Supplementary Information

Structural insights into recognition of c-di-AMP by the *ydaO* riboswitch

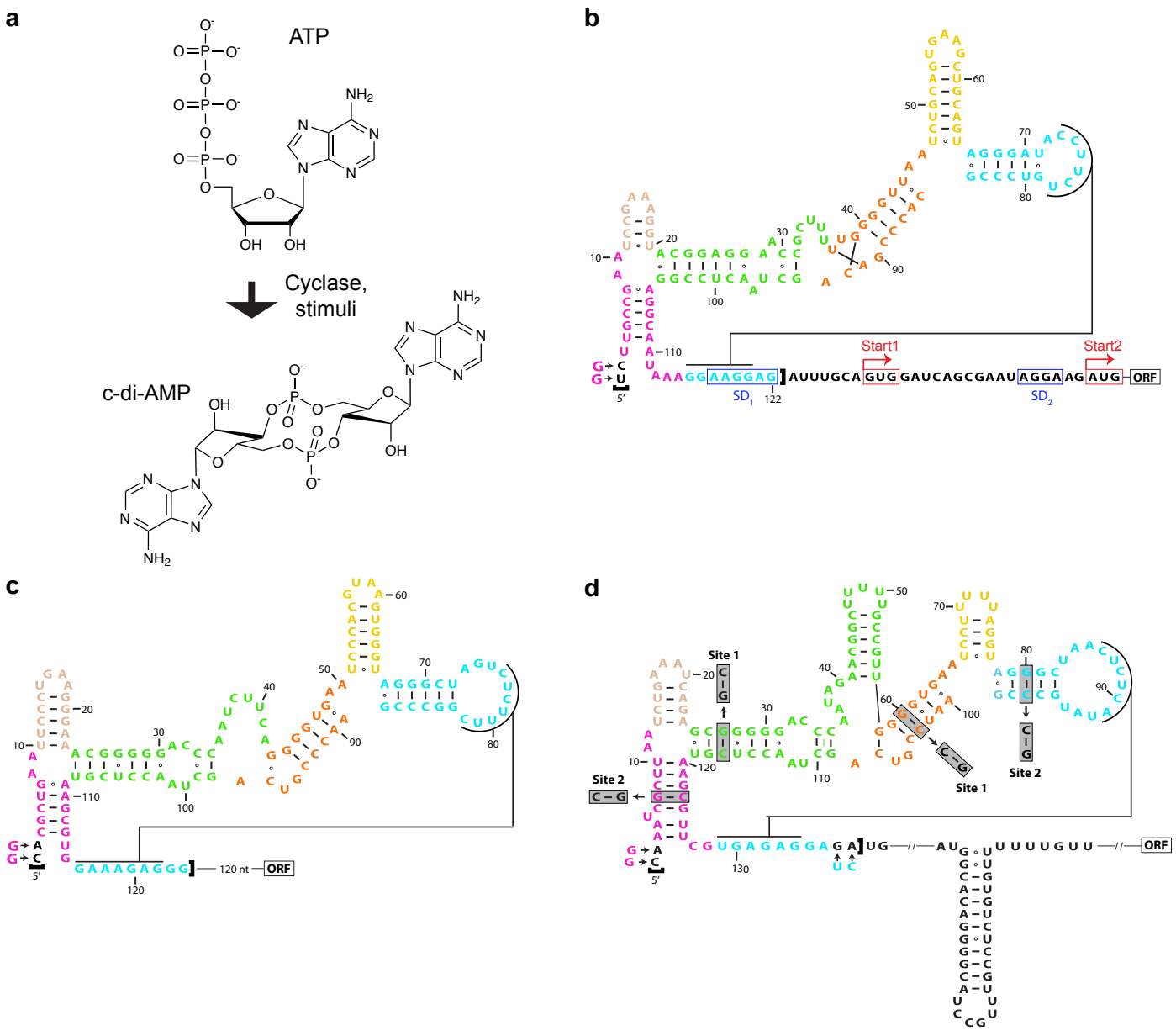
Ang Gao and Alexander Serganov

Supplementary Results

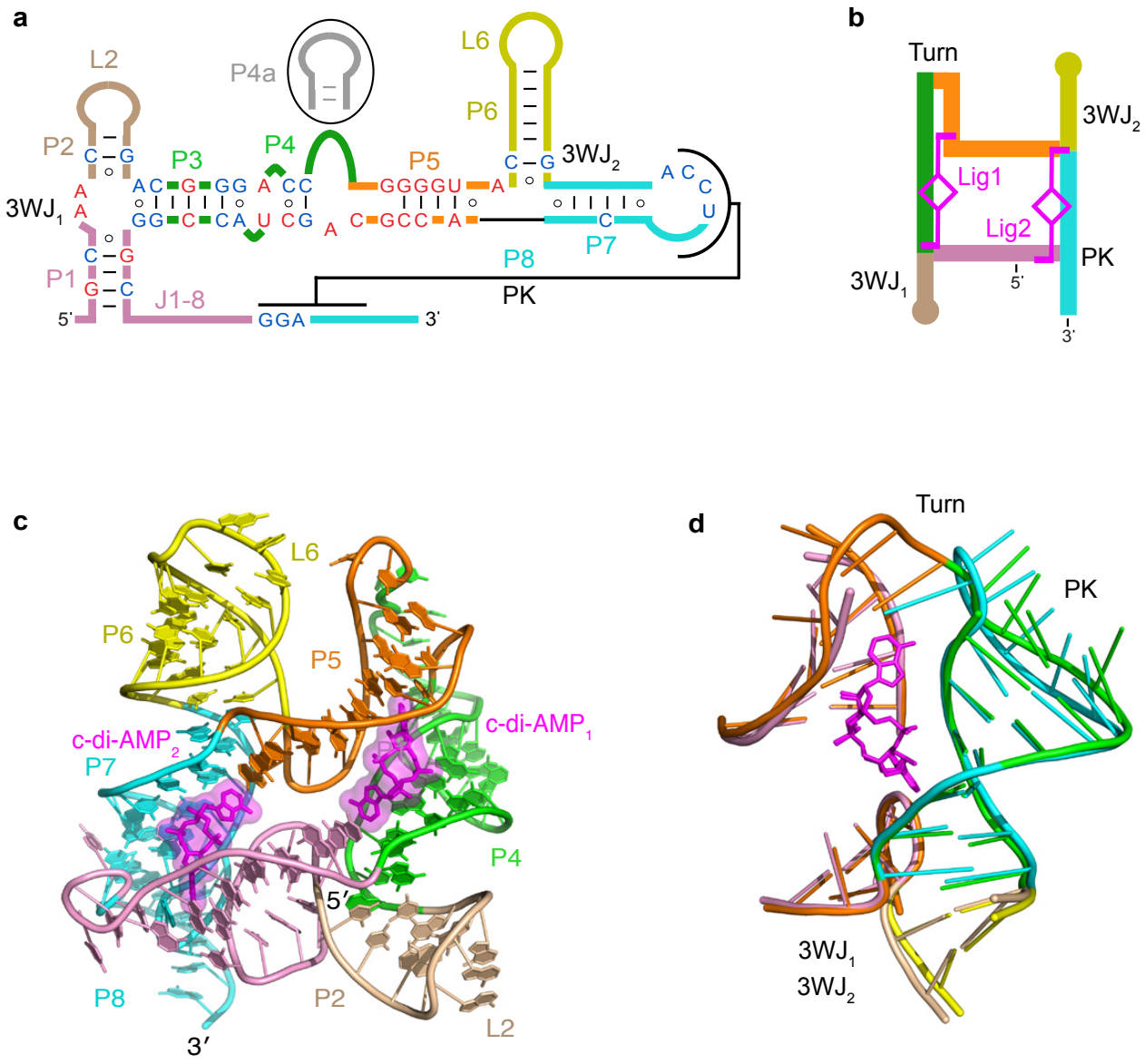
Supplementary Table 1 Data collection, phasing and refinement statistics

