Discovery of the Universal tRNA Binding Mode for the TsaD-like Components of the t6A tRNA Modification Pathway

Boguslaw Stec

Supplementary Materials

Description of the refinement of TsaBDE and Qri7 proteins used in the study

Refinement of the TsaBDE complex from T. maritima

As described in the body of the paper, the models together with the diffraction data associated with them were retrieved from PDB: *T. maritima* Tsa proteins: BDE 2:2:2 complex (6FPE) [2], Qri7 (3WUH) [16], and the Kae1 complex [17] (PDB structures 3ENH, 3EN9, 4WW5, 4WXA) [5,12,13,17]. The models and the data were ported to the CCP4 suit of programs and subjected to refinement. Several cycles of refinement using Refmac 5.8 were needed to obtain desirable quality, and at the end of every macro-cycle, usually consisting of 20 LSQ cycles, the electron density maps were inspected and a model corrected manually.

Careful examination of the individual proteins on the background of retested ED maps showed several areas ready for improvement, as indicated by the bad fit to the electron density and excessive difference densities. Additional motivation for reviewing the original model [2] of the TsaBDE complex from *T. maritima* was provided by departures from the previously obtained crystallographic models of homologous components and direct comparison to the newer model [11]. Originally, the structure of the complex was solved and refined at a 3.1Å resolution [2]. A closer inspection of the model revealed internal asymmetry and other structural problems (some mentioned above) as compared to similar models, particularly [11]. The main inconsistency was a noticeable asymmetry in the otherwise symmetric model in the regions 280-320 of TsaD subunits. The excessive differences in electron densities in several regions were additional indicators of an incomplete model.

The statistics of TsaBDE dimer from *T. maritima* and Qri7 from yeast are presented in table T1. Both models showed reasonable improvements, such as a better fit to the data that improved the R factor, a better fit to the electron density, and improvements in overall stereochemistry, despite modeling more residues in areas of weaker electron density. The progress was also judged by the increasing similarity and convergence of all structurally known homologous proteins. The models for the KEOPS proteins used in this study were only slightly refined to provide a reliable electrostatic representation needed for docking; therefore, they were not submitted to the PDB.

The existing models for all components (TsaB/D/E) were compared to the previously solved and deposited models, such as 3ZEU [26] for the TsaBD complex, 2A6A [35] for the TsaB dimer, or 1HTW for TsaE [36]. The final model contains more than 50 newly refined residues and more than 190 existing residues relocated to new positions (Fig. S1). This accounts for more than 15% of the structure in error. The newly refined elements improved the standard R factor (0.196 from the original 0.223) while leaving the Rfree at almost the same level (0.298 versus 0.294 for the original model). The stereochemistry of the model has been improved in more than 30% of residues, in reference to their original conformers. The corrections improved fit to the electron density and allowed for tentative identification of fully occupied threonylcarabamoyl-AMP (TC-AMP) in both active sites of the TsaD subunits (Fig. 1).

In both TsaB type subunits, the two regions 112-137 and 190-206 were corrected. The first region (112-137) shows a single amino-acid shift as compared with the final model and the previously published structures. The second region (190-206) shows misplaced or undefined residues where the difference maps allowed for more complete modeling (Fig. S2). These errors were clearly identifiable when the initial structure was compared to the published structure of the TsaB dimer (2A6A) [35].

TsaD showed larger errors with a surprising asymmetry between both subunits, particularly in the region 294-327. A comparison of both subunits shows a shift of one residue in the region 297-309. The comparison of both models superposed and covered with the final 2Fo-Fc and difference maps is shown in

supplemental Fig. S2. Additionally, when compared to the final model, both subunits showed a shift of two residues in the 309-327 region, and as a result, the model had two residues from 325-327 missing. Both TsaD subunits had missing residues 31-49 that were identified and refined (Fig. S3). These residues undergo conformational transition upon binding of the TsaE element, as can be inferred by comparing them to homological fragments in other TsaD-type models, like in 3ZEU [26].

The subunit TsaE was refined relatively well, but the quality of the electron density was not as high as for other subunits. However, at closer inspection, the entire C-terminal helix was not properly oriented, and some additional electron density was located near the N-terminus for residues 145-162. The closer analysis of the hydrophobic moments, combined with the comparison to the homologous domains, suggested an alternative position by rotating the helix by a single position. The correct placement of the helix allowed for the avoidance of the clustering of charged residues and the proper placement of hydrophobic residues.

