Portland State University
PDXScholar

Chemistry Faculty Publications and Presentations | Chemistry

8-2015

NMR-based Structural Analysis of Threonylcarbamoyl-AMP Synthase and Its Substrate Interactions

Kimberly A. Harris
North Carolina State University

Benjamin G. Bobay
North Carolina State University

Kathryn L. Sarachan
University at Albany, State University of New York

Alexis F. Sims
University at Albany, State University of New York

Yann Bilbille
North Carolina State University

See next page for additional authors

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/chem_fac


Part of the Biochemistry Commons, and the Chemistry Commons

Citation Details

Harris, K. A., Bobay, B. G., Sarachan, K. L., Sims, A. F., Bilbille, Y., Deutsch, C., … Agris, P. F. (2015). NMR-based Structural Analysis of Threonylcarbamoyl-AMP Synthase and Its Substrate Interactions. The Journal of Biological Chemistry, 290(33), 20032–43. http://doi.org/10.1074/jbc.M114.631242

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Chemistry Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Authors
Kimberly A. Harris, Benjamin G. Bobay, Kathryn L. Sarachan, Alexis F. Sims, Yann Bilbille, Christopher Deutsch, Dirk Iwata-Reuyl, and Paul F. Agris

This post-print is available at PDXScholar: http://pdxscholar.library.pdx.edu/chem_fac/120

NMR-based structural analysis of threonylcarbamoyl-AMP synthase and its substrate interactions*

Kimberly A. Harris[‡,§,1], Benjamin G. Bobay[‡,1], Kathryn L. Sarachan[§1], Alexis F. Sims[§], Yann Bilbille[‡], Christopher Deutsch[¶], Dirk Iwata-Reuyl[¶] and Paul F. Agris[§,2]

[‡] Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA;
[§] The RNA Institute, Department of Biological Sciences and Department of Chemistry, University at Albany, Albany, NY 12222, USA;
[¶] Department of Chemistry, Portland State University, Portland, Oregon 97207, USA

*Running title: *Threonylcarbamoyl-AMP synthase substrate interactions*

[2]To whom correspondence may be addressed: Paul F. Agris, The RNA Institute, University at Albany 1400 Washington Ave., Albany, NY 12222, USA, Tel.: (518) 437-4448; FAX: (518) 437-4456; E-mail: pagris@albany.edu.
Keywords: RNA modification; Transfer RNA (tRNA); *Escherichia coli*; nuclear magnetic resonance; ligand binding protein; protein structure

---

Background: Threonylcarbamoyl-AMP synthase catalyzes formation of the biosynthetic intermediate of a critical tRNA modification, t^6A_{37}.
Results: Structural analyses provide insight into the interaction between the substrates ATP and L-threonine and the *E. coli* enzyme.
Conclusion: The threonylcarbamoyl-AMP synthase binds L-threonine and ATP cooperatively; L-threonine is required for positioning of ATP.
Significance: Mechanistic insights into t^6A_{37} biosynthesis provide an understanding of this complex enzymatic pathway.

ABSTRACT

The hypermodified nucleoside *N*^{6}-threonylcarbamoyladenosine (t^6A_{37}) is present in many distinct tRNA species and has been found in organisms in all domains of life. This post-transcriptional modification enhances translation fidelity by stabilizing the anticodon-codon interaction in the ribosomal decoding site. The biosynthetic pathway of t^6A_{37} is complex and not well understood. In bacteria, four proteins have been discovered to be both required and sufficient for t^6A_{37} modification: TsaC, TsaD, TsaB and TsaE. Of these, TsaC and TsaD are members of universally conserved protein families. Although TsaC has been shown to catalyze the formation of L-threonylcarbamoyl-AMP (TC-AMP), a key intermediate in the biosynthesis of t^6A_{37}, the details of the enzymatic mechanism remain unsolved. Therefore, the solution structure of *Escherichia coli* TsaC was characterized by NMR in order to further study the interactions with ATP and L-threonine, both substrates of TsaC in the biosynthesis of TC-AMP. Several conserved amino acids were identified that create a hydrophobic binding pocket for the adenine of ATP. Additionally, two residues were found to interact with L-threonine. Both binding sites are located in a deep cavity at the center of the protein. Models derived from the NMR data and molecular modeling reveal several sites with considerable conformational flexibility in TsaC that may be important for L-threonine recognition, ATP activation and/or protein-protein interactions. These observations further the understanding of the enzymatic reaction catalyzed by TsaC, a threonylcarbamoyl-AMP synthase, and provide structure-based insight into the mechanism of t^6A_{37} biosynthesis.

An essential aspect of RNA maturation is the post-transcriptional modification of nucleosides, which adds chemical complexity to the four major nucleosides and permits a greater range of functionality. Modified nucleosides are particularly abundant in transfer RNA (tRNA), where over 90 distinct modifications are found across all phylogenetic domains (1). One of the sites with the

highest frequency of modification in tRNA and some of the most chemically complex modifications found in any RNA is that of the conserved purine 3′-adjacent to the anticodon at position 37 (2). More than 70% of tRNA species are modified at this site (3). Of these, the universally conserved N^6-threonylcarbamoyl-adenosine (t^6A_{37}) modification and its derivatives, 2-methylthio-N^6-threonylcarbamoyladenosine ($ms^2t^6A_{37}$) and N^6-methyl-N^6-threonylcarbamoyl-adenosine ($m^6t^6A_{37}$), are found at the 37 position in nearly all tRNAs that read ANN codons (3,4) (Fig. 1A-B). Each of these tRNAs contains a strictly conserved sequence of $U_{36}A_{37}A_{38}$ in the anticodon stem-loop (3). Both U_{36} and A_{37} are essential for modification and A_{38} enhances the rate of the modification reaction (5). The t^6A_{37} modification has been shown to have an important role in ribosome-mediated codon binding for several tRNA species (6-8), mainly due to its ability to enhance the stability of the anticodon-codon base-pairing by creating cross-strand base stacking interactions with the first position of the codon, as depicted in structural analyses of $tRNA^{Lys}_{UUU}$ decoding at the ribosome A-site (9,10). This stabilization is necessary to overcome the low enthalpy of binding for U-A base-pairs, and $tRNA^{Lys}_{UUU}$ in particular with three U-A base-pairs (11). The size and location of t^6A_{37} on the Watson-Crick face of the nucleoside negates intra-loop base-pairing with the invariant U_{33}, which is critical for the U-turn backbone structure of tRNA. This creates an open, structured loop that is more favorable for entry into the ribosome and the binding of the mRNA codon (9,12). By strengthening the codon-anticodon interaction and ribosome entry, t^6A_{37} is believed to participate in maintenance of the translational reading frame. This is supported by the increase in translational frameshifts in cells lacking the ability to form t^6A_{37} (13-15).

Recent studies of the t^6A_{37} biosynthesis pathway have identified four proteins in *Escherichia coli* that are necessary and sufficient for *in vitro* formation of the modification: TsaC, TsaD, TsaE and TsaB, also known as YrdC, YgjD, YjeE and YeaZ, respectively (16). Two of the four proteins essential to the enzyme complex are unique to bacteria, TsaB and TsaE. The other two, TsaC and TsaD, are members of universal families and are associated with the t^6A pathway in several organisms. The biosynthesis of t^6A_{37} has been reconstituted *in vitro* in the bacterial species *E. coli* and *Bacillus subtilis* (17), and the eukaryotic and archaeal species, *Saccharomyces cerevisiae* and *Pyroccocus abyssi*, respectively (18,19). The biosynthesis in *S. cerevisiae* and *P. abyssi* requires L-threonine, ATP and CO_2/HCO_3^-, analogous to bacteria, but five proteins are necessary: Sua5 (the TsaC homolog), Kae1 (the TsaD homolog), Bud32, Pcc1 and Cgi121 (18). The latter four proteins comprise the KEOPS complex, a protein complex associated with a variety of physiological phenomena (15,20-23). Though both divalent and monovalent cations are suspected to be essential for this reaction, no systematic analysis has determined their requirement.

While the proteins involved in t^6A_{37} biosynthesis have been identified and experimental evidence suggests that they function together in a hetero-multimeric complex, little is known about the structure of the complexes or the mechanistic contributions of each protein. TsaC, which is central to the biosynthesis of t^6A_{37}, is a member of the TsaC/Sua5 protein family. This family has been found in all sequenced genomes to date. However, the essentiality varies between organisms. TsaC is essential in *E. coli*, but Sua5 is not essential in *S. cerevisiae* (24), even though both appear to have similar roles in t^6A_{37} biosynthesis. *E. coli* TsaC selectively binds both ATP and L-threonine (24,25) and is the sole protein of the t^6A-synthase complex capable of L-threonine-dependent conversion of ATP to AMP (16). More recently, both TsaC and its *B. subtilis* homolog, YwlC, have been shown to catalyze the formation of the activated precursor L-threonylcarbamoyl-AMP (TC-AMP) from L-threonine, ATP and CO_2/HCO_3^- (17) (Fig. 1C). This strongly suggests that TsaC is the enzyme responsible for the first step in the mechanism. Subsequently, TC-AMP is transferred to the A_{37} of the tRNA substrate by TsaB, TsaD and TsaE (YeaZ, YgjD and YjeE). Consistent with this model is the observation that the *B. subtilis* homologs of TsaBDE are capable of t^6A_{37} biosynthesis in the presence of TC-AMP and tRNA and the absence of YwlC, the TsaC homolog (17). The argument is strengthened by the ability of *P. abyssi* Kae1, the TsaD homolog, to bind TC-AMP and catalyze the transfer to tRNA (26). However, several studies have shown that *E. coli* TsaC selectively binds hypomodified tRNA (24,25,27), indicating TsaC may have a more substantial, diverse role in

bacteria than in eukaryotes since *S. cerevisiae* Sua5 does not exhibit RNA-binding activity (18).

Within the TsaC/Sua5 family four structures have been solved by X-ray crystallography, including *E. coli* YciO (28), *Sulfolobus tokodaii* Sua5 (29), *E. coli* HypF (30) and *E. coli* TsaC (27). Each of these structures contains a TsaC domain with a unique folding pattern and the ability to catalyze similar chemistries. However, neither HypF nor YciO are associated with the t^6A_{37} pathway. Unfortunately, the TsaC crystal structure lacks several of the C-terminal residues (presumably due to low electron density). In addition, the crystal-packing induced dimerization appears to alter the conformation of functionally relevant regions in the binding of L-threonine, ATP, and protein-protein interactions (27). Here, we sought to characterize the dynamics and structural interactions of TsaC with its substrates in order to provide insight into the mechanism of action of TsaC. Therefore, structural characterization through solution nuclear magnetic resonance (NMR) was employed to observe full-length TsaC under the native reaction conditions. As such, we report the high-resolution structure of the monomer and interaction studies with L-threonine and ATP in solution using NMR. We were able to identify several key residues in TsaC that interact with L-threonine and ATP using chemical shift perturbation analyses. These analyses were used to guide molecular docking to provide the first models of interactions between *E. coli* TsaC and ATP and L-threonine. These studies provide insight into how TsaC catalyzes the formation of TC-AMP.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification – Protein samples were expressed and purified as previously described (16). For isotope-labeled TsaC, M9 minimal media was supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose (ISOTEC) (31). The TsaC construct lacks an N-terminal Met1 due to a Factor Xa cleavage site used to remove the His_6-tag. All TsaC samples were prepared in a buffer of 90:10% $H_2O/^2H_2O$, 20 mM potassium phosphate, pH 7.0, 100 mM NaCl unless otherwise noted.