	T. pseudethanolicus $[Ir(NH_3)_6]^{3+}$-soaked	*T. pseudethanolicus* Native	*T. lienii* Native
Data collection			
Space group	$P3_121$	$P3_121$	$P2_13$
Cell dimensions			
a, *b*, *c* (Å)	116.0, 116.0, 114.1	114.9, 114.9, 114.7	110.3, 110.3, 110.3
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Resolution (Å)	30.00-3.20 (3.31-3.20) *	30.00-3.05 (3.16-3.05)	30.00-3.00 (3.11-3.00)
R_{sym} or R_{merge}	0.12 (0.71)	0.07 (0.52)	0.11 (0.55)
$I/\sigma I$	37.3 (3.0)	34.7 (3.5)	40.4 (6.0)
Completeness (%)	100.0 (99.7)	99.6 (99.9)	99.8 (100.0)
Redundancy	17.7 (12.1)	6.0 (6.1)	11.5 (12.2)
Refinement			
Resolution (Å)	30.00-3.20	30.00-3.05	30.00-3.00
No. reflections	28,200	16,997	9,173
R_{work}/ R_{free}	17.1 / 19.7	18.1 / 19.5	18.2 / 22.4
No. atoms			
RNA	2,578	2,626	2,516
Ligand/ion	130	98	99
Water	2	4	0
B-factors			
RNA	99.4	90.9	57.2
Ligand/ion	84.9	63.0	43.3
Water	56.7	56.1	----
R.m.s deviations			
Bond lengths (Å)	1.049	1.051	0.987
Bond angles (°)	0.005	0.005	0.005

*Highest resolution shell is shown in parenthesis.