Overall improvements to the model, particularly around the active site, resulted in the appearance of the electron difference densities of a large ATP-like molecule. In the original model [2], only one site was partially modeled with glycerol and a short PEG. The final model was refined with full occupancy of TC-AMP at both active sites of subunit TsaD. After modeling the substrate TC-AMP, some residual electron density suggested the presence of partially occupied metal ions. Zinc ions were refined with 40% occupancy at both active sites. The only fragments at both active sites that were weakly defined by electron density were residues 291-294. These residues show more extended conformation than the same region in homologous proteins that attained alpha helical conformation. It appears this transition accommodates a more twisted/open conformation of the substrate binding domain of the TsaD, as visualized in Fig 2. This domain shows a 17° rotation as compared to the reference structure of the *E. coli* TsaB-TsaD heterodimer (4YDU) [25].

This is the first observation by X-ray crystallography of the TC-AMP, which is unstable in aqueous solutions. The full occupancy of this moiety is quite an important observation, as it has only 3.5 minutes of half-lifetime in solution [6]. The presence of this compound in the crystal structure would suggest that this compound was protected by binding to the protein and survived the entire crystallization and structure determination process. The real burning question is the origin of this compound. The original paper clearly presented the method of obtaining the samples of the proteins. Missoury *et al.* [2] described the process as the production of active DBBD dimers *in vivo* and purifying the complex. This clearly suggests that TC-AMP, synthesized in vivo by native *E. coli* TsaC, could have been captured by a DBBD dimer, survived the purification and crystallization processes, and was detected in the final refined model. Another interesting fact is the correlation between the occupancy of this substrate and the presence of the metal ion. Fractional occupancy of the metal ion suggests that its presence is indispensable for the transfer of the TC moiety to the A37, but not necessary for binding of the TC-AMP molecule. That creates a very interesting situation in which the binding of the metal ion correlates with the presence of ATP, and its hydrolysis by the TsaE element controls the activity of the entire complex.

Refinement and main structural changes in Qri7 dimers obtained in the refinement.

The original model 3WUH represented an accidental dimer not related to the biochemical function. The first operation was to transform it so it was reminiscent of the BD dimer arrangement in bacteria. Subsequently, the missing fragments 70-80, 240-250, and 270-280 were modeled in. Despite their relatively weak electron density, the fragments were placed with sufficient certainty to allow for unambiguous biochemical conclusions to be drawn. This reliable model allowed for the explanation of the significant difference in electron density features at the interfaces of the dimer. These were tentatively identified as either individual nucleotides or pieces of DNA/RNA bound in the cleft flanking the active sites. These are suggestive of the tRNA binding site, which was partially confirmed by the final tRNA binding mode. The loops 70-80 and 270-280 turned out to be particularly important for tRNA binding and are visible flanking the tRNA in Fig 4. The analogous region in the *T. maritima* BDE complex (30-50) was observed to change conformation upon TsaE binding, as described above.

Modeling of the substrates and transition states in the TsaC/C2 and in TsaD.

The corrected and refined model derived from 6FPE by proper tracing and filling in missing pieces resulted in the model 6NAK, which represented a BDE dimer and was used for docking of tRNA. Before docking, the E element was removed from the model, and in one of the D subunits, the region 30-50 was

modeled in a conformation corresponding to the structure of a BD dimer from bacteria (4YDU) in a more extended conformation. Subsequently, the threonyl modification was removed from the tRNA extracted from the 1QF6 model. The pose/decoy 13497 was used as the most representative of the tRNA binding to the TsaD. Upon small manual adjustments, the appropriate model of A37 placed in direct proximity to the TC-AMP was obtained and is presented in Fig. 5 and Fig. S6.

The model of the TsaC2 with all three substrates at the active site was obtained in the following manner. The model was generated starting from the model of TsaC2 with ATP and Thr (3AJE). This model was compared/superposed with the model of the TsaC2 with the product of the reaction TC-AMP (4E1B). This comparison revealed the conformational/rotational change/movement of the alpha phosphate of the ATP. Modeling this motion created room for the third substrate, CO_3, sandwiched between the Thr and the alpha phosphate of the ATP. In placing the CO_3, we used guidance from the model (6Z81). The final model was created by manually adjusting the initial locations of CO_3 and ATP for optimal contacts and a possible change in the location of the metal ion driving the reaction. This metal ion migration is suggested by the conformational differences in the location of the Oγ group of the Ser144 in the 4E1B structure. The resulting model is shown in Fig. 6B and Fig. S7.

