Structure of the N-Terminal Domain of *Escherichia coli* Glutamine Synthetase Adenylyltransferase

Yibin Xu,[1,*] Rongguand Zhang,[4] Andrzej Joachimiak,[4]
Paul D. Carr,[2] Thomas Huber,[3]
Subhash G. Vasudevan,[1,5] and David L. Ollis[2]
[1]Department of Biochemistry and Molecular Biology
James Cook University
Townsville, Queensland 4811
Australia
[2]Research School of Chemistry
Australian National University
Building 35, Science Road
Canberra, ACT0200
Australia
[3]Department of Mathematics
The University of Queensland
Brisbane, Queensland 4072
Australia
[4]Structural Biology Center
Argonne National Laboratory
Argonne, Illinois 60439

Summary

We report the crystal structure of the N-terminal domain of *Escherichia coli* adenylyltransferase that catalyzes the reversible nucleotidylation of glutamine synthetase (GS), a key enzyme in nitrogen assimilation. This domain (AT-N440) catalyzes the deadenylylation and subsequent activation of GS. The structure has been divided into three subdomains, two of which bear some similarity to kanamycin nucleotidyltransferase (KNT). However, the orientation of the two domains in AT-N440 differs from that in KNT. The active site of AT-N440 has been identified on the basis of structural comparisons with KNT, DNA polymerase β, and polyadenylate polymerase. AT-N440 has a cluster of metal binding residues that are conserved in polβ-like nucleotidyl transferases. The location of residues conserved in all ATase sequences was found to cluster around the active site. Many of these residues are very likely to play a role in catalysis, substrate binding, or effector binding.

Introduction

Glutamine synthetase adenylyltransferase (EC 2.7.7.42) (ATase; systematic name, ATP:[L-glutamate:ammonia ligase {ADP-forming}] adenylyltransferase) is part of the regulatory machinery of glutamine synthetase, GS. This latter enzyme is ubiquitous and catalyzes the ATP-dependent conversion of glutamate to glutamine. In most bacteria, this reaction can be used to regulate ammonia assimilation, and they have evolved rather sophisticated machinery to regulate GS activity. This regu-

*Correspondence: yibin.xu@jcu.edu.au
[5]Present address: Novartis Institute for Tropical Diseases, 1 Science Park Road, #04-14 The Capricorn, Singapore Science Park II, Singapore 117528.

lation is best characterized in *Escherichia coli*, where ATase was first identified. The regulatory machinery in *E. coli* consists of a series of proteins that sense the nitrogen level in the cell and regulate both the transcription and catalytic activity of GS (Magasanik, 1993). ATase is the component of this machinery that directly affects the activity of GS. It is a bifunctional enzyme that catalyzes the adenylylation and deadenylylation of GS. A detailed review of the importance of GS in the regulation of nitrogen uptake along with a description of the regulatory proteins can be found elsewhere (Merrick and Edwards, 1995; Ninfa and Atkinson, 2000; Arcondeguy et al., 2001). The following discussion will only refer to those components of the regulatory machinery that are relevant to ATase activity.

The two activities of ATase are carefully regulated by the PII protein as well as the low molecular mass cofactors glutamine and 2-oxoglutarate. The PII protein is the nitrogen indicator of the cell, with PII-UMP indicating low nitrogen levels while unmodified PII indicates nitrogen excess. PII is modified by uridylyltransferase (UTase), which can both add UMP to, or remove UMP from, PII. In the cell, glutamine accumulates in high-nitrogen conditions while 2-oxoglutarate accumulates under low-nitrogen conditions. The adenylylation activity of ATase is stimulated by PII and glutamine, while its deadenylylation activity is stimulated by PII-UMP and 2-oxoglutarate (Engleman and Francis, 1978).

Full-length ATase has 945 amino acids containing two duplicated homologous domains that share nearly 25% sequence identity when residues 1–411 are aligned with residues 523–945 (Jaggi et al., 1997). The two activities of ATase reside on two separate domains (Jaggi et al., 1997). The C-terminal domain (AT-C) catalyzes the adenylylation of GS, decreasing its activity. The activity of isolated AT-C is independent of PII. The N-terminal domain (AT-N440) catalyzes the deadenylylation of GS-AMP and results in an increasing activity of GS. The AT-N440 domain, which is the subject of this work, consists of residues 1 through 440. The expression and activity of this domain have been described in a previous manuscript (Xu et al., 2004). The domain appears to be stable in solution and has a lower level of the deadenylylation activity than the intact ATase. In addition, AT-N440 does not appear to be regulated by PII-UMP or 2-oxoglutarate. The reaction catalyzed by AT-N440 is a phosphorolysis reaction and is similar to that catalyzed by AT-C, as shown below:

$$GS\text{-}AMP + P_i \leftrightarrow GS + ADP \ (AT\text{-}N440)$$

$$GS + ATP \leftrightarrow GS\text{-}AMP + P\text{-}P_i \ (AT\text{-}C).$$

The reaction catalyzed by AT-N440 does not give the starting reagents of the AT-C reaction. ADP is produced by AT-N440, while ATP is the starting material of AT-C. The reaction catalyzed by AT-C is clearly similar to that catalyzed by DNA polymerases, with the difference that DNA polymerases use DNA as a substrate, while AT-C uses the protein GS. In the same sense, the reaction

Table 1. Crystallographic Data for the AT-N440 Structure Analysis

Parameters	Se Remote	Se Peak	Se Inflection
Wavelength (Å)	0.95372	0.97844	0.97855
Resolution (Å)	50–2.0	50–2.0	50–2.0
Total reflections	1304799	1028551	1295596
Unique reflections	36424	36083	36358
Completeness (%)	100 (100)[a]	99.7 (99.7)	99.9 (100)
R_{merge} (%)[b]	8.3 (51.7)	9.9 (50.2)	10.2 (51.8)
I/σ (I)	8.9 (5.2)	8.2 (3.7)	8.2 (4.6)
Refinement			
Resolution (Å)	50–2.0		
Reflections (working/test)	34466/1815		
R_{work}/R_{free} (%)	22.69/26.75		
Rmsd			
Bond lengths (Å)	0.011		
Bond angles (°)	1.072		