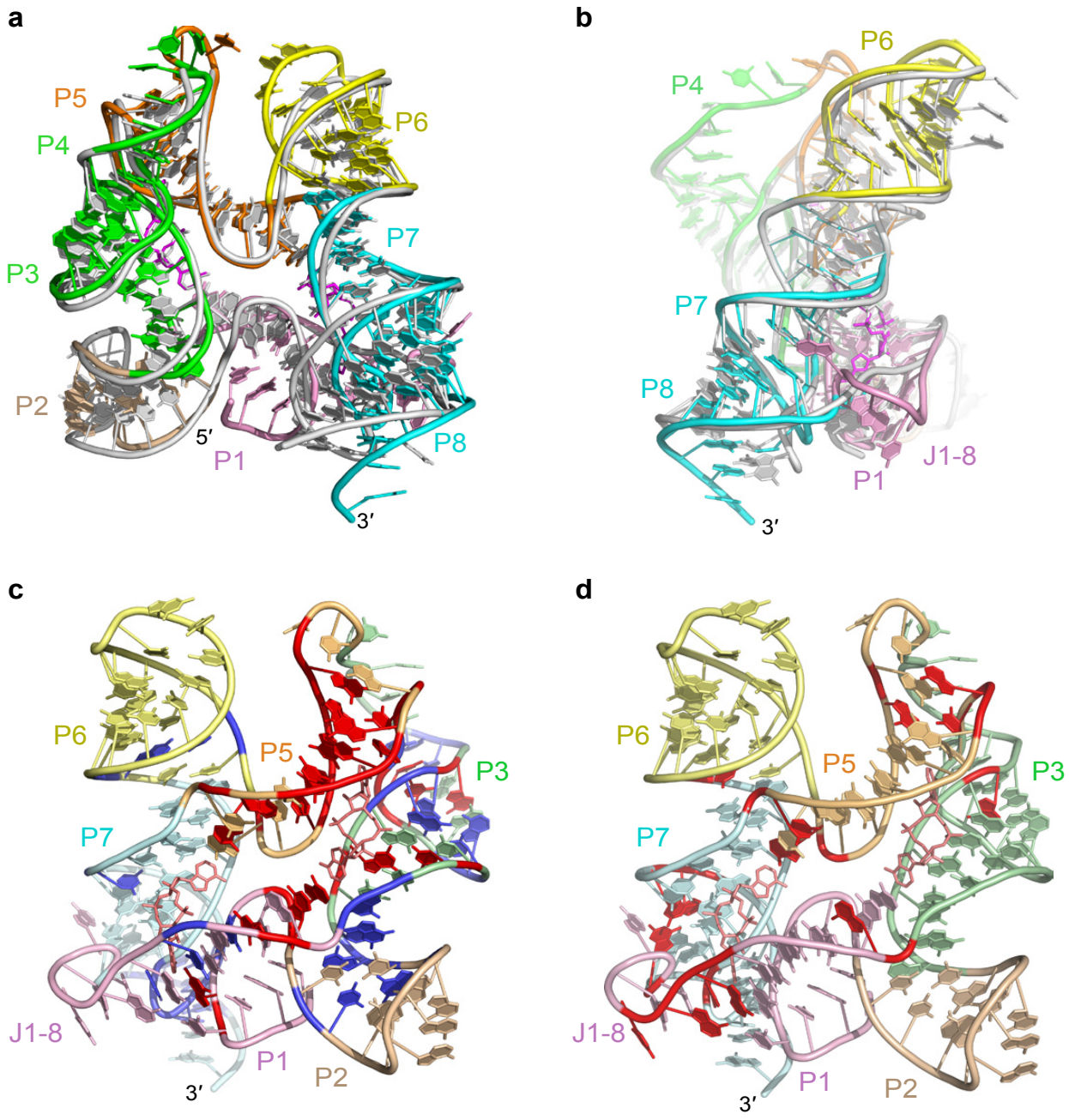


Supplementary Figure 1 | Schematics of c-di-AMP biosynthesis and *ydaO* riboswitches used in the structural and biochemical studies.
Riboswitch sequences are shown together with the RNA constructs used in the study. The RNA constructs are bracketed. Nucleotide changes in the constructs are shown with arrows. Secondary structure schematics are based on the crystal structures and the consensus secondary structure[12]. Nucleotides are color-coded according to the structural elements observed in the crystal structures. (**a**), Biosynthesis of c-di-AMP from two molecules of ATP. (**b**), The *T. pseudethanolicus* riboswitch. The riboswitch is located upstream of the gene encoding for a NLP/P60 protein of unknown function from the family of cell wall-associated hydrolases. The protein is predicted to begin from start codon 2; however, the Shine-Dalgarno sequence (SD_2) that precedes this start codon is short and is located at the unfavorably close 2-nt distance from AUG. Another potential in-frame start codon 1 is preceded by a much longer Shine-Dalgarno sequence (SD_1) positioned at the ideal 7-nt distance. Therefore, the protein translation may start from the non-canonical GUG codon located upstream of the predicted start codon. Since SD_1 is positioned within a riboswitch region that base-pairs upon c-di-AMP binding, the riboswitch may repress initiation of the protein translation via sequestering this Shine-Dalgarno sequence. (**c**), The *T. lienii* riboswitch sequence. The riboswitch is located 120 nt upstream of the gene encoding for GCN5-related N-acetyltransferase and is anticipated to be controlled via a transcription termination mechanism. (**d**), The *B. subtilis* riboswitch. The riboswitch forms a transcription terminator and aborts transcription termination upon ligand binding[12]. The site 1 and 2 mutations are shown in gray squares.

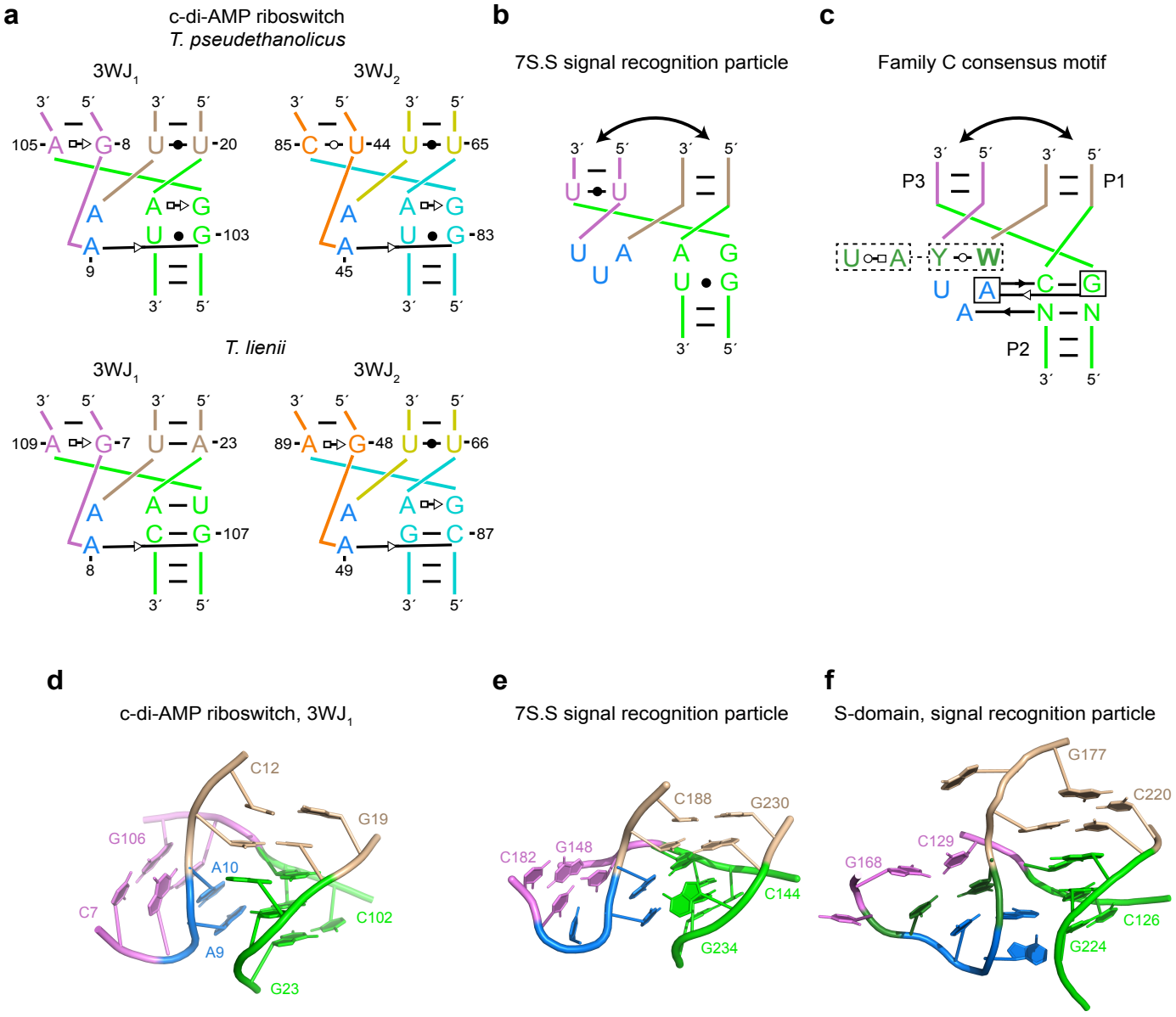


Supplementary Figure 2 | Overall structure and schematics of the *ydaO* riboswitch.

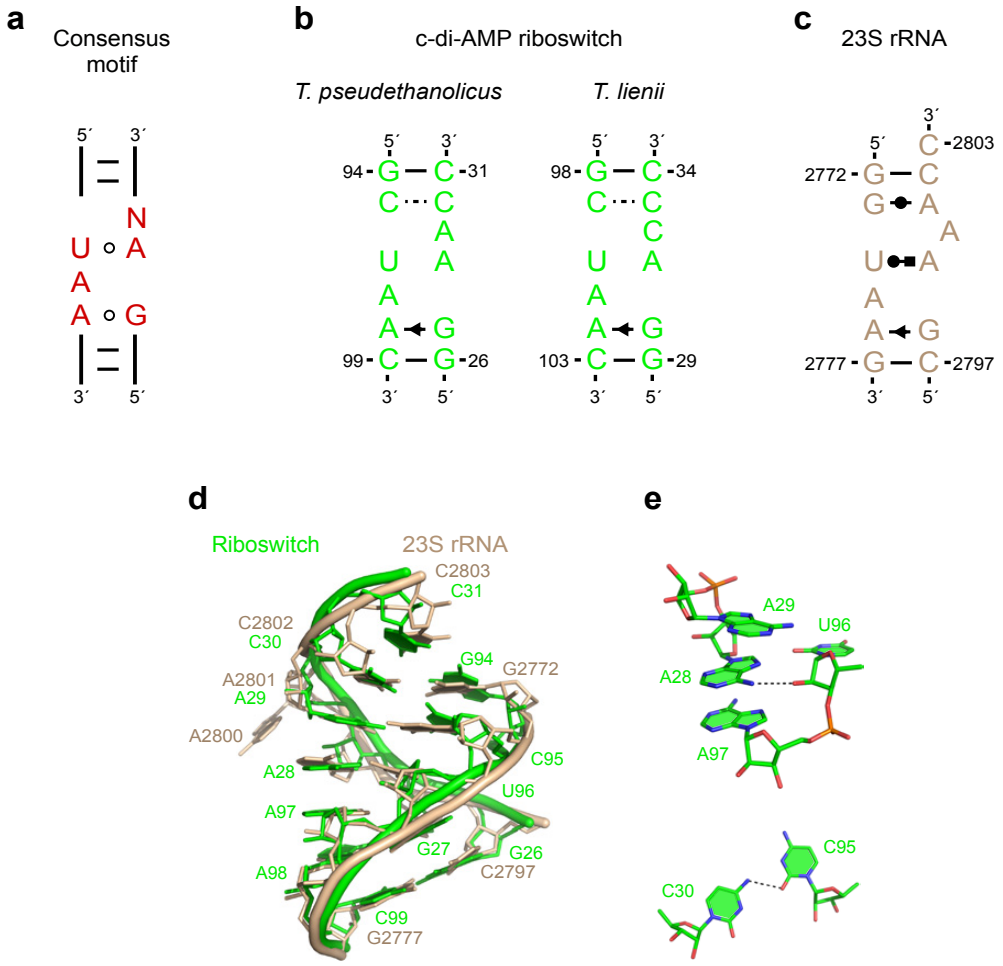
(**a**), Updated consensus secondary structure of the *ydaO* riboswitch modified according to the crystal structures. Red and blue nucleotides are conserved in more than 97% and 90% of species, respectively. An additional hairpin P4a, present in some species, is shown in gray. 3WJ, three-way junction; PK, pseudoknot. (**b**), Structure-based schematic of the 'square-shaped' riboswitch fold. The bound c-di-AMP molecules (Lig) are in magenta. (**c**), Overall structure of the *T. pseudethanolicus* c-di-AMP riboswitch in a view rotated by ~180° in respect to the Fig. 1b view. Ligands are in stick and surface representations. (**d**), Superposition of two halves of the *T. pseudethanolicus* riboswitch aligned by the c-di-AMP molecules.



