Published in final edited form as:
Proteins. 2008 February 1; 70(2): 311–319. doi:10.1002/prot.21511.

Crystal structure and putative function of small Toprim domain-containing protein from *Bacillus stearothermophilus*

Pavlína Řezáčová[1], **Dominika Borek**[1], **Shiu F. Moy**[2], **Andrzej Joachimiak**[2], and **Zbyszek Otwinowski**[1,*]

1*Department of Biochemistry, UT Southwestern Medical Center, Dallas, Texas*

2*Structural Biology Center and Midwest Center for Structural Genomics, Biosciences Division, Argonne National Laboratory, Argonne, Illinois*

Abstract

The crystal structure of the Midwest Center for Structural Genomics target APC35832, a 14.7-kDa cytosolic protein from Bacillus stearothermophilus, has been determined at 1.3 Å resolution by the single anomalous diffraction method from a mercury soaked crystal. The APC35832 protein is a representative of large group of bacterial and archeal proteins entirely consisting of the Toprim (topoisomerase-primase) domain. This domain is found in the catalytic centers of many enzymes catalyzing phosphodiester bond formation or cleavage, but the function of small Toprim domain proteins remains unknown. Consistent with the sequence analysis, the APC35832 structure shows a conserved Toprim fold, with a central 4-stranded parallel β-sheet surrounded by four α-helixes. Comparison of the APC35832 structure with its closest structural homolog, the catalytic core of bacteriophage T7 primase, revealed structural conservation of a metal binding site and isothermal titration calorimetry indicates that APC35832 binds Mg^{2+} with a sub-millimolar dissociation constant (K_d). The APC35832-Mg^{2+} complex structure was determined at 1.65 Å and reveals the role of conserved acidic residues in Mg^{2+} ion coordination. The structural similarities to other Toprim domain containing proteins and potential function and substrates of APC35832 are discussed in this article.

Keywords

magnesium binding; *isothermal titration calorimetry*; *RNase*; *structural genomics*

Introduction

The Toprim (topoisomerase-primase) domain is present in a wide range of bacterial, archeal, and eukaryotic proteins, catalyzing the formation or cleavage of a phosphodiester bond.[1] The domain consists of about 100 amino acid residues and has two strictly conserved sequence motifs of acidic residues (motifs IV and V). The importance of these residues for catalytic

*Correspondence to: Zbyszek Otwinowski, Department of Biochemistry, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75390-8816. E-mail: E-mail: zbyszek@work.swmed.edu.

activity in a metal-assisted phosphodiester bond cleavage or formation has been documented for many enzymes.[2-5] The Toprim domain is found in DnaG-like primases, DNA topoisomerases (bacterial IA, II, and VI), OLD family nucleases, and RecR family of DNA repair proteins. It is embedded within their catalytic domains and always found in combination with other functional domains, reflecting the different substrate binding requirements and functions of these enzymes.[1] In a number of bacteria and archea, small proteins entirely consisting of the Toprim domain are found. They have the conserved motifs mentioned earlier (Fig. 1), and it was proposed that they may represent a novel class of nucleotidyl transferases or nucleases.[1]

Here we report a crystal structure at 1.3 Å resolution of a small Toprim domain-containing cytosolic protein from *Bacillus stearothermophilus*, target APC35832 of the Midwest Center for Structural Genomics. Based on the high structural similarity with the catalytic core of bacteriophage T7 DNA primase,[7] we recognized a conserved metal-binding site in APC35832. We established the binding of Mg^{2+} to APC35832 by isothermal titration calorimetry (ITC) and determined the APC35832-Mg^{2+} complex structure at 1.65 Å resolution.

MATERIALS AND METHODS

Protein expression and purification

The APC35832 gene was cloned in the pMCSG7 vector[8] and overexpressed in *Escherichia coli* BL21 (DE3). The vector encodes an N-terminal leader sequence consisting of a hexahistidine affinity tag followed by the tobacco etch virus (TEV) protease recognition site.

The cells were grown using Terrific Broth II medium (Q-Biogene) with 0.8% (v/v) glycerol and 20 μg/mL ampicillin at 37°C. Protein expression was induced with 1 m*M* isopropyl-β-D-thiogalactopyranoside at OD_{600} nm ~ 1. After induction, the cells were cultured overnight at 30°C and harvested by centrifugation. Cell paste was resuspended in 5 volumes of lysis buffer (50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 10 m*M* imidazole, and 5% (v/v) glycerol) and cells were lysed by sonication after the addition of protease inhibitors (1 m*M* PMSF, 1 m*M* benzamide). The lysate was clarified by centrifugation and loaded onto Ni^{2+}-nitrilotriacetic acid (Ni-NTA) Superflow resin (Qiagen). The His_6-tagged protein was eluted by a buffer containing 250 m*M* imidazole, 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 10 m*M* imidazole, and 5% (v/v) glycerol.

The affinity tag was then removed during incubation overnight at 25°C with recombinant TEV protease under dialysis against 50 m*M* HEPES pH 7.5, 500 m*M* NaCl, 10 m*M* imidazole, and 5% (v/v) glycerol. In the second Ni-NTA metal-affinity chromatography step the flowthrough fraction containing APC35832 protein minus the affinity tag was collected, dialyzed into 20 m*M* HEPES pH 7.5, and concentrated to 4.0 mg/mL, using a Vivaspin Concentrator (ISC BioExpress), flash-frozen in liquid nitrogen and stored at -8°C.

Protein crystallization

The protein crystals were initially obtained with Wizard I and II crystallization screens (Emerald Biostructures), using the vapor-diffusion technique in hanging drops. After optimization, high quality crystals were obtained under the following conditions: 2 μL of the protein solution was mixed with 1 μL of reservoir solution containing 100 m*M* MES pH 6.5, 26% (w/v) PEG 3350, 200 m*M* ammonium sulfate, and 18% (v/v) glycerol. Crystallization seeds were introduced by streak seeding drops with a cat whisker after touching an APC35832 crystal grown spontaneously under similar conditions. The drop was then equilibrated at 20° C over 1 mL of reservoir solution. Crystals with dimensions 0.2 × 0.1 × 0.1 mm^3, which typically appeared within 3 days, were flash-cooled in liquid nitrogen for data collection.