[a]Values in parentheses are for the outer shell of data 2.07–2.0 Å.
[b]$R_{merge} = \Sigma|I - <I>|/\Sigma I$, where I is the observed intensity and $<I>$ is the average intensity obtained from multiple observations of symmetry-related reflections, and the summations are over all Miller indices.

catalyzed by AT-N440 is similar to the reverse reaction that can also be catalyzed by the DNA polymerases; however, this reaction occurs less frequently in vivo due to the low concentration of PP_i resulting from the action of inorganic pyrophosphatase. Given the mechanistic similarities, it is not surprising that active site residues of DNA polymerase β (Polβ) are highly conserved in kanamycin nucleotidyltransferase (KNT) and other protein nucleotidyltransferases such as ATase and UTase (Holm and Sander, 1995a; Aravind and Koonin, 1999). ATase, like DNA polymerase, is capable of removing a nucleoside monophosphate. DNA polymerases often possess a hydrolytic (nuclease) domain or subunit to edit the polymerase reaction. The reaction catalyzed by AT-N440 is unusual in the sense that it is mechanistically similar to a polymerase reaction (adenylyl transfer) but functionally equivalent to an editing protein (nuclease).

Previous studies have detected a low level of sequence similarity (∼25%) between AT-N440 and AT-C, and the sequences of both domains can be aligned with DNA polymerase β. The sequence similarity is concentrated in the active site and is not significant over the entire sequences. A similar level of sequence similarity has been found between DNA polymerase β and UTase. There are a number of other proteins that show a similarity to DNA polymerase β (Holm and Sander, 1995a). Of these, structures are available for kanamycin nucleotidyltransferase (PDB code 1kny) (Pedersen et al., 1995), DNA polymerase β (PDB code 1bpy) (Sawaya et al., 1997), and polyadenylate polymerase (PAP; PDB code 1vfr) (Bard et al., 2000; Martin et al., 2000). The structural and sequence similarity is concentrated in a domain that is known as the "palm" domain in various polymerase structures. This is domain 2 of the current structure.

We have undertaken the structure determination of the AT-N440 domain in order to better understand how it functions. The structure is described and compared with other proteins that also contain the nucleotidylyl transferase fold. The active site is also examined with a view of identifying the residues important for binding substrates and facilitating the phosphorolysis reaction. We have scanned a sequence database and used sequence alignments to identify highly conserved residues. The location of these residues is used to better understand how the protein might function.

Results

Molecular Structure

The structure of AT-N440 was determined by the MAD phasing technique and refined to 2.0 Å resolution. The refinement statistics and model quality parameters are listed in Table 1. Not all amino acids could be built into electron density. As a consequence, there is a break in the center of the molecule between residues 182 and 191, and the last three residues, 438–440, are also not observed. In terms of linear sequence, the break in the center of the molecule is close to key catalytic residues, but this section of the peptide does not appear to form part of the active site.

The overall shape of AT-N440 is discoidal with major axes of approximately 63 × 54 × 26 Å. A schematic representation of the structure is shown in Figure 1. The molecule is predominantly helical with a small amount of β structure that forms one side of a cleft in the center of the molecule. This cleft is thought to be part of the active site of the molecule, an idea that is consistent with the structural comparisons described below. For the purposes of these comparisons, it is convenient to divide the structure into three subdomains: the first is formed by residues 1 through 110, while the second is formed by residues 111 through to 283, and the third is formed by residues 284 to 437. As can be seen in Figure 1, each of the three domains interacts with the other two to form a single globular structure.

The first four helices of domain 1 are relatively short and form a left-handed, supercoiled helix. The supercoiling is relatively loose, with lengths of random coil separating α1 and α2 as well as α3 and α4. Helices α2 and α3 are linked by Ser39, which forms a bend between the two helices that allows them to form the supercoiled helix. The side chain of Ser39 forms stabilizing links with other sections of the protein. Specifically, the OG atom hydrogen bonds with the amino nitrogen atom of Val42
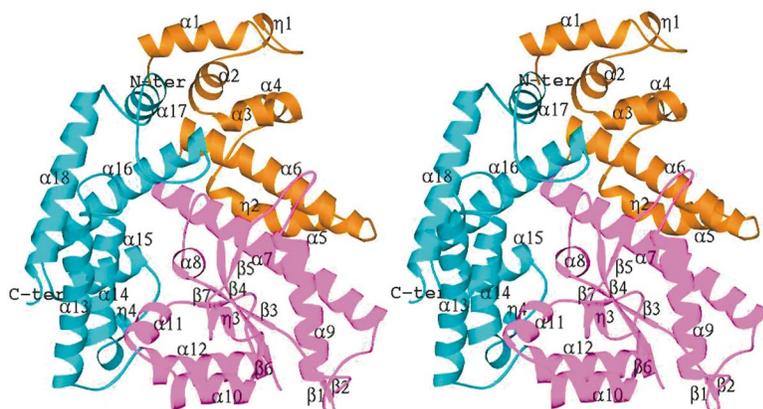
Figure 1. Schematic Diagram of AT-N440 with the Three Subdomains Colored Orange, Magenta, and Cyan

The secondary structures were also labeled: α for α helix, β for β strand, η for 3₁₀ helix.

and the carboxyl oxygen atom of Ser386. These helices are followed by a pair of longer antiparallel helices (α5 and α6).