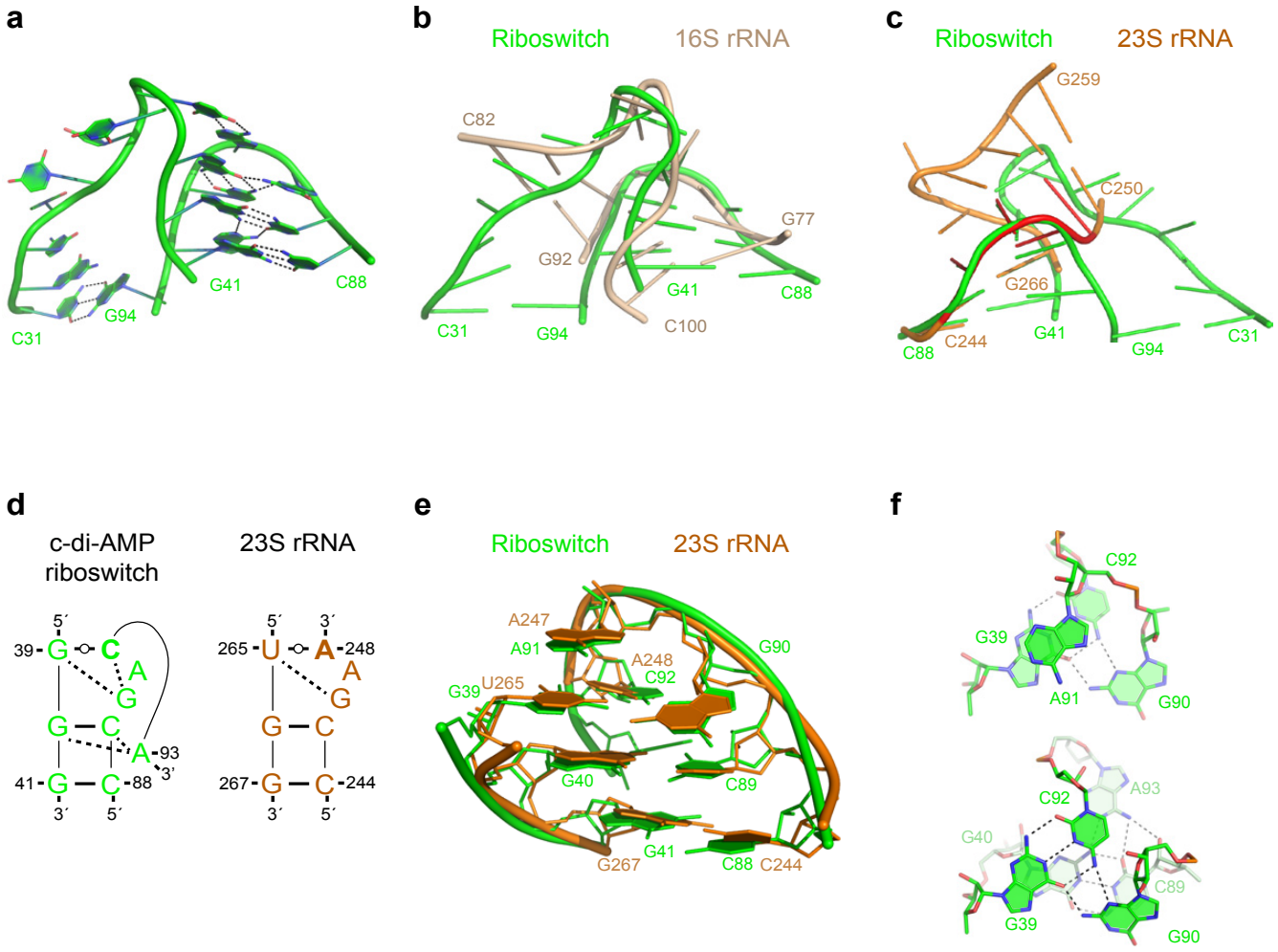
Supplementary Figure 3 | Comparison of c-di-AMP riboswitch structures, nucleotide conservation and effects of c-di-AMP binding.
(**a,b**), Superposition of the riboswitch structures from *T. pseudethanolicus* (colored) and *T. lienii* (gray). The structures were superimposed manually based on the c-di-AMP molecules and are shown in front (a) and side (b) views, respectively. (**c**), Nucleotide conservation projected on the structure of the *T. pseudethanolicus* riboswitch. Red and blue nucleotides are conserved in more than 97% and 90% of species, respectively[12]. The colors designating structural elements in this view are lighter than in other panels of the figure but follow the same color code. (**d**), Projection of the *B. subtilis* riboswitch in-line probing data on the structure of the *T. pseudethanolicus* riboswitch. Nucleotides in red show reduced scission upon c-di-AMP binding[12].