Crystals of the APC35832-Mg^{2+} complex were soaked in crystallization solution containing 25 m*M* magnesium chloride for 6 h. Heavy atom derivatives were prepared by soaking crystals for 1 h in crystallization solution supplemented with 10 m*M* heavy atom compounds.

Data collection

Diffraction data were collected at 100 K at 19-ID and 19-BM beamlines of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory. The native and single anomalous diffraction (SAD) data were collected at wavelength 0.975 Å on the 19-ID beam line. The native protein crystal diffracted X-rays to a d_{min} of 1.3 Å and the crystal soaked in mercury chloride diffracted X-rays to a d_{min} of 1.6 Å. Data for the protein-Mg^{2+} complex were collected at wavelength 0.979 Å at 19-BM and the crystal diffracted X-rays to a d_{min} of 1.65 Å. All data were processed and scaled within the HKL2000 suite[9]; crystal parameters and data collection statistics are given in Table I.

Structure determination and refinement

The structure was determined by SAD phasing utilizing the anomalous signal from Hg atoms to a d_{min} of 2.0 Å within the program suite HKL3000_PH.[10] The initial model was built automatically using the program ARP/wARP,[11] with 67% of total residues built. This model was then used for molecular replacement against the higher resolution native data using the program MOLREP.[12] Automatic model building within ARP/wARP using the phases obtained from the molecular replacement search yielded 95% of total residues built. Model refinement was carried out using the program REFMAC 5.2[13] from the CCP4 package[14] and interlaced with manual building, using the program Coot.[15] The stereochemistry of the structure was checked with the program Molprobity.[16] The final refinement statistics for the APC35832 metal-free and Mg^{2+}-bound structures are given in Table II. Atomic coordinates and experimental structure factors of APC35832 and the APC35832-Mg^{2+} complex have been deposited to the PDB and are accessible under the codes 2FCJ and 2I5R, respectively.

Isothermal titration calorimetry

Thermodynamic analysis by ITC was performed using VP-ITC (Microcal). Protein APC35832 was extensively dialyzed into dialysis buffer containing 20 m*M* HEPES pH 7.5 and 100 m*M* NaCl. At each injection (a total of 35 times), 8 μL of 5 m*M* $MgCl_2$ (dissolved in dialysis buffer) was injected into a sample cell containing 1.5 mL of 0.2 m*M* APC35832.

Association constants and thermodynamic parameters were fitted to these data, using the program Origin (version 7.0, Microcal). The data reported are the average of two measurements.

RESULTS AND DISCUSSION

Overall structure

The asymmetric unit of the APC35832 crystal structure contains three protein molecules. While most of the 119 amino acid protein chain is well ordered, some of the protein ends are disordered. Following residues are missing in the final model: residues 115-119 of the chain A, residue 1 of chain B, and residues 1-2 and 119 of chain C. The solvent-accessible surface area buried by the contacts between crystallographically-related monomers is in the range of 520-590 $Å^2$ (below 9% of the overall monomer surface area); therefore it is likely that the monomer is the biologically relevant form.

Consistent with the sequence analysis, APC35832 structure shows a conserved Toprim fold that resembles a classic nucleotide-binding fold, with a central 4-stranded parallel β-sheet surrounded by four α-helixes [Fig. 2(A)]. Structural homologs in the Brookhaven Protein Data

Bank identified using a DALI search[17] are catalytic domains of bacterial and bacteriophage primases and bacterial and archeal DNA topoisomerases.

The closest structural homolog, the Toprim domain of bacteriophage T7 primase (PDB code 1NUI),[7] superimposes on the APC32832 structure with an rmsd of 2.4 Å for 100 C-α-atoms. Invariant acidic residues from motifs IV and V superimpose very nicely in both structures [Fig. 2(B)]. These residues are located in loops connecting secondary structure elements, and cluster into an acidic surface patch.

In T7 primase, this patch is part of the RNA-synthesis active site located in the shallow cleft between the Toprim and an adjacent N-terminal part of the RNA-binding domain. In the single domain APC35832, the stand-alone acidic patch forms a concave surface pocket about 9 Å long and 7.5 Å wide, with a protruding Glu88 side chain on one side [Fig. 2(C)]. Interestingly, during the model building, non-protein density was observed near this acidic surface patch in one monomer of the three present in the asymmetric unit. The high quality of the electron density helped to identify the ligand as a molecule of 2-(*N*-Morpholino)-ethanesulfonic acid (MES), which was used as a buffer during crystallization in 100 m*M* concentration. When crystals were prepared in the absence of MES buffer (replaced by sodium acetate), this electron density was not present.

Mg^{2+} Binding

Side chains of conserved acidic residues of motifs IV and V coordinate metal ions, presumably Mg^{2+}, in the T7 primase crystal structure,[7] *E. coli* DnaG primase,[18] and archeal topoisomerase VI.[19] To determine if the metal-binding function is also conserved in APC35832, the binding of Mg^{2+} was analyzed using ITC. ITC experiments were carried out at pH 7.5, in 20 m*M* HEPES, 100 m*M* NaCl at 27°C. A typical binding profile is shown in Figure 3. The ion binding event was shown to be endothermic, as the upward peaks [Fig. 3(A)] indicate heat uptake on the injection of ligand into the protein solution. The exothermic (downward) peaks reflect the dilution heat of magnesium chloride [Fig. 3(B)] and were omitted during data analysis. Weak enthalpy of Mg^{2+} binding accounted for ITC data with low signal-to-noise ratio, nevertheless, the integrated data could be fitted to a single site model of binding [Fig. 3(C)]. The fit provided estimate of dissociation constant in sub-millimolar range (K_d = 106 ± 31 μ*M*) for binding of one Mg^{2+} by APC35832 monomer.

The binding of one Mg^{2+} was further confirmed by the crystal structure of APC35832-Mg^{2+} complex. APC35832 crystals were soaked with 25 m*M* $MgCl_2$ and the structure of the APC35832-Mg^{2+} complex was determined at a resolution of 1.65 Å. In this structure, two out of the three molecules present in the asymmetric unit bind a single Mg^{2+} ion each. The binding of Mg^{2+} by the third molecule is obstructed by the binding of the MES molecule and the contacts with neighboring, crystal symmetry-related molecules.