Nuclear Magnetic Resonance – All NMR experiments were performed at 298 K on either a Varian Inova 600 MHz or a Bruker Avance II 700 MHz spectrometer, both equipped with cryoprobes. Backbone chemical shifts were assigned in a sequential manner from the following experiments: 2D ^1H-^{15}N-HSQC, HNCA, HN(CO)CA, HN(CO)CACB, HNCACB, HNCO, and HN(CA)CO. TALOS was used to determine coupling constants for assigning backbone ψ and φ angles (32). Side chain proton and carbon chemical shifts were assigned using the following experiments: CC(CO)NH (ncyc 2 and 3), H(CCO)NH (ncyc 3 and 4), and a ^{15}N-TOCSY-HSQC with a 50 ms mixing time. NOEs were obtained from 3D ^{15}N-NOESY-HSQC experiments with 120 and 150 ms mixing times and 3D ^{13}C-NOESY-HSQC experiments with 100 and 120 ms mixing times. For residual dipolar couplings, ^{15}N-labeled TsaC was aligned in bacteriophage Pf1 (ASLA Biotech) to a final concentration of 12 mg/ml. $^1D_{HN}$ couplings were obtained from a two-dimensional ^1H-^{15}N HSQC- IPAP (in-phase/antiphase) experiment (33). Data were processed using NMRPipe (34) and analyzed using SPARKY (35) and NMRViewJ (36).

Protein backbone dynamics were determined using steady-state heteronuclear ^1H-^{15}N NOE experiments as previously described (37) at 298 K. Values were measured from the peak heights in 2D ^1H-^{15}N-HSQC spectra, normalized for background noise and graphed by I_{sat}/I_{unsat}.

Structure Calculation – CYANA 2.1 was first used for automatic NOE assignment, and one hundred structures were calculated with the standard simulated annealing protocol, using NOE and hydrogen bonding distance restraints and ψ and φ dihedral angle restraints from TALOS+ predictions (38). The final round of one hundred structure calculations was performed in XPLOR-NIH using the same NOE, hydrogen bonding and dihedral restraints together with residual dipolar coupling restraints (39, 40). The ensemble of twenty lowest energy structures was assessed using the Protein Structure Validation Software (PSVS) Suite (41). Molecules were visualized and aligned with PyMol (42).

Substrate titrations observed by NMR – The binding of L-threonine and ATP by TsaC were observed at 298 K on a Bruker Avance III 500 MHz spectrometer equipped with an ultra-sensitive triple resonance cryoprobe capable of applying pulsed

field gradients along the z-axis. The TsaC protein sample was concentrated to 100 μM in a buffer of 90:10% $H_2O/^2H_2O$, 50 mM potassium phosphate, pH 7.5, 150 mM KCl. ATP (Sigma-Aldrich) was titrated into samples that provided the following molar ratios of ligand to protein: 0:1, 0.5:1, 1:1, 2:1, 4:1 and 8:1. L-threonine (Sigma-Aldrich) was added to the protein in ligand to protein ratios of 0:1, 1:1, 2:1 and 4:1. A 2D ^1H-^{15}N-HSQC experiment was collected at each titration point. Data were processed using NMRPipe (34) and analyzed using SPARKY (35). The total shift change, $\Delta\delta_{N-H}$, for each peak was calculated by $\Delta\delta_{N-H}=\sqrt{(\Delta\delta_{HN})^2+0.2(\Delta\delta_N)^2}$ where $\Delta\delta_{HN}$ and $\Delta\delta_N$ are the chemical shift differences for ^1H and ^{15}N, respectively. NMR spectra of the protein in the absence of any ligands were the same though collected under buffer conditions for structure determination (90:10% $H_2O/^2H_2O$, 20 mM potassium phosphate, pH 7.0, 100 mM NaCl) that were slightly different in pH and ionic strength from conditions for the protein's titration with ATP and L-threonine followed by NMR and ITC (90:10% $H_2O/^2H_2O$, 50 mM potassium phosphate, pH 7.5, 150 mM KCl).

Isothermal titration calorimetry – The interaction of TsaC with ATP was characterized by ITC experiments conducted at 4 °C (MicroCal VP-ITC). TsaC was prepared at 100 μM in 50 mM potassium phosphate, pH 7.5, 150 mM KCl. ATP at a concentration of 1 mM in the same buffer was titrated in 10 μL injection volumes into the experimental cell containing TsaC. Titration curves were baseline-subtracted and analyzed by non-linear least-squares fitting using the MicroCal Origin 5.0 software.

HADDOCK Docking Procedure – Default HADDOCK (43) parameters were used throughout all docking procedures. Active and passive residues (Table 3) with solvent accessibility >50% calculated by NACCESS (44) were assigned from TsaC NMR titration studies. Passive residues were defined as all other residues with 50% solvent accessibility. The adenosine moiety of the ATP molecule was defined as the active portion, while the phosphate closest to the adenosine moiety was defined as passive. The remaining phosphate moieties were not actively involved in the interaction between ATP and TsaC. One thousand structures were generated per iteration, and the 200 lowest energy structures were water refined. Each docking attempt was performed ten times and the solution with the lowest HADDOCK score was retained. The root-mean-square deviation (RMSD) values of the complexes were calculated using the McLachlan algorithm (45) as implemented in ProFit (46). A cluster analysis was performed on the final docking solutions using a minimum cluster size of four. The cut-off for clustering was manually determined for each docking run. The RMSD matrix was calculated over the backbone atoms of the interface residues.

RESULTS

TsaC has been shown to interact with three small substrates, L-threonine (25), ATP (24), CO_2/HCO_3^- (17), as well as hypomodified tRNA (24,25,27), TsaD and TsaB (16). The synthesis of TC-AMP, the central role of TsaC in t^6A_{37} biosynthesis, requires a number of intermolecular interactions and enzymatic activity. These interactions have yet to be fully characterized structurally and mechanistically, and consequently, the precise details of this enzymatic biosynthesis are unknown. In order to provide some understanding of the mechanism of TC-AMP synthase, we characterized the structure and dynamics of the functional TsaC monomer and its substrate interactions. Based on this structural information, molecular docking was performed between TsaC and ATP and L-threonine.

Structure of E. coli TsaC

The monomeric structure of *E. coli* TsaC in solution was verified by NMR because there was the possibility of dimerization as had been observed in the crystal structure, but contrary to monomers observed in solution by dynamic light scattering (27). The application of a pulsed field gradient method for analysis of diffusion (47) confirmed that the protein was a monomer under NMR conditions (data not shown). The ^{15}N-^{13}C-TsaC displayed excellent spectral dispersion of resonances for a 20.6 kDa protein in a 2D ^1H-^{15}N- transverse relaxation optimized spectroscopy (TROSY) spectrum (Fig. 2). Therefore, conventional heteronuclear multidimensional NMR methods were utilized to assign ^1H, ^{15}N and ^{13}C backbone and side chain resonances of isotope-labeled *E. coli*

TsaC. Most amide peaks were present in the ^{1}H-^{15}N-HSQC and were assigned with the exception of Asn2, Asn3, Val118 and Ser139. In total, the molecule was 87% assigned with 98% assignment of the backbone, including HN, N, Cα and C', and 79% assignment of the side chains. Using these assignments, sequential connectivities and distances between intramolecular protons were determined for construction of a high-resolution solution structure of TsaC.

The twenty lowest energy, water-refined structures were determined using a total of 1869 distance restraints with 656 sequential, 444 medium range, and 379 long range restraints. The structures contained zero NOE and hydrogen bond violations (Table 1; Fig. 3). The TsaC structures have an average backbone RMSD of 0.79 Å and an average heavy atom RMSD of 1.35 Å for residues comprising regions of secondary structure. The Ramachandran analysis indicates the reported structure of TsaC is of good quality; of all the residues in the ensemble of twenty structures, 95.3% are in the allowed Ramanchandran space or better. The residues that are in the generously allowed (4.1%) or the disallowed (0.6%) regions reside either in loop areas of the protein or regions of no RDC restraints due to spectral overlap (Table 1).

The *E. coli* TsaC structure has an α/β twisted open-sheet structure with parallel and antiparallel adjacent β-strands, consistent with other members of the family (Fig. 3A). The structure comprises seven α-helices and seven β-strands connected by thirteen loop regions. The β-strands are aligned in the center of the protein and helically twist 180° from β1 to β7. The C-terminus exhibits considerable conformational flexibility, suggesting an enzymatically-relevant role. At the center of the protein there is a deep cavity of hydrophobic character lined with positive surface potential, providing potential binding surfaces (27) (Fig 3B-C). Comparison of the crystal structure (27) and the NMR solution structure reveals the areas of greatest difference based on Cα RMSD (Fig. 3D). There is considerable divergence between the two structures in the malformed dimerization region from the crystal structure (dotted box), suggesting that crystal packing greatly affected the structure in this region, which is important for substrate- and protein-protein interactions.

The L-threonine binding site

E. coli TsaC and *S. tokodaii* Sua5 have been shown to exhibit specificity for L-threonine over other amino acids (25,50), and *E. coli* TsaC has been shown to catalyze TC-AMP formation (17). Therefore, we sought to characterize the TsaC:L-threonine binding interface to provide insight into structural aspects of the enzymatic mechanism. Using the fully-assigned ^{1}H-^{15}N-HSQC spectrum (Fig. 2), amide chemical shift changes within the protein were monitored by ^{1}H-^{15}N-HSQC experiments with protein:ligand ratios of 1:0, 1:1, 1:2, and 1:4. We observed two distinct and significant changes in the spectra for Thr27 and Ser176. There were no pronounced up- or downfield movements in any resonance. However, the amide peaks for Thr27 and Ser176 broaden to the point of disappearance (Fig. 4A-B). This is likely due to the line broadening caused by intermediate exchange between the free and bound forms and is indicative of L-threonine interaction at these sites. Thr27 is located in the loop between β1 and β2 strands, and Ser176 is within the loop connecting α7 and β7. Both are situated in close proximity to each other within the putative ligand-binding active site and are likely to coordinate L-threonine during TC-AMP formation.

Data from titration of the protein with L-threonine were incorporated into molecular modeling studies for the interaction between L-threonine and TsaC. HADDOCK (High Ambiguity Driven Docking) (43) was used to develop a structural model of the binding of L-threonine from the NMR structure of TsaC, and active and passive residues with solvent accessibility >50% were assigned from TsaC NMR titration studies. (Fig. 4C-D; Table 3). A plot of the E_{inter}, the sum of restraint, van der Waals, and electrostatic energy terms, as a function of backbone RMSD from the lowest energy model reveals that the models converge to a Cα (protein) and Cα (L-threonine) RMSD of 0.5 ± 0.7 Å at the defined protein-L-threonine interface with an average buried surface area of 215 ± 79Å^2. Three clusters of structures with low RMSD and energy were obtained for all calculated models based on a minimum cluster size of four models and a Cα RMSD of 7.5 Å. The lowest energy structure displays direct interactions between TsaC Ser176 and Thr27 and L-threonine (Fig. 4D). The distances between the heavy atoms of the hydroxyls of Ser176 and Thr27 and the

substrate are suggestive of hydrogen bonding. The hydroxyl oxygen of Ser176 is 2.7 Å from the carboxyl oxygen of L-threonine. The distance between the hydroxyl oxygen of Thr27 and the hydroxyl oxygen of the substrate threonine is 3.3 Å.

TsaC interaction with ATP

E. coli TsaC has been shown to bind ATP and with L-threonine to produce TC-AMP (16,24,25). The interaction of TsaC with ATP was characterized by isothermal titration calorimetry (ITC; Fig. 5A). The binding was shown to have a 1:1 stoichiometry and considerable affinity with a dissociation constant (K_d) of 14.8 ± 2.0 μM (Table 2). The Gibbs free energy (ΔG) of the binding event was -6.12 ± 0.01 kcal/mol, indicating a favorable interaction. These ITC results could represent a non-productive mode of binding for ATP in the absence of the other substrates.

To identify the amino acids involved in the binding of ATP, amide chemical shift changes of ^{15}N-labeled TsaC were monitored by observing ^{1}H-^{15}N-HSQC spectra. Spectra were collected at six different concentrations of ATP resulting in protein:ligand ratios of 1:0 to 1:8 (Fig. 5B). The amino acid resonances found to be the most strongly affected by the addition of ATP were those of Arg188, Gly109, Ile59, Leu114 and Ala115. They exhibited maximum total chemical shift changes of $\Delta\delta_{N,H}$ of 1.03, 0.96, 0.59 and 0.58 ppm, respectively, at a protein:ATP ratio of 1:8 (Fig. 5C). These resonances among others of the most affected residues are located in three distinct regions of the TsaC structure. Arg188 is at the C-terminus, Gly109 is found in the loop between β4-β5, Ile59 is located in β3, and Leu114 and Ala115 are within β5, but are all situated on one side of the central protein cavity (Fig. 5D). The perturbations of these chemical shifts with the addition of ATP suggest that these amino acids are in direct contact with ATP, or are indirectly affected by the environmental changes produced by ATP binding or through conformational rearrangement of the protein.