The first helix of the second domain feeds into two antiparallel strands that form a small sheet. The second strand connects to the first of four strands that form a larger antiparallel sheet in the center of what is thought to be the active site. The first two strands of this sheet, β3 and β4, are connected by a helix, as are β4 and β5. A turn connects β5 and β6, both of which hydrogen bond to β4. Strand β6 is followed by two helices. These are followed by β7, which connects to the final helix of this domain. This final helix, α12, interacts with α10 to form the base of the molecule as shown in Figure 1.

Domain 3 consists entirely of α helices, the first two of which run antiparallel to each other, with the second, α14, bordering the active site. The link between β13 and β14 contains Lys317, which may be involved in catalysis, as described below. In addition to α14, α16 is close to the active site. The final helix, α18, extends from 414 to 434. In a previous study, a shorter fragment, AT-N423, was used but exhibited poor solubility (Jaggi et al., 1997). Secondary structure predictions indicated that this may have resulted in helix α18 being truncated (Xu et al., 2004) and that this was responsible for the poor solubility exhibited by this initial fragment. This idea is consistent with the structure. Residues 424–436 make a number of interactions with residues of α14 (323–337) and α16 (364–388). These three helixes form a helical bundle that has a hydrophobic interface between them. Since AT-N423 does not have residues 424–440, the hydrophobic faces of the helices α14 and α16 are exposed to solution in this molecule.

Structural Comparisons

A search for structural similarity was carried out using DALI (Holm and Sander, 1995b), with the entire AT-N440 molecule as the target. This search gave matches that were low in both sequence identity and structural similarity. The DALI searches were more successful when the protein was broken down into domains the boundaries of which have been assigned as a result of comparisons with related molecules. For the first domain, a similar arrangement of the first four helices are observed in the human hyperplastic disc protein (PDB 1i2t) that is an ortholog of polyadenylate binding protein (Deo et

al., 2001). Inspection of Figure 2 reveals a low level of similarity between domain 2 and the central catalytic domain of Polβ and PAP. Domains 2 and 3 taken together look somewhat similar to KNT (Pedersen et al., 1995), as is evident from the topology diagram shown in Figure 2. However, comparison of domains 2 and 3 with KNT are somewhat problematic because the relative orientation of the two domains differs in the two structures. Taken together, these two domains gave a Dali Z score of 4.2 when compared with KNT, while a
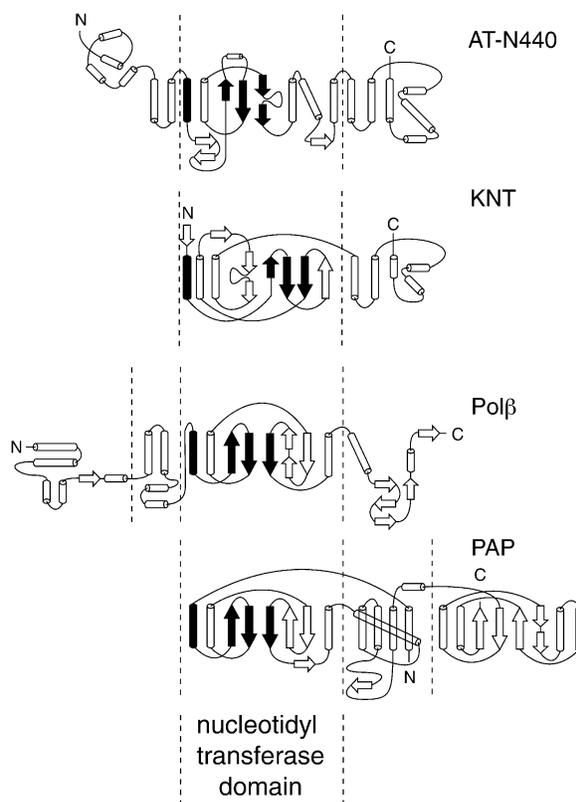
Figure 2. Topology of Nucleotidyl Transferase Proteins

