stem III are actually highly conserved, implying functional importance. Both flanks of stem III are purine rich and conform to a motif of 5'-RRRxG-stem III-AA-3', where 'x' denotes any nucleotide (Fig. 1b,c and Extended Data Fig. 1). Remarkably, appending stem III and its flanking sequences to the binding-incompetent antiterminator conferred robust tRNA binding (Extended Data Fig. 2). We further show that the conserved purines are required for tRNA binding, as their deletions or

substitutions with uridines abolished tRNA binding. Crucially, this binding is selective for uncharged tRNA, as both binding and transcription readthrough were inversely correlated with the molecular volume of the tRNA 3' end, supporting a proposed steric occlusion mechanism¹⁴ (Extended Data Fig. 2). Using isothermal titration calorimetry (ITC), we further showed that the mere addition of a phosphate group (approximately the same size as glycine) to the tRNA 3' end completely abolished binding (Fig. 1d). We conclude that stem III and its flanking purines collaborate with the adjacent antiterminator to form a single functional unit that selectively binds uncharged tRNAs and discriminates against charged tRNAs, and we thus named the region the 'T-box discriminator'.

Overall structure of the T-box discriminator–tRNA complex.

Next, we elucidated how the T-box discriminator, merely 66 nts in size, senses tRNA aminoacylation without proteins. We cocrystallized the *G. kaustophilus glyQ* T-box discriminator in complex with its cognate uncharged tRNA^{Gly} and solved the structure at 2.66-Å resolution (Table 1, Extended Data Fig. 3 and Supplementary Movie 1). The discriminator consists of three RNA helices (stem III, helix A1 and helix A2) and three single-stranded regions (both flanks of stem III and the T-box bulge; Fig. 1e,f). The combined stability of helices A1 and A2 dictates the transcriptional outcome, because the 3' strands of both helices also form the 5' strand of the terminator hairpin. This arrangement is similar to most small molecule-binding riboswitches, wherein the stability of a proximal 'P1' helix determines the genetic outcome^{32–35}. The difference is that with conventional riboswitches, small-molecule metabolites directly bind and modulate P1 stability, whereas in T-boxes, it is the binding of a macromolecule—a tRNA—that controls the stability of the antiterminator. Our cocrystal structure revealed Watson–Crick base pairing between the tRNA 3'-UCCA terminus and the 5'-AGGU tetranucleotide of the T-box bulge (Fig. 1b,e,f), consistent with previous covariation, genetic and biochemical analyses^{15,16,24}. Crucially, this intermolecular helix stacks coaxially with the tRNA acceptor stem and helix A1 of the antiterminator (Fig. 1f), thus allowing the tRNA to stabilize the antiterminator, driving transcription readthrough. This functionally critical tRNA–antiterminator stacking is further supported by the nearly complete fluorescence quenching of a 2-aminopurine positioned at the tRNA 3'-end and of a pyrrolocytosine located on helix A1 on the T-box side¹⁴. Helix A2 of the discriminator is positioned at an ~80° angle from the tRNA–helix A1 RNA stack, such that its terminal base pair abuts the tRNA 3'-OH and is primarily responsible for sensing tRNA aminoacylation (Fig. 1f, red base pair). Finally, the base of stem III is attached to the proximal region (as opposed to the distal loop region) of helix A2, such that its flanking conserved purines closely track the extended minor groove and make key stabilizing interactions (Fig. 1e).

A tandem adenosine-minor groove 'latch' reinforces the tRNA–discriminator interface. The 5' end of the T-box discriminator features two conserved adenosines (A128 and A129) and tracks the minor groove of the tRNA–antiterminator duplex (Fig. 2a–d). Both adenosines make hydrogen bonds across the groove, thus stabilizing the duplex (Fig. 2b,c). A128 and A129 are not only covalently linked but also stack with each other (Fig. 2a). Because each adenosine interacts with one side of the tRNA–T-box stacking interface, the A128–A129 stacking interaction acts as a latch or staple to reinforce the functionally crucial stacking (Fig. 2d). Thus, this pair of adenosines provide both lateral and axial stabilization to the tRNA–mRNA interface. Furthermore, A129 packs its nucleobase against the ribose of G161 to lend additional stability (Fig. 2a). A similar nucleobase–ribose packing interaction was observed between a cytosine in stem I (C64) and tRNA U20 near the elbow²². Substituting either A128 or A129 with a uridine drastically reduced

Table 1 | Data collection and refinement statistics.

	SAD Phasing	Model Complex (PDB 6PMO)	Native
Data collection			
Space group	C222 ₁	C222 ₁	C222 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	95.01, 118.41, 134.09	44.85, 139.91, 135.99	45.05, 139.55, 134.97
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	0.9218	1.0000	0.9793
Resolution (Å) ^a	74.10–4.53 (5.06–4.53)	35.0–2.66 (2.75–2.66)	67.49–2.84 (2.94–2.84)
<i>R</i> _{merge} (%) ^a	21.3 (90)	6.2 (115)	7.6 (80)
<i>I</i> / σ (<i>I</i>) ^a	24.0 (12.0)	16.9 (1.4)	11.9 (1.8)
CC _{1/2}	0.995 (0.993)	0.998 (0.698)	1.000 (0.465)
CC*	0.999 (0.998)	1.000 (0.907)	1.000 (0.797)
Completeness (%) ^a	100.0 (100.0)	99.5 (97.6)	99.8 (100.0)
Redundancy ^a	56.8 (58.5)	6.0 (5.9)	6.1 (6.0)
Refinement			
Resolution (Å) ^a		35.0–2.66 (2.75–2.66)	69.78–2.84 (2.94–2.84)
No. reflections ^b		12724 (1215)	10437 (1032)
<i>R</i> _{work} / <i>R</i> _{free} (%) ^a		22.7 (37.9) / 27.2 (36.5)	25.3 (38.9) / 27.1 (52.4)
No. atoms		2,797	2,880
RNA		2,762	2,848
Ions		17	32
Water		18	0
Mean <i>B</i> factors (Å ²)		89.2	98.3
RNA		88.9	97.5
Mg and Ir ions		176.6	170.1
Water		56	–
R.m.s. deviations			
Bond lengths (Å)		0.004	0.003
Bond angles (°)		0.77	1.50
Maximum likelihood coordinate precision (Å)		0.47	0.48

^aValues in parentheses are for the highest-resolution shell. ^bValues in parentheses are for the cross-validation set. The SAD phasing dataset contains merged data from nine crystals of the wild-type T-box sequence. The single datasets of the model complex and native crystals were from crystals of the A156C::U189G variant T-box, which exhibited improved diffraction limits.

tRNA-mediated readthrough (Fig. 2h and Extended Data Fig. 4). In stark contrast, replacing four immediately preceding adenosines (A124–127) not involved in interactions with uridines had no effect (Fig. 2h). Tandem, stacked adenosines are frequently observed to laterally stabilize RNA helices and heteroduplexes across their minor grooves³⁶. One prominent example is the highly conserved ribosomal A1492–A1493 dinucleotide in the A site, which, in conjunction with G530, latches the codon–anticodon duplex and triggers 30S closure to ensure decoding fidelity^{37,38} (Extended Data Fig. 5). The unusual placement of the A128–A129 latch across an important tRNA–mRNA interface locally bolsters RNA–RNA interactions to rheostat gene expression (Fig. 2d).

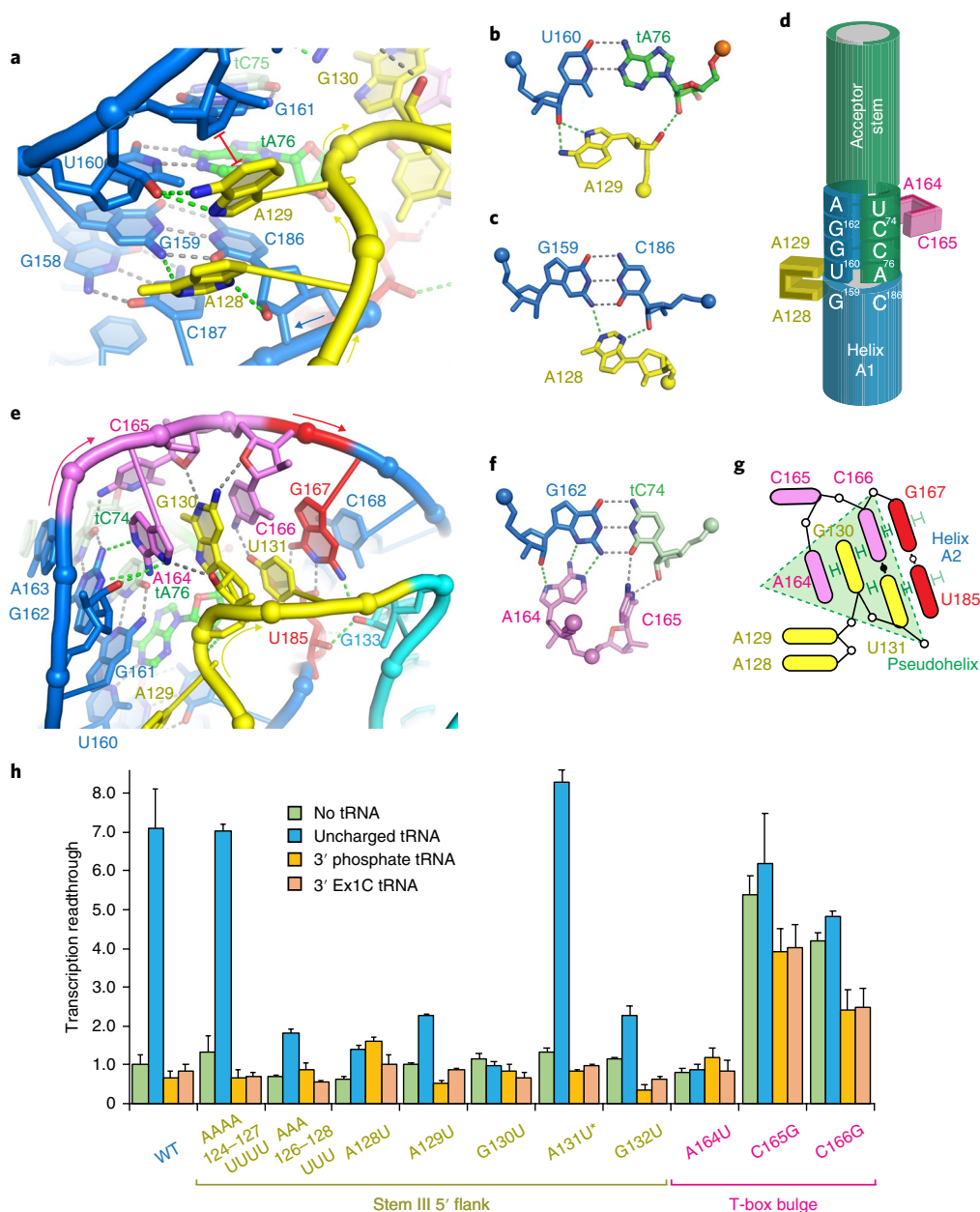


Fig. 2 | The A-minor latch and pseudohelix stabilize tRNA-T-box interactions. **a**, The A128-A129 latch (yellow) reinforces tRNA (green)-helix A1 (blue) stacking interface. **b,c**, Hydrogen-bonding patterns in **a**. **d**, Cartoon illustrating how the A128-A129 axial latch and A164-C165 lateral latch reinforce the tRNA-T-box duplex. **e**, A pseudohelix stacks with and modulates helix A2 stability. **f**, A164 and C165 interact across the minor groove of the T-box-tRNA duplex. **g**, Cartoon diagram of the pseudo-helix structure (green shaded triangle). Stacking interactions are indicated by green connectors. **h**, In vitro transcription readthrough analysis of wild-type and mutant T-boxes (mean and s.d., $n=3$ biologically independent samples). *A131U is where the *B. subtilis* (A) and *G. kaustophilus* (U) sequences naturally diverge.

A long-range pseudohelix dictates the T-box response. Immediately 3' to the tandem A-minor latch, the RNA backbone bends to pair U131 with C166 of the T-box bulge, forming an unusual *cis* pyrimidine-pyrimidine pair. G130 intercalates between A164 and the U131-C166 pair and pushes out C165, forming a long-range, parallel-stranded pseudohelix (Fig. 2e-g). To cap this three-layered pseudohelix, A164 flips its nucleobase inward in a rare (~4% occurrence in functional RNAs³⁹), energetically unfavorable *syn* conformation, whereas the regular *anti* conformation would disengage it from the pseudohelix (Fig. 2f). To offset the energetic penalty of the unusual *syn* conformation (~1.2 kcal mol⁻¹ at 35 °C)⁴⁰, numerous hydrogen bonds and stacking with G130 hold

the spring-loaded A164 in place (Fig. 2e,f). This in turn allows the A164-C165 dinucleotide to act as a lateral latch (Fig. 2d) to stabilize the tRNA-T-box duplex via five hydrogen bonds to its minor groove (Fig. 2f). Thus, the inward flipping of A164 and outward flipping of C165 allow the dinucleotide to make extensive nucleobase-specific contacts across the groove for stabilization. This explains the strict sequence conservation of the T-box bulge^{15,16} and highlights the frequent use of compact *syn* nucleotides in critical functional centers of diverse RNAs³⁹.

Because the pseudohelix stacks against helix A2 at its base (with the G167•U185 pair; Fig. 2e,g), the stability of the pseudohelix impacts that of its neighboring helix A2, which in turn

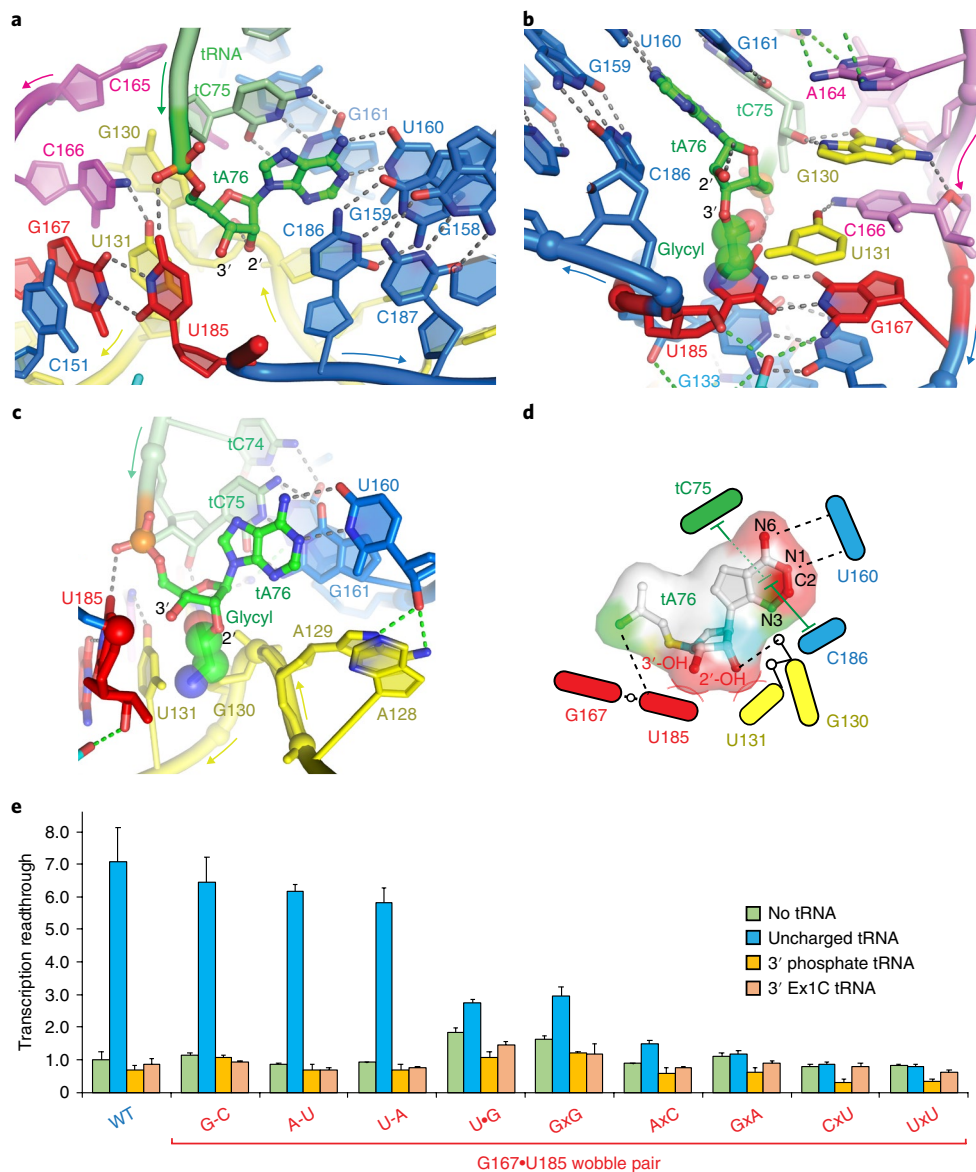


Fig. 3 | Structural basis of tRNA aminoacylation sensing by the T-box discriminator. **a**, Environment and interactions of tA76 at the tRNA 3' end. **b**, Location of a modeled 3'-glycyl moiety (spheres) and overlap with the U185 nucleobase. **c**, Location of a modeled 2'-glycyl moiety (spheres) and overlap with the A-minor latch or pseudohelix. **d**, Cartoon diagram of the environment of tA76 buried inside the T-box discriminator. 2'- and 3'-aminoacyl moieties are rejected by the G130-U131 dinucleotide (part of pseudohelix; Fig. 2e,g) and G167•U185 wobble pair, respectively. **e**, In vitro transcription readthrough analysis of wild-type and mutant T-boxes at the G167•U185 wobble pair (mean and s.d., $n=3$ biologically independent samples).