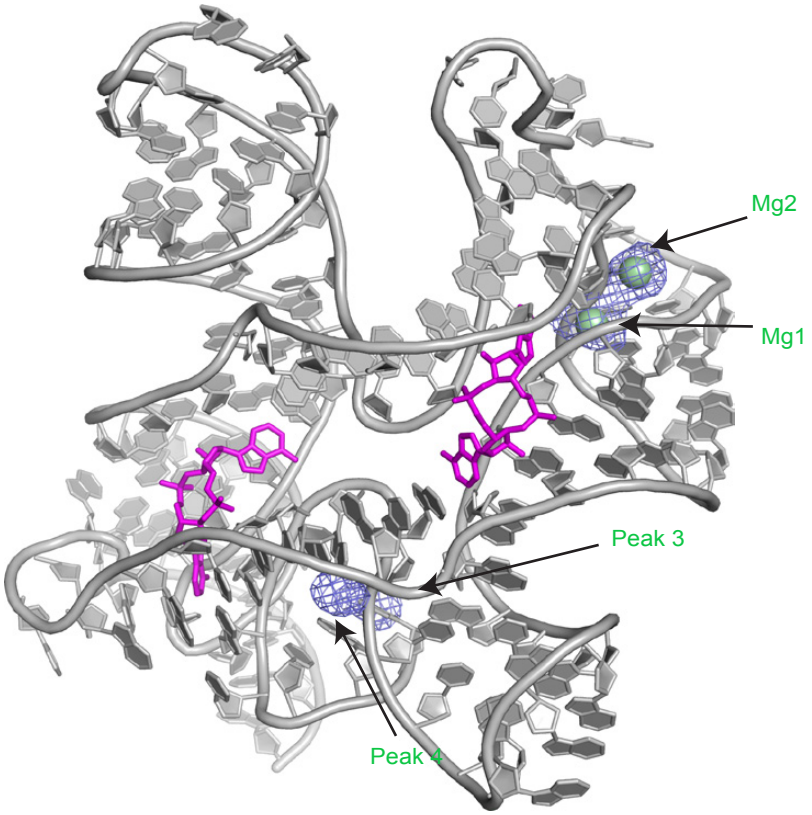
Supplementary Figure 4 | Three-way junctions from the c-di-AMP riboswitch structures and comparison with other junctions.
(**a**), Secondary structures of the three-way junctions from the *T. pseudethanolicus* (top) and *T. lienii* (bottom) riboswitches shown according to the base pair nomenclature of Leontis & Westhof. The junctions belong to family C[15] but contain an atypical short two-adenine junctional loop (blue) positioned along the minor groove of two base pairs that participate in stacking of two helical stems. (**b**), The 7S.S RNA junction from the structure of the SRP19 -7S.S signal recognition particle complex (PDB ID 1LNG)[16]. This junction also contains a short junctional loop and therefore resembles the c-di-AMP riboswitch junctions. The junctional stems are connected by a protein molecule on the top (double arrow, the protein not shown). (**c**), Consensus schematic of family C junctions. The motif is characterized by a junctional tri-loop closed by a *trans* base pair (dark green color). The tri-loop forms multiple tertiary interactions with helix P2. In almost all cases, helices P3 and P1 are connected by long-range tertiary interactions (double arrow). (**d**), Structure of $3WJ_1$ of the *T. pseudethanolicus* c-di-AMP riboswitch. (**e**), Structure of the junction from the SRP19-7S.S signal recognition particle complex (PDB ID 1LNG). (**f**), Structure of a typical family C junction from the S domain of the signal recognition particle complex (PDB ID 1MFQ).



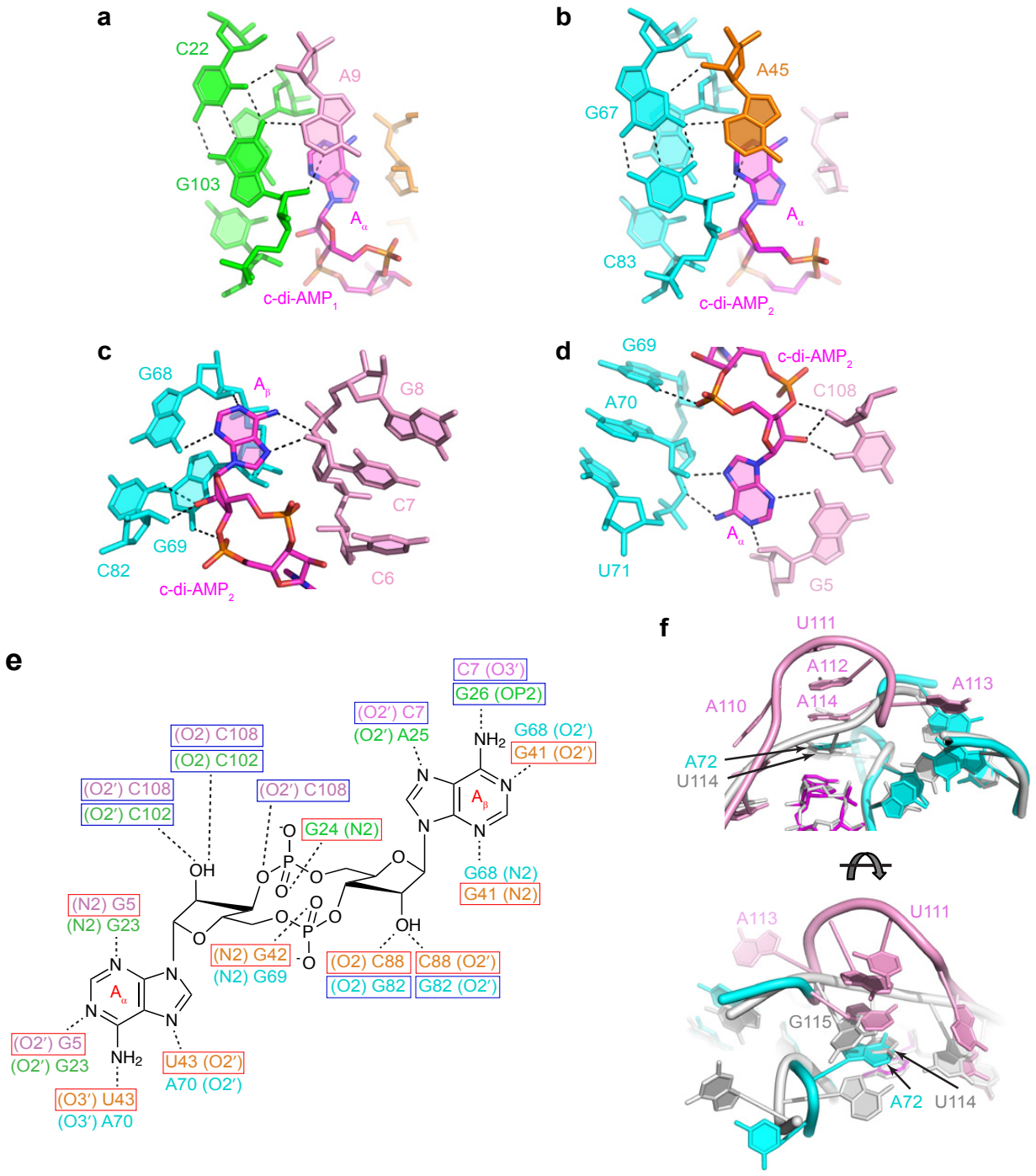
Supplementary Figure 5 | The UAA/GAN motif from the c-di-AMP riboswitch and comparison with other UAA/GAN motifs.
(**a**), The UAA/GAN consensus motif[17]. (**b**), The UAA/GAN motif from the *T. pseudethanolicus* (left) and *T. lienii* (right) c-di-AMP riboswitches. (**c**), A typical UAA/GAN motif from 23S rRNA (PDB ID 1FFK). In contrast to the UAA/GAN motif from 23S rRNA, the motif from the c-di-AMP riboswitches does not feature a looped-out adenine, is flanked by a non-canonical C-C base pair, and instead of a non-canonical U•A base pair contains U96 and A28 which interact through sugar-base hydrogen bonding. (**d**), Superposition of the *T. pseudethanolicus* riboswitch and rRNA motifs. The motifs are best superposed (r.m.s.d. 1.7 Å) when nucleotides of the 5′ strands at the top parts of the motifs are aligned with a register shift so that C31 is superposed with C2803, C30 with C2802, and A29 with A2801. (**e**), Structural details of the *T. pseudethanolicus* riboswitch motif. Zoomed-in views of the triple adenine stack with U96-A28 interactions (top) and non-canonical C30•C95 base pair (bottom).