As with the TsaC:L-threonine modeled complex, the ATP titration data were used to direct molecular modeling of the NMR-derived structure of *E. coli* TsaC with ATP using HADDOCK (Table 3). The E_{inter} was plotted as a function of backbone RMSD from the lowest energy model. The models converged to a Cα (protein) and C+P (ATP) RMSD of 0.6 ± 1.3 Å at the defined protein-ATP interface with an average buried surface area of 481 ± 168 Å^2. Nine clusters of structures with low RMSD and energy were obtained for all calculated models based on a minimum cluster size of four models and a Cα (TsaC) and C+P (ATP) RMSD of 7.5 Å. Of the resulting models in each cluster, the first cluster contained 62% of the total structures calculated indicating a high degree of convergence. The lowest energy structure in this cluster depicts the ATP-bound TsaC in directcontact only with the adenosine, with the phosphates outside of the binding pocket. As seen in the ^{1}H-^{15}N-HSQC experiments, the docking results depict aliphatic residues in the region of β3 and β5 strands create a hydrophobic pocket for adenosine.

Lastly, the TsaC:L-threonine:ATP complex was modeled using the lowest energy TsaC:L-threonine structure and docking ATP (Table 3). In plotting the E_{inter} as a function of backbone RMSD (Fig. 6) from the lowest energy model, the models converged to a Cα (protein) and C+P (ATP) RMSD of 0.3 ± 0.5 Å at the defined protein-ATP interface with an average buried surface area of 644 ± 74 Å^2. Three clusters of structures (Fig. 6) with low RMSD and energy were obtained for all calculated models based on a minimum cluster size of four models and a Cα (TsaC) and C+P (ATP) RMSD of 7.5 Å Of the resulting models in each cluster, the lowest energy structure in the third cluster best characterizes the interaction (Fig. 7A). However, this model differs noticeably from the TsaC:ATP structure in the positioning of the ligand. The adenosine of ATP is in a similar conformation, but the phosphates are now flipped into the binding pocket. So, it appears that the presence of L-threonine provides a more favorable environment for the phosphates to enter the binding pocket, indicating that L-threonine may be required for ATP binding and AMP formation. In the comparison of the TsaC:L-threonine:ATP modeled complex to the co-crystal structure of *S. tokodaii* Sua5 with L-threonine and AMP-PNP (Fig. 7B), several similarities and differences can be observed. For instance, the location of ATP in the binding pocket is similar and agrees quite well, but the orientation is slightly different. The difference in orientation could be the result of a comparison between AMP-PNP in the crystal and ATP of the NMR study. In both structures, the functionally

important amino group of the substrate L-threonine is positioned near ATP in the same orientation.

Backbone dynamics of TsaC

The potential for the protein's binding of L-threonine and ATP to include conformational changes prompted us to investigate the structural dynamics of TsaC. Backbone RMSD values were calculated for individual residues amongst the ten lowest energy structures using MOLMOL (51) to identify regions of localized high and low RMSD values (Fig 8A). Several regions of high local RMSD were observed. Loop regions and both termini exhibited high RMSD values compared to the average, which is common for all protein structures due to the range of motion at the termini. However, residues 186 through 190, at the C-terminus, had considerably higher values than typically observed. This is reflected in the ensemble of lowest energy structures and the sparse data for the C-terminus in the NMR spectra (Fig. 3A). Pro165-Glu177, located in the long loop that connects the α7 helix to β7 strand, had the highest individual RMSD values of the internal residues. This region contains the putative binding site for L-threonine and is one of the regions that displays the largest difference between the crystal and NMR structures (Fig. 3D). This difference is suggestive of a dynamic role in binding that is observable by NMR.

In order to verify that the high local RMSD values result from inherent dynamics and are not due to sparse restraints, TsaC backbone dynamics were determined using steady-state heteronuclear ^{1}H-^{15}N-NOE experiments (35). On the nanosecond timescale, the heteronuclear NOE values were indicative of a lack of backbone motion for most regions (Fig. 8B). The C-terminus exhibited the greatest range of motion with residues Arg188 and Gln189 having two of the lowest NOE values. The Gly190 signal completely disappeared resulting in a negative NOE. The N-terminus residues Leu5-Arg7 also had a lower average NOE, indicative of backbone dynamics. Additionally, residues Gly167 and Asn174 exhibited increased conformational flexibility. Both of these are located in the long loop of TsaC that connects α7 to β7, flanking the location of the L-threonine binding site (Fig. 4C).

The backbone dynamics correlate with the RMSD values that were calculated for the NMR-derived structures. This indicates that the most dynamic regions, specifically the C-terminal residues and the α7-β7 loop, may be important for protein function such as the coordination of L-threonine binding with ATP binding. To test this experimentally, a heteronuclear NOE experiment was conducted in which TsaC was titrated with ATP (Fig. 8C). The addition of ATP induced noteworthy changes to a few amino acids. Overall, the C-terminus remained dynamic upon the addition of ATP, but several other resonances were affected. Ile59, Ala63, Tyr131, Arg110, and Gly144 exhibited greater conformational dynamics, whereas Asn174, Phe111, Ala115, Glu47, and Gly167 became more structured in the presence of ATP. Of these, Ile59 and Ala115 were the two most affected by the titration with ATP in this experiment. This is consistent with the amide chemical shift changes observed in the ^{1}H-^{15}N-HSQC spectra. Combining results of these datasets provides evidence that Ile59 and Ala115 could be essential for the binding of ATP by TsaC.

DISCUSSION

The *N*^{6}-threonylcarbamoyladenosine (t^{6}A) nucleoside modification, found 3′-adjacent to the anticodon in ANN-decoding tRNAs across all domains of life, is essential to translational fidelity. Of the four proteins in *E. coli* that have been found to be required for the biosynthesis of t^{6}A, TsaC has been the most extensively studied in the context of this modification pathway, and is essential to t^{6}A$_{37}$ formation and cell viability (24). More recently, TsaC has been shown to function as a threonylcarbamoyl-AMP synthase by catalyzing TC-AMP formation from L-threonine, CO_2/HCO_3^- and ATP (17). Here, insights into the mechanism of TsaC function have been achieved by utilizing an NMR-derived structure, biochemical data and molecular modeling.

The high-resolution solution structure of the full-length protein is present in solution as a monomer with a large, hydrophobic binding pocket (Fig. 3). The structure is consistent with all protein structures solved to date in the TsaC/Sua5 family, such as YciO (28), Sua5 (29) and HypF (30). They all have an α/β twisted open-sheet topology with parallel and antiparallel adjacent β-strands and a central concave cavity lined with positive electrostatic potential. In relation to the TsaC crystal structure, the NMR-derived structure of *E.coli* TsaC is generally homologous to one subunit

of the homodimer (27). Some differences, likely caused by the crystal packing, are observed and obscure structural information in the substrate-binding regions in the crystal structure (Fig. 3D). In particular, the dimerization interface affects the C-terminus of TsaC, which appears to be important for substrate- and protein-protein interactions.

The mechanism by which TsaC catalyzes the formation of TC-AMP appears to include conformational changes in the protein with the binding of the substrates. Using NMR titration analyses we were able to observe and track these changes upon binding of L-threonine and ATP. The titration data with TsaC and L-threonine suggest that Thr27 and Ser176 in *E.coli* TsaC are involved in substrate binding. Both of these residues are conserved in the TsaC/Sua5 protein family (Fig. 9), suggesting that they may be important for function. In the co-crystal structure of *S. tokodaii* Sua5 with L-threonine, the corresponding residues, Thr34 and Ser182 both interact with L-threonine (50). *S. tokodaii* Sua5 Thr34 forms a hydrophobic environment for the γ-carbon of L-threonine, and Ser182 forms two hydrogen bonds with the carboxyl oxygen (50). From our molecular docking studies, we observe Ser176 coordinating L-threonine in a similar fashion. However, TsaC Thr27 appears to be making a hydrogen bond with the hydroxyl of the substrate (Fig. 4). This could be significant to differences in coordination of L-threonine or possibly an artifact of the molecular docking process.

The binding of ATP by TsaC observed in the ^{1}H-^{15}N-HSQC spectra indicates a binding site within the putative catalytic pocket. The residues with the greatest changes in amide chemical shifts were the hydrophobic amino acids Gly109, Leu114, Ala115 and Ile59 (Fig. 5). The *E. coli* TsaC residues Lys56, Leu58-Ile59-Leu60 and Ser113-Leu114-Ala115-Val116-Arg117 are conserved across this family (Fig. 9). These residues have been shown to surround the adenine-binding site in co-crystal structures of both *E. coli* HypF (middle domain, residues 188-378) and *S. tokodaii* Sua5 with the non-hydrolysable ATP analog AMP-PNP (30,50). For HypF, the adenine is buried in a hydrophobic environment between Leu277 and Pro249, and forms weak hydrogen bonds with Glu296 and Arg372 (30). Pro249 near the ATP binding site in HypF is not conserved in this family of proteins and not present in TsaC. Sua5 coordinates the adenosine with hydrophobic residues Ile66, Val101, Ala120, and Ile184 (50). When TsaC is aligned with the HypF and Sua5 structures, Ile59, Ala115 and Arg188 (resonances that were all greatly affected during the titration with ATP) correspond to the residues in the other two proteins that are important for the binding of the adenosine. To examine this further, HADDOCK was used to dock ATP to the *E. coli* TsaC structure to model the binding mode. The lowest energy TsaC:ATP structure resulting from the docking simulation placed the adenosine in the hydrophobic region of Leu114, Ala115 and Ile59. Thus, the observation that these residues in TsaC are affected by titration with ATP appears to be caused by the binding of adenosine. Gly109 in TsaC was also affected by ATP binding, yet it is located in the loop between β4 and β5, a region of the protein not in direct contact with ATP. Therefore, we postulate that ATP binding may cause a conformational change in this loop. However, no significant conformational change is observed in the lowest energy docking model.

It is possible that the TsaC forms contacts with ligands additional to those observable by NMR using the ^{1}H-^{15}N-HSQC analysis. The KxR/SxN ATP-binding motif present in TsaC is conserved throughout the TsaC/Sua5 family (24). This motif is important for the coordination of the phosphates of ATP in *S. tokodaii* Sua5 (50), but no significant changes in chemical shifts were detected for Lys50, Arg52 or Asn141 (Ser139 was unobservable) in our experiments with *E. coli* TsaC. ATP hydrolysis, dynamics, or H_2O exchange are reasonable explanations for not observing the binding of the phosphates. Therefore, we repeated the titration using the non-hydrolysable ATP analog, AMP-PNP, used in the crystallography of *S. tokodaii* Sua5 and *E. coli* HypF. However, TsaC residues Lys50, Arg52 and Asn141 remained unaffected in our NMR studies (data not shown). This indicates that the inability to observe interactions of TsaC with the phosphates of ATP was not due to hydrolysis of the phosphates. To further investigate the interaction of TsaC with the ATP phosphates, the TsaC:ATP interaction was modeled and it was revealed that the phosphate atoms were not in contact with any part of TsaC. The lack of contact between the two entities explains the unaffected KxR/SxN residues in the ^{1}H-^{15}N-HSQC experiments.