The Mg^{2+} in APC35832 is coordinated by side chains of two invariant residues of motif V, Asp58, Asp60, and by a conserved Glu88. In addition, three water molecules complete the octahedral coordination of the Mg^{2+} ion [Fig. 4(A,B)]. The acidic residues of motif IV are not part of the inner sphere coordination of the Mg^{2+}, however the side chain of Glu11 forms a hydrogen bond with one of the water molecules that coordinate the Mg^{2+} ion.

The structures of the unliganded and the Mg^{2+}-bound proteins are highly similar, with an rmsd of 0.54 Å for 1422 main-chain atoms and 1.26 Å for 1459 side-chain atoms. Structural changes are primarily in the vicinity of the metal ion-binding site. The bound Mg^{2+} induces a local rearrangement in side-chain positions of residues that coordinate it, associated with a minor positional change of main-chain atoms of residues 57-59 and 81-89.

In the T7 DNA primase structure, conserved acidic residues coordinate two Mg^{2+} ions [Fig. 4 (C)]. The first Mg^{2+} is coordinated by Glu157 (motif IV) and two aspartates of motif V. The same conserved triad coordinates two metal ions in *E. coli* DnaG primase[18] and a single metal ion in archeal topoisomerase VI[19] structures. APC35832 coordinates a Mg^{2+} ion with two conserved aspartates from motif V and the conserved Glu88. In T7 primase, the residue at this position is an aspartate (Asp237) that is conserved in the primase, but not in all Toprim domains. This aspartate binds a second Mg^{2+} in T7 primase [Fig. 4(C)] and DnaG primase.[21,22] The two ions in the primases are reminiscent of those required in the two metal-ion catalytic mechanism for nucleotide polymerization.[23]

As opposed to primases, our structure shows the presence of only one metal-binding site, composed of motif V aspartates and an acidic residue analogous to the primase motif VI aspartate. Also, the results of ITC experiments suggested that the Toprim domain of APC35832 by itself binds a single Mg^{2+} ion.

The putative catalytic site and ligands

A single crystal structure of the Toprim domain-containing protein from the thermophilic bacteria *Aquifex aeolicus* has been deposited in the PDB by the Midwest Center for Structural Genomics, with the accession code 1T6T. The sequence identity with APC35832 is only 16% [sequence alignment in Fig. 5(D)], but the structures superimpose quite well, with an overall rmsd of 3.5Å. The overall shape of both molecules is similar, but the *Aquifex aeolicus* structure has an extra N-terminal helix occupying the space of the C-terminal helix in APC35832. In addition, the β4 strand of APC35832 is replaced by a shorter strand followed by an α-helix in the 1T6T structure. In spite of the differences, the conserved residues of motifs IV and V superimpose noticeably well in both structures [Fig. 5(C)].

A projection of residue conservation among 44 small, Toprim domain-containing protein sequences from bacteria and archea onto the APC35832 surface shows that the most conserved residues are grouped around the surface acidic patch composed of acidic residues of the primase motif IV and V and the Mg^{2+} binding site [Fig. 5(A)]. Thus this region of the protein represents the putative catalytic site.

T7 primase binds ATP or CTP in two nucleoside triphosphate binding sites,[24,25] and the acidic residues of motifs IV and V were shown to be involved in nucleotide binding.[26] We found in this crystal form that APC35832 can accommodate a MES molecule in the putative catalytic site [Fig. 2(C)]. Interestingly, three conserved residues of motif IV and V (Glu11, Asp58, and Asp60) are involved in direct or water-mediated contacts with the buffer molecule. We might therefore presume that the MES molecule, structurally very similar to a ribose-5-phosphate molecule, can mimic the nucleotide substrate.

We took the position of the MES molecule as a model for a ribose-5-phosphate moiety and docked various nucleotides into the putative catalytic site of APC35832. An example of a very good fit for an adenine monophosphate molecule in the conserved pocket is demonstrated in Figure 5(B). The position of the modeled nucleotide is such that the phosphate group might be able to coordinate the Mg^{2+} ion [white spheres in Fig. 5(B)].

To support our model with structural data, we soaked crystals of APC35832 with various nucleotides (mono-phosphates and triphosphates) in the presence or absence of $MgCl_2$, but no binding was observed. We also tested nucleotide binding to APC35832 in solution by ITC, but we have not obtained any notable interaction heat changes for any of the nucleotides.

The putative function

Comparison of the APC35832 structure with its closest structural homolog, the catalytic core of bacteriophage T7 primase,[7] revealed conservation of residues involved in catalytic activity, including the metal binding site. We have shown that APC35832 binds a Mg^{2+} ion that may possibly be used in catalysis, but we have not detected any binding of a nucleotide substrate to APC35832. So far, the Toprim domain in all characterized enzymes was found exclusively in combination with other domains, embedded in different structural environments reflecting the different substrate binding requirements and catalytic functions of these enzymes. A standalone Toprim domain would most likely need a protein interaction partner to function as a nucleotidyl transferase. However, one should also consider the possibility of folded RNA being a structural partner of small, Toprim domain proteins.

Among homologous proteins with characterized function, the one with the most similar sequence to APC35832 is RNase M5, a somewhat larger protein with an extra C-terminal region [Fig. 5(D)]. This 190-amino acid enzyme catalyzes the maturation of 5S ribosomal RNA in low G + C gram-positive bacteria, and its N-terminal domain is predicted to have Toprim domain structure similar to the DnaG core.[27] Residues critical for enzymatic activity of RNase M5 are also conserved in APC35832. In fact, most of them belong to residues conserved amongst all the members of the Toprim domain family (motifs IV and V), but some are specific for RNases M5 [Fig. 5(D)]. In addition to a Toprim domain, RNase M5 has a C-terminal domain with unknown structure, whose arginine-rich region (around residue 140) was shown to participate in RNA binding.[27] This domain is not present in APC35832, but all other residues involved in RNA binding by RNase M5 (Gln10, Gly61, Ile64, and Arg65) have their equivalents in APC35832 (Gln11, Gly63, Leu66, and Arg67). Another similarity of APC35832 and RNase M5 is the binding of Mg^{2+}, which is required for ribonuclease function. The binding site of Mg^{2+} in RNase M5 and its role in the cleavage reaction is not yet known, but many RNases use Mg^{2+} as a direct participant in the cleavage reaction. For example, the structure of RNase III[28] revealed single metal ion in the catalytic site coordinated in a very similar way as the Mg^{2+} in the APC35832 structure, and the proposed RNA cleavage mechanism involves formation of the second Mg^{2+} binding site upon substrate binding.[29]

Thus, a potential function of ACP35832 as a ribonuclease seems reasonable. If small Toprim domain proteins recognize RNA molecules of a particular shape, this could explain the preservation of the overall protein shape, even in very divergent proteins, both in sequence and evolutionary distance. Structural comparison with a member of a divergent branch of a small, Toprim domain protein family showed strong preservation of the protein shape despite a very low sequence homology.