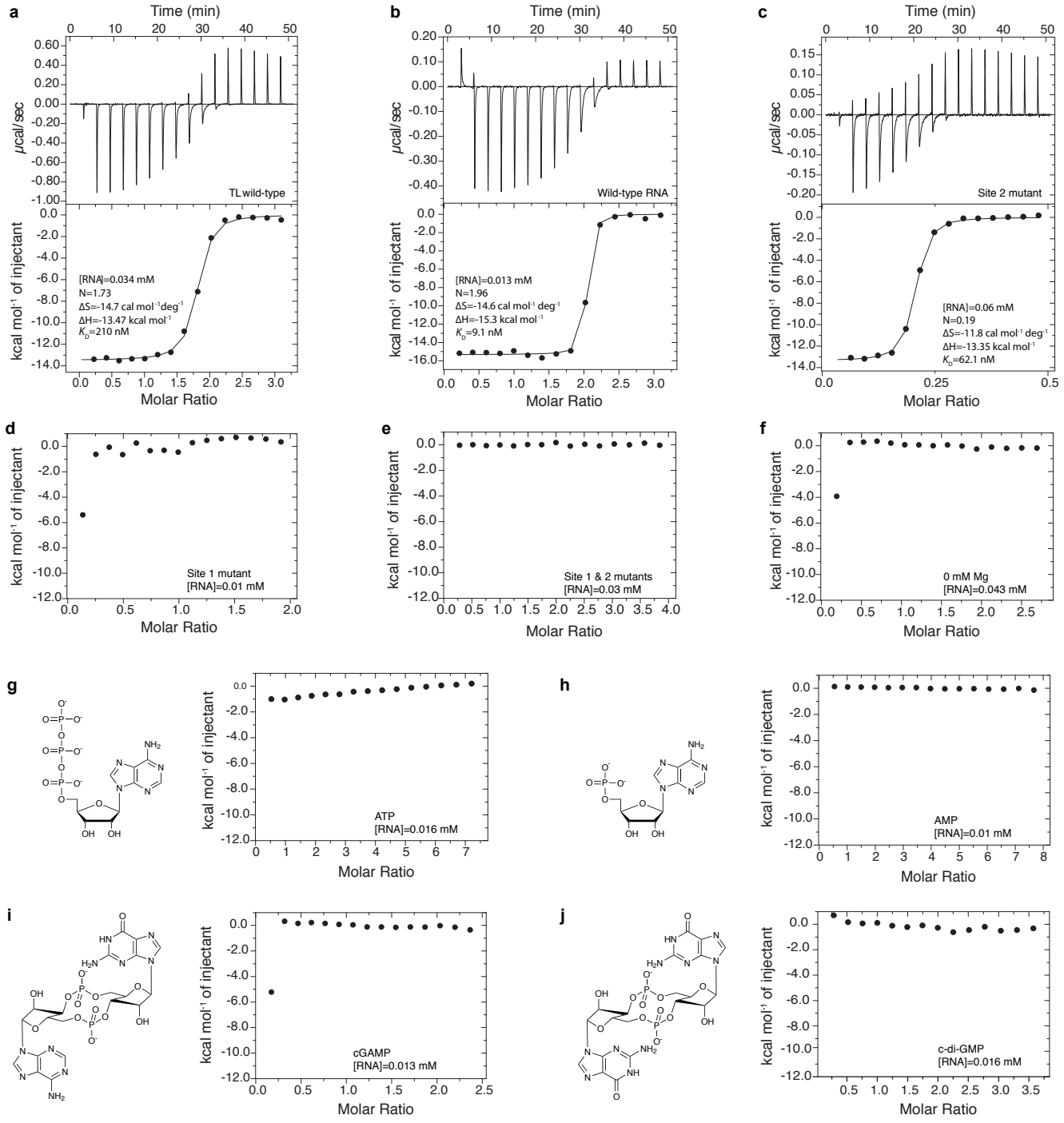
Supplementary Figure 6 | The P4-P5 turn from the *T. pseudethanolicus* c-di-AMP riboswitch and comparison with other turns.
(**a**), The crystal structure of the P4-P5 turn from the *T. pseudethanolicus* c-di-AMP riboswitch shown with hydrogen bond interactions. The turn contains several non-paired uridines in the 5′ half and several base pairs and triples in the 3′ half. The turns from the *T. lienii* (not shown) and *T. pseudethanolicus* riboswitch are similar in the regions visible in the structures. (**b**), Manual superposition of the *T. pseudethanolicus* riboswitch turn (green) with the kink-turn motif kt-7[19] (beige) from the 16S rRNA showing overall resemblance of the backbone conformations in both turns; the characteristic features of kink-turns are not present in the c-di-AMP riboswitch turn. (**c**), Manual superposition of the riboswitch turn (green) with the turn (orange) that contains an omega motif[20] (red) from 23S rRNA (PDB ID 1S72). The view is rotated by 180 degrees with respect to the view in panel b. (**d**), Schematics of the riboswitch and 23S rRNA turns. (**e**), Superposition of the 3′ half of the riboswitch turn with the corresponding part of the 23S rRNA turn. Despite large differences in overall conformations, the riboswitch and 23S rRNA turns share several features in the 3′ halves including unpaired adenine (A91 and A247) stacked on the major groove triple (G39•C92•G90 in the riboswitch and U265•A248•G246 in the 23S rRNA). (**f**), Zoomed-in views of stacking and hydrogen bonding in the riboswitch turn. The unpaired adenine A91 stacks on the non-canonical G39•C92 (U265•A248 in the rRNA) base pair formed by the nucleotides in a parallel strand orientation with C92 (A248 in rRNA) adopting *syn* conformation. The G39•C92•G90 triple is stacked on the 'A-amino kissing' triple[21] G40-C92•A93 that involves base-paring of the Watson-Crick edge of A93 with the minor groove of the G40-C92 base pair. In the 23S rRNA, the counterpart of A23 is paired within the 5′ half of the turn, while in the riboswitch, A93 does not form pairing interactions in the 5′ half of the turn and instead reinforces the 3′ half of the turn through the formation of a base triple.