Perhaps TsaC requires the L-threonine to properly coordinate ATP in the active site. To test this idea, and further understand the mechanism employed by TsaC for the synthesis of TC-AMP, we added L-threonine to the TsaC:ATP NMR sample. No additional ^{1}H-^{15}N-HSQC chemical shift changes were observed (data not shown). Therefore, the ability of TsaC to bind both substrates simultaneously, and possibly cooperatively, was unclear. The binding of ATP could require the binding of L-threonine to occur first, providing the contact surface area for ATP to bind in the correct conformation. As seen in the Sua5 structure (50) contacts are observed between ATP and the L-threonine with the L-threonine buried further into the Sua5 binding cavity (Fig. 7B). To investigate this possibility, the TsaC:L-threonine modeled structure was docked to ATP (Fig. 7A). The lowest energy structure of this modeled interaction is remarkably similar to that seen in the Sua5 crystal structure (Fig. 7) (45). In fact, the phosphates of ATP are positioned in close enough proximity to K50 and S139 of the KxR/SxN ATP-binding motif to suggest coordination. This compellingly suggests that the recognition of ATP is dependent on the presence of L-threonine. Indeed, it has been shown that TsaC forms AMP only in the presence of L-threonine and bicarbonate (16). However, the binding order and kinetics of this reaction are unknown and more investigation into this is necessary.

E. coli TsaC contains sites that are highly dynamic, particularly the C-terminus. Both the heteronuclear NOEs and the RMSD values from the ten lowest energy structures depict considerable flexibility for G167-G190 (Fig. 8). The G167-G190 region may function in protein-protein interactions with the other $t^{6}A$-synthase subunits, TsaD, TsaE and TsaB, or as an arm capable of folding into the concave binding pocket. The presence of two consecutive glycines at 170-171 indicates that the G167-G190 region may function as a hinge enabling the C-terminus to act as a gate during substrate binding. Both Arg188 and Gln189 display a high degree of conformational motion relative to the other residues and also are affected by ATP binding. The conformational dynamics of Arg188 and Gln189 are consistent with the hypothesis that the C-terminus acts a flexible linker or arm that may participate in the binding of ATP or interact with other proteins. HypF is composed of domains homologous to TsaC and TsaD/Kae1. In fact, the TsaC domain of HypF (residues 188-378) is linked directly to the TsaD domain (residues 379-746) (30) suggesting a direct interaction of the C-terminus of TsaC with TsaD in $t^{6}A_{37}$ biosynthesis. There are clear structural differences between the TsaC crystal and NMR structures with the greatest difference at this exact location (Fig 3). Since we wished to visualize the C-terminus of TsaC in our studies, it was important to use the NMR-derived structure of TsaC and perform a blind docking of the TsaC:L-threonine:ATP modeled interaction.

In conclusion, the study presented here provides valuable information about the mechanism of TsaC in $t^{6}A_{37}$ biosynthesis. The identification of TsaC residues that are affected by the binding of ATP and L-threonine provide sites at which a future in depth mutational analysis coupled to ITC and NMR experiments could confirm the importance of individual amino acids, and/or their properties, for binding. *E. coli* TsaC binds both ATP and L-threonine adjacent to each other with conserved residues located within the large, concave cavity at the center of the structure. The adenosine of ATP is inserted into a hydrophobic environment created by residues within β3 and β5. The proper coordination of ATP in the binding pocket seems to require the presence of L-threonine, suggesting a cooperative binding event between these substrates, in which the binding of L-threonine occurs first. The C-terminal amino acids appear to be important for ATP binding and possibly catalytic activity. A central and compelling issue in the mechanism of TC-AMP formation, not addressed in this work, is whether TsaC utilizes CO_2 or HCO_3^-, and how it reacts with L-threonine to form intermediate L-threonine carbamate. Further analyses of TsaC with CO_2/HCO_3^-, and the structural interactions of the binding of TsaC with tRNA, TsaD, TsaB and TsaE, especially in revealing the protein-RNA and protein-protein interfaces, will provide additional and significant insight into this complex enzymatic process.

Acknowledgements

We thank Drs. Manal Swairjo and Kyla Frohlich for insightful discussions, Dr. Alex Shekhtman for NMR assistance and the Research IT team at the University at Albany.

This work was supported by the National Science Foundation (MCB-0548602; MCB-1101859) and the National Institutes of Health (5R01-GM23037-25) to PFA.

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 2MX1. The chemical shift assignments are available at the Biological Magnetic Resonance Data Bank under accession number 25381.

[1] The three authors contributed equally to this work.

[2] To whom correspondence may be addressed: Paul F. Agris, The RNA Institute, University at Albany, 1400 Washington Ave., Albany, NY 12222, USA, Tel.: (518) 437-4448; FAX: (518) 437-4456; E-mail: pagris@albany.edu

[3] The abbreviations used here are: t^6A, N^6-threonylcarbamoyladenosine; HSQC, heteronuclear single quantum coherence; TOCSY, total correlation spectroscopy; TROSY, transverse relaxation optimized spectroscopy; RMSD, root-mean-square deviation; ITC, isothermal titration calorimetry; AMP-PNP, adenosine-5'-(β,γ-imido)triphosphate

Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author Contributions

PFA and DIR designed the study. KAH and KLS performed NMR experiments, analyzed data and solved the structure. KAH and AFS conducted binding experiments and analyzed data. BB performed *in silico* docking experiments. BB, YB and CD provided technical assistance in NMR spectroscopy and protein expression. KAH, KLS and PFA wrote the paper. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., Vendeix, F. A. P., Fabris, D., and Agris, P. F. (2010) The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res* **39**, D195-D201
2. Auffinger, P., and Westhof, P. (1998) Location and Distribution of Modified Nucleotides in tRNA. in *Modification and Editing of RNA* (Grosjean, H., and Benne, R. eds.), ASM Press, Washington, D.C. pp 569-576
3. Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., and Pütz, J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**, D159-D162
4. Grosjean, H., Sprinzl, M., and Steinberg, S. (1995) Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. *Biochimie* **77**, 139-141
5. Morin, A., Auxilien, S., Senger, B., Tewari, R., and Grosjean, H. (1988) Structural requirements for enzymatic formation of threonylcarbamoyladenosine (t6A) in tRNA: An in vivo study with *Xenopus laevis* oocytes. *RNA* **4**, 24-37
6. Miller, J. P., Hussain, Z., and Schweizer, M. P. (1976) The involvement of the anticodon adjacent modified nucleoside N-(9-(BETA-D-ribofuranosyl) purine-6-ylcarbamoyl)-threonine in the biological function of *E. coli* tRNAIle. *Nucleic Acids Res* **3**, 1185-1201
7. Weissenbach, J., and Grosjean, H. (1981) Effect of threonylcarbamoyl modification (t6A) in yeast tRNA Arg III on codon-anticodon and anticodon-anticodon interactions. A thermodynamic and kinetic evaluation. *Eur J Biochem* **116**, 207-213
8. Yarian, C., Townsend, H., Czestkowski, W., Sochacka, E., Malkiewicz, A., Guenther, R., Miskiewicz, A., and Agris, P. F. (2002) Accurate Translation of the Genetic Code Depends on tRNA Modified Nucleosides. *J Biol Chem* **277**, 16391-16395
9. Murphy, F. V., 4th, Ramakrishnan, V., Malkiewicz, A., and Agris, P. F. (2004) The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat Struct Mol Biol* **11**, 1186-1191
10. Vendeix, F. A. P., Murphy, F. V., Cantara, W. A., Leszczyńska, G., Gustilo, E. M., Sproat, B., Malkiewicz, A., and Agris, P. F. (2012) Human tRNA(Lys3)UUU Is Pre-Structured by Natural Modifications for Cognate and Wobble Codon Binding through Keto–Enol Tautomerism. *Journal of Molecular Biology* **416**, 467-485
11. Agris, P. F. (1991) Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* **73**, 1345-1349
12. Nishimura, S. (1972) Minor Components in Transfer RNA: Their Characterization, Location, and Function. *Prog Nucleic Acid Res Mol Biol* **12**, 49-85
13. Lin, C. A., Ellis, S. R., and True, H. L. (2010) The Sua5 protein is essential for normal translational regulation in yeast. *Mol Cell Biol* **30**, 354-363
14. El Yacoubi, B., Hatin, I., Deutsch, C., Kahveci, T., Rousset, J.-P., Iwata-Reuyl, D., G Murzin, A., and de Crécy-Lagard, V. (2011) A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification. *EMBO J* **30**, 882-893
15. Daugeron, M. C., Lenstra, T. L., Frizzarin, M., El Yacoubi, B., Liu, X., Baudin-Baillieu, A., Lijnzaad, P., Decourty, L., Saveanu, C., Jacquier, A., Holstege, F. C. P., de Crecy-Lagard, V., van Tilbeurgh, H., and Libri, D. (2011) Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t6A modification of tRNAs. *Nucleic Acids Res* **39**, 6148-6160
16. Deutsch, C., El Yacoubi, B., de Crécy-Lagard, V., and Iwata-Reuyl, D. (2012) Biosynthesis of Threonylcarbamoyl Adenosine (t6A), a Universal tRNA Nucleoside. *J Biol Chem* **287**, 13666-13673
17. Lauhon, C. T. (2012) Mechanism of N6-threonylcarbamoyladenonsine (t6A) biosynthesis: Isolation and characterization of the intermediate threonylcarbamoyl-AMP. *Biochemistry* **51**, 8950-8963
18. Perrochia, L., Crozat, E., Hecker, A., Zhang, W., Bareille, J., Collinet, B., van Tilbeurgh, H., Forterre, P., and Basta, T. (2013) *In vitro* biosynthesis of a universal t6A tRNA modification in Archaea and Eukarya. *Nucleic Acids Res* **41**, 1953-1964
19. Wan, L. C. K., Mao, D. Y. L., Neculai, D., Strecker, J., Chiovitti, D., Kurinov, I., Poda, G., Thevakumaran, N., Yuan, F., Szilard, R. K., Lissina, E., Nislow, C., Caudy, A. A., Durocher, D., and Sicheri, F. (2013)

Reconstitution and characterization of eukaryotic N6-threonylcarbamoylation of tRNA using a minimal enzyme system. *Nucleic Acids Research* **41**, 6332-6346
20. Downey, M., Houlsworth, R., Maringele, L., Rollie, A., Brehme, M., Galicia, S., Guillard, S., Partington, M., Zubko, M. K., Krogan, N. J., Emili, A., Greenblatt, J. F., Harrington, L., Lydall, D., and Durocher, D. (2006) A Genome-Wide Screen Identifies the Evolutionarily Conserved KEOPS Complex as a Telomere Regulator. *Cell* **124**, 1155-1168
21. Kisseleva-Romanova, E., Lopreiato, R., Baudin-Baillieu, A., Rousselle, J. C., Ilan, L., Hofmann, K., Namane, A., Mann, C., and Libri, D. (2006) Yeast homolog of cancer-testis antigen defines a new transcription complex. *EMBO J* **25**, 3576-3585
22. Mao, D. Y. L., Neculai, D., Downey, M., Orlicky, S., Haffani, Y. Z., Ceccarelli, D. F., Ho, J. S. L., Szilard, R. K., Zhang, W., Ho, C. S., Wan, L., Fares, C., Rumpel, S., Kurinov, I., Arrowsmith, C. H., Durocher, D., and Sicheri, F. (2008) Atomic Structure of the KEOPS Complex: An Ancient Protein Kinase-Containing Molecular Machine. *Mol Cell* **32**, 259-275
23. Srinivasan, M., Mehta, P., Yu, Y., Prugar, E., Koonin, E. V., Karzai, A. W., and Sternglanz, R. (2011) The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. *EMBO J* **30**, 873-881
24. El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., Williamson, J. R., Schimmel, P., Swairjo, M. A., and de Crécy-Lagard, V. (2009) The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. *Nucleic Acids Res* **37**, 2894-2909
25. Harris, K. A., Jones, V., Bilbille, Y., Swairjo, M. A., and Agris, P. F. (2011) YrdC exhibits properties expected of a subunit for a tRNA threonylcarbamoyl transferase. *RNA* **17**, 1678-1687
26. Perrochia, L., Guetta, D., Hecker, A., Forterre, P., and Basta, T. (2013) Functional assignment of KEOPS/EKC complex subunits in the biosynthesis of the universal t6A tRNA modification. *Nucleic Acid Res* **41**, 1953-1964
27. Teplova, M., Tereshko, V., Sanishvili, R., Joachimiak, A., Bushueva, T., Anderson, W. F., and Egli, M. (2000) The structure of the *yrdC* gene product from *Escherichia coli* reveals a new fold and suggests a role in RNA binding. *Protein Sci* **9**, 2557-2566
28. Jia, J., Lunin, V. V., Sauvé, V., Huang, L. W., Matte, A., and Cygler, M. (2002) Crystal structure of the YciO protein from *Escherichia coli*. *Proteins* **49**, 139-141
29. Agari, Y., Sato, S., Wakamatsu, T., Bessho, Y., Ebihara, A., Yokoyama, S., Kuramitsu, S., and Shinkai, A. (2007) X-ray crystal structure of a hypothetical Sua5 protein from *Sulfolobus tokodaii* strain 7. *Proteins* **70**, 1108-1111
30. Petkun, S., Shi, R., Li, Y., Asinas, A., Munger, C., Zhang, L., Waclawek, M., Soboh, B., Sawers, R. G., and Cygler, M. (2011) Structure of hydrogenase maturation protein HypF with reaction intermediates shows two active sites. *Structure* **19**, 1773-1783
31. Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J., and Li, Q. (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci* **18**, 936-948
32. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: A hybrid method fro predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213-223
33. Ottiger, M., Delaglio, F., and Bax, A. Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **131**, 373-378
34. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293
35. Goddard, T. D., and Kneller, D. G. (2008) SPARKY 3. University of California, San Francisco
36. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-352
37. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phophopeptide-complexed Src Homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984-6003
38. Güntert, P. (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-378

39. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**, 66-74
40. Schwieters, C. D., Kuszewski, J. J., and Clore, G. M. Using Xplor-NIH for NMR molecular structure determination. *Progr NMR Spectroscopy* **48**, 47-62
41. Bhattacharya, A., Tejero, R., and Montelione, G. T. (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* **66**, 778-795
42. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3.
43. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-1737
44. Hubbard, S. J., and Thornton, J. M. (1993) NACCESS. University College London
45. McLachlan, A. D. (1982) Rapid Comparison of Protein Structures. *Acta Cryst* **A38**, 871-873
46. Martin, A. C. R., and Porter, C. T. (1996) ProFit. University College London
47. Ferrage, F., Zoonens, M., Warschawski, D. E., Popot, J.-L., and Bodenhausen, G. (2003) Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. *J Am Chem Soc* **125**, 2541-2545
48. Clore, G. M. and Garrett, D. S. *R*-factor, free *R*, and complete cross-validation for dipolar coupling refinement of NMR structures. *J Am Chem Soc* **121**, 9008-9012
49. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477-486
50. Kuratani, M., Kasai, T., Akasaka, R., Higashijima, K., Terada, T., Kigawa, T., Shinkai, A., Bessho, Y., and Yokoyama, S. (2011) Crystal structure of *Sulfolobus tokodaii* Sua5 complexed with L-threonine and AMPPNP. *Proteins* **79**, 2065-2075
51. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graphics* **14**, 51-55

FIGURE LEGENDS

Figure 1. Requirements for threonylcarbamoylation of tRNA. (A) The structure of N^6-threonylcarmbamoyladenosine (t^6A). (B) Consensus sequence for t^6A-modifed tRNA species, which all recognize ANN codons in mRNA. (C) L-threonylcarbamoyl-AMP (TC-AMP) formation requires L-threonine, CO_2/HCO_3^-, ATP and TsaC.

Figure 2. *E. coli* TsaC ^1H-^{15}N-TROSY spectrum with amino acid amide assignments. All expected amides were present and assigned with the exception of Asn2, Asn3, Val118 and Ser139. The two unlabeled downfield peaks are the Trp89 and Trp106 NHε resonances.

Figure 3. TsaC NMR solution structure. (A) The twenty lowest energy structures of *E. coli* TsaC characterized by NMR. (B) Hydrophobic residues (orange) of the TsaC NMR solution structure. (C) Electrostatic potential of the TsaC NMR solution structure, where blue and red denote positive and negative charges, respectively. (D) Comparison of the TsaC crystal (PDB code 1HRU) and NMR-solved TsaC structures. The loop width is directly proportional to the differences in Cα RMSD between the two structures. Small to large differences in Cα RMSD are colored red to white to blue. Spheres denote ATP (solid box) and L-threonine (dotted box) binding sites. The solid box also represents the crystal packing dimerization interface of the crystal structure. All structures are in the same orientation.

Figure 4. Binding of L-threonine and TsaC in ^1H-^{15}N-HSQC spectra. Spectra of the various protein:ligand ratios are overlaid: 1:0 (gray), 1:1 (yellow), 1:2 (purple) and 1:4 (black). (A) The ^1H-^{15}N-HSQC region containing Ser176. At a 1:4 ratio, the resonance has completely disappeared. (B) The ^1H-^{15}N-HSQC displaying the signal for Thr27 weakens due to exchange with each addition of L-threonine. (C) Lowest energy modeled structure of L-threonine (magenta) with the *E. coli* TsaC structure from cluster 1. Residues Thr27, Ser176 (cyan), Asn174 and Gly167 (green) are highlighted. (D) Local structure of the L-threonine binding site with Thr27 and Ser176 (cyan).

Figure 5. The interaction of TsaC with ATP. (A) ITC of TsaC with ATP. The baseline-corrected data of the exothermic reaction (top) and the integrated heat changes (bottom) obtained upon the titration of ATP into TsaC. The solid line represents the data fit to a non-linear least squared isotherm fitting using a 1:1 binding model. (B) Overlaid ^1H-^{15}N-HSQC spectra of TsaC titrated with ATP. The chemical shifts with the greatest changes are labeled and marked with colored arrows corresponding to the chemical shift perturbation profile. (C) Chemical shift perturbation profile of the combined amide chemical shift changes for each TsaC residue at the final titration point (1:8 molar ratio of TsaC:ATP). (D) Local structure of *E. coli* TsaC with residues Ile59, Gly109, Leu114, Ala115, Arg188 and Gln189 highlighted in colors corresponding to the graph.

Figure 6. HADDOCK was used to develop a structural model of the L-threonine:ATP and TsaC using the NMR structure of TsaC and the NMR titration data. (A) A plot of the E_{inter} (sum of restraint, van der Waals, and electrostatic energy terms) as a function of backbone r.m.s.d from the lowest energy model. (B-D) The lowest energy structure from each cluster, cluster 1-3, respectively.

Figure 7. Data-driven molecular docking of *E. coli* TsaC with L-threonine and ATP. (A) The lowest energy modeled structure of the most populated cluster for the modeled interaction of TsaC with L-threonine (magenta) and ATP (green). (B) For comparison, the co-crystal structure of *S. tokodaii* Sua5 (green, PDB code 3AJE) with L-threonine and AMPPNP is aligned with the *E. coli* TsaC structure shown in A (cyan).

Figure 8. Backbone dynamics of TsaC. (A) Backbone RMSD values per residue calculated using the ten lowest energy structures. (B) ^1H-^{15}N heteronuclear NOE relaxation data. (C) The difference in ^1H-^{15}N

heteronuclear NOE relaxation values with and without ATP present. Negative and positive values represent more and less dynamic residues, respectively.

Figure 9. TsaC/Sua5 conserved amino acids. Alignment of *E. coli* TsaC with *E. coli* HypF, *S. tokodaii* Sua5, *B. subtilis* YwlC and *E. coli* YciO was completed using the ClustalW2 sequence alignment server. L-threonine-binding residues, Thr27 and Ser175 (orange) are moderately conserved. Adenosine-binding residues (blue), Ile59, Leu114, Ala115, Leu178, are conserved, with Arg188 highly conserved; however, Gly109 is not conserved. The KxR/SxN ATP-binding motif is highlighted in purple. Gray and black indicate moderately and highly conserved residues, respectively.

TABLES

Table 1. TsaC NMR and refinement statistics

NMR experimental restraints		
Distance restraints		
	Total NOE	1869
	Intra-residue	390
	Inter-residue	1479
	Sequential ($\|i-j\| = 1$)	656
	Medium range ($\|i-j\| < 4$)	444
	Long range ($\|i-j\| > 5$)	379
	Hydrogen bonds	99
Total dihedral angle restraints		
	φ	98
	ψ	89
^{15}N-^{1}H residual dipolar couplings		66
Structure statistics		
RMSD from solution NMR restraints		
	Distance restraints (Å)	0.061 ± 0.001
	Dihedral angle restraints (°)	0.599 ± 0.066
	RDC restraints (Hz)	0.111 ± 0.020
RMSD from idealized covalent geometry		
	Bonds (Å)	0.006 ± 0.0001
	Angles (°)	0.746 ± 0.007
	Impropers (°)	0.504 ± 0.010
RDC quality factor[a]		0.182 ± 0.006
Precision of atomic coordinates over secondary structure[b]		
	Backbone (N, Cα, C', O) (Å)	0.792 ± 0.148
	Heavy atoms (Å)	1.355 ± 0.132
Ramachandran analysis[c]		
	Most favored	77.4 %
	Additionally allowed	17.9 %
	Generously allowed	4.1 %
	Disallowed	0.6 %

[a] RDC quality factor calculated according to the method of Clore and Garrett (48).

[b] The secondary elements used were as follows: residues 4-19, 40-48, 65-71, 79-86 105-109, 123-131, 151-158, 23-25, 32-35, 59-62, 94-97,114-117, 135-138 and 177-180.

[C] Ramachandran statistics calculated for the ensemble using PROCHECK-NMR (49).

Table 2. Thermodynamic parameters of TsaC binding

	ATP
Stoichiometry (n)	1.07 ± 0.022
K_a (M^{-1})	$(6.83 \pm 0.92) \times 10^4$
K_d (μM)	14.8 ± 2.0
ΔH (cal/mol)	-440.6 ± 13.6
ΔS [cal/(K*mol)]	20.5
TΔS (kcal/mol)	5.678
ΔG (kcal/mol)	-6.12 ± 0.01

Table 3. Active and passive residues for all docking

	Ligand[a]		TsaC	
	Active	**Passive**	**Active**	**Passive**
TsaC:L-**threonine**	1	N/A	27,176	25,28-31,59,91-93,121-123,167-170,174,175,177,187-190
TsaC:**ATP**	Adenosine moiety	Phosphate group closest to the adenosine moiety	59,109,114,115,188	25,28-31,59,91-93,121-123,167-170,174,175,177,187,189,190
TsaC:L-threonine:**ATP**	Adenosine moiety	Phosphate group closest to the adenosine moiety	59,109,114,115,188	38,10,14,17,20, 28,30,38,40,41, 44,48,5152,64,66, 69-70,73,75,77-80,83,86-87,91, 100-103,110-111, 120-121,125,129, 131-132,139,141, 143-146,149-150, 152,155-158, 161-162,167-168, 170-173,175,180, 182-183,185-187,189, 190