controls the level of transcription readthrough. As a result, the purine content of the pseudohelix (that is, the number of purines within the region, which stack more stably than pyrimidines due to their larger, bicyclic nucleobases) strongly modulates the T-box response. This effect is so dominant that it overrides the effect of tRNA stabilization. Specifically, replacing A164 or G130 with a uridine abrogated tRNA-mediated readthrough, and substituting C165 or C166 with a guanosine rendered the T-box constitutively active (Fig. 2h and Extended Data Fig. 4). The latter likely resulted from overstabilization of helix A2 from excessive stacking between the purines in the pseudohelix, obviating the need for tRNA stabilization (Fig. 2g). Thus, the pseudohelix is a key regulatory element through which external signals such as regulatory proteins, metabolites, or antibiotics could bind and dictate the T-box response, consistent with the strong modulatory effects of ribosome inhibitors on T-boxes^{41–43}.

Structural basis of tRNA aminoacylation sensing by a mRNA.

Direct sensing of tRNA aminoacylation is the central function of the T-box discriminator. Our cocrystal structure reveals that the compact RNA domain accomplishes this by capturing and burying the tRNA 3'-end deep inside the discriminator core and juxtaposing it with a steric barrier intolerant of an aminoacyl group (Fig. 3 and Extended Data Fig. 6).

The tRNA 3' end (tA76) is paired with U160, stacks against tC75 on the tRNA side and C186 of helix A1 on the T-box side and further engages two hydrogen bonds via its 2'-OH and phosphate oxygen (Figs. 2b and 3a). These interactions collectively bury 500 Å² of solvent-accessible interface (Extended Data Fig. 6) and position the tRNA 3'-OH merely 3.2 Å from a G167•U185 wobble pair located at the base of helix A2 (Fig. 3a,b). This base pair acts as the primary steric barrier against 3'-aminoacyl tRNAs. Binding by a 3'-aminoacyl tRNA would place its universal amino moiety in direct steric

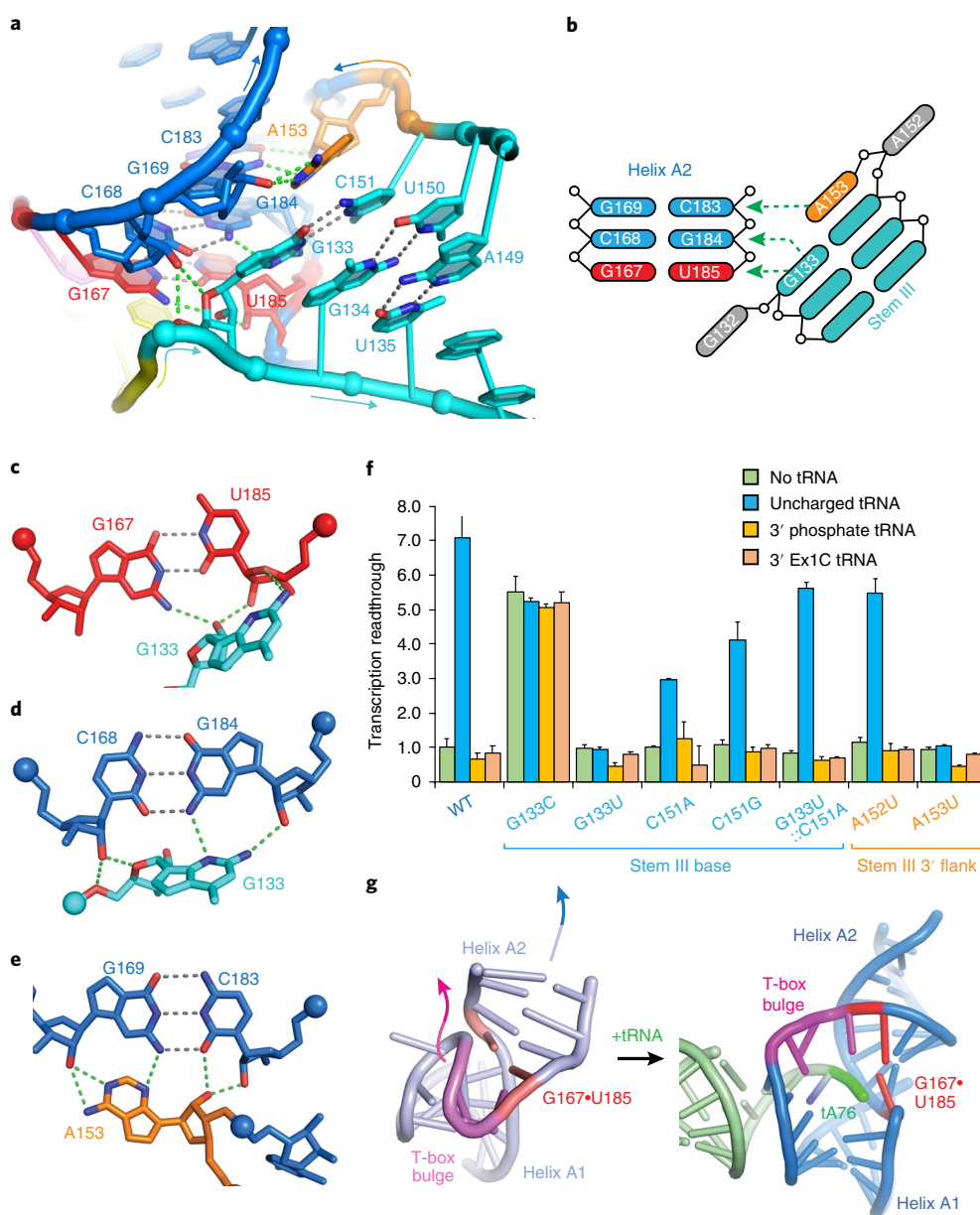


Fig. 4 | Docking of stem III with helix A2 reinforces and rigidifies complex structure. **a**, Stem III uses three consecutive purine-minor groove interactions and two nucleobase-ribose packing contacts to anchor itself to helix A2. **b**, Cartoon diagram illustrating the interactions in **a**. **c–e**, Hydrogen bonding patterns in **a** that involve G133 (**c,d**) or A153 (**e**). **f**, In vitro transcription readthrough analysis of T-box stem III mutants (mean and s.d., $n = 3$ biologically independent samples). **g**, tRNA binding dramatically remodels the discriminator. Left, NMR structure of an isolated *B. subtilis* *tyrS* T-box antiterminator (discriminator without stem III; PDB 1N53)⁴⁹. Arrows denote conformational transitions up to 28 Å induced by tRNA binding. Right, cocrystal structure of the T-box discriminator with bound tRNA. Stem III is omitted for clarity.

clash with the nucleobase of U185 (Fig. 3b). Interestingly, the G•U wobble shifts U185 outward into the major groove, ~2 Å closer to the tRNA 3'-OH than a Watson–Crick pair (Extended Data Fig. 7), potentially amplifying the steric conflict. As a result, the T-box discriminator has constructed a steric pocket even more encompassing and restrictive than the ribosome–RelA complex^{11,12}, which positions the $\beta 5$ strand of its TGS domain ~5 Å away from the tRNA 3'-OH to exert similar steric selection (Extended Data Fig. 7). Unlike the ribosome, which only receives 3'-aminoacyl tRNAs preselected by EF-Tu⁴⁴, the T-box functions independently of EF-Tu¹⁴ and encounters both 2'- and 3'-aminoacyl tRNAs that undergo rapid regioisomerization (~5 s⁻¹)⁴⁵. Reassuringly, a 2'-aminoacyl group would clash with the tandem A-minor latch or the pseudohelix,

both of which are essential for antiterminator stabilization (Fig. 3c,d). Indeed, interface analysis revealed that the 2'- and 3'-OH of tA76 are completely buried by the discriminator (Extended Data Fig. 6). Thus, a singular discriminator structure tightly clutches tA76 and sterically rejects both 2'- and 3'-aminoacyl tRNAs. We conclude that steric sensing of tRNA aminoacylation is a conserved mechanism to detect nutrient limitation and can be effectively implemented by either a protein or an RNA device.

The G167•U185 wobble pair, which acts as the steric filter, occurs in 58% of the 390 transcription-regulating T-boxes analyzed. Other naturally occurring configurations include A-U (25%) and G-C (17%) but essentially no others. Curiously, all of the 39 translation-regulating T-boxes in Actinobacteria exclusively contain the

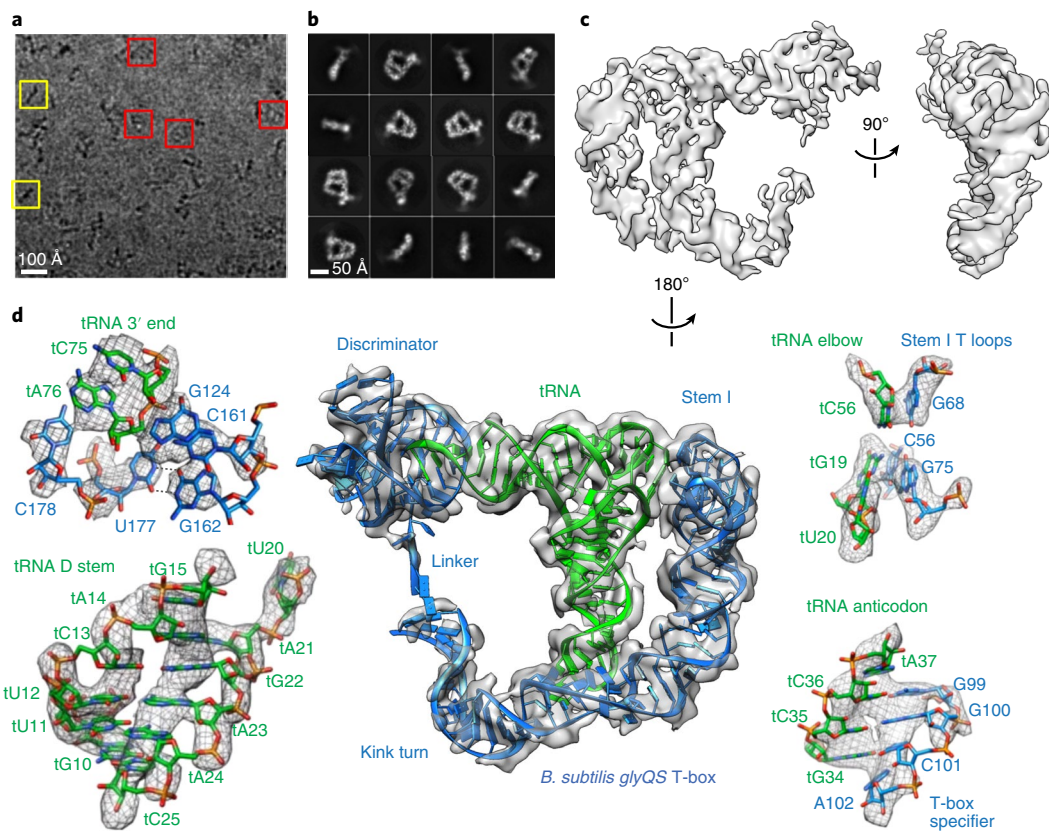


Fig. 5 | 4.9-Å resolution cryo-EM map and structure of a full-length *B. subtilis* glyQS T-box-tRNA^{Gly} complex. **a**, Raw micrograph with top view (red boxes) and side view (yellow boxes) of the T-box–tRNA complex particles. **b**, 2D averages showing high-resolution RNA features. **c**, Front and side views of the final cryo-EM reconstruction at 4.9-Å resolution. **d**, Cryo-EM density overlaid with final structural model (middle) with zoom-in views of the interface between the tRNA 3' end and the discriminator (upper left), tRNA D stem with resolved ribose and phosphate backbone (lower left), interface between the tRNA elbow and stem I distal interdigitated T-loops (upper right) and interface between the tRNA anticodon and the T-box specifier (lower right).

G•U wobble pair^{26,27}. The clear preference for wobble pairs may stem from their unique geometric and energetic characteristics. The G•U wobble pair underwinds the helix by 14°, significantly enhancing nucleobase overlap and stacking with the penultimate C-G pair, reminiscent of the tRNA T-stem loop G49•U65 wobble pair^{46,47} (Extended Data Fig. 8). Despite the clear G•U preference, Watson–Crick pairs adequately support uncharged tRNA-mediated readthrough and rejection of tRNAs that carry a 3' phosphate or an extra cytosine *in vitro*¹⁴ (Ex1C; Fig. 3e). *In vivo* substitutions with Watson–Crick pairs, however, led to a two- to three-fold reduction in downstream gene expression, corroborating the enhanced stacking stability from the G•U wobble pair⁴⁸ (Extended Data Fig. 4). Notably, the reversed U•G wobble pair caused a significant defect (Fig. 3e). Because a U-A pair is nearly fully functional (Fig. 3e), the large defect from U•G is likely attributable to the reversed geometric effects of the wobble. We also tested the possibility of forming an A⁺•C wobble pair, which is isosteric with the G•U wobble pair. The A-C configuration was as defective as other mismatches, suggesting that neighboring interactions and solute conditions do not sufficiently promote adenosine protonation to form the A⁺•C wobble pair in this context (Fig. 3e). Taken together, the extraordinary use of a strategically positioned G•U wobble base pair as both a steric filter and a geometric optimizer expands the functional repertoire of wobble pairs.

Unexpected structural role of T-box stem III and flanking purines. Immediately adjacent to the central G167•U185 pair are two nearly invariant G-C/C-G pairs (Extended Data Fig. 1).

Perplexingly, their substitution with isosteric C-G/G-C pairs produced a dramatic 57-fold reduction of gene expression *in vivo*⁴⁸. Our structure revealed that the pairs provide a sequence-specific docking site for the base of stem III via three consecutive sets of purine-minor groove interactions from G133 and A153 and attendant nucleobase-ribose packing interactions (Fig. 4a–e). Consequently, substitutions of G133 or A153 strongly impacted transcription readthrough (Fig. 4f). Thus, instead of the previously proposed coaxial stacking with the base of helix A1 (ref. 28), the terminal pair of stem III and its flanking purines form an extended latch along the minor groove and act to fasten and stiffen the entire discriminator (Figs. 1e and 4a,b). Comparison with an NMR structure of an isolated antiterminator (discriminator without stem III and its flanking purines)⁴⁹ revealed that drastic rearrangements are required to bind tRNA: the extensively stacked nucleotides in the free T-box bulge must unstack and unravel to accommodate the tRNA 3' end (Fig. 4g). Thus the compact T-box discriminator captures the universal tRNA 3'-NCCA end, employing a cooperative network of energetically coupled contacts that include four intermolecular base pairs, two instances of intermolecular coaxial stacking, six sets of purine-minor groove interactions and three sugar-nucleobase packing interactions (Fig. 1b, e). These multimodal RNA-RNA contacts create a stiff, snug pocket that fashions a steric barrier from a wobble pair to select for uncharged tRNAs.