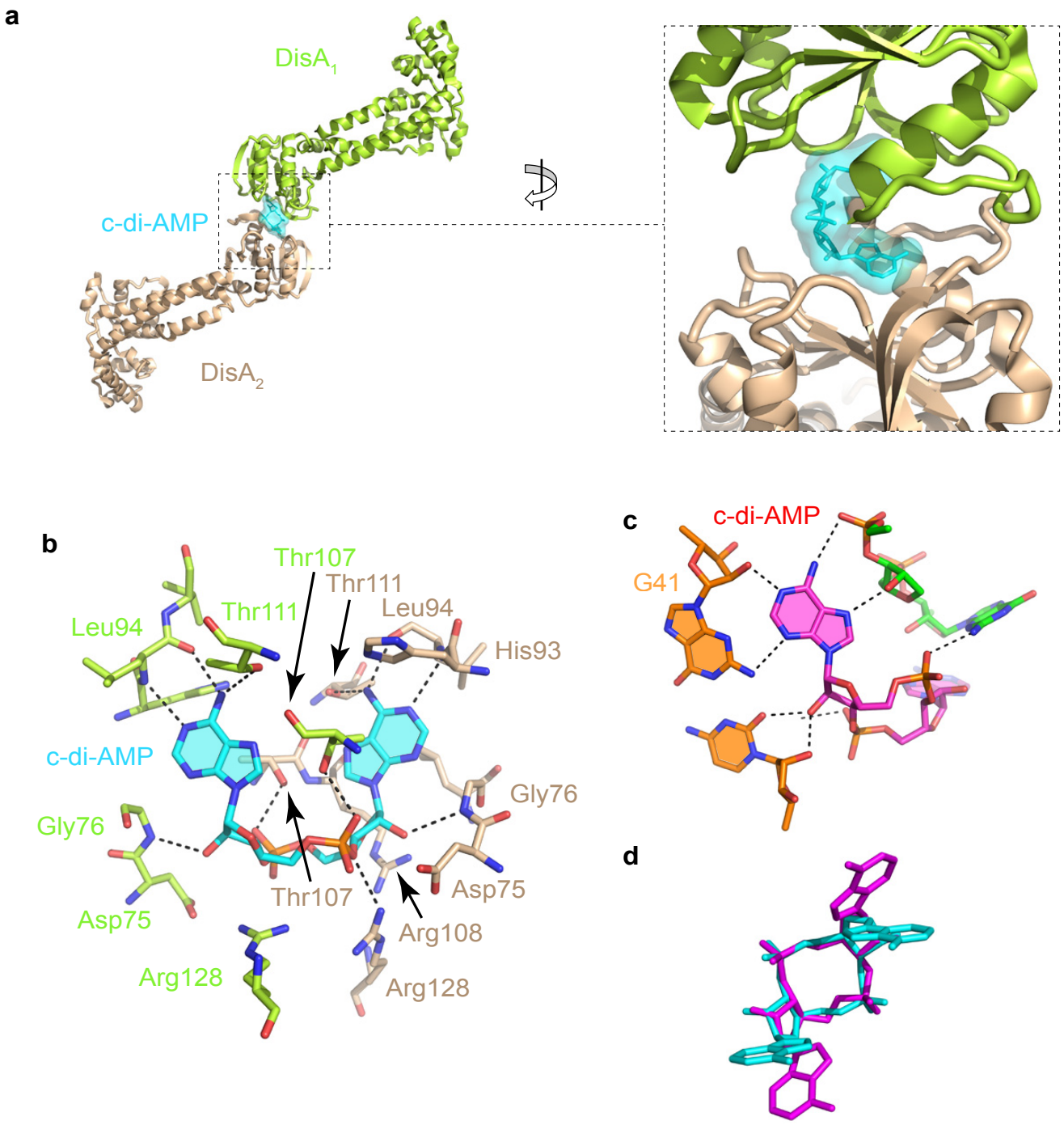
Supplementary Figure 7 | Mn^{2+} binding sites in the *T. pseudethanolicus* c-di-AMP riboswitch. The refined riboswitch model is superposed with the Mn^{2+} anomalous map (blue) contoured at the 5 σ level. The anomalous data were collected at 1.54 Å wavelength. Four peaks have been identified in the anomalous density map. Two peaks, 1 and 2, correspond to the Mg^{2+} cations (Mg1 and Mg2, green spheres) identified in the structures of both *T. pseudethanolicus* and *T. lienii* riboswitches. Peak 3 corresponds to a Mg^{2+} cation which interacts with the 5´ tri-phosphate moiety. This cation has not been added to the final structure models because of conformational heterogeneity of the tri-phosphate moiety and difficulties in interpretation of the map. Peak 4 corresponds to either a Mn^{2+}-specific site or a loosely bound Mg^{2+} cation which does not provide a substantial $2F_o$-F_c density map. This cation has not been added to the structure.