The conserved nucleotidylyl transferase motif is shown in black. The conserved aspartates are located on the C-terminal pair of β strands of the conserved motif.
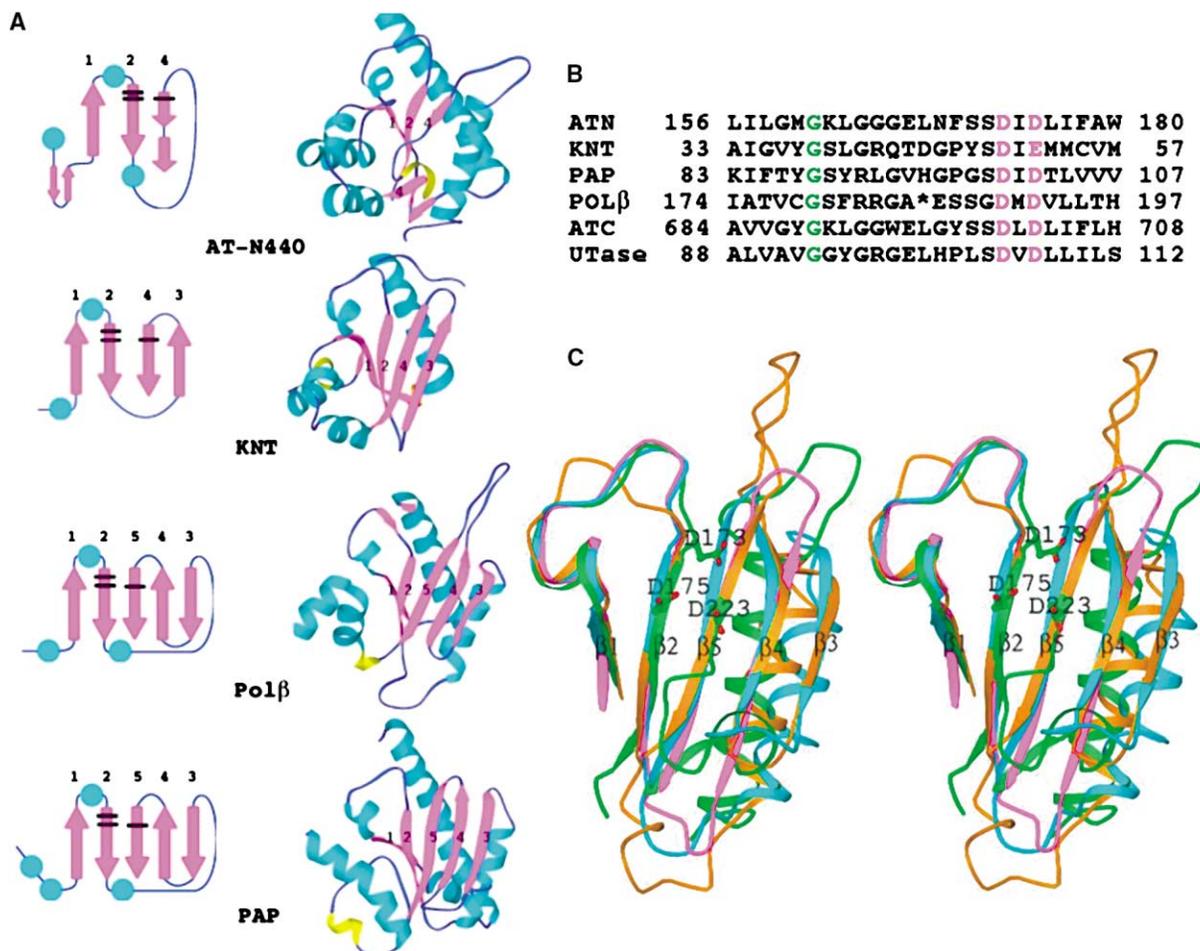
Figure 3. Comparison of the Common Structure Motif of AT-N440, KNT, Polβ, and PAP

(A) Topology and structure diagrams of the four enzymes. Three conserved aspartate residues are shown in black horizontal lines, cyan circles represent helix, and 3₁₀ helix is indicated in yellow.

(B) Sequence alignment of active regions of nucleotidyl transferase. The conserved residues are shown in color, and an asterisk indicates a deletion.

(C) Stereo view of overlay Polβ conserved nucleotidylyl transferase motif. AT-N440, green; KNT, magenta; Polβ, orange; and PAP, cyan. Three conserved aspartate residues are also shown.

value of 69.4 was obtained when AT-N440 was overlaid upon itself. The two domains gave rise to a root-square-mean deviation (rsmd) of 4.2 Å when aligned with only 128 (of a possible 326) Cα positions. Most of the amino acids used for alignment were from domain 2. Domain 2 is much more similar to KNT than domain 3; it gave a Dali Z score of 7.1 and an rmsd of 3.1 Å for 106 (of a possible 173) Cα positions.

Domain 2 contains the active site, with a motif found in the catalytic domains of the nucleotidylyl transferases. Figure 3A shows topology diagrams for this domain along with the three-dimensional structures of the nucleotidyl transferase motifs of AT-N440, KNT, and the polymerases. The nucleotidyl transferase domains have similar, although not identical, topologies, as can be seen in Figures 2 and 3A. Nucleotide phosphate binding and subsequent transfer occur via interactions with metal ions which bind to three conserved carboxyl side chains in this domain (Sawaya et al., 1997; Bard et al., 2000; Martin et al., 2000). The structural motif involved

in the metal binding consists of three adjacent β strands, the first of which is antiparallel to the other two. At the end of the first strand (β3 of AT-N440) is a glycine residue that initiates a helical turn in the loop connecting to the second strand. This glycine (Gly161 in AT-N440) is conserved in all nucleotidyl transferase domains for which structures are known, as shown in Figure 3B. However, the assignment of this glycine shown in Figure 3B is based on structural comparisons and differs from that of Holm and Sander (1995a), whose assignment was based only on sequence comparisons. The residue following this glycine is a serine or glycine in other nucleotidylyl transferases, while in AT-N440 it is a lysine (residue Lys162) that is conserved in other ATase sequences. This difference in sequence alignments, coupled to the fact that three residues downstream of Gly161 are a Gly-Gly combination, explains why Holm and Sander assigned the latter as the start of the phosphate binding loop. Twelve residues after Gly161, at the start of the second strand (β4 of AT-N440), are found
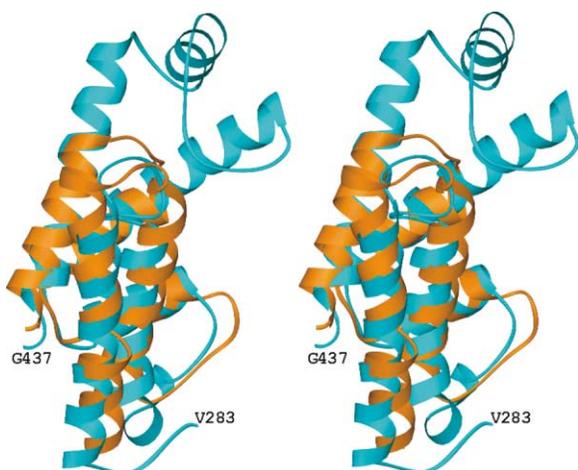
Figure 4. Ribbon Diagram of Overlay Domain 3 and KNT

The beginning and last residues of the domain 3 were labeled: AT-N440, cyan; KNT, orange.