A full-length T-box mRNA forms a molecular clamp vise. Our cocrystal structure reveals the detailed interactions between the tRNA 3' region and the discriminator, yet it remains unknown

Table 2 | Cryo-EM data collection, refinement and validation statistics

	<i>B. subtilis</i> glyQS T-box–tRNA ^{Gly} complex (EMD-20416, PDB 6POM)
Data collection and processing	
Magnification	165,000
Voltage (kV)	300
Electron exposure (e ⁻ /Å ²)	38
Defocus range (μm)	–0.4 to –3.5
Pixel size (Å)	0.82
Symmetry imposed	C1
Initial particle images (no.)	1,962,681
Final particle images (no.)	189,361
Map resolution (Å)	4.9
FSC threshold	0.143
Map resolution range (Å)	4.1–6.0
Refinement	
Initial model used (PDB code)	4LCK, 4MGN, 6PMO
Model resolution (Å)	5.2
FSC threshold	0.5
Model resolution range (Å)	–
Map sharpening <i>B</i> factor (Å ²)	–353
Model composition	–
Nonhydrogen atoms	4,937
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.808
Validation	
MolProbity score	2.08
Clashscore	1.75

how stem I and discriminator domains coordinate their binding to opposite flanks of tRNAs in a complete T-box. Multiple conflicting structural models have been proposed based on SAXS envelopes^{28,29} and modeling^{14,50,51}, and it was proposed that stem I must release the tRNA elbow contact in order for the antiterminator to bind the tRNA 3' end²⁹.

Cryo-EM single-particle analysis (SPA) is now routinely used to resolve large protein and protein-stabilized ribonucleoprotein structures at near-atomic resolutions. However, despite enhanced contrast compared to proteins, imaging protein-free structured RNAs to resolutions that provide biochemical insights (better than 4–5 Å) remains a significant technical challenge. This is likely due to RNA's inherent structural flexibility, frequent conformational heterogeneity and the relatively small sizes of most known RNA domains. These challenges are exemplified by a recent cryo-EM analysis of a 30 kDa HIV-1 dimerization initiation site (DIS) RNA at 9-Å resolution, in which low signal-to-noise ratio and RNA flexibility or dynamics hampered further resolution improvement⁵².

In order to distinguish among competing T-box structural models and extend the applicability of single-particle analysis to medium-sized RNA-only structures, we subjected a full-length *B. subtilis* glyQS T-box complexed with its cognate uncharged tRNA^{Gly} to cryo-EM SPA. The RNA complexes embedded in vitrified thin ice were unambiguously recognized with identified top and side views (Fig. 5a). Two-dimensional class averages revealed salient high-resolution features, and the final reconstruction achieved 4.9-Å

overall resolution, despite its unusually small size and flexible structure (78 kDa; Fig. 5b,c, Table 2 and Extended Data Fig. 9). The central tRNA encased by the peripheral T-box was better resolved to ~4.0 Å resolution, according to the local-resolution map (Extended Data Fig. 10), in which we started to see resolved density of the ribose and phosphate backbone (Fig. 5d). The discriminator-tRNA and stem I-tRNA cocystal structures were rigidly fitted into the cryo-EM map, resulting in a full-length T-box–tRNA model (Fig. 5d and Supplementary Video 2). The model was flexibly fitted and optimized to the cryo-EM map, with r.m.s.d. of 2.1 Å and 1.9 Å from each crystal structure. The resulting cryo-EM model captured all three principal tRNA-mRNA contacts observed in the two cocystal structures, suggesting that all three contacts are engaged concurrently in the full complex in the readthrough configuration (Fig. 5d and Supplementary Video 2).

The full-length *B. subtilis* cryo-EM structure revealed a stem I trajectory distinct from and intermediate between those of the stem I-tRNA cocystal structures from *Oceanobacillus ihenyensis*²² and *G. kaustophilus*²³ (Fig. 6a). To adapt to structural divergences among the three stem I domains, the same tRNA^{Gly} flexes about its t26•t44 hinge up to 22° to maintain nearly identical contacts to its anticodon and elbow (Fig. 5d). When the stem I distal interdigitated T-loops are superimposed, the tRNA elbow to which the T-loops bind overlay among the three structures (Fig. 6b), suggesting remarkable conservation of the interface across three species despite the lack of base-pairing interactions^{21,22,53}. Indeed, mutation or deletion of the interdigitated T-loops was reported to slow tRNA binding and to drastically reduce complex stability^{31,51}. Thus, the tripartite T-box interactions with the tRNA anticodon, elbow, and 3' end persist concurrently in the same complex. We did not observe the previously proposed 90° pivoting motion that would disengage the stem I distal region from the tRNA elbow²⁹, which may occur in transient intermediates not detectable in our analysis. On the 3' side, while the tRNA and discriminator each superimpose well with the cryo-EM structure (RMSDs of 1.5 and 1.6 Å, respectively), the tRNA body is rotated 8° between the cocystal and cryo-EM structures, enabled by apparent flexibility of the tRNA acceptor stem (Fig. 6c). To confirm the overall architectures of both T-box complexes in solution, we performed SAXS analysis of both complexes and observed scattering profiles highly congruent with the cocystal and cryo-EM structures (χ^2 ~1.7 and 0.7, respectively; Fig. 6f,g).

The simultaneous binding of stem I and discriminator regions to both flanks of tRNA create a molecular vise that clamps the top half of the uncharged tRNA (Fig. 6d,e). Remarkably, contiguous coaxial stacking traverses three intermolecular interfaces, forming a central spine of 31 or 32 layers of stacked pairs (Fig. 6d,e). This spine provides the requisite stability—in conjunction with the kinetic advantage of being transcribed first—that preclude the formation of a transcription terminator or Shine–Dalgarno sequester. In amino acid abundance, the presence of an esterified amino acid on the tRNA 3' end interrupts the tRNA-helix A1 stacking, which severs the central spine and prevents the aminoacyl-tRNA from stabilizing the antiterminator. Without tRNA stabilization, the thermodynamically more stable terminator or sequester forms and shuts off downstream amino acid genes, completing a negative feedback loop.

Discussion

Together, our cocystal, cryo-EM, SAXS, and mutational analyses delineate how the T-box mRNA envelopes a cognate uncharged tRNA through sequential, tripartite interactions to construct a remarkably stable complex (k_{off} ~0.0002 s⁻¹)⁵¹ to permit conditional gene expression (Fig. 7). Our analyses distinguish among a number of proposed structural models^{14,28,29,50,51} and establish how mRNA encapsulates a tRNA, sterically probes its aminoacylation status, and switches conformations based on this readout. Contoured by

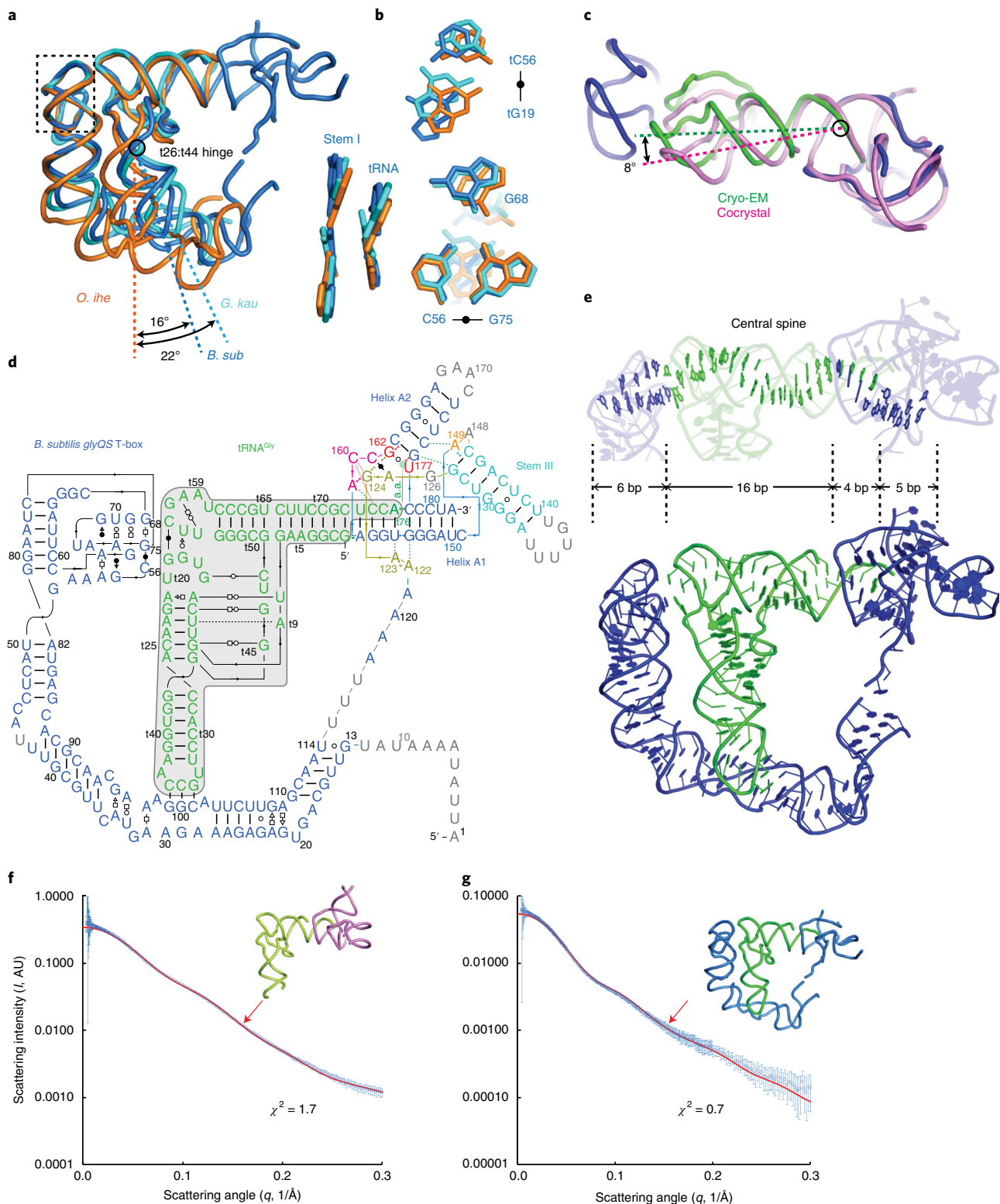


Fig. 6 | The T-box central spine and structural comparisons. **a**, Superposition of the stem I distal interdigitated T-loops (boxed) in the cryo-EM structure (blue) with those in the stem I-tRNA cocystal structures from *O. ihenyensis* (orange)²² and *G. kaustophilus* (cyan)²³. **b**, Comparison of the stem I-tRNA stacking interface among the three structures. Numbering is based on the *B. subtilis* cryo-EM structure (blue). **c**, Superposition of the discriminator structures in the cocystal (magenta) and cryo-EM (blue and green) structures reveals an 8° rotation of the tRNA acceptor stem. **d**, Secondary structure diagram of the full-length *B. subtilis* glyQS T-box-tRNA^{Gly} complex. Disordered residues are shown in gray. **e**, Coaxial stacking at three tRNA-T-box interfaces align and assemble a -31- or 32-layered central spine that stabilizes the antiterminator conformation. **f,g**, Overlay of back-calculated SAXS scattering curves (red lines) computed from the cocystal structure of the T-box discriminator-tRNA complex (**f**) or cryo-EM structure of the full-length T-box-tRNA complex (**g**) with CRY SOL⁶¹ with experimental scattering profiles in solution (blue circles; mean and s.d. of 30 measurements of one sample; Methods). χ^2 values of the fits are indicated.

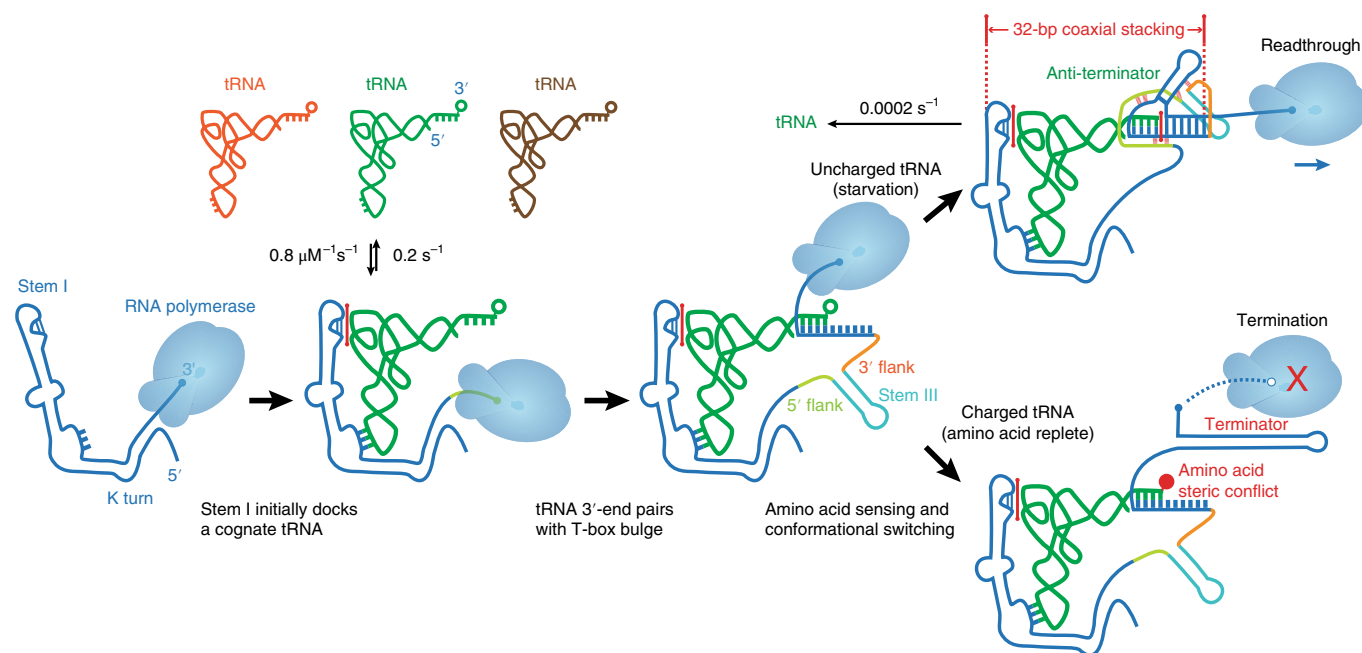


Fig. 7 | Mechanistic model of a cotranscriptionally acting T-box riboswitch. A transcriptional T-box initially uses its stem I to select a cognate tRNA regardless of its 3' aminoacylation state, based primarily on anticodon-specifier complementarity. k_{on} and k_{off} for the cognate tRNA are derived from single-molecule fluorescence measurements^{19,31,51}. The anticodon-specifier interaction engages first and is reinforced by the stacking interaction at the tRNA elbow (red stick). Subsequent elongation of the T-box transcript exposes stem III and the 5' half of the antiterminator, which then forms four base pairs with the 3'-NCCA end of the docked tRNA. Finally, the 3' strand of the antiterminator, transcribed last, will attempt to anneal with its 5' counterpart. In starvation, the annealing is successful, forming a 32-layered, stacked central spine. This spine stabilizes the antiterminator conformer and allows RNAP to traverse this region to transcribe downstream genes. In amino acid abundance, the presence of an esterified amino acid on the tRNA blocks annealing of the 3' strand of the antiterminator due to a steric conflict with the G167•U185 wobble pair, leading to terminator formation and termination of gene expression, completing a negative feedback loop.

subtle bends (for example, the stem I C loop^{50,53}), sharp kinks (for example, the stem I K turn^{54,55}), and flexible joints (for example, the interdomain linker), the U-shaped T-box scaffold employs a linear arrangement of strategically positioned RNA-interaction modules (for example, the specifier loop, interdigitated T-loops and T-box bulge) to capture tRNA with nanomolar affinity and digital specificity (as they decipher the anticodon). By covalently linking a steric filter (G167•U185 wobble) with a conditional stacking conductor (C186::tA76), the 66-nt discriminator achieves tunable gene expression control conditional upon tRNA aminoacylation. This universal, parsimonious design enables Gram-positive bacteria to rapidly respond to fluctuating nutrient availability in their environment and constitutes an attractive drug target to curb widespread antibiotic resistance^{20,41,42}. Further, the discovery and visualization of highly defined T-box–tRNA complex structures and interfaces permit rational design and targeting by small-molecule antibiotics.