[a]The molecule defined as the ligand is shown in bold

FIGURE 1

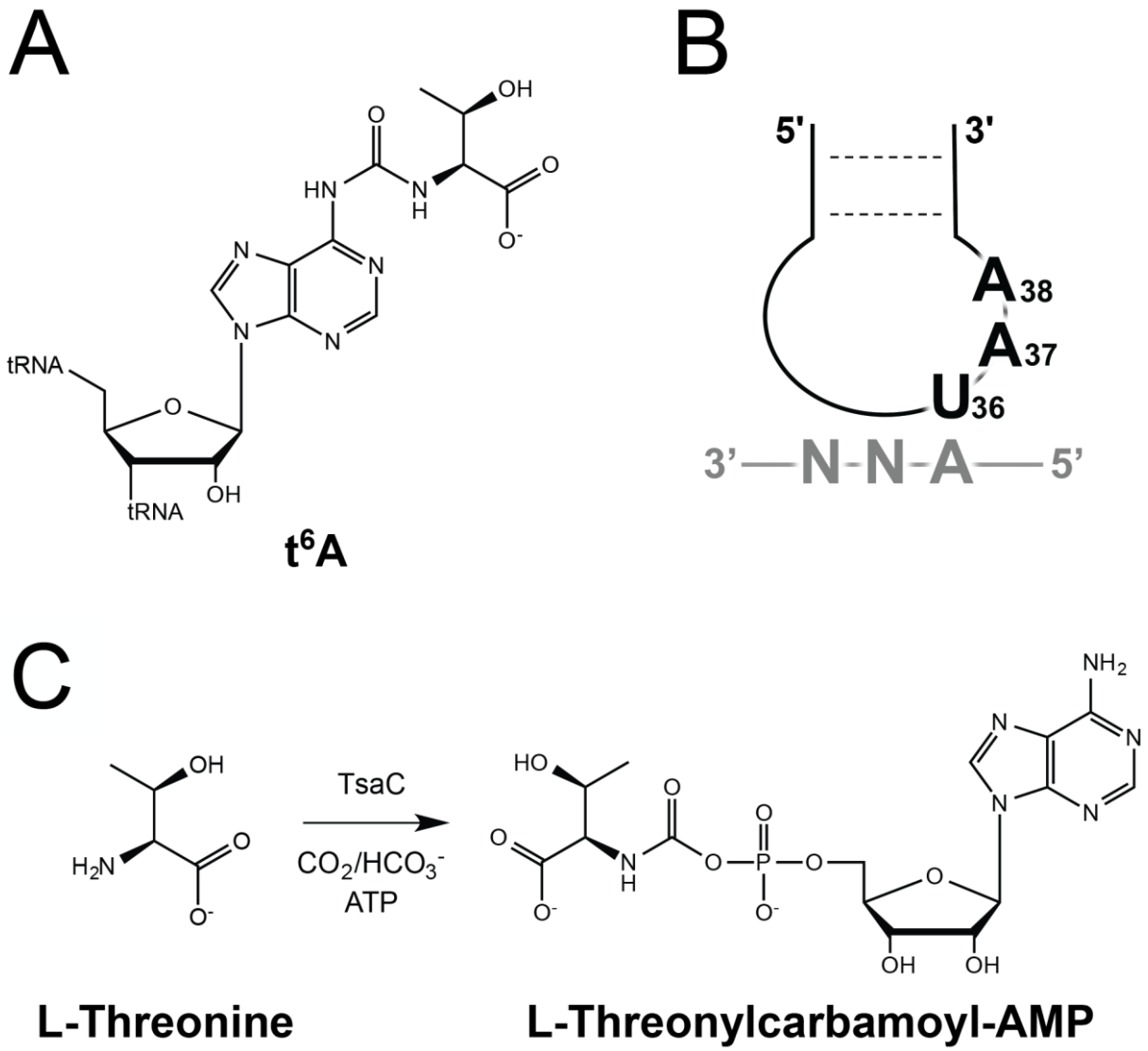

A
OH
O
HN
N
H
O
O-
N
N
N
N
tRNA
O
OH
tRNA
t6A
B
5'
3'
A38
A37
U36
3'—N-N-A—5'
C
OH
H2N
O
O-
TsaC
CO2/HCO3-
ATP
HO
O
O
N
H
O
P
O
O-
O
O
NH2
N
N
N
N
OH
OH
L-Threonine
L-Threonylcarbamoyl-AMP