CONCLUSIONS

This is the first extended structural study of a small, Toprim domain-containing protein, a member of a family present both in bacteria and archea. The structures of protein APC35832 from B. *stearothermophilus* reveal the binding of Mg^{2+} ion in its putative catalytic site. Phylogenetic analysis, structural comparisons, metal binding and the position of conserved acidic residues imply that the APC35832 protein has catalytic activity involved in phosphodiester bond formation or cleavage, with the most likely substrate being an RNA molecule. A potential protein or RNA interaction partner, and substrates of the APC35832 protein, are yet to be identified.

Our work sets the base for functional and biochemical characterization of a large group of bacterial and archeal proteins with a well-known fold but of as yet unknown function.

ACKNOWLEDGMENTS

The authors wish to thank members of the Structural Biology Center at Argonne National Laboratory for their help in data collection at 19-ID and 19-BM beamlines and to Celestine Thomas for his help with ITC experiments.

Grant sponsor: National Institutes of Health; Grant number: GM074942.

REFERENCES

1. Aravind L, Leipe DD, Koonin EV. Toprim—a conserved catalytic domain in type IA and II topoisomerases. DnaG-type primases OLD family nucleases and RecR proteins. Nucleic Acids Res 1998;26:4205–4213. [PubMed: 9722641]
2. Strack B, Lessl M, Calendar R, Lanka E. A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the alpha protein of the *Escherichia coli* satellite phage P4. J Biol Chem 1992;267:13062–13072. [PubMed: 1618804]
3. Dracheva S, Koonin EV, Crute JJ. Identification of the primase active site of the herpes simplex virus type 1 helicase-primase. J Biol Chem 1995;270:14148–14153. [PubMed: 7775476]
4. Chen SJ, Wang JC. Identification of active site residues in *Escherichia coli* DNA topoisomerase I. J Biol Chem 1998;273:6050–6056. [PubMed: 9497321]
5. Zhu CX, Roche CJ, Papanicolaou N, DiPietrantonio A, Tse-Dinh YC. Site-directed mutagenesis of conserved aspartates, glutamates and arginines in the active site region of *Escherichia coli* DNA topoisomerase I. J Biol Chem 1998;273:8783–8789. [PubMed: 9535856]
6. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res 2003;31:365–370. [PubMed: 12520024]
7. Kato M, Ito T, Wagner G, Richardson CC, Ellenberger T. Modular architecture of the bacteriophage T7 primase couples RNA primer synthesis to DNA synthesis. Mol Cell 2003;11:1349–1360. [PubMed: 12769857]
8. Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr Purif 2002;25:8–15. [PubMed: 12071693]
9. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
10. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr 2006;62(Part 8):859–866. [PubMed: 16855301]
11. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 1999;6:458–463. [PubMed: 10331874]
12. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr 2000;56(Part 12):1622–1624. [PubMed: 11092928]
13. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53(Part 3):240–255. [PubMed: 15299926]
14. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50 (Part 5):760–763. [PubMed: 15299374]
15. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60(Part 12 No 1):2126–2132. [PubMed: 15572765]
16. Lovell SC, Davis IW, Arendall WB III, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins 2003;50:437–450. [PubMed: 12557186]
17. Holm L, Sander C. Searching protein structure databases has come of age. Proteins 1994;19:165–173. [PubMed: 7937731]
18. Keck JL, Roche DD, Lynch AS, Berger JM. Structure of the RNA polymerase domain of *E. coli* primase. Science 2000;287:2482–2486. [PubMed: 10741967]

19. Nichols MD, DeAngelis K, Keck JL, Berger JM. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. EMBO J 1999;18:6177–6188. [PubMed: 10545127]

20. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 1995;8:127–134. [PubMed: 7630882]

21. Godson GN, Schoenich J, Sun W, Mustaev AA. Identification of the magnesium ion binding site in the catalytic center of *Escherichia coli* primase by iron cleavage. Biochemistry 2000;39:332–339. [PubMed: 10630993]

22. Urlacher TM, Griep MA. Magnesium acetate induces a conformational change in *Escherichia coli* primase. Biochemistry 1995;34:16708–16714. [PubMed: 8527445]

23. Steitz TA. A mechanism for all polymerases. Nature 1998;391:231–232. [PubMed: 9440683]

24. Kusakabe T, Richardson CC. Gene 4 DNA primase of bacteriophage T7 mediates the annealing and extension of ribo-oligonucleotides at primase recognition sites. J Biol Chem 1997;272:12446–12453. [PubMed: 9139692]

25. Frick DN, Kumar S, Richardson CC. Interaction of ribonucleoside triphosphates with the *gene 4* primase of bacteriophage T7. J Biol Chem 1999;274:35899–35907. [PubMed: 10585475]

26. Lee SJ, Richardson CC. Acidic residues in the nucleotide-binding site of the bacteriophage T7 DNA primase. J Biol Chem 2005;280:26984–26991. [PubMed: 15917241]

27. Allemand F, Mathy N, Brechemier-Baey D, Condon C. The 5S rRNA maturase, ribonuclease M5, is a Toprim domain family member. Nucleic Acids Res 2005;33:4368–4376. [PubMed: 16077031]

28. Blaszczyk J, Tropea JE, Bubunenko M, Routzahn KM, Waugh DS, Court DL, Ji X. Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. Structure 2001;9:1225–1236. [PubMed: 11738048]

29. Sun W, Pertzev A, Nicholson AW. Catalytic mechanism of *Escherichia coli* ribonuclease III: kinetic and inhibitor evidence for the involvement of two magnesium ions in RNA phosphodiester hydrolysis. Nucleic Acids Res 2005;33:807–815. [PubMed: 15699182]

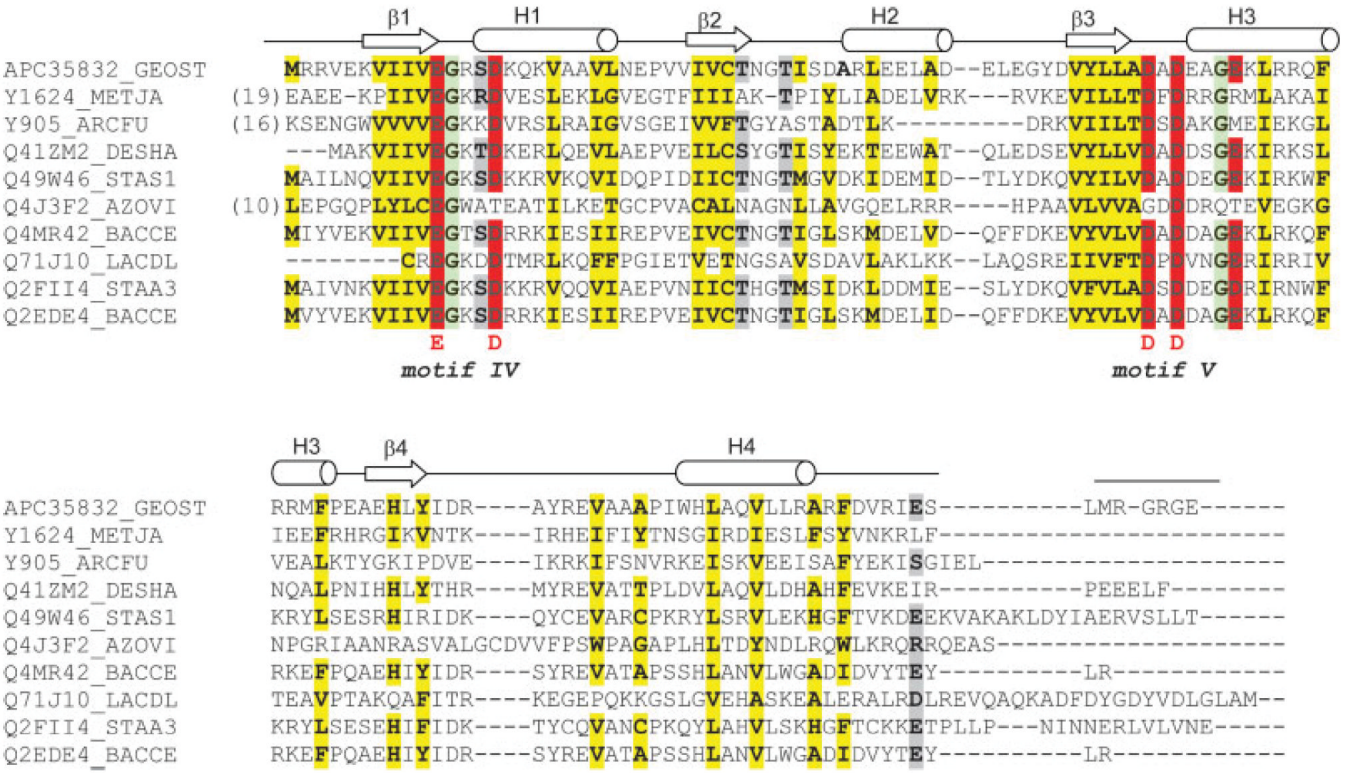

Figure 1.

Multiple sequence alignment of small, Toprim domain-containing proteins. Sequences from 44 proteins from the Swiss-Prot database[6] were aligned, and the 10 most similar to APC35832 are shown. Conserved acidic residues from motifs IV and V are marked. The color coding for residues that are conserved in at least 80% of all aligned sequences is red for negatively charged, yellow for hydrophobic, green for small, and gray for polar amino acids. The secondary structure elements are derived from the APC35832 structure. The protein identifiers are followed by an abbreviation of source organism name: *GEOST*, Geobacillus stearothermophilus; *METJA*, Methanococcus jannaschii; *ARCFU*, Archaeoglobus fulgidus; *DESHA*, Desulfitobacterium hafniense; *STAS1*, Staphylococcus saprophyticus; *AZOVI*, Azotobacter vinelandii; *BACCE*, Bacillus cereus G9241; *LACDL*, Lactobacillus delbrueckii *subsp*. lactis; *STAA3*, Staphylococcus aureus (*strain USA300*).

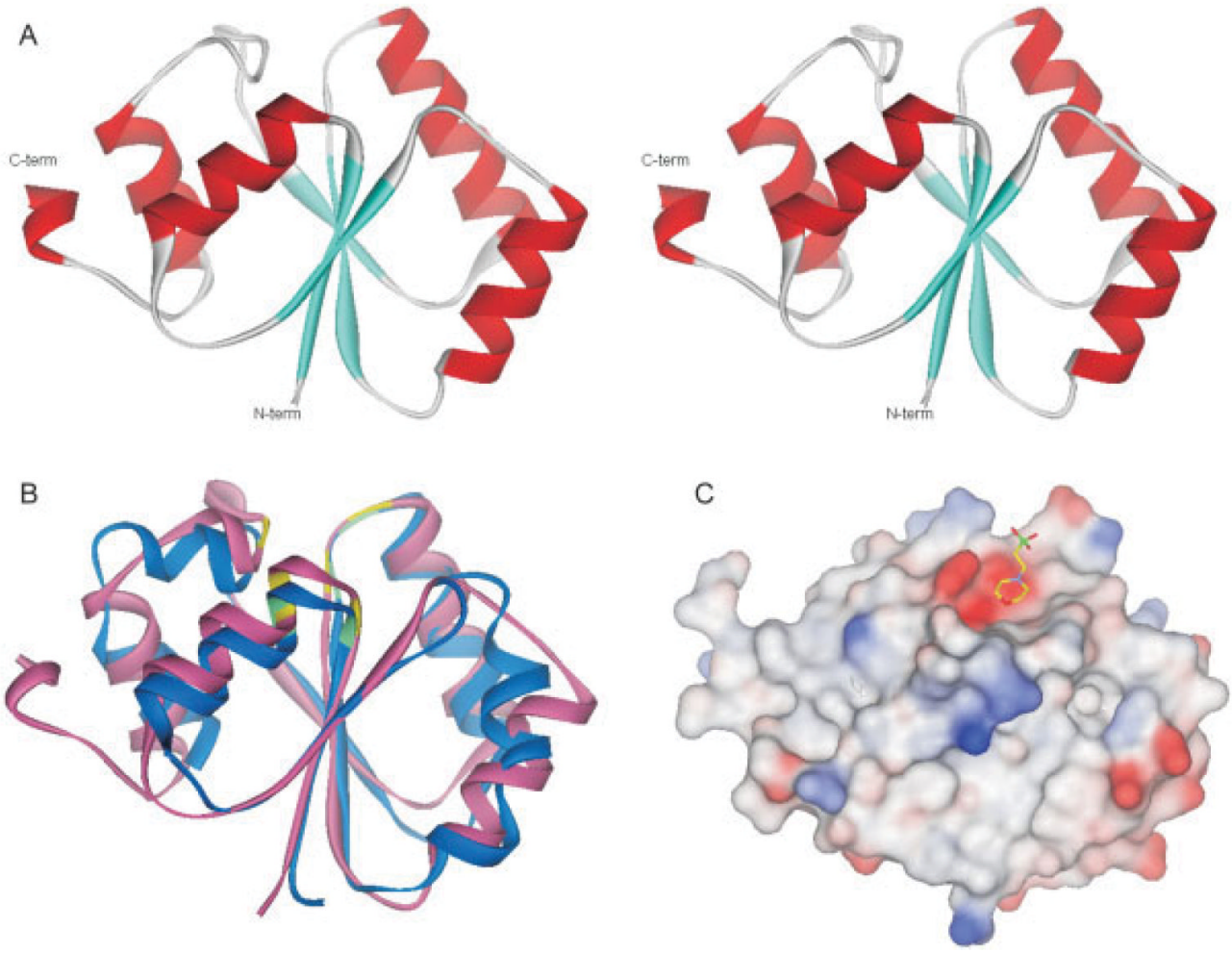

Figure 2.

A: A stereo diagram showing the overall APC35832 structure. **B:** Superimposition of APC35832 and the Toprim domain of T7 DNA primase (PDB entry 1NUI), showing conserved acidic residue positions. APC35832 is colored pink, 1NUI is blue and conserved acidic residues are highlighted in green and yellow. **C:** The solvent-accessible surface colored according to electrostatic potential (blue positive, red negative, calculated with DS ViewerPro 6.0 [Accelerys Software]). The acidic surface pocket is visible, with the MES molecule shown as sticks.

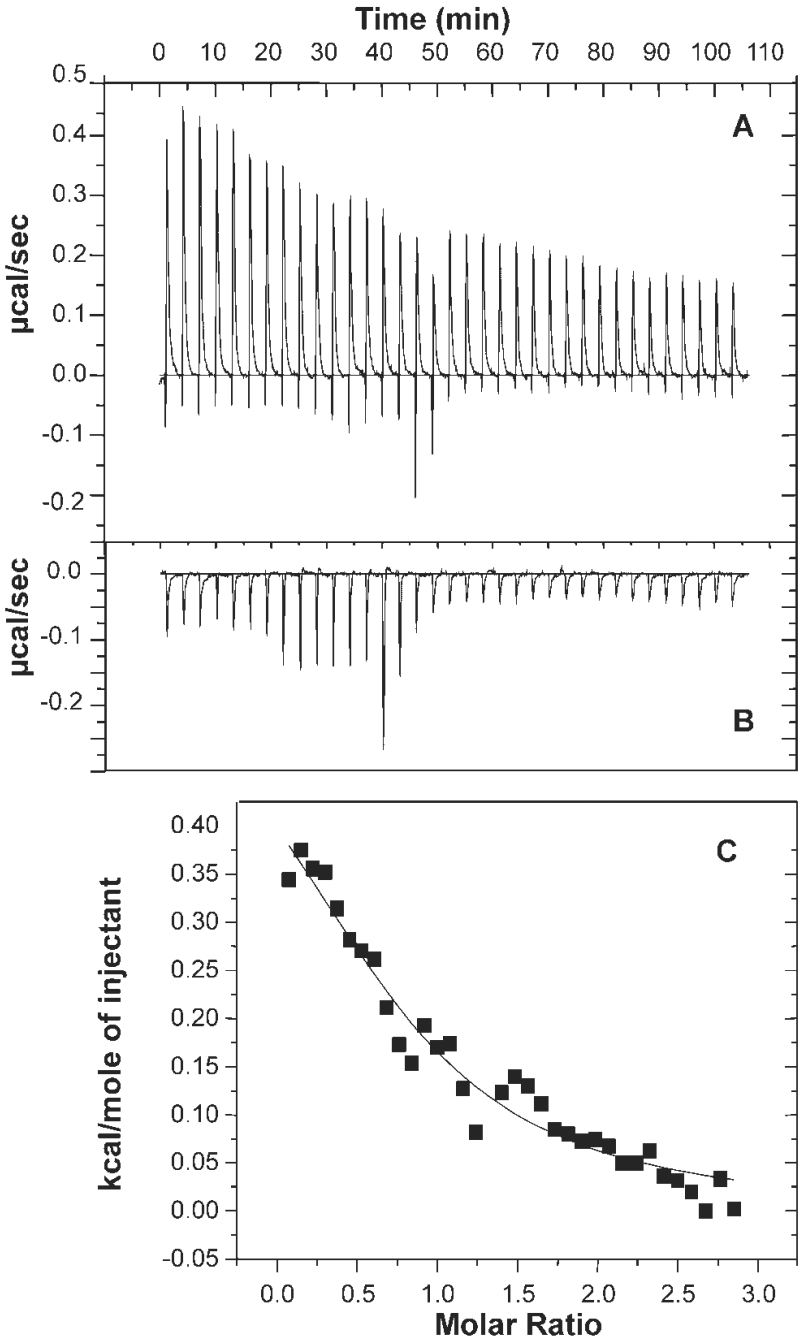

Figure 3.

Isothermal titration of APC35832 with Mg^{2+} ion. **A:** Experimental data for titration of 200 µM APC35832 with 3 mM Mg^{2+}. **B:** Experimental data for negative control experiment, injections of 3 mM Mg^{2+} into protein buffer. **C:** Integrated ITC data injection (full squares) with the fit to a one-site model (line).

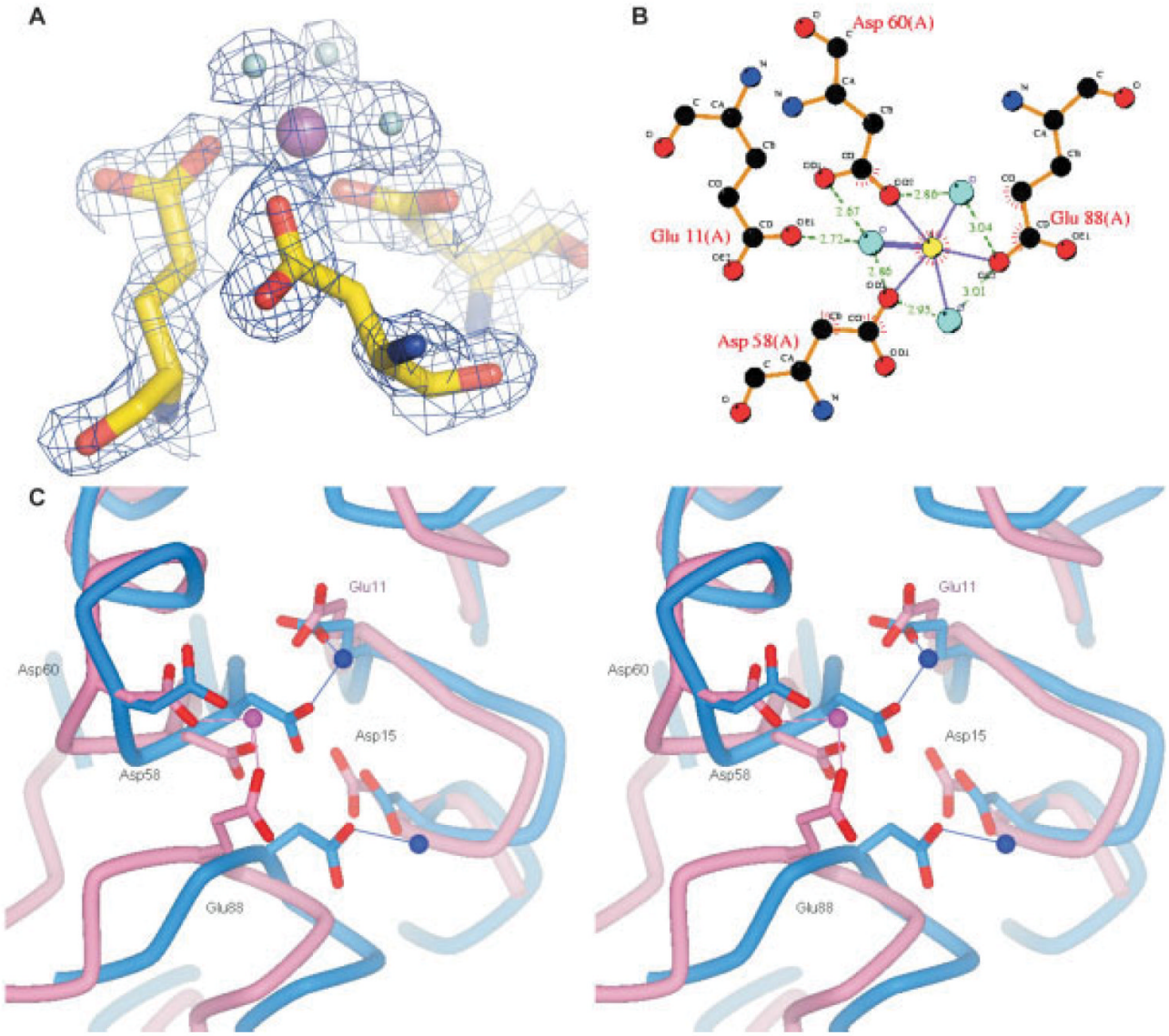

Figure 4.

A: Detail of octahedral coordination of Mg^{2+} ion with $2F_o$-F_c electron density map contoured at the 1.3σ level of the map. **B:** Schematic of Mg^{2+} ion binding by APC35832. Mg^{2+} is shown as a yellow sphere, water molecules are represented with turquoise spheres, ion coordination is shown by violet lines, and hydrogen bonds as green dotted lines with their length inÅ. Figure was prepared using the program Ligplot.[20] **C:** A stereo view of the superimposed Mg^{2+} binding sites in APC35832 (pink) and T7 primase, PDB entry 1NUI (blue), Mg^{2+} ions are colored accordingly. Residues numbers are for the APC35832 protein.

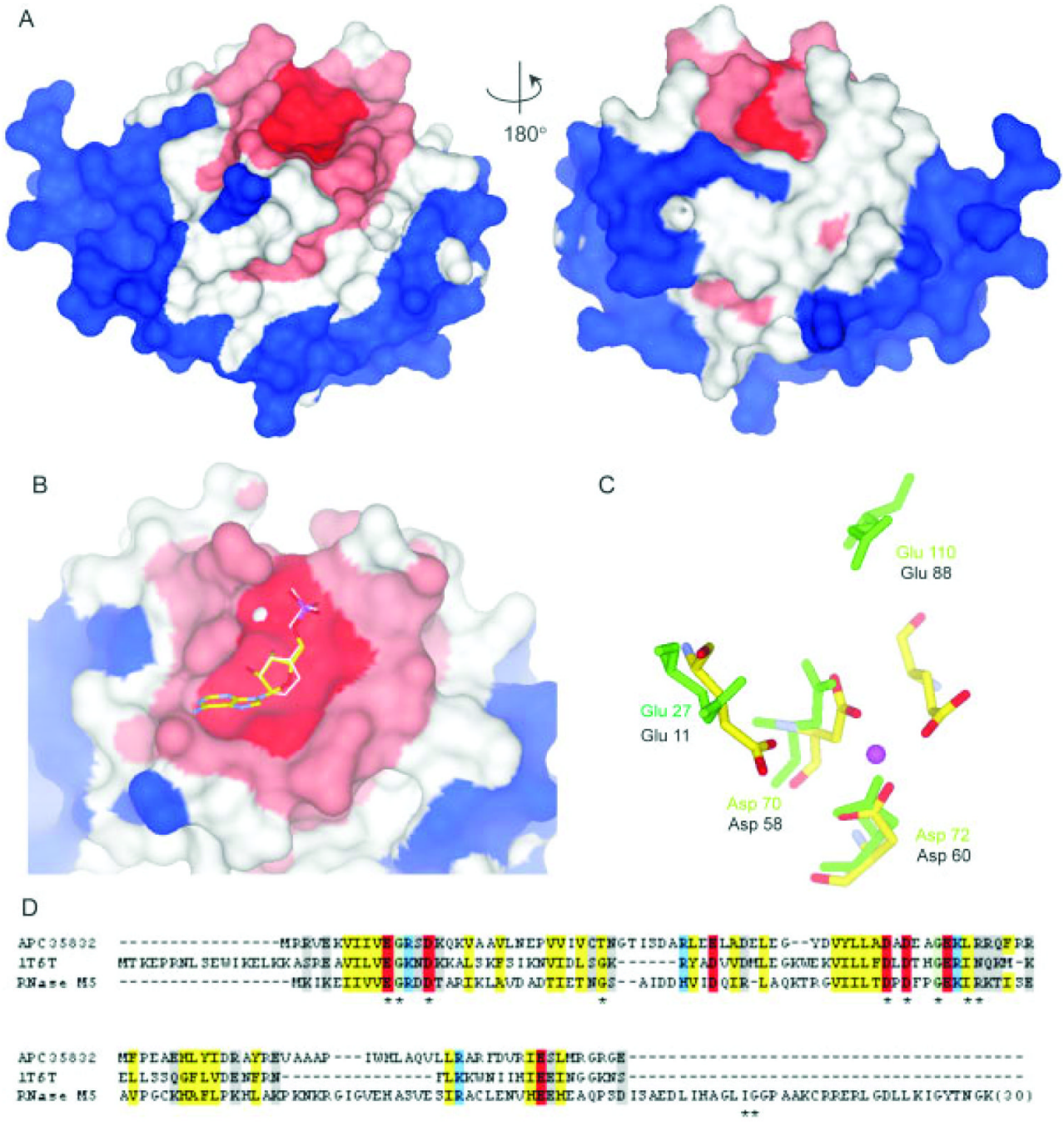

Figure 5.

A: Solvent-accessible surface of the APC35832 Toprim domain showing distribution of conserved residues among small, Toprim domain-containing proteins. The surface is colored from blue to red, with blue representing sequence identity 40% or lower and red corresponding to 90% sequence identity. **B:** A stick model of AMP in the APC35832 putative catalytic site. The position of ribose-5-phosphate is based on the position of an MES molecule found in the crystal structure (white line). The Mg^{2+} ion from the crystal structure is represented by a white sphere. The protein surface is colored by residue conservation and electrostatic potential, as calculated within DS ViewerPro 6.0 (Accelerys Software, red for negative, blue for positive). **C:** Positions of conserved residues in APC35832 (yellow carbons) and the 1T6T structure

(green). The Mg^{2+} ion bound in the APC35832 structure is shown as a magenta sphere, black numbers designate residues of APC35832 and green numbers designate residues of 1T6T. **D:** Sequence alignment of APC35832, 1T6T and B. subtilis RNase M5. Residues required for RNase M5 function are marked with asterisks below the sequence.

Table I

Data Collection Statistics

Crystal	**Native**	**$HgCl_2$ soaked**	**Mg^{2+} complex**
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a, b, c (Å)	36.6, 72.7, 116.3	36.8, 72.9, 115.7	36.6, 73.0, 115.0
Molecules per asymmetric unit	3	3	3
Wavelength (Å)	0.975	0.975	0.979
Resolution range (Å)	50.0-1.3 (1.32-1.30)[a]	50.0-1.6(1.66-1.60)[a]	50.00-1.65(1.71-1.65)[a]
No. of unique reflections	77,810	42,018	38,256
Completeness (%)	99.5 (94.4)[a]	98.8 (95.7)[a]	99.4 (96.6)[a]
R-merge[b]	3.2 (53.6)[a]	4.1 (36.8)[a]	4.8 (37.9)[a]
Average //σ(/)	76.8 (2.0)[a]	39.4 (2.4)[a]	27.7 (2.3)[a]

[a] In parentheses are statistics for the highest resolution shell.

[b] *R*-merge = $(|I_{hkl} - \langle I \rangle|)/I_{hkb}$ where the average intensity $\langle I \rangle$ is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for any given reflection.

Table II

Refinement Statistics

Structure	APC35832	APC35832-Mg^{2+} complex
Resolution range (Å)	50.0-1.3 (1.33-1.30)[a]	50.0-1.65 (1.7-1.65)[a]
No. of reflections in working set	75,438 (5,303)[a]	36143 (2,602)[a]
No. of reflections in test set	1,538 (95)[a]	1903 (141)[a]
R value (%)[b]	19.2 (27.6)[a]	17.1 (19.7)[a]
R-free value (%)[c]	21.4 (31.5)[a]	20.8 (25.0)[a]
RMSD deviation from ideal geometry		
Bond length (Å)	0.012	0.012
Angle (°)	1.16	1.41
Number of atoms in au	3352	3263
Protein atoms	2924	2991
Water molecules	363	211
Mg^{2+} ions	N/A	2
Sulfate ions	2	2
MES	1	1
Glycerol	2	6
Mean *B* value ($Å^2$)	23.61	21.24
Ramachandran plot statistics		
Residues in favored regions	98.60%	98.55%
Residues in allowed regions	100.00%	100.00%

[a] In parentheses are statistics for the highest resolution shell.

[b] *R*-value = $||F_O|- |F_C||/|F_O|$, where F_O and F_C are the observed and calculated structure factor, respectively.

[c] *R*-free is equivalent to *R* value but is calculated for 5% of the reflections chosen at random and omitted from the refinement process.