The T-box paradigm provides general insights into how individually weak interactions coordinately construct stable RNA–RNA complexes through multivalency, while maintaining required selectivity. Most tRNAs share the same discriminator base (mostly A or G in *B. subtilis*) and thus a nearly identical 3'-NCCA end. However, the T-box discriminator cannot allow non-cognate tRNAs to interact and interfere with specific tRNA aminoacylation sensing. Our measured $2.0 \mu\text{M}$ K_d is appropriately weak and comparable to that of the anticodon-specifier interaction alone²². Stem I distal T-loops cannot interact with the tRNA elbow without the prior engagement of the anticodon-specifier contact²². Thus the requisite total binding affinity to overcome the terminator or sequesterator is distributed among three weak contacts. This distributive mode of tRNA–mRNA interaction is advantageous as it effectively rejects any local non-cognate

interactions and requires a global fit of the intended tRNA substrate into the T-box's U-shaped concave (Fig. 6e). Among the three principal tRNA–T-box contacts, only seven base pairs are utilized. This limited sequence complementarity is augmented by extensive structure-specific interactions consisting of two components: (1) three stacking interactions that align and assemble the central stacking spine traversing the entire length of the complex, and (2) six sets of purine-minor groove interactions and three sugar-nucleobase packing contacts, both of which are tertiary contacts that the discriminator employs to reinforce the tRNA 3' end–T-box discriminator interaction.

Multivalent tRNA binding by the modular T-box mRNA may further extend to the enigmatic stem II domain absent from the glyQ T-boxes under study but present in most non-glycyl T-boxes^{15,16}. Secondary structure analyses indicated an elongated stem II hairpin followed by a potential pseudoknot in this region. This structurally complex domain is proposed to make additional tRNA contacts near its elbow region, based on a cross-linking analysis^{26,27}. However, detailed structural analysis is required to clarify the structure and function of the stem II domain.

Similar distributive binding mechanisms and extensive utilization of stacking and tertiary contacts to supplement limited sequence complementarity likely also operate in several types of RNA and ribonucleoprotein condensates or granules, including those containing pathogenic repeat expansion RNAs^{56–58}. In these systems, limited sequence complementarity (for example, between short CAG or CUG near-palindromic repeats) first seeds local RNA–RNA interactions. Subsequently, multivalency drives their self-assembly into RNA networks or foci⁵⁸. Similarly, the type of multivalent interactions employing intra- and intermolecular stacking and extensive tertiary contacts found in the T-box–tRNA complexes may also

drive the assembly of long noncoding RNA domains or viral RNA genomes⁵⁹.

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 data and code availability are available at <https://doi.org/10.1038/s41594-019-0326-7>.

Received: 24 June 2019; Accepted: 7 October 2019;

Published online: 18 November 2019

References

- Chantranupong, L., Wolfson, R. L. & Sabatini, D. M. Nutrient-sensing mechanisms across evolution. *Cell* **161**, 67–83 (2015).
- Saxton, R. A. et al. Structural basis for leucine sensing by the Sestrin2-mTORC1 pathway. *Science* **351**, 53–58 (2016).
- Chantranupong, L. et al. The CASTOR proteins are arginine sensors for the mTORC1 Pathway. *Cell* **165**, 153–164 (2016).
- Antson, A. A. et al. Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* **401**, 235–242 (1999).
- Serganov, A. & Patel, D. J. Amino acid recognition and gene regulation by riboswitches. *Biochim. Biophys. Acta* **1789**, 592–611 (2009).
- Roth, A. & Breaker, R. R. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **78**, 305–334 (2009).
- Pang, Y. L., Poruri, K. & Martinis, S. A. tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA* **5**, 461–480 (2014).
- Bullwinkle, T. J. & Ibba, M. Emergence and evolution. *Top. Curr. Chem.* **344**, 43–87 (2014).
- Beuning, P. J. & Musier-Forsyth, K. Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers* **52**, 1–28 (1999).
- Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751–758 (1981).
- Loveland, A. B. et al. Ribosome*RelA structures reveal the mechanism of stringent response activation. *eLife* **5**, e17029 (2016).
- Brown, A., Fernandez, I. S., Gordiyenko, Y. & Ramakrishnan, V. Ribosome-dependent activation of stringent control. *Nature* **534**, 277–280 (2016).
- Dong, J., Qiu, H., Garcia-Barrio, M., Anderson, J. & Hinnebusch, A. G. Uncharged tRNA activates GCN2 by displacing the protein kinase moiety from a bipartite tRNA-binding domain. *Mol. Cell* **6**, 269–279 (2000).
- Zhang, J. & Ferré-D'Amaré, A. R. Direct evaluation of tRNA aminoacylation status by the T-box riboswitch using tRNA-mRNA stacking and steric readout. *Mol. Cell* **55**, 148–155 (2014).
- 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).
- Gutierrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C. & Merino, E. Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiol. Mol. Biol. Rev.* **73**, 36–61 (2009).
- Saad, N. Y. et al. Two-codon T-box riboswitch binding two tRNAs. *Proc. Natl Acad. Sci. USA* **110**, 12756–12761 (2013).
- Apostolidi, M. et al. A glyS T-box riboswitch with species-specific structural features responding to both proteinogenic and nonproteinogenic tRNA^{Gly} isoacceptors. *RNA* **21**, 1790–1806 (2015).
- Suddala, K. C. & Zhang, J. An evolving tale of two interacting RNAs—themes and variations of the T-box riboswitch mechanism. *IUBMB Life* **71**, 1167–1180 (2019).
- Frohlich, K. M. et al. Discovery of small-molecule antibiotics against a unique tRNA-mediated regulation of transcription in gram-positive bacteria. *ChemMedChem* **14**, 758–769 (2019).
- Lehmann, J., Jossinet, F. & Gautheret, D. A universal RNA structural motif docking the elbow of tRNA in the ribosome, RNase P and T-box leaders. *Nucleic Acids Res.* **41**, 5494–5502 (2013).
- Zhang, J. & Ferré-D'Amaré, A. R. Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* **500**, 363–366 (2013).
- Grigg, J. C. & Ke, A. Structural determinants for geometry and information decoding of tRNA by T box leader RNA. *Structure* **21**, 2025–2032 (2013).
- Grundy, F. J. & Henkin, T. M. tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell* **74**, 475–482 (1993).
- Grundy, F. J., Winkler, W. C. & Henkin, T. M. tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proc. Natl Acad. Sci. USA* **99**, 11121–11126 (2002).
- Sherwood, A. V., Frandsen, J. K., Grundy, F. J. & Henkin, T. M. New tRNA contacts facilitate ligand binding in a Mycobacterium smegmatis T box riboswitch. *Proc. Natl Acad. Sci. USA* **115**, 3894–3899 (2018).
- 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).
- Chetnani, B. & Mondragon, A. Molecular envelope and atomic model of an anti-terminated glyQS T-box regulator in complex with tRNA^{Gly}. *Nucleic Acids Res.* **45**, 8079–8090 (2017).
- Fang, X., Michnicka, M., Zhang, Y., Wang, Y. X. & Nikonowicz, E. P. Capture and release of tRNA by the T-Loop receptor in the function of the T-Box riboswitch. *Biochemistry* **56**, 3549–3558 (2017).
- Grundy, F. J. & Henkin, T. M. Kinetic analysis of tRNA-directed transcription antitermination of the Bacillus subtilis glyQS gene in vitro. *J. Bacteriol.* **186**, 5392–5399 (2004).
- Zhang, J. et al. Specific structural elements of the T-box riboswitch drive the two-step binding of the tRNA ligand. *eLife* **7**, e39518 (2018).
- Stagno, J. R. et al. Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242–246 (2017).
- Nozinovic, S. et al. The importance of helix P1 stability for structural pre-organization and ligand binding affinity of the adenine riboswitch aptamer domain. *RNA Biol.* **11**, 655–656 (2014).
- Peselis, A. & Serganov, A. ykkC riboswitches employ an add-on helix to adjust specificity for polyanionic ligands. *Nat. Chem. Biol.* **14**, 887–894 (2018).
- Ren, A. et al. Structural and dynamic basis for low-affinity, high-selectivity binding of L-glutamine by the glutamine riboswitch. *Cell Reports* **13**, 1800–1813 (2015).
- Doherty, E. A., Batey, R. T., Masquida, B. & Doudna, J. A. A universal mode of helix packing in RNA. *Nat. Struct. Biol.* **8**, 339–343 (2001).
- Ogle, J. M. et al. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
- Loveland, A. B., Demo, G., Grigorieff, N. & Korostelev, A. A. Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **546**, 113–117 (2017).
- Sokoloski, J. E., Godfrey, S. A., Dombrowski, S. E. & Bevilacqua, P. C. Prevalence of syn nucleobases in the active sites of functional RNAs. *RNA* **17**, 1775–1787 (2011).
- Hemmes, P. R., Oppenheimer, L. & Jordan, F. Ultrasonic relaxation evaluation of the thermodynamics of syn-anti glycosidic isomerization in adenosine. *J. Am. Chem. Soc.* **96**, 6023–6026 (1974).
- Dar, D. et al. Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**, aad9822 (2016).
- Stamatopoulou, V. et al. Direct modulation of T-box riboswitch-controlled transcription by protein synthesis inhibitors. *Nucleic Acids Res.* **45**, 10242–10258 (2017).
- Liu, J. et al. Identification of spermidine binding site in T-box riboswitch antiterminator RNA. *Chem. Biol. Drug Des.* **87**, 182–189 (2016).
- Pleiss, J. A. & Uhlenbeck, O. C. Identification of thermodynamically relevant interactions between EF-Tu and backbone elements of tRNA. *J. Mol. Biol.* **308**, 895–905 (2001).
- Tajiri, M., Yokoyama, S. & Miyazawa, T. Transacylation rates of (aminoacyl) adenosine moiety at the 3'-terminus of aminoacyl transfer ribonucleic acid. *Biochemistry* **22**, 3220–3225 (1983).
- Varani, G. & McClain, W. H. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* **1**, 18–23 (2000).
- Masquida, B. & Westhof, E. On the wobble GoU and related pairs. *RNA* **6**, 9–15 (2000).
- Grundy, F. J., Moir, T. R., Haldeman, M. T. & Henkin, T. M. Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: disparity between conservation and functional requirements. *Nucleic Acids Res.* **30**, 1646–1655 (2002).
- Gerdeman, M. S., Henkin, T. M. & Hines, J. V. Solution structure of the Bacillus subtilis T-box antiterminator RNA: seven nucleotide bulge characterized by stacking and flexibility. *J. Mol. Biol.* **326**, 189–201 (2003).
- Grigg, J. C. et al. T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *Proc. Natl Acad. Sci. USA* **110**, 7240–7245 (2013).
- Suddala, K. C. et al. Hierarchical mechanism of amino acid sensing by the T-box riboswitch. *Nat. Commun.* **9**, 1896 (2018).
- Zhang, K. et al. Structure of the 30 kDa HIV-1 RNA dimerization signal by a hybrid Cryo-EM, NMR, and molecular dynamics approach. *Structure* **26**, 490–498.e3 (2018).
- Zhang, J. & Ferré-D'Amaré, A. R. Structure and mechanism of the T-box riboswitches. *Wiley Interdiscip. Rev. RNA* **6**, 419–433 (2015).
- Winkler, W. C., Grundy, F. J., Murphy, B. A. & Henkin, T. M. The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* **7**, 1165–1172 (2001).

55. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).
56. Van Treeck, B. & Parker, R. Emerging roles for intermolecular RNA-RNA interactions in RNP assemblies. *Cell* **174**, 791–802 (2018).
57. Langdon, E. M. et al. mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922–927 (2018).
58. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).
59. Dethoff, E. A. et al. Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl Acad. Sci. USA* **115**, 11513–11518 (2018).
60. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).
61. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).

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

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply. 2019

Methods

Conservation analysis of the T-box discriminator region. T-box sequences were retrieved from the Rfam database⁶². The search motif used to localize the T-box discriminator in each sequence was 5'-UGGNANC-3', where N denotes any nucleotide (Extended Data Fig. 1). Sequences of 49 residues were aligned, and duplicates were removed. All remaining sequences ($N=709$) were folded with RNAfold⁶³ while forcing the three residues upstream and the three residues downstream of the 5'-UGGNANC-3' motif to base pair with downstream residues and prohibiting the residues within the T-box motif and the last eight residues of the sequences from base pairing. Only the lowest free-energy folds that satisfy these requirements were selected ($N=390$). Because the length of helix A2 is not conserved, sequence alignment had to be established separately on either side of helix A2. Aligned sequences were used to generate sequence logos with Weblogo⁶⁴ (Extended Data Fig. 1). To determine a coherent conservation pattern on the 5' side of highly variable stem III, only sequences with the highly conserved adenosine doublet (AA) in nt 152–153 (six residues upstream of the 5'-UGGNANC-3' motif) were selected from the Rfam database. The size of these sequences was limited to 38 nucleotides to limit folding artifacts with the upstream segment. Duplicates were then removed. The remaining sequences ($N=373$) were folded with RNAfold while enforcing base pairing to 4 residues upstream of the adenosine doublet. After the first round of folding, additional constraints were applied to select folded sequences in order to prevent alternate folds from occurring on the 5' side of the sequence. Sequences that did not meet these requirements were discarded. To determine the base conservation pattern, all folded sequences were subsequently aligned to the 5' strand of stem III (whose stem length ranges from 3 to about 20 base pairs). Because the overall conservation pattern shows that the functionally important G130 is not always a guanosine (upper panel, Extended Data Fig. 1), the conservation pattern was split for the sequences bearing a G at that position (middle panel, Extended Data Fig. 1; $n=205$ sequences) and for the sequences that do not have a G (lower panel, Extended Data Fig. 1; $n=46$ sequences).

RNA preparations. *B. subtilis* tRNA^{Gly} used for in vitro termination-readthrough assays was transcribed in vitro by T7 RNA polymerase via 'GMP priming', where a 20 mM GMP and 2 mM GTP mixture was used to install predominantly (~91%) 5'-monophosphate^{14,65}. The RNAs were purified by electrophoresis on 10% polyacrylamide (29:1 acrylamide:bisacrylamide), 8 M urea TBE (Tris-borate-EDTA) gels, electroeluted, washed once with 1 M KCl, desalted via ultrafiltration, and stored at -80°C before use. tRNA variants carrying 3'-azido and 3'-N-pentenoyl-glycine modifications were prepared using the Klenow fragment and flexizyme^{14,66}. The 3'-phosphate tRNA was generated by in vitro transcription with T7 RNA polymerase, then cleaved by a transacting Varkud satellite ribozyme⁶⁷. The cleaved RNA was treated with 10 mM HCl at room temperature for 30 min to open the 2',3'-cyclic phosphate to yield a 3' phosphate⁶⁸. The HCl-treated tRNA was washed three times with diethyl pyrocarbonate (DEPC)-treated H₂O and stored at -80°C before use. To prepare the T-box discriminator-tRNA complex for co-crystallization and SAXS analysis, wild-type (crystal 1–9; Table 1) or A156C::U189G mutant (crystal 10–11; stabilized helix A1) *G. kaustophilus* glyQ T-box discriminator RNA (residues 126–191) was co-transcribed with excess *B. subtilis* (or *G. kaustophilus*, identical sequences) tRNA^{Gly}_{GCC} in vitro by T7 RNA polymerase using PCR products as templates as described²². The complex was isolated from excess tRNAs by size exclusion chromatography in a buffer consisting of 10 mM Tris-HCl, pH 7.4, 100 mM KCl and 20 mM MgCl₂. Fractions containing the complex were pooled and concentrated to 8 mg/ml.