the first two metal binding carboxylate residues (Asp173 and Asp175) separated by a single residue. The third strand (β5 of AT-N440) of the nucleotidyl transferase motif contains the last of the metal binding residues (Asp223) and is connected to the second strand by a section of peptide that contains a helix (α9 of AT-N440). As can be seen in Figure 3A, this connection is different in AT-N440, KNT, and the polymerases. In KNT this helix is missing, while in the polymerases, a helix also follows the second strand, but the connection to the third strand goes via an additional two β strands. The length of the third strand of AT-N440 (β5) is much shorter than the corresponding strands in KNT, PAP, and Polβ. The peptide following the third conserved acid residue (Asp223) folds into a turn that resembles a single turn of 3_{10} helix. As noted below, the residues making up this turn are conserved (Leu226, Arg227, and Pro228) in all ATase sequences, and the turn makes a very close approach to the putative ADP binding site. As can be seen from Figure 3C, there is a high degree of structural similarity in the nucleotidyl transferase motif found different proteins. The similarity covers the first two strands and the loop that connects them. The connection between strands 2 and 3 differs between the proteins as does the length of the strand 3.

The structure of domain 3 appears to be most similar to the C-terminal domain of KNT (Pedersen et al., 1995), despite the fact that the two domains can only be aligned with 5% sequence identity. Shown in Figure 4 is an overlay of domain 3 with KNT that gives a Dali Z score of 3.5 and an rmsd of 3.6 Å for 91 (of a possible 154) Cα atoms. A similar level of similarity was observed in comparisons with a domain in adenylsuccinate lyase (PDB code 1f1o): Dali Z score 5.4, rmsd 3.9 Å, based on 83 Cα positions (Toth et al., 2000). In this latter case, the two peptides share only 7% sequence identity. Both comparisons are characterized by low sequence identity, large insertions, deletions, and conformational differences, indicating a significant distance in evolutionary space between these structures.

Conserved Residues

A search for sequences showing homology to ATase was carried out as described in the Experimental Procedures. A total of 64 sequences were collected on the basis of their similarity to *E. coli* ATase, with only 11 sequences between 20% and 25% sequence identity. The sequences that were detected in this search were found to belong to four bacterial subdivisions and were used to produce the phylogenetic tree given in Figure 5. There were no sequences detected among numerous other organisms that possess the PII protein. For example, PII sequences were found among the archaea, the cyanobacteria, and the low G+C Gram positives, but ATase sequences could not be found in the same organisms.

The aligned ATase sequences gave 35 absolutely conserved residues and 24 highly conserved residues that were different in at most two sequences. Within the AT-N440 fragment, the conserved residues are found in domains 2 and 3 and occur in stretches that are concentrated around the active site, as shown in Figure 6A. For the most part, the highly conserved residues also clustered around the active site, as shown in Figure 6B. One exception to this rule is Ser39, which was referred to in the description of domain 1. An alignment of the AT-N440 and AT-C sequences is also shown in Figure 6A. In AT-N440, residues 161–169 are highly conserved and cover the phosphate binding loop. The same stretch in AT-C contains only two conserved residues. Residues 172–228 cover the strands containing the metal binding ligands. These three aspartate residues are conserved in both domains. Residues 245–260 are part of a crevice that is also made up of the conserved residues that follow strand 5. Again, this feature is conserved in AT-C. Residues 317–322 are found in a loop between α13 and α14. Within this loop, the side chain of Lys317 points into the active site. Of the four conserved residues in this loop, only two, Lys317 and Gly321, are conserved in AT-C. The conserved residues in helix α14 occur along the face that forms part of the active site and are not conserved in AT-C. The conserved residues in the range 374–385 are found in helix α16 and are not conserved in AT-C. These residues make contact with conserved residues 167–169 and may be important in terms of stabilizing sections of the phosphate binding loop.

Discussion

The Active Site Regions

Although no metal ions were observed in the structure of AT-N440, a comparison of the conserved aspartate residues of the various nucleotidyl transferase structures suggests that the metal ion binding and subsequent nucleotide binding will be similar in AT-N440, PAP, KNT, and Polβ. The metal binding ligands are conserved in all the AT-N440 sequences, as are the corresponding residues in AT-C. In the case of the polymerases, one metal ion (α) is bound to all three aspartates while the other (β) just to one, in our case Asp173. There is one water molecule (Wat 77) that possibly mimics the metal ion (α) in the AT-N440 structure. Wat 77 forms hydrogen bonds with the side chain carboxyl oxygen atoms of

Gammaproteobacteria (39:0)
Escherichia coli O157:H7 EDL93, NP_289628, 946
Escherichia coli K12, NP_417525, 946
Shigella flexneri 2a str. 301, NP_708863, 983
Shigella flexneri 2a str. 2457, NP_838572, 946
Escherichia coli CFT073, NP_755672, 983
Salmonella enterica subsp. ent, NP_457592, 947
Salmonella typhimurium LT2, NP_462116, 947
Yersinia pestis KIM, NP_670822, 960
Yersinia pestis CO92, NP_404291, 951
Photorhabdus luminescens subsp, NP_931169, 956
Vibrio vulnificus CMCP6, NP_759601, 950
Vibrio vulnificus YJ016, NP_933374, 950
Vibrio parahaemolyticus RIMD 2, NP_796802, 947
Vibrio cholerae O1 biovar elto, NP_232067, 948
Haemophilus somnus 129PT, ZP_00122523, 986
Pasteurella multocida, NP_245312, 986
Haemophilus influenzae Rd KW20, NP_438242, 981
Actinobacillus pleuropneumonia, ZP_00135043, 931
Haemophilus ducreyi 35000HP, NP_874080, 957
Shewanella oneidensis MR-1, NP_719297, 990
Pseudomonas syringae pv. syrin, ZP_00125184, 985
Pseudomonas syringae pv. tomat, NP_794737, 985
Pseudomonas fluorescens PfO-1, ZP_00085607, 979
Pseudomonas putida KT2440, NP_742507, 977
Pseudomonas aeruginosa PAO1, NP_253701, 982
Azotobacter vinelandii, ZP_00091355, 1036
Microbulbifer degradans 2-40, ZP_00067688, 979
Neisseria meningitidis Z2491, NP_282889, 896
Neisseria meningitidis MC58, NP_273281, 896
Chromobacterium violaceum ATCC, NP_901765, 894
Nitrosomonas europaea ATCC 197, NP_841378, 929
Ralstonia metallidurans, ZP_00024377, 1017
Ralstonia solanacearum, NP_520777, 955
Burkholderia fungorum, ZP_00030666, 948
Bordetella bronchiseptica RB50, NP_889769, 941
Bordetella pertussis Tohama I, NP_880032, 960
Bordetella parapertussis, NP_884139, 941
Xanthomonas axonopodis pv. cit, NP_640903, 941
Xanthomonas campestris pv. cam, NP_638945, 937

Deltaproteobacteria (3:0)
Geobacter metallireducens, ZP_00081497, 1207
Geobacter sulfurreducens PCA, NP_954418, 1017
Magnetococcus sp. MC-1, ZP_00043553, 882

Alphaproteobacteria (11:0)
Agrobacterium tumefaciens str., NP_531679, 988
Agrobacterium tumefaciens str., NP_354005, 1010
Sinorhizobium meliloti, NP_385130, 986
Brucella melitensis 16M, NP_540244, 938
Brucella suis 1330, NP_697628, 983
Mesorhizobium loti, NP_107961, 986
Bradyrhizobium japonicum USDA, NP_769773, 995
Rhodopseudomonas palustris, ZP_00011510, 990
Caulobacter crescentus CB15, NP_421549, 986
Rhodobacter sphaeroides, ZP_00007815, 1184
Novosphingobium aromaticivoran, ZP_00094684, 902

Actinobacteria (11:0)
Corynebacterium diphtheriae, NP_940010, 1054
Corynebacterium glutamicum ATC, NP_601431, 1045
Corynebacterium efficiens YS-3, NP_738736, 1050
Bifidobacterium longum NCC2705, NP_695974, 1076
Bifidobacterium longum DJO10A, ZP_00120298, 1076
Mycobacterium tuberculosis H37, NP_216737, 994
Mycobacterium tuberculosis CDC, NP_336750, 994
Mycobacterium leprae, NP_302122, 1004
Streptomyces avermitilis MA-46, NP_827149, 1000
Streptomyces coelicolor A3(2), NP_733564, 999
Thermobifida fusca, ZP_00058435, 1052

Figure 5. Evolutionary Distance Dentogram of ATase Amino Acid Sequences

Branching supported with bootstrap values less than 80% is rooted to the last highly supported ancestor.

Asp173 (2.7 Å), Asp175 (2.8 Å), and Asp223 (2.8 Å), respectively.

A model of AT-N440 complexed with adenosine dinucleotide (ADP) was produced by using adenosine triphosphate (ATP) and metal ion coordinates from PAP rotated into the active site of AT-N440 by overlaying the conserved aspartates and then truncating the nucleotide to ADP by removal of the γ-phosphoryl group (Figure 6B). The model does not allow detailed predictions of interactions between the nucleotide and the protein, particularly when it is considered that the conformation of the protein is likely to change on binding substrate. However, some indication of the location of the various components of the ADP can be inferred with this model and the location of conserved residues that are given in Figure 6B.

The base of the ADP fits into a crevice such that it could, with some side chain movement, interact with Trp253 and Tyr245 (conserved in 56 of the 57 sequences) as shown in Figure 6B. The residue Leu226 is close enough to interact with the base and sugar of ADP, suggesting that the model requires minor modification. The next residue is Arg227, whose side chain makes a number of hydrogen bonds with backbone atoms, suggesting that it functions in stabilizing the structure of the active site. All the residues that cluster around the base are conserved in both AT-N440 and AT-C. The phosphates appear to interact with a number of basic residues as shown in Figure 6B. One such residue is Lys162, which sits at the start of the helical turn that follows the first strand. As mentioned earlier, in Polβ, the residue at this position is a serine (Ser180) (Sawaya et al., 1997) that interacts with the γ-phosphate of the nucleoside triphosphate. In AT-N440, Lys162 is hydro-

A

```
                    α1                    η1         α2           α3           α4
AT-N    1 MKPLSSPLQQ YWQTVVERLP EPLAEESLSA QAKSVLTFSD FVQDSVIAHP EWLTELESQP
AT-C  528 ITALLVGIVT RTTYLELLSE FPAALKHLIS LCAASPMIAS QLARYPLLLD ELLDPNTLYQ
           η2         α5                   α6                                  α7
AT-N   61 PQADEWQHYA AWLQEEALCNV SDEAGLMREL RLFRRRIMVR IAWAQTLALV TEESILQQLS
AT-C  588 PTATDAYRDE LRQYLLRVPE DDEEQQLEAL RQFKQAPVLR IAAADIAGTL PVMKVSDHLT
              α7                          β1      β2     β3       α8          β4
AT-N  121 YLAETLIVAA RDWLYDACCR EWGTPCNA-Q GEAQPLLILG MGKLGGGELN FSSDIDLIFA
AT-C  647 WLAEAMIDAV VQQANVQMVA RYGKPNHLNE REGRGFAVVG YGKLGGWELG YSSDLDLIFL
                                       α9                        β5           η3
AT-N  180 WPEHGC---- TQGGRRELDN AQFFTRMGQR LIKVLDQPTQ DGFVYRVDMR LRPFGESGPL
AT-C  667 ---HDCPMDA MTDGEREIDG RQFYLRLAQR IMHLFSTRTS SGILYEVBAR LRPSGAACML
          β6        α10         α11        β7                         α12
AT-N  236 VLSFAALEDY YQEQGRDWER YAMVKARIM- GDSE-----G VYANELRAML RPFVFRRYID
AT-C  765 VTSAEAFADY QKNEAWTMEH QALVRARVVY GDPQLTAHFD AVRREIMTLP RE---GKTLQ
                     α13                                     α14          η4
AT-N  290 FSVIQSLRNM KGMIAREVRR RGLTDNIKLG AGGIREIEFI VQVFQLIRGG REPSLQSRS-
AT-C  821 TEVREMREKM RAHLGNKHRD RF---DIKAD EGGITDIEFI TQYLVLRYAH EKPKLTRWSD
                    α15               α16                                    α17
AT-N  349 LLPTLSAIAE LHLLSENDAE QLRVAYLFLR -RLENL-LQS INDEQTQTLP SDELNRARLA
AT-C  879 NVRILELLAQ NDIMEEQEAM ALTRAYTTLR DELHHLALQE LPGHVSEDCF TAERELVRAS
           α17                   α18
AT-N  407 WAMDFADWPQ LTGALTAHMT NVRRVFNELI GDDE
AT-C  938 WQKWLVEE
```

B

Figure 6. Conserved Residues and Active Site

(A) Sequence alignment of AT-N440 and AT-C. The secondary structure features based on the AT-N440 structure reported here are indicated above its sequence. Conserved residues in both domains based on 57 aligned ATase sequences are colored: red, 57/57; yellow, 56/57; and green, 55/57.

(B) Residues are shown around the active site of AT-N440 and colored by functional groups: green, metal binding strands; cyan, base binding site; and magenta, phosphate binding loop. ADP is displayed in blue, and Mg^{2+} ions are gray.

gen bonded to Glu327 of helix α14. Also found on helix α14 is Arg324, and nearby is Lys317; both of these residues could interact with phosphates of ADP. During catalysis, one of the phosphoryl oxygens of the adenylylated Tyr396 of GS (GS-AMP) would be coordinated to the metal ions. We would expect the attacking phosphate to be close to the location of the β phosphate of the ADP complex, and hence, residues Lys317, Lys162, and Arg324 may serve to orient this ion for nucleophilic attack. Helix α14 contains a number of residues that are

conserved in all the AT-N440 sequences, but which are not conserved in the corresponding sequences of AT-C. The side chains of these conserved residues face the active site, suggesting that they have functional significance, possibly interacting with GS or PII.

Since the two activities of ATase are mechanistically similar to one another, we would expect these active sites to be very similar. As noted above, Lys317, Lys162, and Arg324 are in a position to interact with the phosphates of ADP, and only one Lys corresponding to

Lys317 is very highly conserved in the AT-C sequences, suggesting that other two basic residues may be responsible for the difference in activities of AT-N440 and AT-C. However, other residues may also be important in distinguishing the activities of ATase. AT-N440, unlike AT-C, should not be able to bind ATP. If one models ATP rather than ADP into the active site, then the side chain of Asn169 comes close and could make contact with one of the phosphoryl oxygens of the γ phosphate. The residue in AT-C corresponding to Asn169 is Gly696. The absence of a side chain in glycine may create sufficient space to accommodate the γ phosphate. However, this glycine is not conserved in all AT-C sequences. If the loop containing Asn169 is important for determining substrate specificity, then the structure of the corresponding loop in AT-C is probably different, at least in some ATase molecules. Although the role of Asn169 cannot be stated with absolute certainty, the degree of sequence conservation in this region suggests that it has some functional importance. As mentioned in the previous section, the conformation of the conserved residues around Asn169 (Figure 6B) appears to be stabilized by interactions with conserved residues on helix α16 (Tyr374, Arg378, and Glu381). For example, the side chain of Arg378 interacts with the backbone atoms of conserved residue Glu167.

Concluding Remarks

AT-N440 is the first reported crystal structure of a nucleotidyltransferase that has a protein as the substrate. Although the structure of AT-N440 is unlike any known protein solved to date, domains 2 and 3 possess some similarity to KNT, albeit subject to a large rigid body movement between the domains. The active site region of domain 2 shows some similarities to the active sites of KNT, Polβ, and PAP despite very low sequence identity. The relative positions of the metal binding residues are highly conserved in all the proteins, but residues that could interact with the phosphates and the base of ADP are clearly different for AT-N440. A number of residues have been identified that could be responsible for the different catalytic functions of AT-N440 and AT-C.

Although AT-N440 has provided a great deal of information, there are many questions that it sheds little light on. For example, how does PII change the activities of ATase? Answers to this question await the structure of the intact ATase protein, and for the moment, we can only speculate how PII might bind and alter ATase. We know that the activity of ATase is regulated by PII, yet the activities of AT-N440 and AT-C are not. In the absence of PII, we think that the two domains of ATase pack together in such a way that neither active site is exposed to solution and accessible to its substrates. We believe that PII separates the two domains in such a way that only one active site can act on GS. How this is achieved remains to be determined, but it is reasonable to suppose that the T loop of PII blocks the AT-N440 active site while the T loop of PII-UMP blocks the AT-C active site. The interaction of PII with ATase must require both domains to explain why its regulatory ability of PII is lost when the domains are expressed separately.

Experimental Procedures

Crystallization and Data Collection

Expression, purification, and crystallization of recombinant AT-N440 were published elsewhere (Xu et al., 2004). Briefly, selenomethionine-labeled AT-N440 protein was overexpressed in *E. coli* JM109 strains by methionine pathway inhibition described by Van Duyne et al., (1993) in M9 medium. Its purification and crystallization were almost identical to the native AT-N440. MAD data were collected at 100K at the Structural Biology Center beamline 19ID at the Advanced Photon Source at Argonne National Laboratory, Chicago. The inflection, peak, and high-energy remote wavelengths were determined from an X-ray fluorescence spectrum collected from the mounted crystal. Integration, scaling, and merging of data were performed with the program HKL 2000 (Otwinowski and Minor, 1997). The data collection statistics are shown in Table 1.

Structure Determination

The program SOLVE (Terwilliger and Berendzen, 1999) was used to locate the selenium sites as well as to calculate the initial phasing maps. Density modification and automatic chain tracing were undertaken using the program RESOLVE (Terwilliger, 2000). 320 out of 440 residues were built by RESOLVE, though 200 residues were fitted as alanine residues. The solvent flipping maps calculated in CNS (Brunger et al., 1998) were used to complete the model with O (Jones et al., 1991). Residues 3–143, 153–180, and 195–436 of AT-N440 were traceable into the original density-modified maps. The model was refined with CNS interspersed with the rounds of manual model building. The final refinement was performed in REFMAC 5.1 of CCP4 (Murshudov et al., 1997) with loose B factor restraints. The final model contained residues 1–182 and 191–437, with amino acids 2, 13, 17, 18, 25, 53, 58, 59, 75, 79, 141, 144, 151, 153, 193, 204, 247, 268, 270, 308, 310, 312, 390, 397, 399, 410, and 429 modeled as alanines. The R factor and R_{free} values of the final model were 22.69% and 26.75%, respectively. A total of 92.4% (354 amino acids) of nonglycine residues were in the most favored regions of Ramachandran plot, with 6.8% (26 amino acids) in additionally allowed regions, residues 18, 144, and 194 in generously allowed regions, and none in the disallowed region. The atomic coordinates and structure factors (code 1V4A) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/). Ribbon diagrams were generated using RIBBONS (Carson, 1997).

Conserved Residues

Amino acid sequences belonging to the ATase family were obtained by BLAST sequence homology search of the *E. coli* sequences against 186 microbial genomes available at NCBI at the time of the analysis. Sequences with BLAST E values less than 1e-25 and a minimum length of 800 residues were aligned to each other using the automatic alignment program clustalW (Thompson et al., 1994). Regions of ambiguous positional homology were removed from the alignment, and an evolutionary distance tree was constructed from the remainder using the neighbor joining method (Saitou and Nei, 1987). The robustness of the tree topology was verified by bootstrap resampling of the tree 1000 times.

References

Aravind, L., and Koonin, E.V. (1999). DNA polymerase b-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. Nucleic Acids Res. *27*, 1609–1618.

Arcondeguy, T., Jack, R., and Merrick, M. (2001). P(II) signal transduction proteins, pivotal players in microbial nitrogen control. Microbiol. Mol. Biol. Rev. *65*, 80–105.

Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., and Pelletier, H. (2000). Structure of yeast poly(A) polymerase alone and in complex with 3′-dATP. Science *289*, 1346–1349.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Carson, M. (1997). Ribbons. Methods Enzymol. *277*, 493–505.

Deo, R.C., Sonenberg, N., and Burley, S.K. (2001). X-ray structure of the human hyperplastic discs protein: an ortholog of the C-terminal domain of poly(A)-binding protein. Proc. Natl. Acad. Sci. USA *98*, 4414–4419.

Engleman, E.G., and Francis, S.H. (1978). Cascade control of E. coli glutamine synthetase. II. Metabolite regulation of the enzymes in the cascade. Arch. Biochem. Biophys. *191*, 602–612.

Holm, L., and Sander, C. (1995a). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. *20*, 478–480.

Holm, L., and Sander, C. (1995b). DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily. Trends Biochem. Sci. *20*, 345–347.

Jaggi, R., van Heeswijk, W.C., Westerhoff, H.V., Ollis, D.L., and Vasudevan, S.G. (1997). The two opposing activities of adenylyltransferase reside in distinct homologous domains, with intramolecular signal transduction. EMBO J. *16*, 5562–5571.

Jones, T., Zou, J., Cowan, S., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Magasanik, B. (1993). The regulation of nitrogen utilization in enteric bacteria. J. Cell. Biochem. *51*, 34–40.

Martin, G., Keller, W., and Doublie, S. (2000). Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. EMBO J. *19*, 4193–4203.

Merrick, M.J., and Edwards, R.A. (1995). Nitrogen control in bacteria. Microbiol. Rev. *59*, 604–622.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Ninfa, A.J., and Atkinson, M.R. (2000). PII signal transduction proteins. Trends Microbiol. *8*, 172–179.

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Pedersen, L.C., Benning, M.M., and Holden, H.M. (1995). Structural investigation of the antibiotic and ATP-binding sites in kanamycin nucleotidyltransferase. Biochemistry *34*, 13305–13311.

Saitou, N., and Nei, M. (1987). A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Sawaya, M.R., Prasad, R., Wilson, S.H., Kraut, J., and Pelletier, H. (1997). Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. Biochemistry *36*, 11205–11215.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. *56*, 965–972.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Toth, E.A., Worby, C., Dixon, J.E., Goedken, E.R., Marqusee, S., and Yeates, T.O. (2000). The crystal structure of adenylosuccinate lyase from Pyrobaculum aerophilum reveals an intracellular protein with three disulfide bonds. J. Mol. Biol. *301*, 433–450.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J. Mol. Biol. *229*, 105–124.

Xu, Y., Wen, D., Clancy, P., Carr, P.D., Ollis, D.L., and Vasudevan, S.G. (2004). Expression, purification, crystallization and preliminary x-ray analysis of the N-terminal domain of Escherichia coli adenylyl transferase. Protein Expr. Purif. *34*, 142–146.

Accession Numbers

The atomic coordinates and structure factors (code 1V4A) have been deposited in the Protein Data Bank.