The epilogue

In this report, we discussed the consequences of structural improvements made to several crystal structures deposited in the PDB for understanding tRNA binding and functioning along the entire pathway of the t^6A_{37} tRNA modification. Particularly, we focused our attention on investigating models of relevant proteins from the *T. maritima* TsaBDE complex described by the French group [2]. In preparation for the submission of our findings to the press [30], we shared our refined model (6NAK) with the French group. Soon after we provided our model, the French group replaced its original model, 6FPE, with a new one, 6S84, that accepted most of our postulated changes [23]. In doing so, they never acknowledged our contribution.

To clarify the issues raised by them, we subjected the new model 6S84 to the same scrutiny as the previous model 6FPE. The results clearly showed that our interpretation of the TC-AMP at the active site is superior to the presence of AMPCPP as judged by three criteria: better quality of electron density, better fit of the temperature factors of the ligand with the environment, and a better outcome of numerical descriptors like the R-factor. A similar conclusion can be extended to the conformation and location of the last helix in subunit TsaE. Therefore, our original corrections were confirmed and retained for modeling, allowing us to draw our conclusions about the mode of tRNA binding. For more details, please refer to Figs. S1–S5, their figure captures, and [30]. Additionally, in a follow-up paper, they reported the structure of the TC-AMP stable analog [24] bound to the TsaBD heterodimer from *E. coli*. This work is clearly inspired by our conclusions about the presence of TC-AMP at the active site of the TsaBDE complex described in this report.

Supplementary Table S1. Refinement statistics of the original and refined model of *T. maritima* TsaB-TsaD-TsaE complex in B2D2E2 composition and Qri7 from yeast.

Protein	***T. maritima* $B_2D_2E_2$ complex**		**Qri7**	
PDB code	**Original model 6FPE**	**Re-refined model 6NAK**	**Original model 3WUH**	**Re-refined model 6NBJ**
Space group	P212121	P212121	P4332	P4332
Unit cell dimensions				
a (Å)	84.31	84.31	180.3	180.3
b (Å)	113.99	113.99	180.3	180.3
c (Å)	177.57	177.57	180.3	180.3
Resolution (Å)	3.14	3.14	2.94	2.94
Average redundancy	4.56	4.56	3.8	3.8
Average I/r (I)	7.5	7.5	8.0	8.0
Completeness (%)	99.1	99.1	99.8	99.8
Total R-merge (%)	16.8	16.8	8.9	8.9
No Refined Reflections	30281	28747	21792	20688
R-value (%)	22.9	19.6	22.6	20.0
Rfree-value (%)	29.4	29.8	27.3	26.9
No of residues	1313	1470	676	752
No of atoms	10496	10993	5279	6084
Rmsd, bond lengths (Å)	0.10	0.09	0.09	0.11
Rmsd, bond angles (deg.)	1.25	1.35	1.2	1.6
Average B for subunitsB	77, 68	87, 77	NA	NA
Average B for subunitsD	89, 76	101, 89	97, 99	103, 103
Average B for subunitsE	99, 73	111, 82	NA	NA
Average B ligands at sitesD	62	102, 93	113, 118	103, 108
Average B ligands at sitesE	96, 72	88, 75	NA	NA
Ligands(D, E)	GOL, APC	TC-AMP, ATP	AMP	AMP, 4-C
Ramachandran (within)	91.4	87.9	93.1	88.6
Ramachandran (cond)	7.5	8.8	5.5	8.2
Ramachandran (outside)	1.1	3.3	1.5	3.2

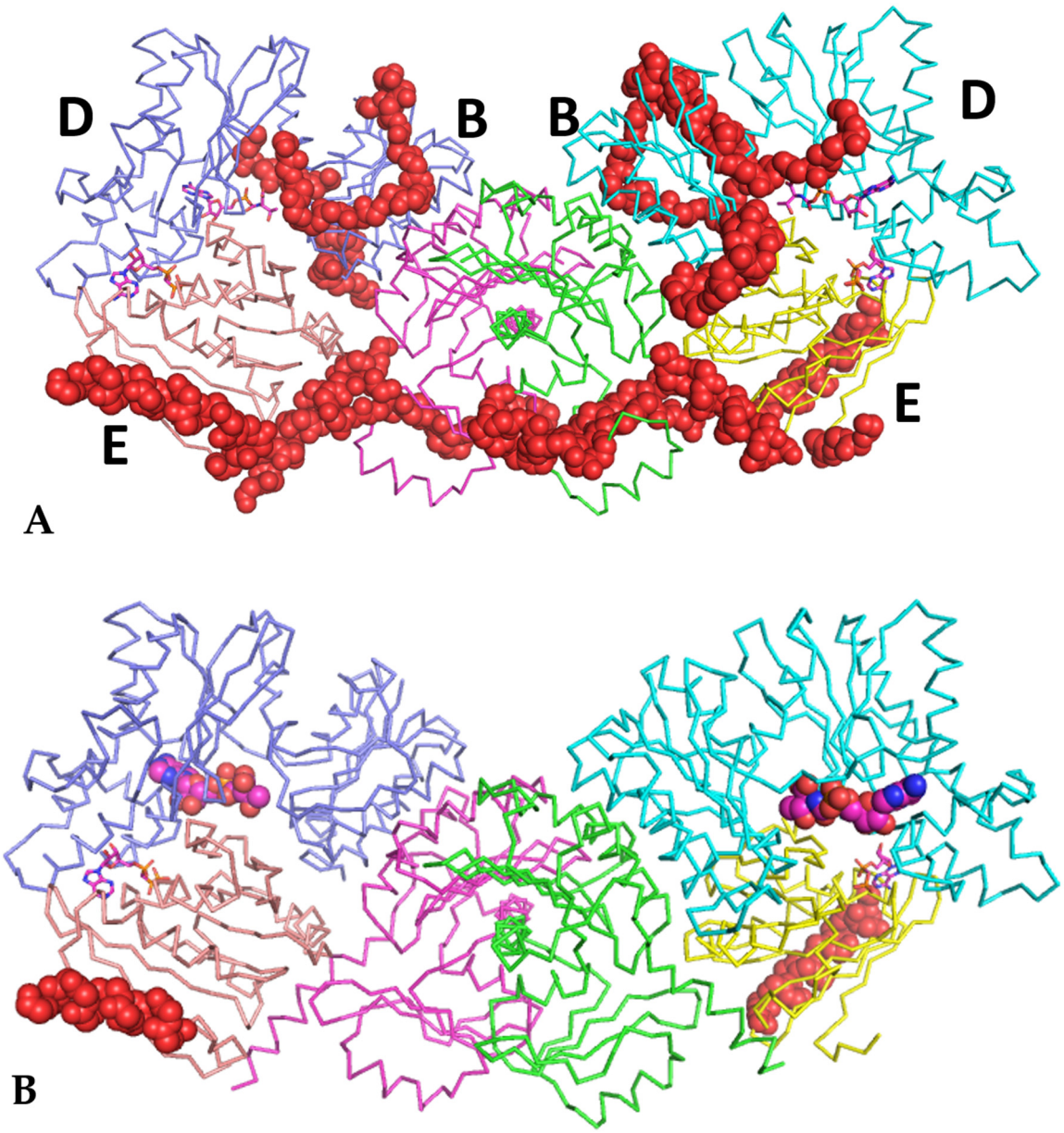


Figure S1. The general view of the final refined $B_2D_2E_2$ complex (6NAK) in C-alpha representation. Different subunits are represented in different colors. The substrates are represented by purple stick models. (A) The fragments of the model that were incorrectly placed in the original model (6FPE) and that constitute one- or two-residue shifts or misplaced or missing residues are represented in red spheres. These encompass approximately 15% of the structure and closely surround the active sites of TsaD. (B) The fragments that remain incorrectly refined in the newest model (6S84) are represented by the space-filling spheres. These are the terminal helix of subunit E (residues 148-163) and the identity of the substrate bound at subunit D.

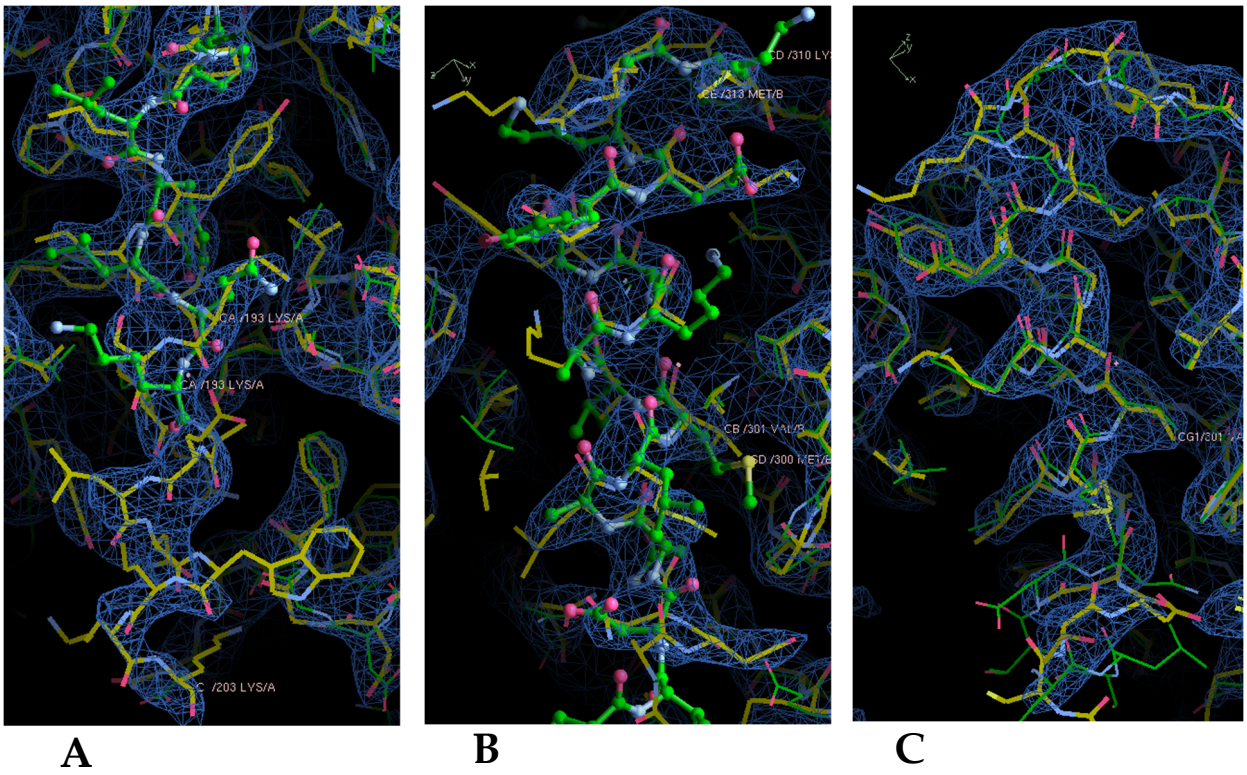


Figure S2. The electron density and model improvements during the refinement. The original model is represented in green, while the final refined model is in yellow. The 2Fo-Fc electron density phased with the final model (in blue) is contoured at 1 σ level. (A) The fragment of the model around residues 190-203 in TsaB, which clearly shows that the original model was misplaced for residues 190-193 and not refined (untraced for residues 193-203). (B) An improperly refined fragment of 295-303 in TsaD with one residue shift is followed by a mistraced fragment that is not shown. (C) The homologous fragment in the symmetric TsaD subunit that was properly traced in 290-302 region but similarly misplaced in 303-323 region.

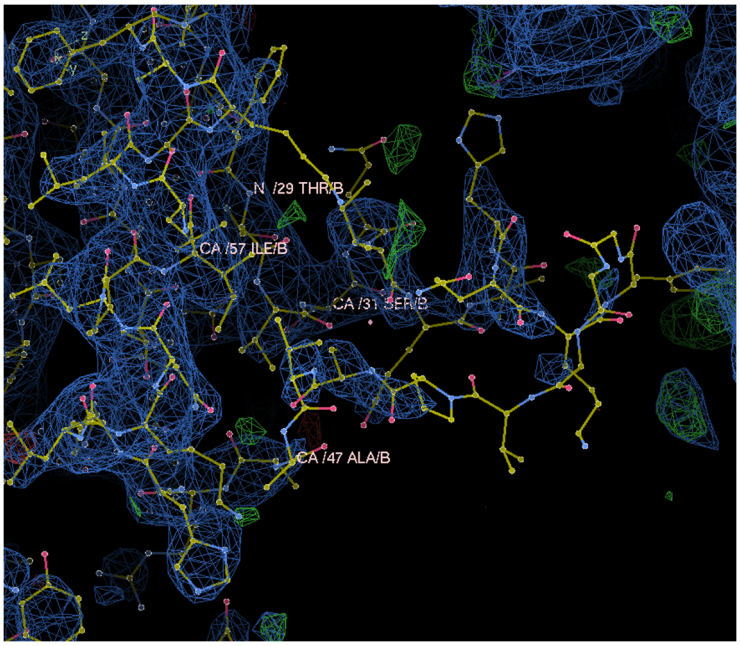
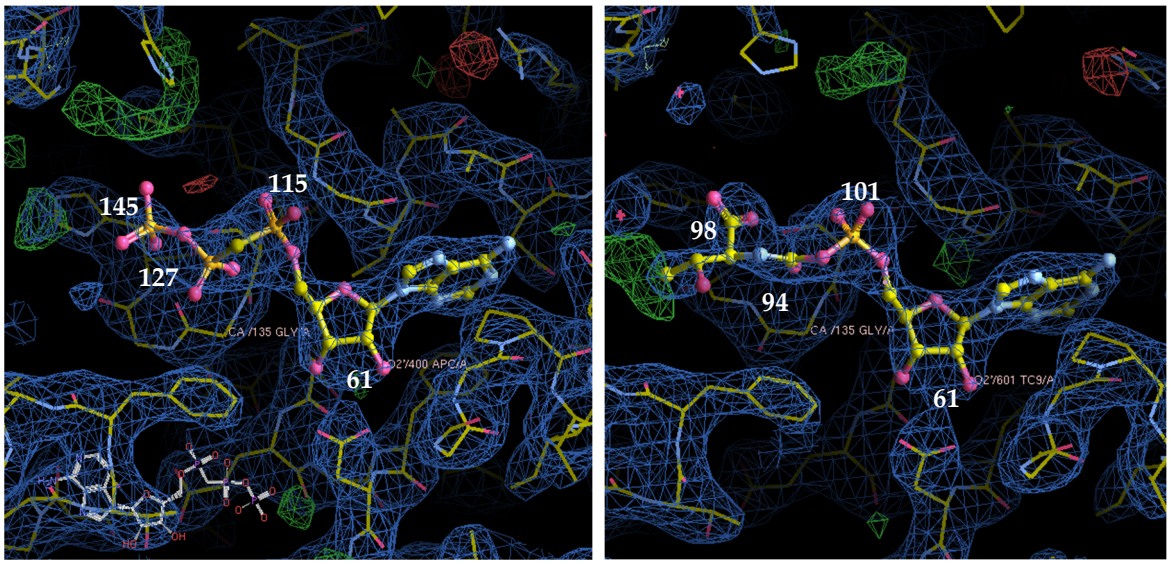
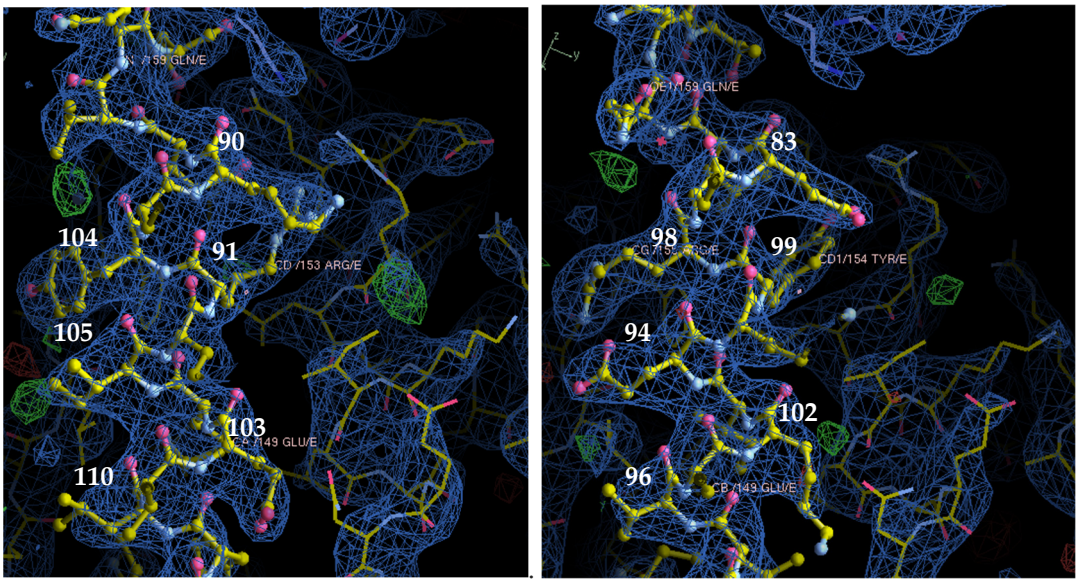


Figure S3. The electron density at 0.9 sigma level and model improvements in the region of 30-50 of the TsaD subunit. The original 6FPE model did not contain the regions 31-47.



A B

Figure S4. Comparison of the refined models of the ligand found at the active site of subunit TsaD. (A) The electron density contoured at the 1.3 σ level phased with model 6S84 that includes the bound AMCPP (diphosphomethylphosphonic acid adenosyl ester) in ball and stick representation. Temperature factors B are marked next to the phosphates. The original temperature factors included in 6S84 for AMPCPP are much higher than those exhibited in our refinement of this model cited above. (B) The electron density contoured at the 1.3 σ level phased with model 6NAK that includes the bound TC-AMP in ball and stick representation. Temperature factors B are marked next to key atoms corresponding to the phosphate groups in AMPCPP. Please note the carbonyl flip in Gly135, which is clearly incorrect in 6S84.



A B

Figure S5. Comparison of the refined models of the C-terminal helix of subunit TsaE. (A) The electron density contoured at the 1 σ level phased with model 6S84. Temperature factors B of selected residues are marked next to the sites. (B) The electron density contoured at the 1 σ level phased with model 6NAK. While both models are very similar to the untrained eye, they have significant differences. The general fit of the model to the density, reflected in the elimination of spurious difference electron densities as well as in the resulting temperature factors, local backbone stereochemistry as well as stereochemistry of the sidechains, and what is even more important, the logic of protein folding principles (hydrophobic directed to the interior and hydrophobic residues to the solvent, not accumulating charged residues together with pairing of ion salts), clearly indicate the superiority of the 6NAK model.

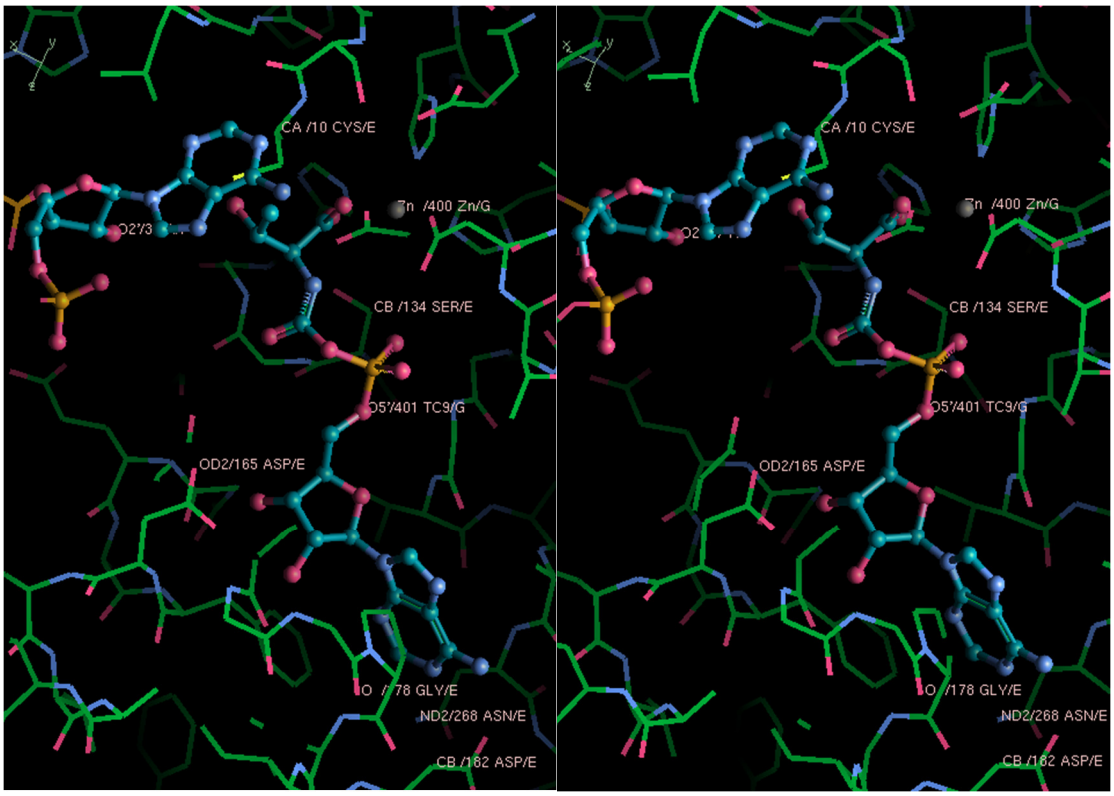


Figure S6. The stereo representation of the active site of TsaD containing refined TC-AMP with docked tRNA. The pose/decoy 13497 with A37 located in the vicinity of TC-AMP is shown without repositioning. Fig. 5A presents a manually adjusted model from this pose to create a plausible model of the transition state.

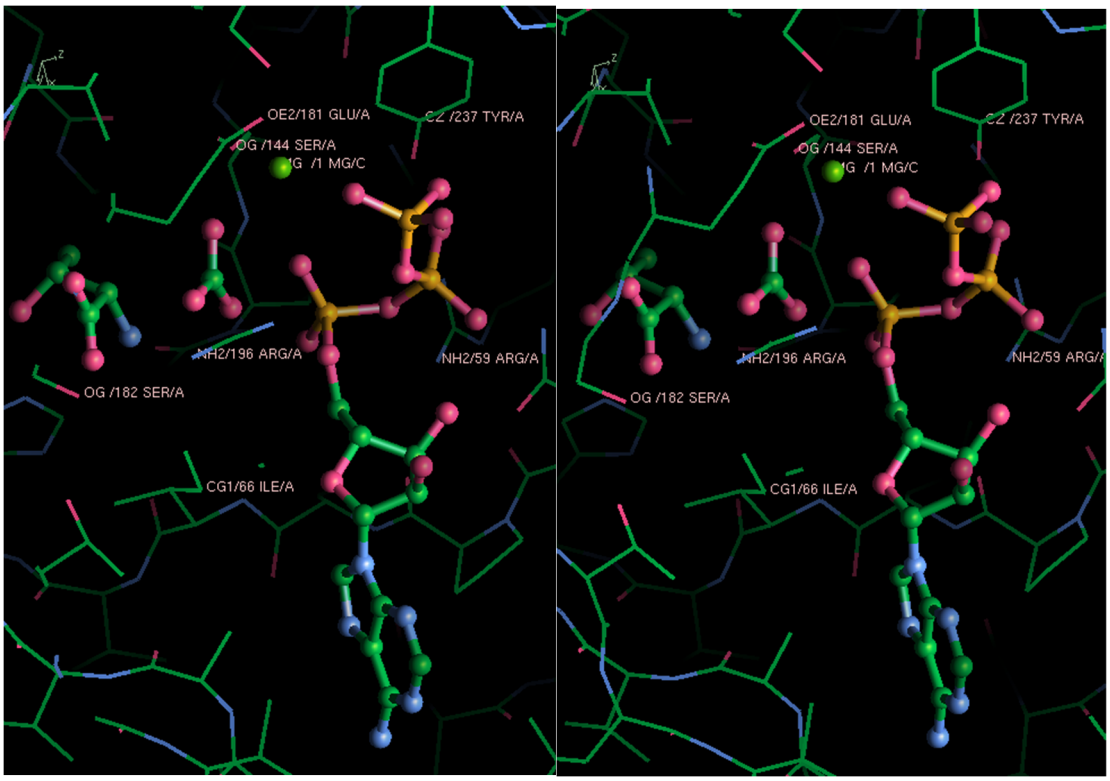


Figure S7. The stereo representation of the active site of TsaC2, containing refined Thr (3AJE), conformationally adjusted ATP, and manually modeled CO_3 and $Mg2^+$. Fig. 6B in the manuscript presents the same model.