To prepare the full-length T-box riboswitch-tRNA complex for cryo-EM and SAXS analyses, *B. subtilis* glyQ T-box RNA (residues 14–182) and *B. subtilis* tRNA^{Gly}_{GCC} were co-transcribed in vitro by T7 RNA polymerase, using PCR products as templates⁶⁹. The RNA complex was purified first using anion exchange chromatography on a Mono Q column, then by size-exclusion chromatography on a Superdex 200 Increase column in a buffer consisting of 10 mM Tris-HCl (pH 7.4), 100 mM KCl and 20 mM MgCl₂. Peak fractions containing the T-box-tRNA complex were pooled and concentrated to 8 mg/ml.

Electrophoretic mobility shift assay. *G. kaustophilus* glyQ T-box discriminator RNA (residues 126–191) was transcribed in vitro by T7 RNA polymerase at 37°C for 3 h. This native transcription product (0.5 μl) was mixed with 2 μM purified tRNA in 10 μl of binding buffer that contains 25 mM Tris-HCl (pH 7.4), 100 mM KCl and 20 mM MgCl₂, incubated at room temperature for 30 min, and electrophoretically separated on a 10% native polyacrylamide gel in Tris-HEPES-EDTA buffer (pH 8.0) supplemented with 10 mM MgCl₂. The gels were stained with ethidium bromide or Gel Red and imaged with a Typhoon Trio system.

Isothermal titration calorimetry. ITC was performed essentially as described in ref. ²². To suppress oligomerization, tRNA^{Gly} was heated to 90°C in water for 3 min and snap-cooled to 4°C over 2 min immediately prior to use. RNA samples for ITC experiments were equilibrated in a buffer comprised of 100 mM KCl, 25 mM Tris-HCl, pH 7.4, 10 mM MgCl₂ via ultrafiltration. ITC experiments were performed in triplicates at 20°C with 25 μM T-box discriminator RNA in the cell and 250 μM uncharged or 3'-phosphate tRNA^{Gly} in the syringe, using a MicroCal

iTC200 microcalorimeter (GE). Data processing was carried out using unbiased integration software NITPIC⁷⁰, then by SEDPHAT⁷¹. The titration shown in Fig. 1d was performed three times using biologically independent samples. For binding by uncharged tRNA^{Gly}, $K_d=2.0\pm 0.6\ \mu\text{M}$, $\Delta H=-11.7\pm 0.9\ \text{kcal mol}^{-1}$, $-\Delta S=15.1\pm 0.8\ \text{kcal mol}^{-1}$. These values are mean, and the errors are s.d.; $n=3$ biologically independent samples. The average binding-incompetent fraction was ~10%. For three independent titrations using 3'-phosphate tRNA^{Gly}, the curves cannot be reliably fit, with an estimated $K_d>50\ \mu\text{M}$ and negligible ΔH (Fig. 1d).

In vitro transcription termination-readthrough assay using a fluorescent reporter. Transcription termination-readthrough assay was performed essentially as described previously, with modifications^{14,25,72}. A hairpin molecular beacon was designed to specifically hybridize to the *B. subtilis* glyQ downstream readthrough transcript to measure T-box-controlled gene expression⁷² (Fig. 1a). It is of the sequence 5'-Cy5-CGCACCuccacuuuucuucauaaTGCC-Dabcyl-3' (Trilink), where upper- and lowercase letters denote deoxyribonucleotides and 2'-O-methyl ribonucleotides, respectively⁷². The 2'-O-methyl modifications were used to enhance hybridization with the target transcript. Multi-round transcription termination-readthrough assays were performed in a transcription buffer composed of 20 mM Tris-HCl, pH 7.4, 40 mM KCl, 6 mM MgCl₂, 1 mM DTT and 5 mM spermidine. Each 36 μl reaction also contained 10 nM PCR-derived linear DNA template, 0.05 U/ μl *E. coli* RNA polymerase holoenzyme (New England Biolabs), 50 nM molecular beacon, and 250 nM snap-cooled tRNA, when specified. Reactions were initiated by addition of nucleoside triphosphates (NTPs) to provide 400 μM ATP, 10 μM UTP, 10 μM CTP, and 10 μM GTP and were allowed to proceed at 37°C for 3 h while fluorescence intensity was monitored every 2 min on a TECAN Infinite 200 Pro microplate reader. An excitation wavelength of 620 nm (bandwidth 20 nm), emission wavelength of 670 nm (bandwidth 25 nm) and a gain setting of 110 were used. The fluorescence intensity increase between 34 and 180 min was nearly linear (Extended Data Fig. 4), and linear regression was used to obtain the rate of transcript production in the presence or absence of various tRNAs. In the presence of robust readthrough transcription, the square of Pearson product-moment correlation coefficient (R^2) was typically >0.9 . In the absence of NTPs, the slope was close to zero and $R^2<0.05$, suggesting minimal transcription-independent interference of the fluorescence signal (that is, nonspecific molecular beacon interaction with the DNA template or RNA polymerase) in the selected time window (Extended Data Fig. 4). After fluorescence measurements, selected reactions were analyzed on a 10% urea-PAGE and visualized with SYBR Gold to validate fluorescence-based measurements (Extended Data Fig. 4). Due to hydrolytic instability of the 3'-aminoacyl tRNA, slightly alkaline conditions (pH 7.4 or higher) in transcription readthrough assays, which exacerbates tRNA deacylation, and extended incubation time (3 h or longer), it was not practical to use 3'-aminoacyl tRNA. Instead, a 3'-phosphate tRNA was used as a stable mimic of 3'-aminoacyl tRNA, as the 3'-phosphate is comparable in size to 3'-glycyl, and tRNAs bearing the two modifications were found to elicit nearly identical transcriptional responses¹⁴. We also used another negative-control tRNA termed the Ex1C tRNA^{14,73}, in which an extra cytosine was appended to the tRNA 3'-CCA end. This tRNA strongly inhibits readthrough, as tC77 competes with C186 for pairing with G159, destabilizing helix A1 and preventing tRNA-helix A1 stacking.

Cocrystallization and diffraction data collection. For crystallization, the purified *G. kaustophilus* glyQ T-box discriminator-tRNA complex was supplemented with 1 mM spermine, and mixed 1:1 with a reservoir solution consisting of 0.1 M Morphes buffer system 2, pH 7.5 (sodium HEPES; MOPS (acid)), 0.09 M halogens (NaF; NaBr; NaI), 10% w/v PEG 20000, and 20% v/v PEG 500 monomethyl ether (MME) and crystallized at 21°C using the hanging-drop vapor diffusion method. Rectangular prism-shaped crystals grew in 2–7 d to maximum dimensions of $300\times 300\times 50\ \mu\text{m}^3$. The crystals were cryoprotected in an artificial mother liquor containing 30% v/v PEG 500MME before vitrification in liquid nitrogen. For de novo phasing with Ir(III), crystals were soaked in 10 mM Tris-HCl (pH 7.4), 100 mM KCl and 20 mM MgCl₂, 0.09 M halogens, 0.1 M Morphes buffer system 2, pH 7.5 (Sodium HEPES; MOPS (acid)), 10% w/v PEG 20000 and 20% v/v PEG 500 MME, 1 mM spermine and 20 mM Ir(III) hexammine for 20 h before vitrification in liquid nitrogen. Single-wavelength anomalous dispersion (SAD) data were collected near the iridium L1 edge (0.9218 Å) at the SER-CAT beamline ID-22 at the Advanced Photon Source (APS). Cocrystals exhibit the symmetry of space group C222₁. Unit cell dimensions are in Table 1.

Crystal structure determination and refinement. Initial SAD phasing using individual data sets from cocrystals of wild-type *G. kaustophilus* glyQ T-box discriminator-tRNA complexes did not correctly identify the Ir substructure. A total of 11 Ir-SAD datasets with diffraction limits extending to 4–5 Å were analyzed by the hierarchical agglomerative program BLEND⁷⁴. A cluster of nine data sets with high isomorphism, as evidenced by a linear cell variation (LCV) value of 0.96% (or 1.5 Å), were merged to produce a high-redundancy (multiplicity 56.8) data set at 4.5-Å resolution. Initial Ir substructure (17 sites) was identified using Phenix.AutoSol⁷⁵, producing an initial electron density map (mean overall figure of merit is 0.53) into which a partial tRNA was first located. After most

visible tRNA and T-box residues were built, the partial model was located in a high-resolution Ir-SAD dataset collected from cocrystals of the A156C::U189G mutant discriminator-tRNA complex (2.66-Å resolution) using PHASER⁷⁶, which allowed identification and completion of the Ir substructure. The A156C::U189G substitution replaced a non-conserved A-U pair in helix A1 with a C-G pair, which resulted in altered unit cell parameters and extended the diffraction limit from 3.4 Å to 2.66 Å (Table 1). Model building was performed in Coot⁷⁷, as guided by iteratively generated MR-SAD maps using an evolving model, corrected by ERRASER⁷⁸, and refined using Phenix.Refine⁷⁹. The refined structure without Ir atoms was located in a native 2.84 Å data set using PHASER (Translational Function Z (TFZ) score = 36.9; log-likelihood gain (LLG) = 2,681) and was also refined. The Ir-containing (model complex) and native structures are nearly identical, with a r.m.s.d. of 0.2 Å. Refinement statistics are summarized in Table 1.

Cryo-EM sample preparation. A total of 3 µl of *B. subtilis* full-length T-box riboswitch-tRNA complex sample (30 µM) was applied onto glow-discharged (30 s) 200-mesh R2/1 Quantifoil Cu grids. The grids were blotted for 3 s in 100% humidity with no blotting offset and rapidly frozen in liquid ethane using a Vitrobot Mark IV (FEI).

Cryo-EM single particle data acquisition and data processing. The aforementioned frozen grids were loaded in a Titan Krios (FEI) operated at 300 kV, with condenser lens aperture 70 µm, spot size 7, parallel beam with illumination area 0.85 µm in diameter. Microscope magnification was at 165,000× (corresponding to a calibrated sampling of 0.82 Å per physical pixel). Movie stacks were collected automatically using EPU software on a K2 direct electron device equipped with a Quantum energy filter operated at 20 eV (Gatan), operating in counting mode at a recording rate of 5 raw frames per second and a total exposure time of 5 s, yielding 25 frames per stack and a total dose of 38 e⁻/Å². A total of 5,600 movie stacks were collected with defocus values ranging from -0.4 to -3.5 µm. These movie stacks were motion corrected using Motioncor2. After CTF determination by CTFFIND4, 5,287 micrographs were subjected to EMAN2.2 for neural network particle picking⁷⁹, with a threshold setting of 0.1 used to maximize inclusion of good particles. Under this threshold, a considerable number of particles from noise were included in the particle extraction, which is sorted out in the latter 2D classification steps. A total of 1,962,681 particles were extracted in Relion 2, with the box size as 224 pixels⁸⁰. After three rounds of 2D classifications, the best classes by visual examination were subjected to EMAN2.2 to build the initial model. A total of 389,305 particles were subjected to 3D classification. Two classes representing intact T-box-tRNA complexes, totaling 189,361 particles, were subjected to autorefinement in Relion 2. A measured B factor of -353 Å² was used for sharpening to yield the final sharpened map at 4.9-Å resolution, estimated by the 0.143 criterion of the FSC curve. A 4.9-Å Gauss low-pass filter was applied to the final 3D map displayed in UCSF Chimera⁸¹.

Cryo-EM model building and refinement. Crystal structures of *O. iheyensis* T-box stem I-tRNA and *G. kaustophilus* discriminator tRNA were first rigidly fitted into the 4.9-Å resolution cryo-EM density using the Chimera Fit in Map tool¹. The nucleobases of docked residues were changed into the sequence of *B. subtilis* glyQS T-box, manually adjusted and rebuilt with Coot as needed. The two fitted models were then combined by eliminating one of the common tRNAs to create the full-length T-box-tRNA model. Hydrogen atoms were added to this complete model, and the model was refined in Phenix.real_space_refine⁸² using a nonbound weight parameter of 1,000, with other parameters remaining as default to minimize clashes. The resulting model-map correlation coefficient (CCmask) was 0.72. Cryo-EM data collection and processing and model refinement statistics are summarized in Table 2.

Small-angle X-ray scattering data collection and analysis. *G. kaustophilus* glyQ T-box discriminator-tRNA complex and *B. subtilis* full-length T-box riboswitch-tRNA complex were purified from potential aggregates by size exclusion chromatography on a Superdex 200 Increase column, in a buffer consisting of 10 mM Tris-HCl (pH 7.4), 100 mM KCl and 20 mM MgCl₂ before SAXS measurements. SAXS data collection, processing and analyses were performed as previously described^{83,84} using an in-house program package NCI-SAXS and the ATSAS⁸⁵ package by Svergun and co-workers (<http://www.embl-hamburg.de/biosaxs/>). X-ray scattering measurements were performed at room temperature at the beamline 12-ID-B of the Advanced Photon Source of the Argonne National Laboratory, adjusted to achieve scattering q values of $0.006 < q < 0.90 \text{ \AA}^{-1}$, where $q = (4\pi/\lambda)\sin\theta$, and 2θ is the scattering angle. Concentration series measurements for the same sample were carried out to remove the scattering contribution due to interparticle interactions and to extrapolate the data to infinite dilution. Thirty 2D images were recorded for each buffer or sample solution using a flow cell, with the exposure time of 0.5–1 s to minimize radiation damage and to yield optimal signal-to-noise ratio. No radiation damage was observed, as confirmed by the absence of systematic signal changes in sequentially collected X-ray scattering images. The 2D images were reduced to one-dimensional scattering profiles using the Matlab software package at the beamlines. Scattering profiles of the RNAs were calculated by subtracting the background buffer contribution from the sample

buffer profile. The solution X-ray scattering curves (Fig. 6f,g) were back calculated for the cocrystal and cryo-EM structures using CRYSOLO⁶⁴, with default parameters. The difference between the back-calculated curve and the experimental data after $q = 0.2 \text{ \AA}^{-1}$ can result from the RNA flexibility in solution, and imperfect simulations of the solvation layer and ion cloud around the RNA.

In vivo T-box-lacZ gene expression assay. The in vivo assay was performed essentially as described previously^{18,42}. A pRB382 plasmid that carries a transcriptional fusion between the *G. kaustophilus* glyQ T-box 5' untranslated region and *lacZ* and a pBAD18 plasmid that conditionally expresses *B. subtilis* tRNA^{Gly} were co-transformed into *E. coli* M5154 competent cells ($\Delta lacZ$; Extended Data Fig. 4). Cells were cultured in the presence of L-arabinose to induce tRNA^{Gly} expression, under glycine starvation (-glycine) and replete (+glycine) conditions. β -galactosidase activities of wild-type and mutant T-boxes under glycine-replete and glycine-starvation conditions were normalized to that of the wild-type T-box-containing strain grown in minimal medium supplemented with glycine (Extended Data Fig. 4). The values and error bars represent mean and s.d., $n = 3$ biologically independent samples.

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

Data availability

Atomic coordinates and structure factor amplitudes for the T-box discriminator in complex with tRNA^{Gly} have been deposited at the Protein Data Bank (PDB) under accession code PDB 6PMO. Cryo-EM structure of the full-length T-box-tRNA complex and map have been deposited to Electron Microscopy Data Bank under EMD-20416 and PDB 6POM. All other data are available upon reasonable request.