FIGURE 2

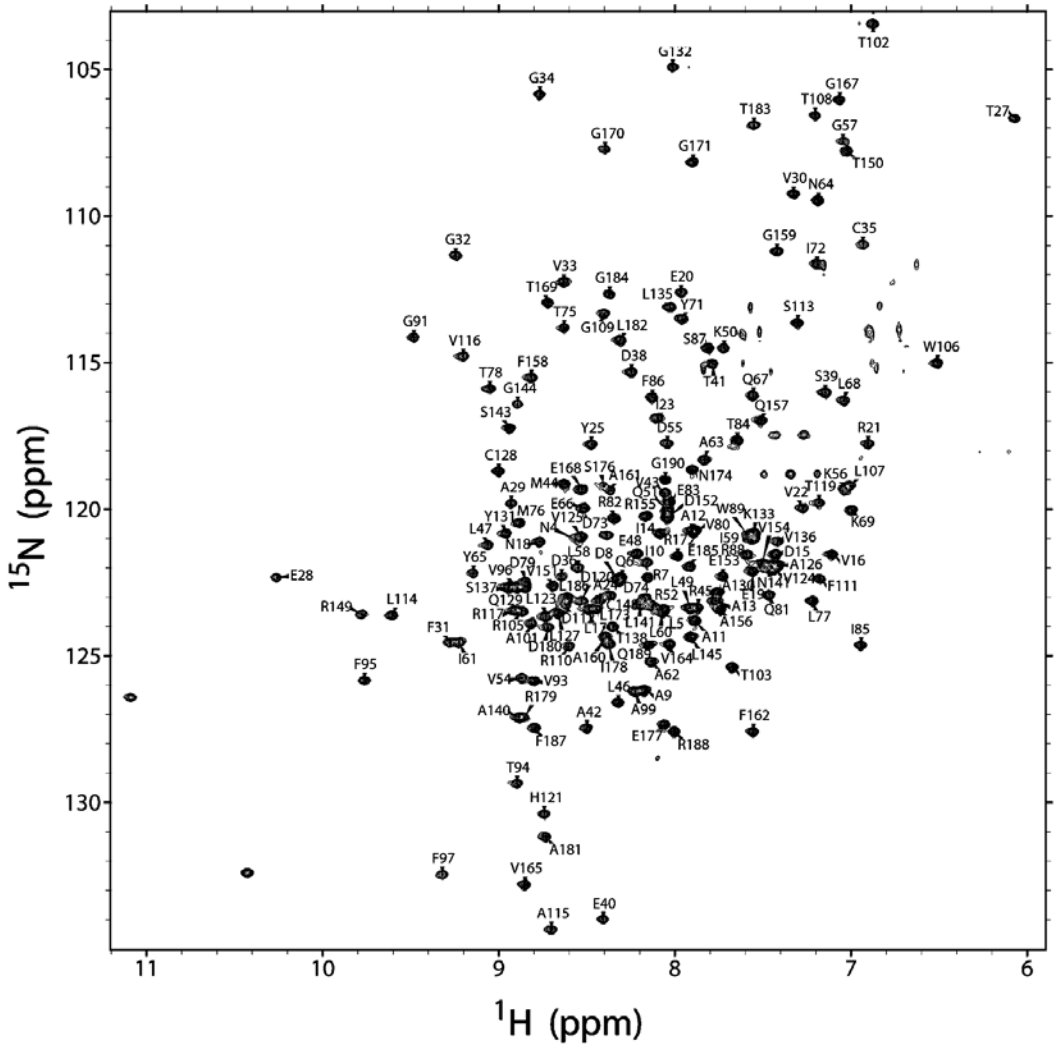


FIGURE 3

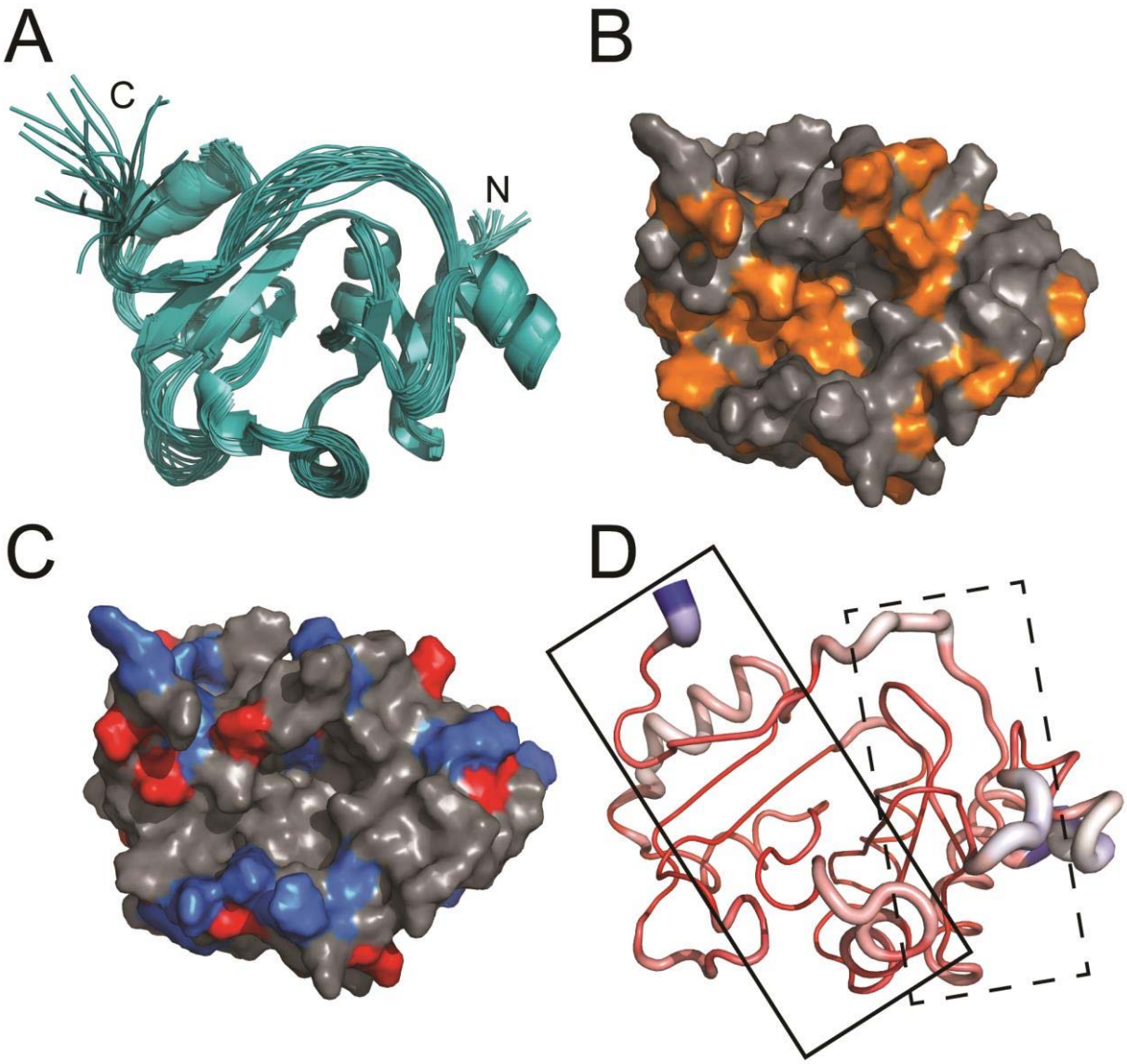


FIGURE 4

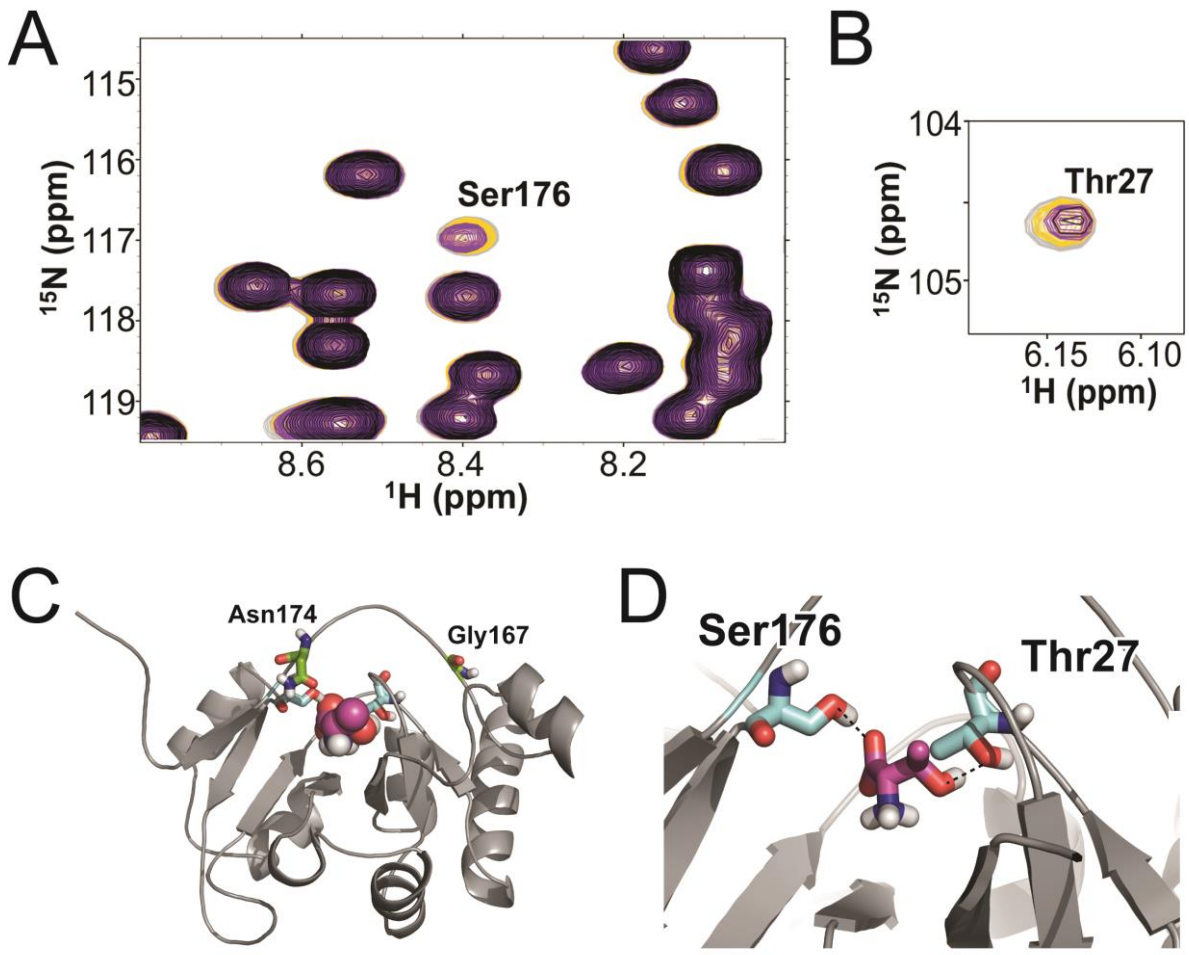

A
115
116
117
118
119
^{15}N (ppm)
Ser176
8.6
8.4
8.2
^{1}H (ppm)
B
104
105
^{15}N (ppm)
Thr27
6.15
6.10
^{1}H (ppm)
C
Asn174
Gly167
D
Ser176
Thr27

FIGURE 5

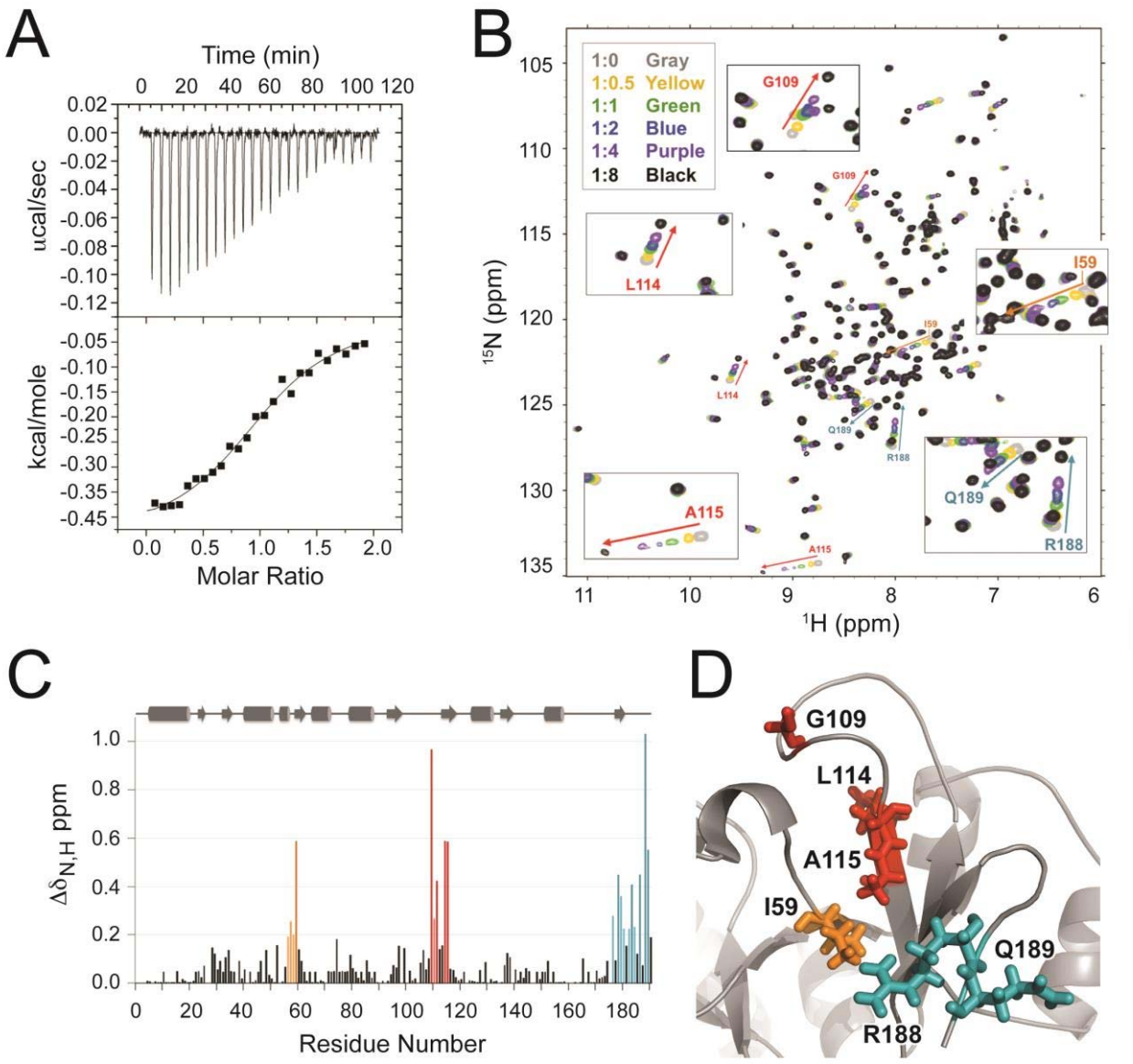

A
Time (min)
0 20 40 60 80 100 120
0.02
0.00
-0.02
-0.04
-0.06
-0.08
-0.10
-0.12
µcal/sec
-0.05
-0.10
-0.15
-0.20
-0.25
-0.30
-0.35
-0.45
kcal/mole
0.0 0.5 1.0 1.5 2.0
Molar Ratio
B
1:0 Gray
1:0.5 Yellow
1:1 Green
1:2 Blue
1:4 Purple
1:8 Black
G109
L114
I59
A115
Q189
R188
105
110
115
120
125
130
135
15N (ppm)
11 10 9 8 7 6
1H (ppm)
C
1.0
0.8
0.6
0.4
0.2
0.0
ΔδN,H ppm
0 20 40 60 80 100 120 140 160 180
Residue Number
D
G109
L114
A115
I59
Q189
R188

FIGURE 6

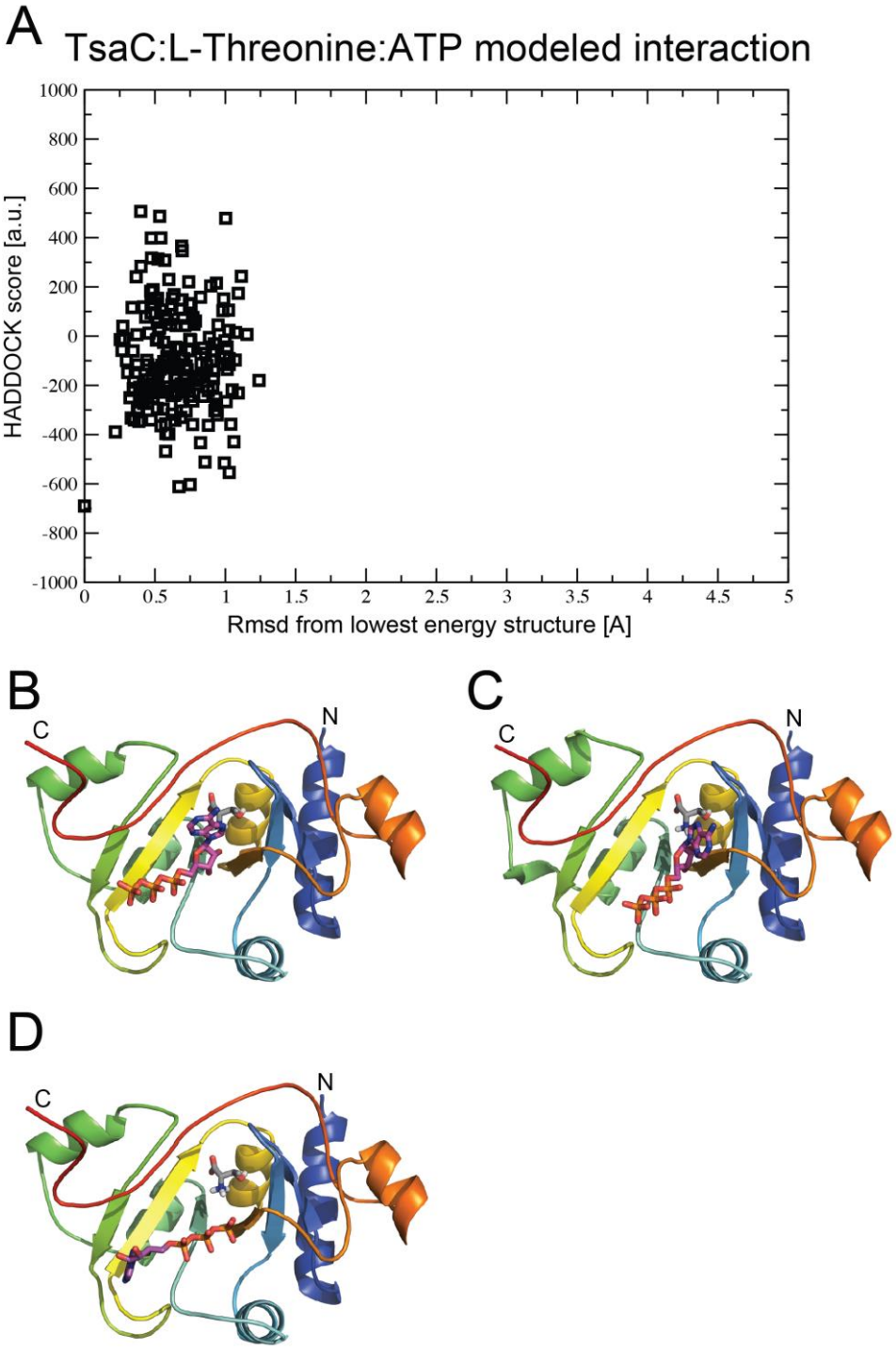

A
TsaC:L-Threonine:ATP modeled interaction
HADDOCK score [a.u.]
1000
800
600
400
200
0
-200
-400
-600
-800
-1000
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Rmsd from lowest energy structure [A]
B
C
N
C
D
C
N