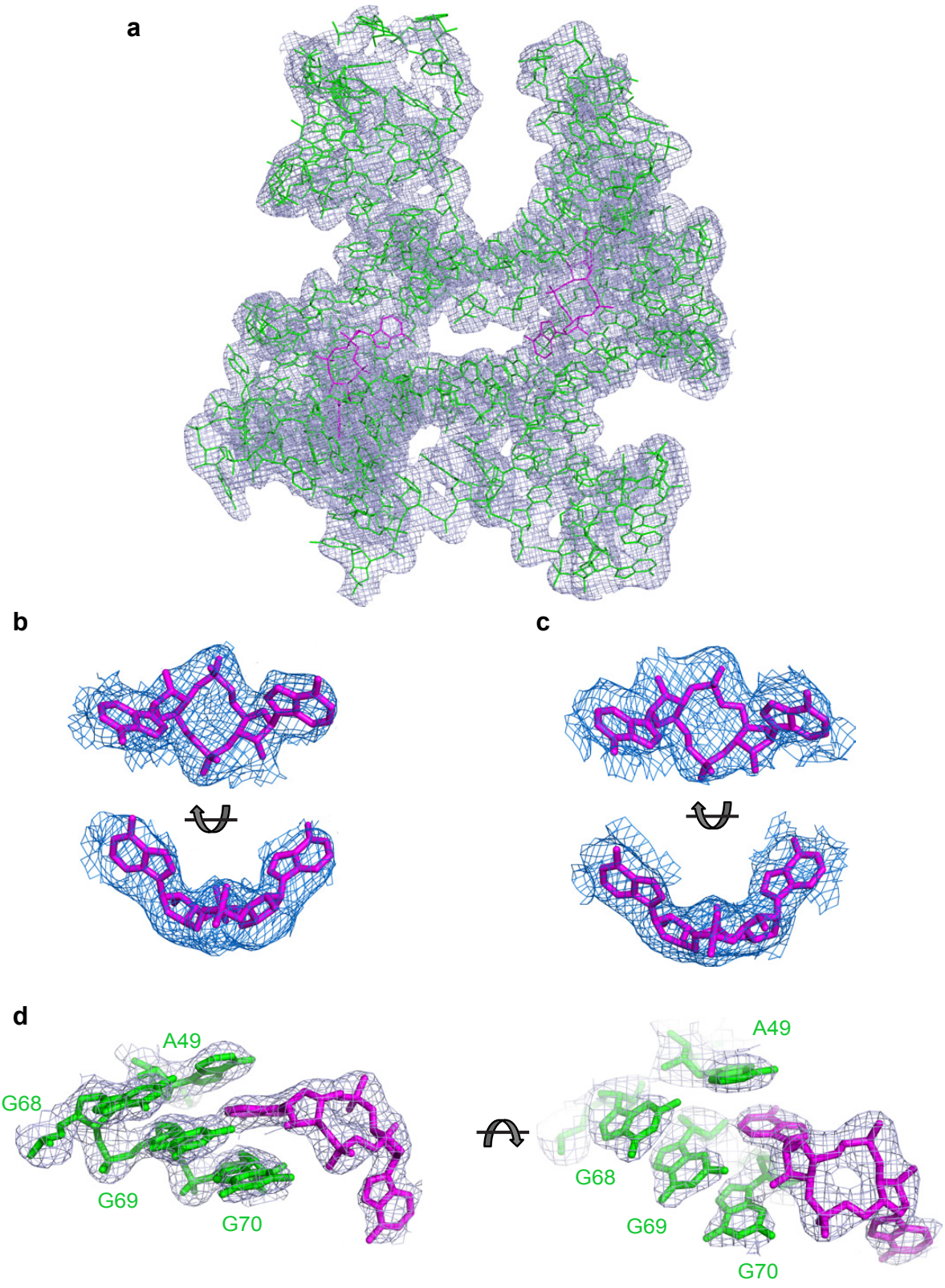
Supplementary Figure 8. Molecular details of c-di-AMP recognition by the c-di-AMP riboswitches. (**a**, **b**), Stabilization of A-minor motif triples in the junctions of the *T. pseudethanolicus* riboswitch through stacking with the adenine bases of c-di-AMP$_1$ (a) and c-di-AMP$_2$ (b). Hydrogen bonds involved in the base triple formation are depicted by dashed lines. (**c**, **d**), Interactions of RNA with the A$_\beta$ (c) and A$_\alpha$ (d) bases of c-di-AMP$_2$ molecules. Hydrogen bond distances between RNA and ligands are shown by dashed lines. (**e**), Comparison of intermolecular interactions between riboswitch modules and c-di-AMP molecules. The schematic is centered on superimposed c-di-AMP molecules. Red and blue squares indicate nucleotides conserved in 97% and 90% of species[12], respectively. Chemical groups are colored according to the structural elements of the riboswitch. (**f**), Zoomed-in views of the J1-8 junctional regions in the *T. pseudethanolicus* (colored) and *T. lienii* (gray) riboswitch structures. Note that a di-nucleotide U114-G115 loop in the *T. lienii* riboswitch replaces a five-nucleotide loop A110-U111-A112-A113-A114 in the *T. pseudethanolicus* riboswitch. In the *T. pseudethanolicus* structure, A113 is positioned in the minor groove of P7 whereas three nucleotides of the loop, U111, A112 and A114, are stacked on A72, which caps the adenine moiety of the ligand. The loop in the *T. lienii* structure does not contain a counterpart of A113 while U114 sits on top of the ligand base thus replacing A72.



Supplementary Figure 9 | Binding of ligands to the c-di-AMP riboswitches measured by ITC. Raw heat changes as a function of time are shown with integrated heats of each injection. Raw heat changes are shown for selected experiments. The data are shown with the concentration of RNA used in the experiment, number of ligand binding sites (N), K_D, and thermodynamic parameters ΔH and ΔS, where applicable. Positive heat changes are the results of ligand dilution and other factors, unidentified at the moment. (**a**), Binding of c-di-AMP to the wild-type *T. lienii* riboswitch at 20 mM Mg^{2+}. (**b**), Binding of c-di-AMP to the wild-type *B. subtilis* riboswitch at 5 mM Mg^{2+}.(**c**), Binding of c-di-AMP to the site 2 mutant of the *B. subtilis* riboswitch at 5 mM Mg^{2+}. (**d**), Binding of c-di-AMP to the site 1 mutant of the *B. subtilis* riboswitch at 5 mM Mg^{2+}.(**e**), Binding of c-di-AMP to the *B. subtilis* riboswitch with mutations in both sites at 5 mM Mg^{2+}. (**f**), Binding of c-di-AMP to the wild-type *B. subtilis* riboswitch in the absense of Mg^{2+}. (**g**), Binding of ATP to the wild-type *B. subtilis* riboswitch at 5 mM Mg^{2+}. The ligand formula is shown on the left. (**h**), Binding of AMP to the wild-type *B. subtilis* riboswitch at 5 mM Mg^{2+}. (**i**), Binding of 3´3´-cGAMP to the wild-type *B. subtilis* riboswitch at 5 mM Mg^{2+}. (**j**), Binding of c-di-GMP to the wild-type riboswitch at 5 mM Mg^{2+}.



Supplementary Figure 10 | Comparison of c-di-AMP binding by dyandenylate cyclase DisA and c-di-AMP riboswitch.
(**a**), Structure of DisA dimer with c-di-AMP (cyan) bound between two protein molecules[3] (PDB ID 3C21). (**b**), Recognition of c-di-AMP by DisA. Amino acids lining the c-di-AMP-binding pocket are shown in stick representation. Putative hydrogen bonds are shown by dashed lines. Note that specific ligand recognition involves only hydrogen bonding of the Watson-Crick edge of c-di-AMP bases with the main chain of Leu94 and the side chain of Thr111. Other amino acids of the protein make hydrogen bonds and electrostatic interactions with the sugar-phosphate backbone of the ligand. (**c**), Recognition of c-di-AMP by the riboswitch shown in the same view as panel (b). Note different patterns of c-di-AMP recognition by the protein and RNA. (**d**), Superposition of c-di-AMP molecules bound to DisA (cyan) and the riboswitch (magenta).



Supplementary Figure 11 | Electron density maps for the c-di-AMP riboswitch structures. (**a**), The refined *T. pseudethanolicus* riboswitch model (green) is shown with the 3.2 Å SAD electron density map (light blue) contoured at the 1.2 σ level. (**b**, **c**), The 3.2 Å SAD map (blue) contoured at the 1.0 σ level shown with the refined models of c-di-AMP$_1$ (b) and c-di-AMP$_2$ (c) from the *T. pseudethanolicus* riboswitch structure. The ligands are in top and side views. (**d**), The refined 3.0 Å $2F_o$-F_c electron density map (light blue) contoured at the 1.0 σ level shown with c-di-AMP$_2$ and 3WJ$_2$ from the *T. lienii* riboswitch structure.