References

- Kalvari, I. et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **46**, D335–D342 (2018).
- Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
- Sampson, J. R. & Uhlenbeck, O. C. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc. Natl Acad. Sci. USA* **85**, 1033–1037 (1988).
- Zhang, J. & Ferre-D'Amare, A. R. A flexible, scalable method for preparation of homogeneous aminoacylated tRNAs. *Methods Enzymol.* **549**, 105–113 (2014).
- Ferré-D'Amaré, A. R. & Doudna, J. A. Use of cis- and trans-ribozymes to remove 5' and 3' heterogeneities from milligrams of in vitro-transcribed RNA. *Nucleic Acids Res.* **24**, 977–978 (1996).
- Xiao, H., Murakami, H., Suga, H. & Ferré-D'Amaré, A. R. Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **454**, 358–361 (2008).
- Hood, I. V. et al. Crystal structure of an adenovirus virus-associated RNA. *Nat. Commun.* **10**, 2871 (2019).
- Keller, S. et al. High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.* **84**, 5066–5073 (2012).
- Zhao, H., Piszczek, G. & Schuck, P. SEDPHAT—a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods* **76**, 137–148 (2015).
- Zeng, C., Zhou, S., Bergmeier, S. C. & Hines, J. V. Factors that influence T box riboswitch efficacy and tRNA affinity. *Bioorg. Med. Chem.* **23**, 5702–5708 (2015).
- Yousef, M. R., Grundy, F. J. & Henkin, T. M. tRNA requirements for glyQS antitermination: a new twist on tRNA. *RNA* **9**, 1148–1156 (2003).
- Foadi, J. et al. Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallogr. D* **69**, 1617–1632 (2013).
- Adams, P. et al. PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Cryst. D* **66**, 213–221 (2010).
- McCoy, A. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Chou, F. C., Sripradkeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76 (2013).
- Chen, M. et al. Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* **14**, 983–985 (2017).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).

83. Wang, J. et al. A method for helical RNA global structure determination in solution using small-angle x-ray scattering and NMR measurements. *J. Mol. Biol.* **393**, 717–734 (2009).
84. Fang, X. et al. An unusual topological structure of the HIV-1 Rev response element. *Cell* **155**, 594–605 (2013).

Acknowledgements

We thank I. Botos for computational support, J.R. Hogg, S. Ranganathan, G. Piszczek, D. Wu, J.C. Lee and M. Watson for support in fluorescence analyses; Y. He and N. Tjandra for fermentation support; R. Levine and D.-Y. Lee for MS support; W. Zhang and J. W. Szostak for a gift of Ir(III) Hexamine; M. Apostolidi for β -galactosidase assay protocols; CAS-Shanghai Science Research Center High-End User Project for preliminary cryo-EM data collection; and S.K. Buchanan, A. Ferré-D'Amaré, N. Baird, M. Lau, K. Suddala, C. Bou Nader, and J. M. Gordon for insightful discussions. This work was supported by the intramural research programs of NIDDK and NCI, an NIH DDIR Innovation Award to J.Z. and Y.-X.W., NIH U54GM103297 and U54AI150470 (Center for HIV RNA Studies, CRNA), P41GM103832, R01GM079429 and S10OD021600 to W.C., and Fundación Santé E515 and project “INSPIRED” (MIS/5002550) implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund), and a Fulbright Scholarship to C.S. V. S. is supported by an IKY-Siemens 2017 Excellence Postdoctoral Fellowship.

Author contributions

S.L. and J.Z. designed experiments. S.L. prepared RNA samples, cocrystals, and samples for cryo-EM and SAXS, with the help of F.E.H, and performed in vitro assays. S.L. and J.Z. collected X-ray diffraction data, solved and refined the crystal structure, and analyzed SAXS data. Z.S. collected cryo-EM data, Z.S., G.D.P., K.Z., M.C., S.J.L. and W.C. performed cryo-EM data processing and modeling. J.L. performed phylogenetic analyses. V.S., N.G. and C.S. carried out in vivo experiments. L.F. and Y.-X.W. collected and processed SAXS data. All authors contributed to the preparation of the manuscript.

Competing interests

Authors declare no competing interests.

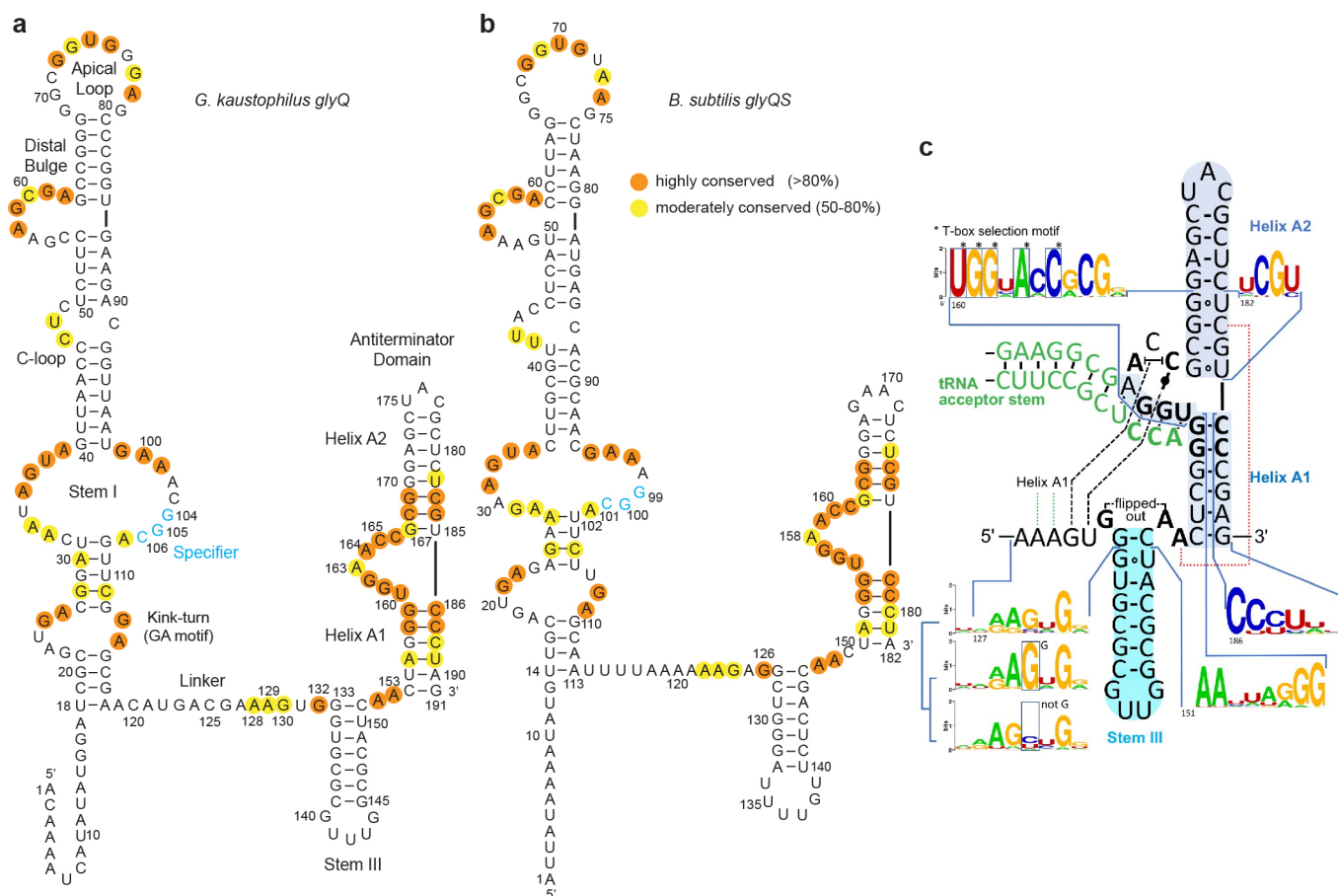
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41594-019-0326-7>.

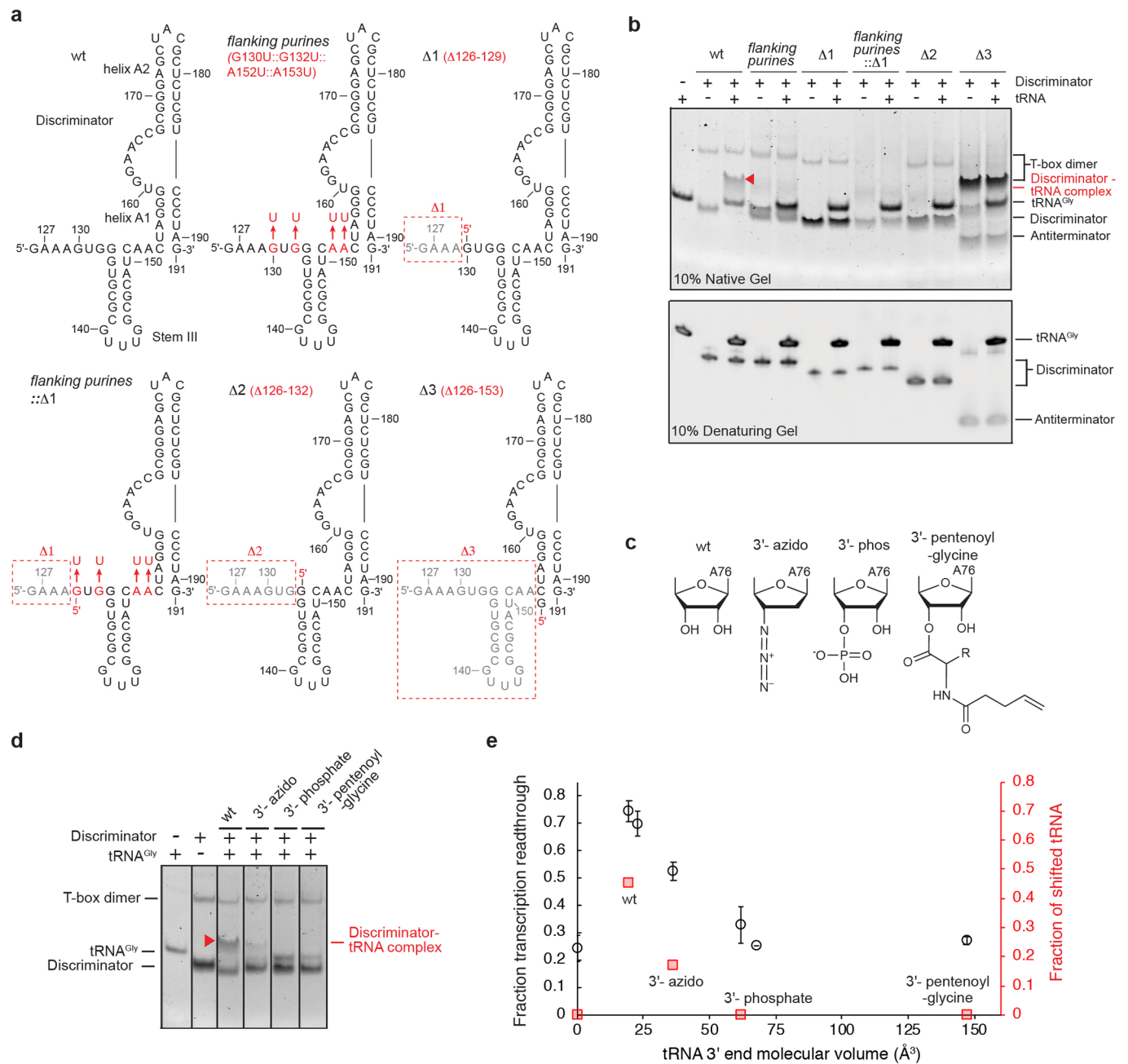
Correspondence and requests for materials should be addressed to W.C. or J.Z.

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.

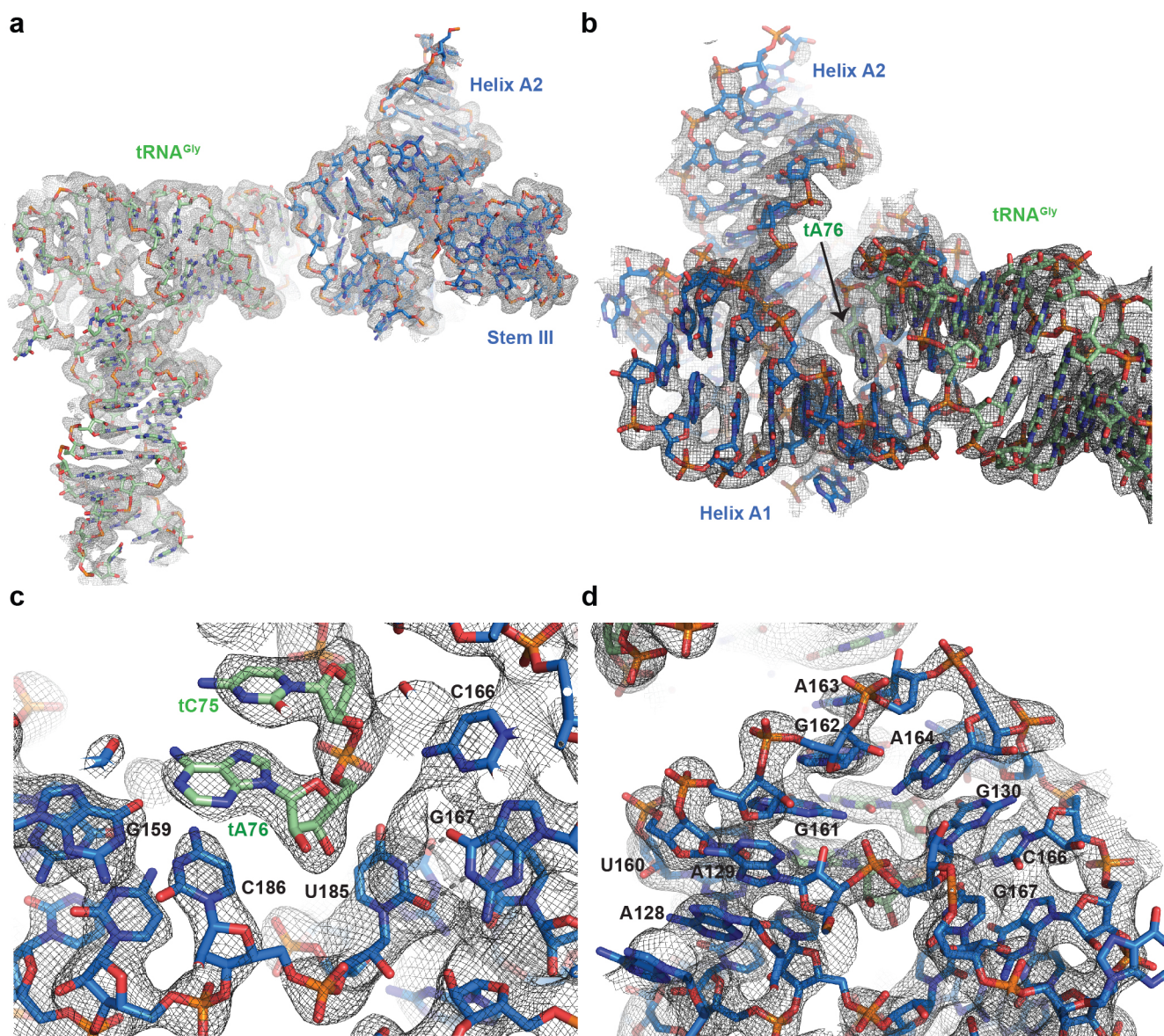
Reprints and permissions information is available at www.nature.com/reprints.