FIGURE 7

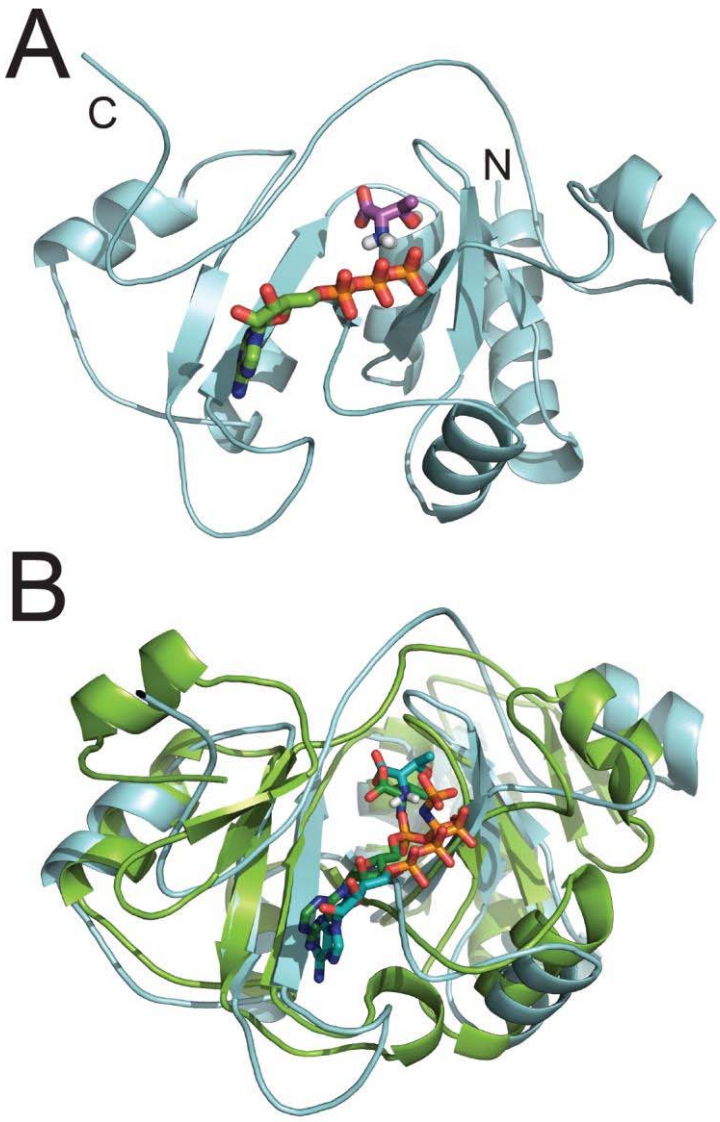


FIGURE 8

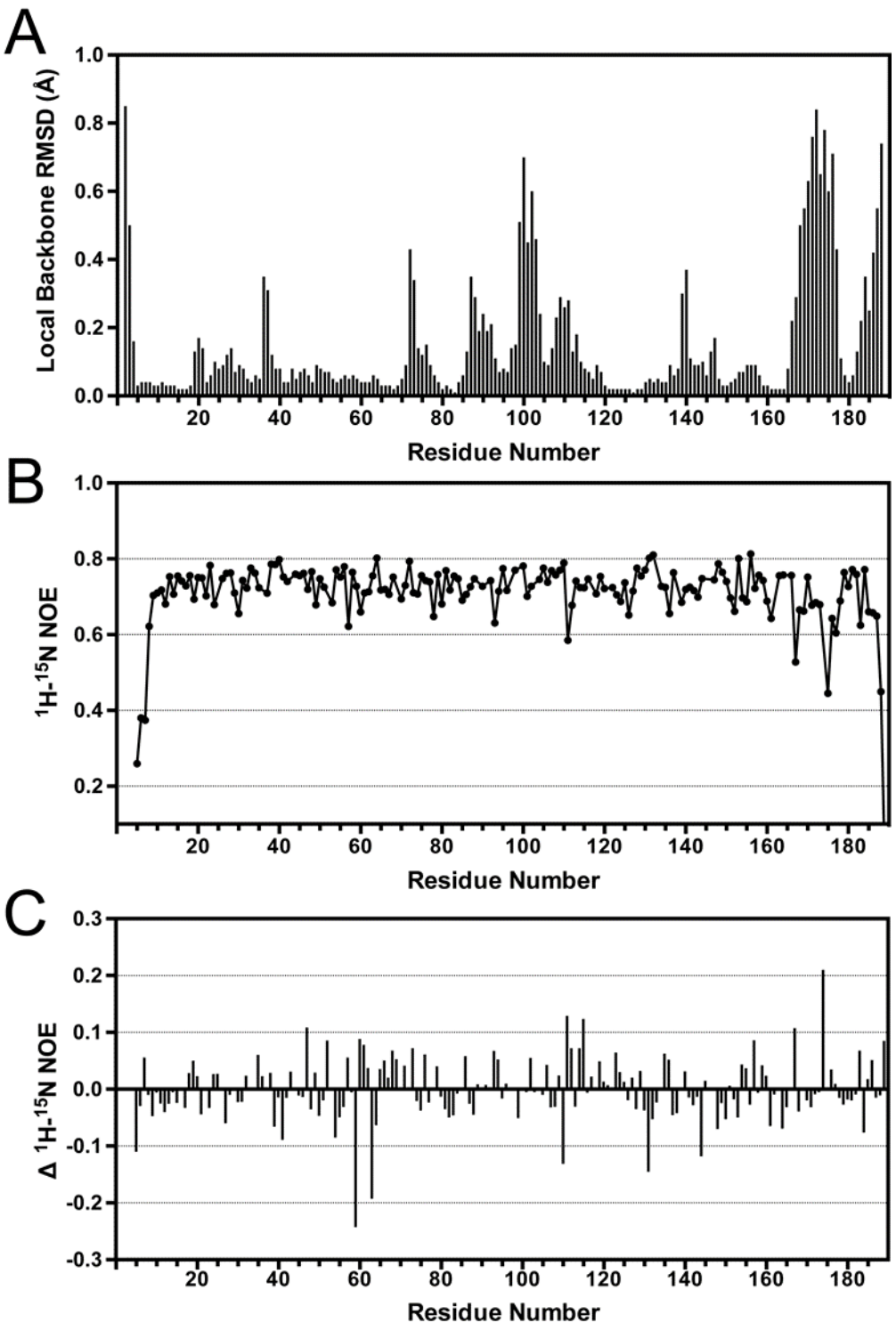
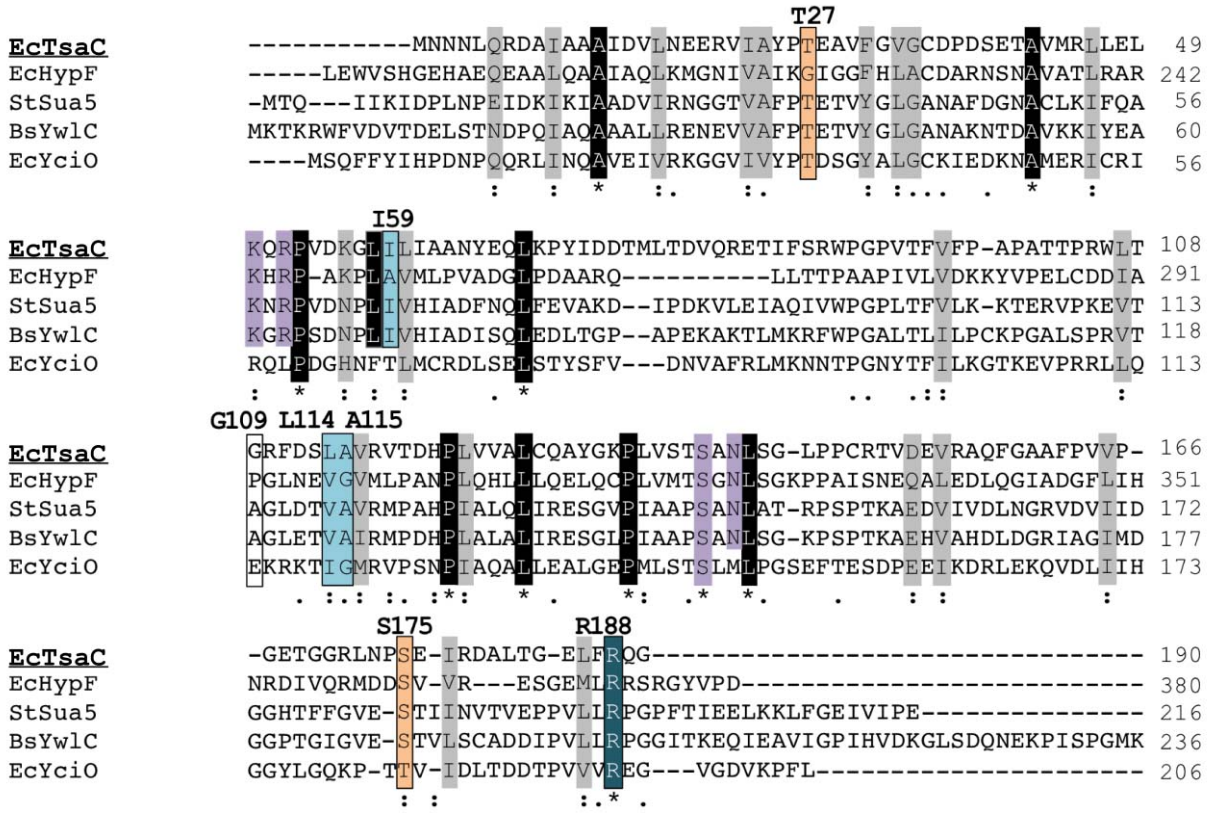

A
Local Backbone RMSD (Å)
1.0
0.8
0.6
0.4
0.2
0.0
20
40
60
80
100
120
140
160
180
Residue Number
B
^{1}H-^{15}N NOE
1.0
0.8
0.6
0.4
0.2
20
40
60
80
100
120
140
160
180
Residue Number
C
Δ ^{1}H-^{15}N NOE
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
20
40
60
80
100
120
140
160
180
Residue Number

FIGURE 9


T27
EcTsaC ----------MNNNLQRDAIAAAIDVLNEERVIAYPTEAVFGVGCDPDSETAVMRLLEL 49
EcHypF -----LEWVSHGEHAEQEAALQAAIAQLKMGNIVAIKGIGGFHLACDARNSNAVATLRAR 242
StSua5 -MTQ---IIKIDPLNPEIDKIKIAADVIRNGGTVAFPTETVYGLGANAFDGNACLKIFQA 56
BsYwlC MKTKRWFVDVTDELSTNDPQIAQAAALLRENEVVAFPTETVYGLGANAKNTDAVKKIYEA 60
EcYciO ----MSQFFYIHPDNPQQRLINQAVEIVRKGGVIVYPTDSGYALGCKIEDKNAMERICRI 56
I59
EcTsaC KQRPVDKGLILIAANYEQLKPYIDDTMLTDVQRETIFSRWPGPVTFVFP-APATTPRWLT 108
EcHypF KHRP-AKPLAVMLPVADGLPDAARQ----------LLTTPAAPIVLVDKKYVPELCDDIA 291
StSua5 KNRPVDNPLIVHIADFNQLFEVAKD--IPDKVLEIAQIVWPGPLTFVLK-KTERVPKEVT 113
BsYwlC KGRPSDNPLIVHIADISQLEDLTGP--APEKAKTLMKRFWPGALTLILPCKPGALSPRVT 118
EcYciO RQLPDGHNFTLMCRDLSELSTYSFV---DNVAFRLMKNNTPGNYTFILKGTKEVPRRLLQ 113
G109 L114 A115
EcTsaC GRFDSLAVRVTDHPLVVALCQAYGKPLVSTSANLSG-LPPCRTVDEVRAQFGAAFPVVP- 166
EcHypF PGLNEVGVMLPANPLQHLLLQELQCPLVMTSGNLSGKPPAISNEQALEDLQGIADGFLIH 351
StSua5 AGLDTVAVRMPAHPIALQLIRESGVPIAAPSANLAT-RPSPTKAEDVIVDLNGRVDVIID 172
BsYwlC AGLETVAIRMPDHPLALALIRESGLPIAAPSANLSG-KPSPTKAEHVAHDLDGRIAGIMD 177
EcYciO EKRKTIGMRVPSNPIAQALLEALGEPMLSTSLMLPGSEFTESDPEEIKDRLEKQVDLIIH 173
S175 R188
EcTsaC -GETGGRLNPSE-IRDALTG-ELFRQG---------- 190
EcHypF NRDIVQRMDDSV-VR---ESGEMLRRSRGYVPD---------- 380
StSua5 GGHTFFGVE-STIINVTVEPPVLLRPGPFTIEELKKLFGEIVIPE---------- 216
BsYwlC GGPTGIGVE-STVLSCADDIPVLLRPGGITKEQIEAVIGPIHVDKGLSDQNEKPISPGMK 236
EcYciO GGYLGQKP-TTV-IDLTDDTPVVVREG---VGDVKPFL---------- 206