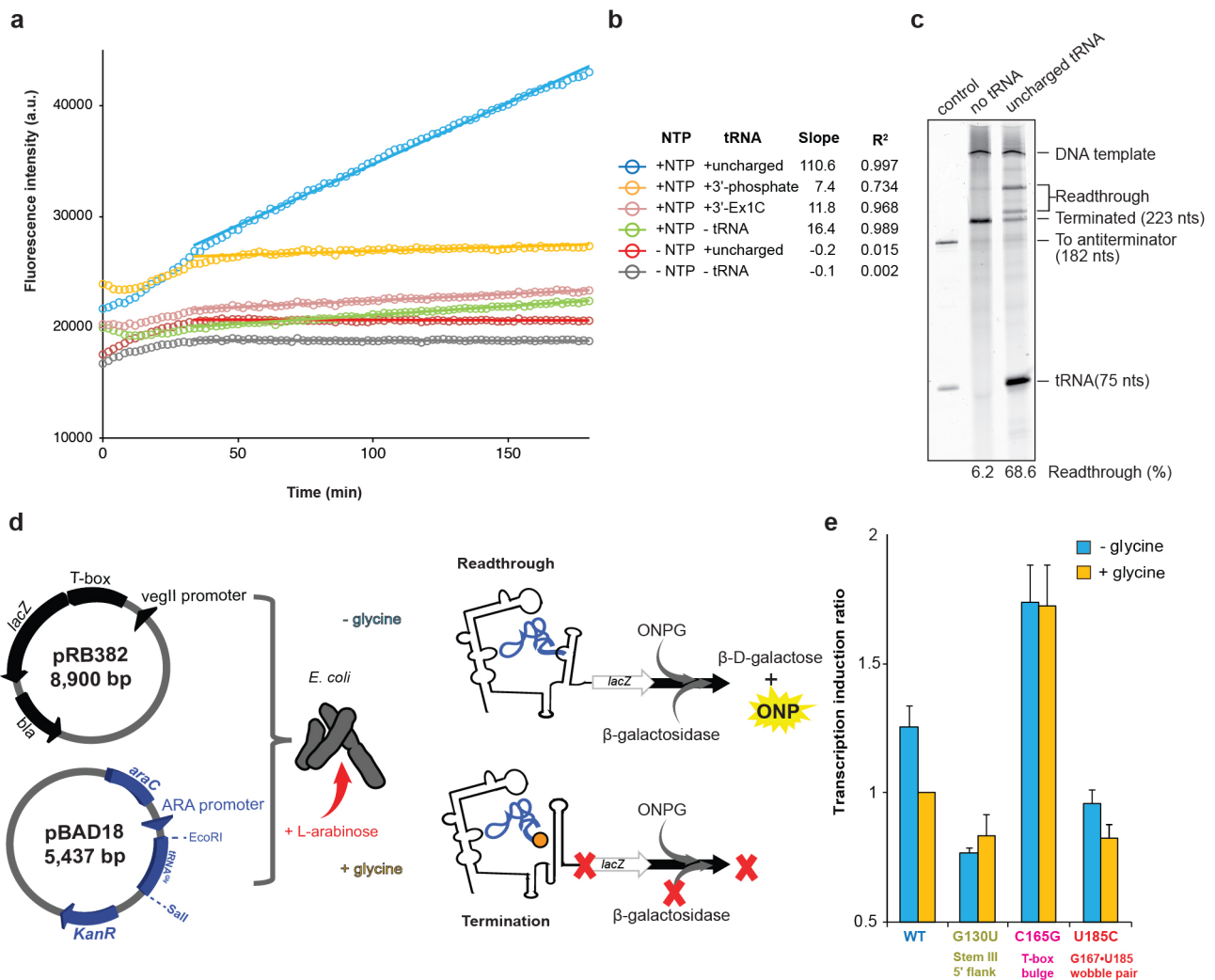
Extended Data Fig. 1 | Secondary structures and conservation analyses of glycyI T-box riboswitches. a, b, Secondary structures of *G. kaustophilus glyQ* and *B. subtilis glyQS* T-box riboswitch. Glycine-specific T-boxes lack the stem II and stem IIA/B pseudoknot structures. Conserved nucleotides are highlighted, based on previous reports supplemented by new phylogenetic analysis (Fig. 1; Methods). Previous sequence annotations of the *G. kaustophilus glyQ* T-box had omitted a 5' ssRNA leader that precedes stem I in all validated T-boxes, which is now restored. The probable transcription start site, 17 nts upstream of stem I, was identified using prokaryotic promoter prediction algorithms. Nucleotide numbering is thus offset by +17 relative to previous reports. **c,** Sequence conservation of the T-box discriminator region based on *G. kaustophilus glyQ* T-box. The split patterns show that the intercalating G130 is at the center of a 5'-AR(U/A)-3' motif (middle). This motif is shifted 1 nt to the left when there is no G in position 130 (bottom). In this case, a moderately conserved G is predominant in position 129, and two pyrimidines are present in positions 130 and 131. Assuming that G129 is the intercalating nucleotide equivalent to G130 in the middle panel, one of these pyrimidines (nt 130 or 131) may adopt an extrahelical conformation to account for the motif shift.



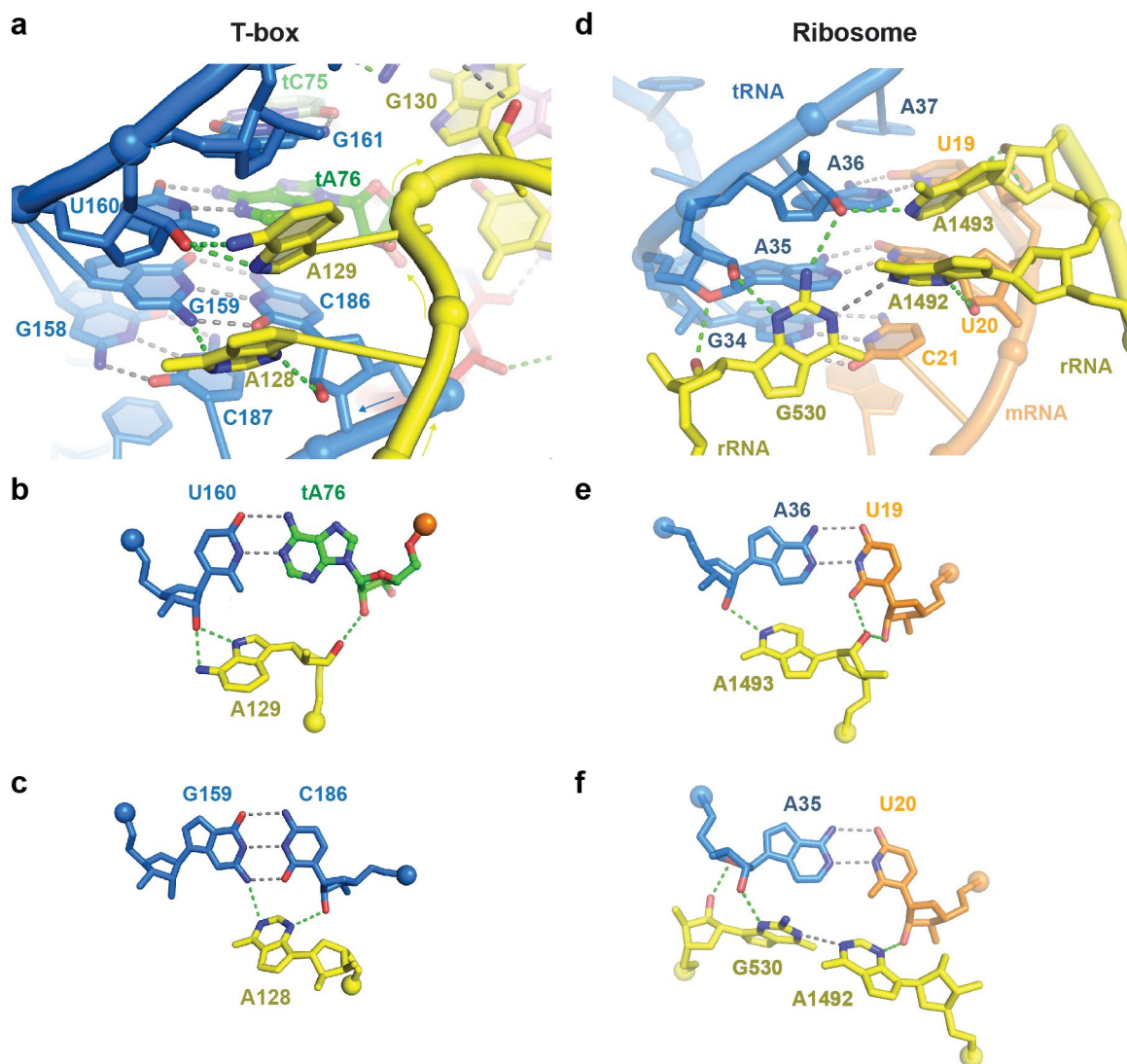
Extended Data Fig. 2 | Mutational analysis of T-box discriminator-tRNA interactions. **a**, Secondary structures of wild-type, mutant, and truncated T-box discriminators. Deletions are indicated by red boxes. **b**, Electrophoretic mobility shift assay (EMSA) analysis of the constructs shown in **a**, showing the requirement of stem III and flanking purines for tRNA binding. The antiterminator (discriminator without stem III and its flanking purines; $\Delta 3$ mutant) is prone to dimerization. **c**, tRNA variants used that carry various 3' chemical modifications. Only the terminal tA76 is shown. **d**, EMSA analysis of constructs in **c**, showing that binding is selective for uncharged tRNA. **e**, Quantitation of **d** and comparison with previously reported in vitro transcription readthrough data of the same tRNA variants. The values and error bars represent mean and s.d., $n=3$ biologically independent samples.



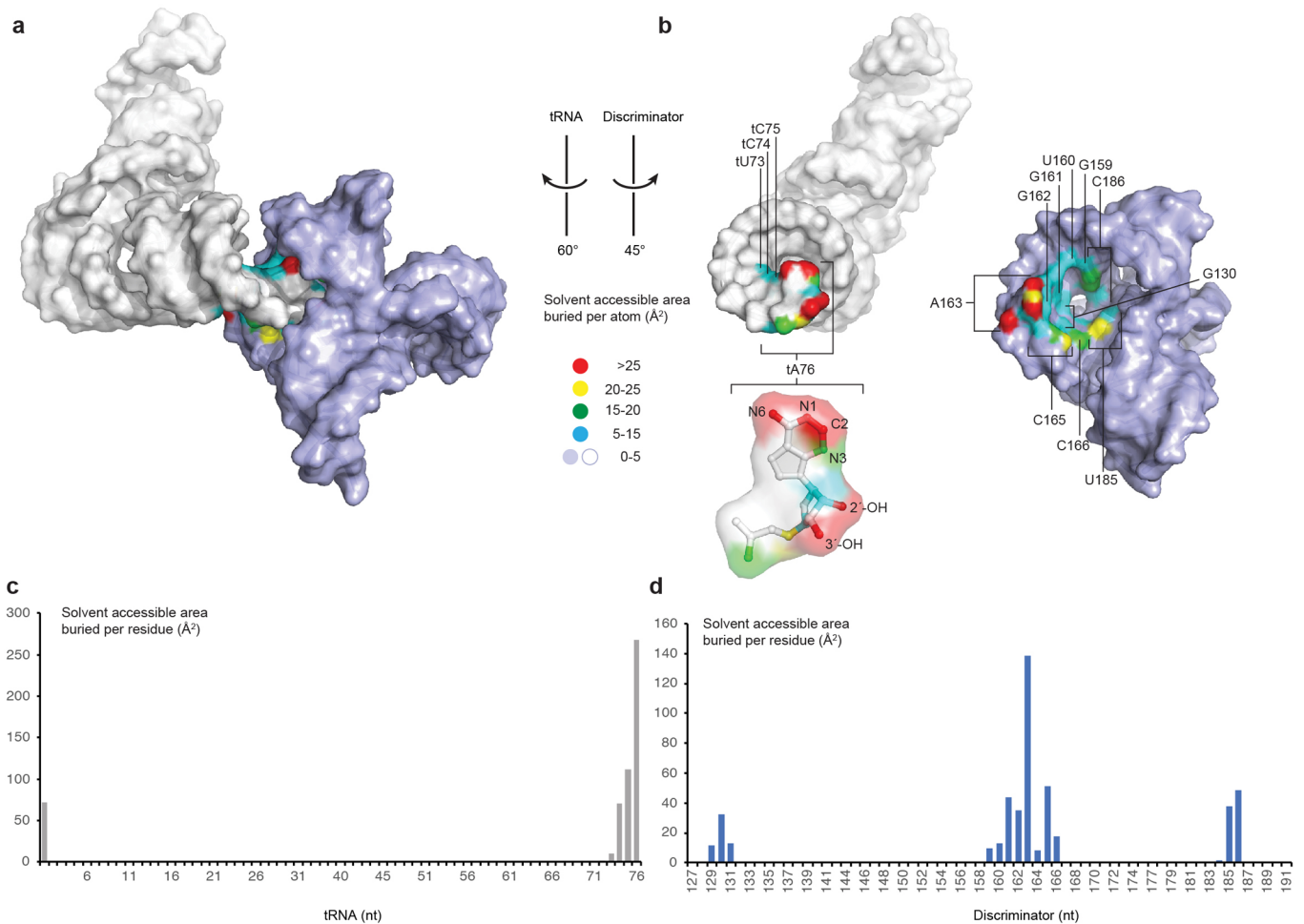
Extended Data Fig. 3 | Representative X-ray crystallographic electron density maps. **a**, Composite simulated anneal-omit $2F_o - F_c$ electron density calculated using the final model (1.0 s.d.) superimposed with the final refined model. **b–d**, Portions of the map showing tRNA-T-box discriminator coaxial stacking (**b**), encapsulation of tRNA 3'-end by the discriminator (**c**), and long-range interactions between stem III 5' flanking purines and the T-box bulge (**d**). Note the density fusion as a result of nucleobase-ribose packing interactions between A129 and G161.



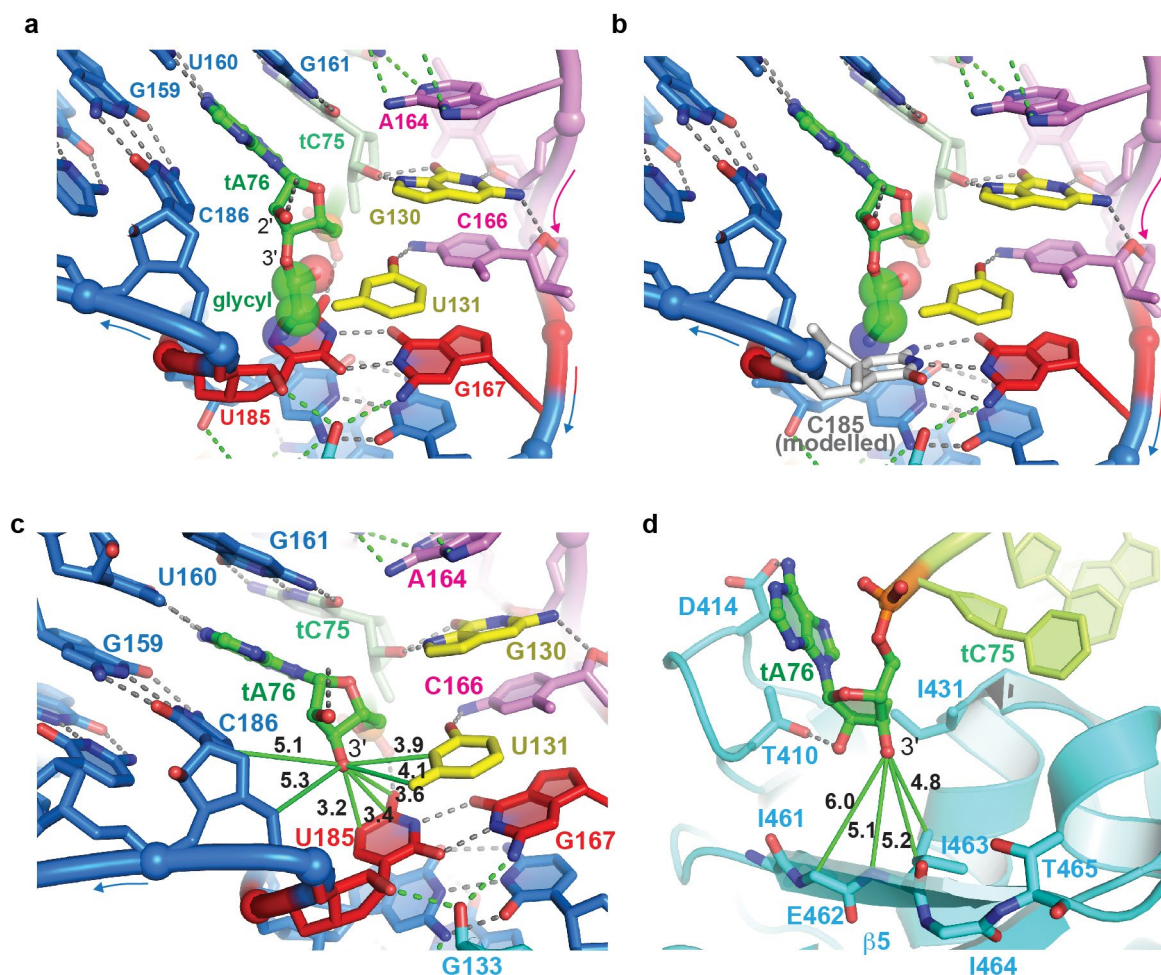
Extended Data Fig. 4 | In vitro transcription termination-readthrough assay and in vivo β -gal assay. **a, Representative raw data of in vitro transcription termination-readthrough assay using wild-type *B. subtilis* T-box riboswitch. The rates of fluorescence increase between 34 and 180 min (segments with trendlines) report the production of readthrough transcripts. **b**, Quantitation of data in **a**. Rates of fluorescence increase (slopes) were subsequently normalized to that of the reference in the presence of NTP but absence of tRNA (green data points) and reported in Figs. 2h, 3e and 4f. **c**, Validation of fluorescence-based readthrough assay in **a** and **b** with subsequent, conventional gel-based analysis of the same samples. Addition of the uncharged tRNA led to significantly increased transcription readthrough. **d**, Scheme of in vivo gene expression assay using the *G. kaustophilus* glyQ T-box riboswitch transcriptionally fused with *lacZ*. **e**, Relative β -gal activity of wild-type and mutant T-boxes under glycine-replete and glycine-starvation conditions, normalized to wild-type T-box-containing strain grown in minimal media supplemented with glycine. The values and error bars represent mean and s.d., $n = 3$ biologically independent samples.**



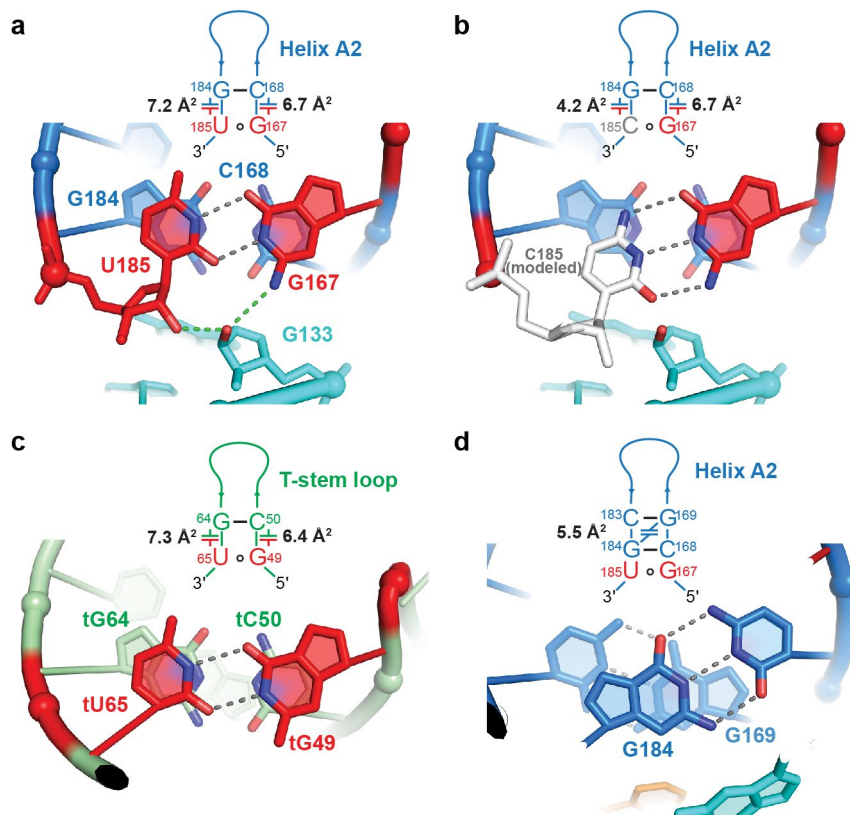
Extended Data Fig. 5 | Comparison of the T-box tandem A-minor latch with the A1492-A1493-G530 latch in the ribosome A site. a, The A128-A129 latch reinforces the functionally important tRNA (green)-helix A1 (blue) stacking interface. A128 and A129 form a continuous adenosine stack. **b and c**, The stacked A128 and A129 engage extensive hydrogen bonds with the minor groove, reinforce tRNA-T-box base-pairing and stacking, and “staple” the two RNAs together. **d**, In the ribosome A site, A1492 and A1493 similarly reinforce the intermolecular codon-anticodon duplex via tandem, stacked A-minor interactions in conjunction with G530. **e and f**, Hydrogen-bond patterns in the ribosome A site resemble those in the T-box (**b and c**).



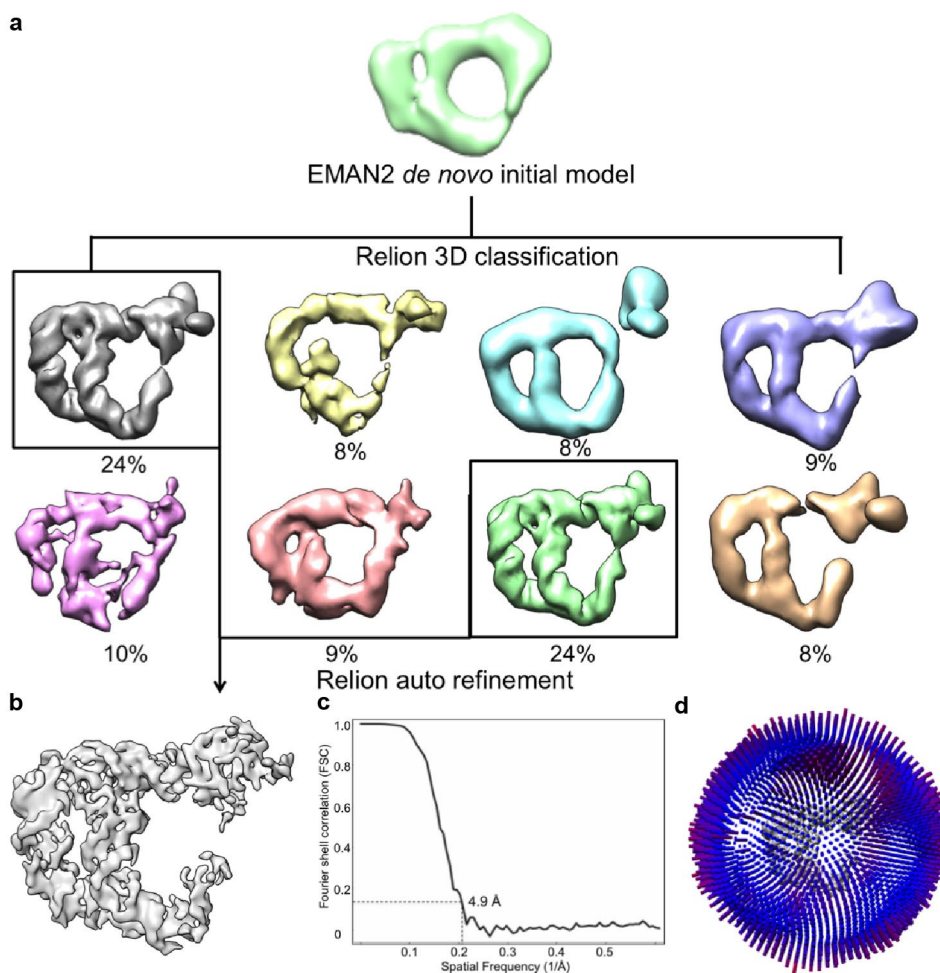
Extended Data Fig. 6 | Intermolecular interface of the T-box discriminator-tRNA complex. **a**, Solvent-accessible surface colored according to area buried from light blue or white (no burial) to red (>25 Å² per atom). **b**, Open-book view of the binding interface. The lower inset shows the extensive burial of tRNA tA76, particularly its Watson-Crick edge (N6-N1-C2) and both 2'-OH and 3'-OH. **c,d**, Solvent-accessible surface area buried per residue for tRNA (**c**) and discriminator (**d**).



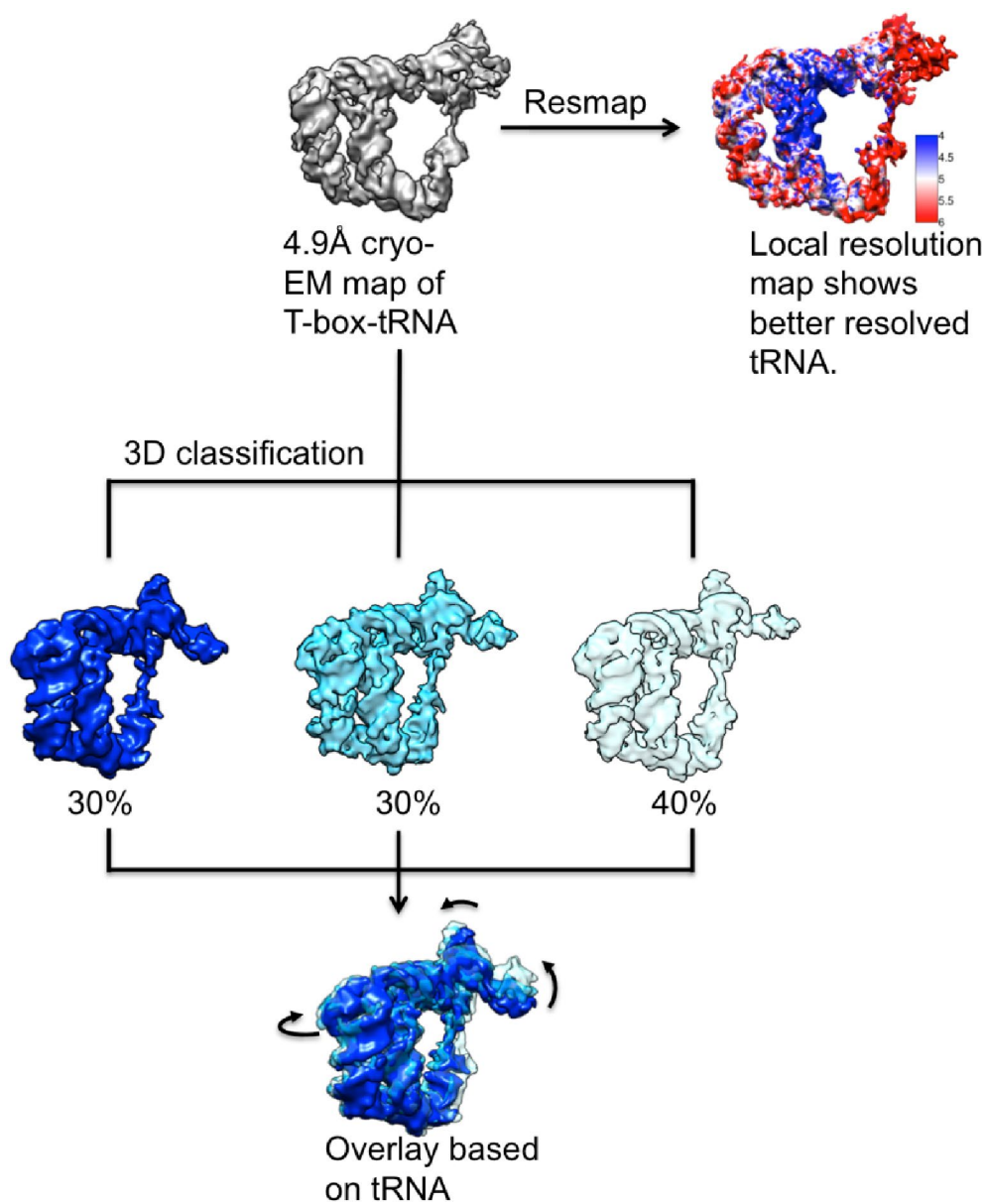
Extended Data Fig. 7 | Effect of the G•U wobble pair on tRNA aminoacylation sensing and comparison of steric sieves in the T-box and ribosome-RelA complex. a, A modeled tRNA 3'-glycyl moiety strongly clashes with the U185 nucleobase of the G•U wobble pair. **b**, Modeled Watson-Crick pair (C185, white) still clashes with the tRNA 3'-glycyl moiety, albeit to a lesser extent than the G•U wobble pair (**a**). **c,d**, Comparison of the steric sieves in the T-box (**c**) and RelA-ribosome complex (**d**). Solid green lines indicate inter-atomic distances in Å. The RelA-ribosome complex structure is based on PDB 5IQR.



Extended Data Fig. 8 | A conserved G•U wobble pair enhances stacking with its neighboring base pair both in the T-box discriminator and in the tRNA T-loop. a, Through local helix underwinding, helix A2 terminal G•U wobble pair produces exceptionally large nucleobase overlap areas and enhances stacking with the penultimate C-G pair. **b**, Reduced nucleobase overlap areas between a modeled G-C pair and the penultimate C-G pair. **c**, The G•U wobble pair is reminiscent of the conserved G49•U65 wobble pair found in the tRNA^{Gly} T-loop in the same complex. **d**, For comparison, the penultimate C-G pair stacks with its neighboring G-C pair with less than half of the total overlap area (5.5 \AA^2 versus 13.9 \AA^2). Overlap areas (in \AA^2) between stacked nucleobases were calculated with 3DNA.



Extended Data Fig. 9 | Cryo-EM single particle analysis (SPA) workflow of full-length *B. subtilis* T-box–tRNA complex. **a**, 3D classification yielded two major classes (black boxes) that were combined for auto refinement. **b**, Final reconstruction. **c**, FSC curve showing 4.9-Å resolution at 0.143 cut-off. **d**, Euler angle distribution of the final reconstruction.



Extended Data Fig. 10 | Relion 3D classification and local resolution of the full-length *B. subtilis* T-box-tRNA complex. 3D classification of the complex converged to three maps. Superposition of the tRNA density in three maps revealed motions of the T-box relative to tRNA as indicated by the arrows. This flexibility of the T-box RNA is the major limitation that prevented cryo-EM reconstruction from achieving higher resolutions. Resmap analysis shows that the tRNA was better resolved at ~4-Å resolution (upper right).

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 & Referees](#) 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

In-house data collection software was provided by SER-CAT beamline ID-22 at the Advanced Photon Source (APS), Argonne National Laboratory (ANL), and by SLAC National Accelerator Laboratory.

Data analysis

The X-ray diffraction data were indexed, integrated, and scaled by XDS. Multi-crystal datasets were analyzed and merged with BLEND. Ir substructure were identified by HKL2map, ShelxD, and Phenix.HySS. De novo phasing was performed using Phenix.AutoSol. Model building was performed using Coot. Refinement was performed using Phenix.Refine and ERRASER. Cryo-EM data were processed using Motioncor2, corrected by CTFFIND4. Particle picking and classifications were performed using EMAN2.2 and refinement was performed using Relion2. Modeling building was performed with Coot and refined with Phenix. SAXS data were analyzed using the ATSAS suite. No custom algorithms or softwares were used.

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

Atomic coordinates and structure factor amplitudes for the T-box discriminator in complex with tRNA-Gly has been deposited at the Protein Data Bank under the accession code 6PMO. Cryo-EM structure of the full-length T-box-tRNA complex and map have been deposited to Electron Microscopy Data Bank under the accession code EMD-20416, and PDB code 6POM. All other data generated or analyzed during this study are included in this article (and its supplementary information files).

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](https://www.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	One RNA crystal sample per X-ray dataset; one solution sample for each Cryo-EM dataset and SAXS dataset; In vitro and in vivo assays were performed in independent triplicates using distinct samples
Data exclusions	n/a
Replication	n/a
Randomization	n/a
Blinding	n/a

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	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging