Crystal Structure of the Tetrameric Cytidine Deaminase from *Bacillus subtilis* at 2.0 Å Resolution[†,‡]

Eva Johansson,[§] Nina Mejlhede,[||] Jan Neuhard,[||] and Sine Larsen*[,§]

Centre for Crystallographic Studies, Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark, and Institute of Molecular Biology, University of Copenhagen, Sølvgade 83H, DK-1302 Copenhagen K, Denmark

Received September 26, 2001; Revised Manuscript Received November 21, 2001

ABSTRACT: Cytidine deaminases (CDA, EC 3.5.4.5) are zinc-containing enzymes in the pyrimidine salvage pathway that catalyze the formation of uridine and deoxyuridine from cytidine and deoxycytidine, respectively. Two different classes have been identified in the CDA family, a homodimeric form (D-CDA) with two zinc ions per dimer and a homotetrameric form (T-CDA) with four zinc ions per tetramer. We have determined the first structure of a T-CDA from *Bacillus subtilis*. The active form of T-CDA is assembled of four identical subunits with one active site apiece. The subunit of D-CDA is composed of two domains each exhibiting the same fold as the T-CDA subunits, but only one of them contains zinc in the active site. The similarity results in a conserved structural core in the two CDA forms. An intriguing difference between the two CDA structures is the zinc coordinating residues found at the N-terminal of two α-helices: three cysteine residues in the tetrameric form and two cysteine residues and one histidine residue in the dimeric form. The role of the zinc ion is to activate a water molecule and thereby generate a hydroxide ion. How the zinc ion in T-CDA surrounded with three negatively charged residues can create a similar activity of T-CDA compared to D-CDA has been an enigma. However, the structure of T-CDA reveals that the negative charge caused by the three ligands is partly neutralized by (1) an arginine residue hydrogen-bonded to two of the cysteine residues and (2) the dipoles of two α-helices.

The pyrimidine salvage pathway enables organisms to utilize exogenous pyrimidine bases and nucleosides, which are not intermediates in the de novo pyrimidine synthesis. The reaction catalyzed by cytidine deaminase (CDA, EC 3.5.4.5)[1] is part of the pyrimidine salvage pathway and catalyzes the hydrolytic deamination of cytidine and deoxycytidine to form uridine and deoxyuridine, respectively (*1*) (Figure 1). Two forms of CDA have been identified, a homotetramer (T-CDA) and a homodimer (D-CDA) represented by the *Bacillus subtilis* and the *Escherichia coli* enzymes, respectively. CDA from *B. subtilis* is a homotetrameric zinc enzyme with a total molecular mass of 4 × 14.9 kDa (*2*). The *E. coli* D-CDA subunit has a molecular mass of 31.5 kDa; it contains three domains: a small N-terminal domain of unknown function, a catalytic domain with a bound zinc ion, and a C-terminal domain that possesses nearly the same tertiary structure as the catalytic domain though it does not contain a zinc ion (*3*). Amino acid sequence alignments have shown that the T-CDA subunit displays sequence similarity with both the catalytic and the C-terminal domain of D-CDA (*3*). Hence, it has been suggested that T-CDA is constructed in the same way as D-CDA, with the structural core formed from four identical subunits placed in relation to each other as the catalytic and C-terminal domains in D-CDA.

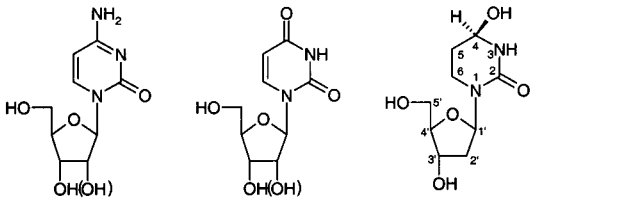

FIGURE 1: Cytidine and deoxycytidine are equally good substrates in the reaction catalyzed by CDA, where uridine and deoxyuridine are formed, respectively. THU is an inhibitor of CDA with a C5−C6 single bond and a hydroxyl group at position 4.

T-CDA as well as D-CDA are found in prokaryotes and eukaryotes, though according to sequence studies it appears that T-CDA is more widely distributed among species (Figure 2). Amino acid sequences of potential T-CDAs have been found in archaea, different types of bacteria, and eukaryotes such as mammals, worms, yeasts, and insects. The D-CDA sequence is found in some proteobacteria and in *Arabidopsis thaliana*. The similarity between the two domains of the D-CDA gene suggests that D-CDA has emerged by a gene

[†] This work was supported by grants from Hellmuth Hertz' Foundation and the Danish National Research Foundation.

[‡] The atomic coordinates and observed structure factor amplitudes have been deposited in the Protein DataBank under accession code 1JTK.

* To whom correspondence should be addressed. Phone: (+45) 35 32 02 82, Fax: (+45) 35 32 02 99, E-mail: sine@ccs.ki.ku.dk.

[§] Centre for Crystallographic Studies.

[||] Institute of Molecular Biology.

[1] Abbreviations: CDA, cytidine deaminase; T-CDA, tetrameric cytidine deaminase; D-CDA, dimeric cytidine deaminase; THU, 3,4,5,6-tetrahydro-2′-deoxyuridine; NCS, noncrystallographic symmetry; rmsd, root-mean-square deviation.

10.1021/bi011849a CCC: $22.00
Published on Web 01/29/2002

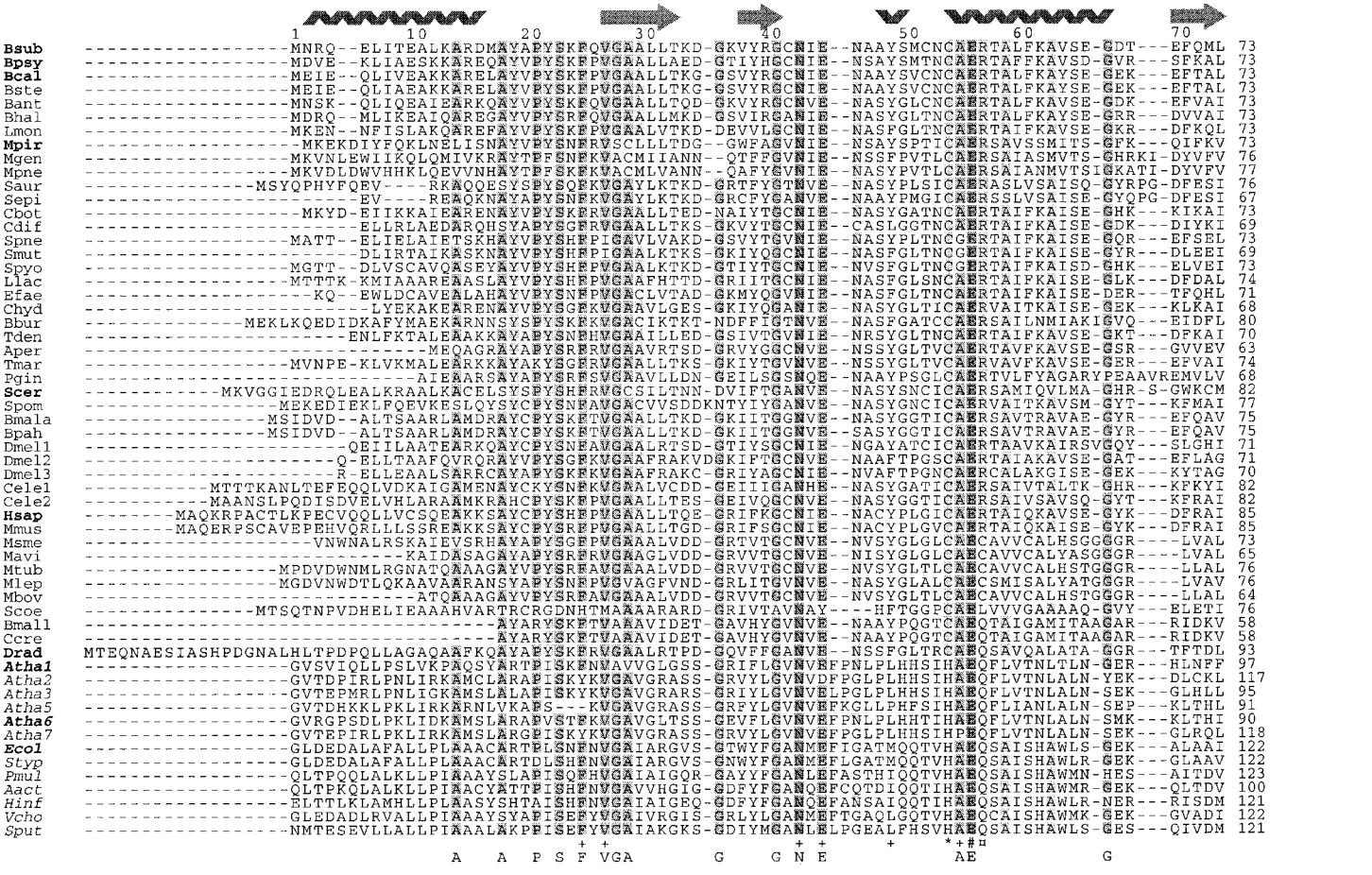
Bsub
Bpsy
Bcal
Bste
Bant
Bhal
Lmon
Mpir
Mgen
Mpne
Saur
Sepi
Cbot
Cdif
Spne
Smut
Spyo
Llac
Efae
Chyd
Bbur
Tden
Aper
Tmar
Pgin
Scer
Spom
Bmala
Bpah
Dmel1
Dmel2
Dmel3
Cele1
Cele2
Hsap
Mmus
Msme
Mavi
Mtub
Mlep
Mbov
Scoe
Bmall
Ccre
Drad
Atha1
Atha2
Atha3
Atha5
Atha6
Atha7
Ecol
Styp
Pmul
Aact
Hinf
Vcho
Sput

FIGURE 2: Sequence alignment of 45 amino acid sequences of potential T-CDAs (roman) and in the bottom 13 sequences of the catalytic domain of potential D-CDAs (italic). The numbering on top of the sequences is according to *B. subtilis* T-CDA. Residues with 100% sequence identity are shown in dark gray boxes, more than 80% identity in gray boxes, and more than 60% identity in light gray boxes. The residues with more than 80% sequence identity are also displayed below the aligned sequences. The secondary structure elements from *B. subtilis* T-CDA are shown above the sequences with a twisted rod for α-helices and an arrow for β-strands. The three zinc ligands are marked with (*), the active site residue Glu55 is marked with (#), the residues interacting with the inhibitor THU are marked with (+), and the charge-compensating residue Arg56 is marked with (○). The names of the organisms with proven CDA activity are in boldface. Organisms abbreviations: T-CDA: Bsub, *Bacillus subtilis* (P19079); Bpsy, *Bacillus psychrophilus* (Q9S3M0); Bcal, *Bacillus caldolyticus* (Q9R2S1); Bste, *Bacillus stearothermophilus*; Bant, *Bacillus anthracis*; Bhal, *Bacillus halodurans* (Q9KD53); Lmon, *Listeria monocytogenes*; Mpir, *Mycoplasma pirum* (P47718); Mgen, *Mycoplasma genitalium* (P47298); Mpne, *Mycoplasma pneumoniae* (P75051); Saur, *Staphylococcus aureus*; Sepi, *Staphylococcus epidermis*; Cbot, *Clostridium botulinum*; Cdif, *Clostridium difficile*; Spne, *Streptococcus pneumoniae*; Smut, *Streptococcus mutans*; Spyo, *Streptococcus pyogenes*; Llac, *Lactococcus lactis* (Q9GFM8); Efae, *Enterococcus faecalis*; Chyd, *Carboxythermus hydrogenoformans*; Bbur, *Borrelia burgdorferi* (O51563); Tden, *Treponema denticola*; Aper, *Aeropyrum pernix* (Q9YD74); Tmar, *Thermatoga maritima* (Q9WZV0); Pgin, *Porphyromonas gingivalis*; Scer, *Saccharomyces cerevisiae* (Q06549); Spom, *Schizosaccharomyces pombe* (Q09190); Bmala, *Brugia malayi* (P90706); Bpah, *Brugia pahangi* (Q93143); Dmel1, *Drosophila melanogaster* I; Dmel2, *D. melanogaster* II; Dmel3, *D. melanogaster* III; Cele1, *Caenorhabditis elegans* I; Cele2, *C. elegans* II (Q20628); Hsap, *Homo sapiens* (P32320); Mmus, *Mus musculus* (P56389); Msme, *Mycobacterium smegmatis*; Mavi, *Mycobacterium avium*; Mtub, *Mycobacterium tuberculosis* (O53367); Mlep, *Mycobacterium leprae* (Q9CBD3); Mbov, *Mycobacterium bovis*; Scoe, *Streptomyces coelicolor* (Q9KYT5); Bmall, *Burkholderia malleii*; Ccre, *Caulobacter crescentus*; Drad, *Deinococcus radiodurans* (Q9RSE7); D-CDA: Atha1, *Arabidopsis thaliana* I (O65896); Atha2, *Arabidopsis thaliana* II (O65571); Atha3, *Arabidopsis thaliana* III (Q9ZT33); Atha4; *Arabidopsis thaliana* IV (Q9S7S2); Atha5, *Arabidopsis thaliana* V (Q9ZT32); Atha6, *Arabidopsis thaliana* VI (Q9SU86); Atha7, *Arabidopsis thaliana* VII (Q9ZT35); Ecol, *Escherichia coli* (P13652); Styp, *Salmonella typhimurium*; Pmul, *Pasteurella multicoda*; Aact, *Actinobacillus actinomycetemcomitans*; Hinf, *Haemophilus influenzae* (P44325); Vcol, *Vibrio cholerae* (Q9KSM5); Sput, *Shewanella putrefaciens*.

duplication of an ancestral T-CDA gene. The C-terminal domain has subsequently evolved and lost its zinc coordinating residues and thereby its catalytic activity.

The zinc ion in the active site plays a central role in the proposed catalytic mechanism of CDA, activating a water molecule to form a hydroxide ion that performs a nucleophilic attack on the substrate (*3*, *4*). An amino acid sequence alignment shows an overall sequence identity of 26% between *B. subtilis* T-CDA and the catalytic domain of *E. coli* D-CDA. The alignment also implies that the zinc liganding residues differ in T-CDA and D-CDA. According to the sequence alignment, the histidine residue that together with two cysteine residues coordinate to the zinc ion in D-CDA is exchanged for a cysteine residue in T-CDA, implying that zinc in T-CDA has three cysteine ligands (Figure 2). This kind of zinc coordination with three negatively charged cysteine residues is quite unusual for a catalytic zinc ion (for reviews, see *5* and *6*), and the undiminished activity of T-CDA as compared to D-CDA has been enigmatic (*7*). From the three-dimensional structure of T-CDA from *B. subtilis*, presented here, we have been able to explain how the zinc ion can maintain a normal level of activity. The refined structure model also includes the CDA inhibitor 3,4,5,6-tetrahydro-2′-deoxyuridine (THU) (Figure 1). The structure has further enabled us to answer some of the questions about the differences and similarities between T-CDA and D-CDA.

MATERIALS AND METHODS

Crystallization. The gene for CDA from *B. subtilis* was cloned and expressed in *E. coli* and purified as reported earlier (*8*). Crystal Screen I from Hampton Research (*9*) was used for screening of crystallization conditions for *B. subtilis* CDA using the vapor diffusion technique. The experiments were performed both with the native enzyme only and in the presence of the CDA inhibitor THU (purchased from Calbiochem, La Jolla, CA). The initial crystallization conditions in solution 1 (30% 2-methyl-2,4-pentanediol, 0.02 M calcium chloride, 0.1 M sodium acetate, pH 4.6) were refined, and the crystal used for data collection was grown from a drop of 2 μL of 4.5 mg/mL CDA and 5 mM THU in 20 mM Tris-HCl, pH 7.6, mixed with 2 μL of mother liquor (26% 2-methyl-2,4-pentanediol, 10 mM calcium chloride, and 0.1 M sodium acetate, pH 4.6). The hanging drops equilibrated over 1 mL of mother liquor at room temperature. At these conditions, crystals appear after a couple of hours and grow to a final size of 0.2 × 0.3 × 0.05 mm in 2 days.

Table 1: Data Statistics[a]

wavelength (Å)	1.5418
resolution (Å)	20−2.03 (2.08−2.03)
*R*sym (%)[b]	7.9 (19.5)
$I/\sigma(I)$	17.4 (5.1)
completeness (%)	98.4 (78.7)
no. of reflections	214942
no. of unique reflections	15867

[a] Values in parentheses are data for the highest resolution shell. [b] $R\text{sym} = \{\sum_{hkl}\sum_i|I_i(hkl) - \overline{I(hkl)}|\}/\{\sum_{hkl}\sum_i I_i(hkl)\}$, where $\underline{I_i(hkl)}$ is the intensity of the *i*th observation of reflection *hkl* and $I(hkl)$ is the average intensity of *i* observations.

Data Collection. Diffraction data were collected at 120 K on a MAR345 image plate detector mounted on a copper rotating anode generator from Rigaku (RU300) operating at 50 kV/80 mA. Auto-indexing, data reduction, and scaling were performed with programs from the HKL suite (*10*). Statistics of the diffraction data are shown in Table 1. The crystals belong to space group *C2* ($a = 75.1$, $b = 66.1$, $c = 55.5$ Å, $\beta = 115.9°$) and diffract X-rays well. However, nonmerohedral twinning of the crystals made the auto-indexing of the data nontrivial and difficult. Two protein chains per asymmetric unit give a reasonable Matthews' coefficient of 2.09 Å^3/Da, corresponding to approximately 41% solvent content.

Structure Determination and Refinement. The three-dimensional structure was determined using the molecular replacement method implemented in the program EPMR (*11*). Several search models based on the catalytic domain of *E. coli* D-CDA (residues 46−184; PDB code 1CTT) were tested. The search model that finally worked contained two copies of the catalytic domain, placed in relation to each other exactly as in the *E. coli* dimer. Both zinc ions were

Table 2: Refinement Statistics[a]

no. of reflections	15599
working set	14822
test set	777
resolution (Å)	19.92−2.04 (2.13−2.04)
no. of atoms	2193
R-factor (%)[b]	20.7 (21.3)
R-free (%)[c]	23.2 (25.0)
average *B*-factor (Å^2)	14.0
average *B*-factor protein atoms (Å^2)	13.1
average *B*-factor zinc ions (Å^2)	9.8
average *B*-factor THU (Å^2)	7.3
average *B*-factor water atoms (Å^2)	25.1
rmsd	
bond length (Å)	0.005
bond angle (deg)	1.2

[a] Values in parentheses are data for the highest resolution shell. [b] *R*-factor $= \{\sum_{hkl}||F_{obs}| - k|F_{calc}||\}/\{\sum_{hkl}F_{obs}\}$. [c] *R*-free $= \{\sum_{hkl,text}||F_{obs}| - k|F_{calc}||\}/\{\sum_{hkl,test}F_{obs}\}$ where F_{obs} and F_{calc} are observed and calculated structure factors, respectively, k is the scale factor, and the test set is 5% of the data chosen randomly.

kept in the search model while some of the loops were cut off, giving 99 core residues out of 139 from the *E. coli* D-CDA catalytic domain. A correlation coefficient of 25.6% and an *R*-factor of 56.6% were obtained for the correct solution. Rigid body refinement of the structure, allowing the two protein chains to move separately, first at low resolution and then at steps of increasing resolution with the program CNS (*12*), gave an *R*-factor of 53.1% (*R*-free 52.5%). The NCS relation was determined, and all residues in the model except from the glycine residues were changed to alanine residues. The zinc atom was also deposed before a step of simulated annealing was performed with CNS (*12*) employing strict 2-fold NCS. As the difference electron density map showed clear density for the zinc ion and its ligands, model building in O (*13*) was started. After a few cycles of simulated annealing refinement with CNS and model building with O, the zinc ion as well as THU and all amino acids, except from the five C-terminal residues, could be introduced into the model. The NCS restraints were relaxed, *B*-values were refined, and water molecules were included in the model. Water molecules related by NCS were found by the program WATNCS (*14*) and added to the NCS restraints list. Model quality was checked during the entire procedure using OOPS (*15*) and PROCHECK (*16*). The final model has an *R*-factor of 20.7% (*R*-free 23.2%) and contains 262 amino acid residues, 2 zinc ions, 2 THU molecules, and 183 water molecules, of which 104 are related by NCS. Refinement statistics are shown in Table 2.

Sequences, Sequence Alignment, and Surface Analysis. The amino acid sequences of 45 different potential T-CDA and the catalytic domain of 13 potential D-CDA were used in the alignment. The sequences were found in the SWISS−PROT and the TrEMBL databases and from ongoing genome projects. The origins of the CDA amino acid sequences are presented in the text of Figure 2. Out of the 58 amino acid sequences, only a few have been proven to have CDA activity. These are *B. subtilis* (*2*), *Bacillus psychrophilus* (*17*), *Bacillus caldolyticus* (*17*), *Mycoplasma pirum* (*18*), *Homo sapiens* (*19*), *Saccharomyces cerevisiae* (*20*), *Deinococcus radiodurans* (J. Neuhard, unpublished results), *E. coli* (*21*), *A. thaliana* I (*22*), and *A. thaliana* VI (S. Vincenzetti, unpublished results). The amino acid sequences were aligned using the program ClustalX (*23*) with the following parameters: gap opening, 5.00; gap extension, 0.5; delay divergent sequences, 50%; protein weight matrix, Gonnet series. The amino acid sequence alignment was imported into the program FarOut (*24*) and edited manually. The magnitudes of the intersubunit interaction surfaces were calculated with a probe of radius 1.4 Å using the program CNS (*12*).

Figures. Figure 2 was prepared with the program FarOut (*24*), Figures 3, 4, and 6 were prepared with the programs MOLSCRIPT (*25*) and Raster3D (*26*), and Figure 5a was prepared with BOBSCRIPT (*27*) and Raster3D (*26*).

Table 3: Buried Surfaces (in Å^2) between Subunits and Domains of T-CDA and D-CDA[a]

	T-CDA			D-CDA		
	C	D	B	C	D	B
A	2822	1159	1842	3269	1174	1663
A + B	7161		−	8714		−
A + C	−	5226		−	3947	

[a] The domains and subunits are labeled A, B, C, and D according to the cartoons in Figure 4.

RESULTS AND DISCUSSION

Overall Structure. The crystal structure of T-CDA from *B. subtilis* has been determined at 2.0 Å resolution by molecular replacement with the catalytic domain of *E. coli* D-CDA as search model. The enzyme crystallizes with two subunits (A and B) of the tetramer in the asymmetric unit. The subunits are related by a pseudo-two-fold axis parallel to the crystallographic *c*-axis. Almost all of the polypeptide chain and the side chains could be fitted into the electron density with the exception of the five C-terminal residues (132−136), which appear to be disordered even at 120 K. The subunit of T-CDA from *B. subtilis* is composed of a mixed β-sheet (β1−5) with one α-helix (α1) on one side and five α-helices (α2−6) on the other side of the β-sheet (Figure 3a). The model has been refined with a final *R*-factor of 20.7% (*R*-free 23.2%), and all residues are found in the allowed region of the Ramachandran plot. Each subunit contains 131 amino acid residues and binds a zinc ion and a THU molecule. The model also includes 183 water molecules of which 104 are related by NCS. There are no major differences between the two subunits; all Cα atoms of the two subunits A and B are superimposed with an rmsd of 0.02 Å using the program O (*13*) with default parameters.

The two subunits A and B in the asymmetric unit, i.e., half of the homotetramer, are related by a pseudo-two-fold symmetry. A perpendicular crystallographic 2-fold axis generates the two other subunits (C and D) and completes the homotetramer, giving an approximate overall *222* symmetry (Figure 3b). This gives rise to a complex set of intersubunit interactions. The entire surface of the tetramer subunit is 7034 Å^2, and the buried surface between subunits A and B, which form the asymmetric unit, is 1842 Å^2 (Table 3 and Figure 4). The buried surface between subunits A and C is 2822 Å^2, and this explains the importance of choosing the two correctly related subunits as a model in the molecular replacement search, since the interactions between the subunits are totally different. A correct solution could not be found using a model of subunits A and C, while it worked with subunits A and B.

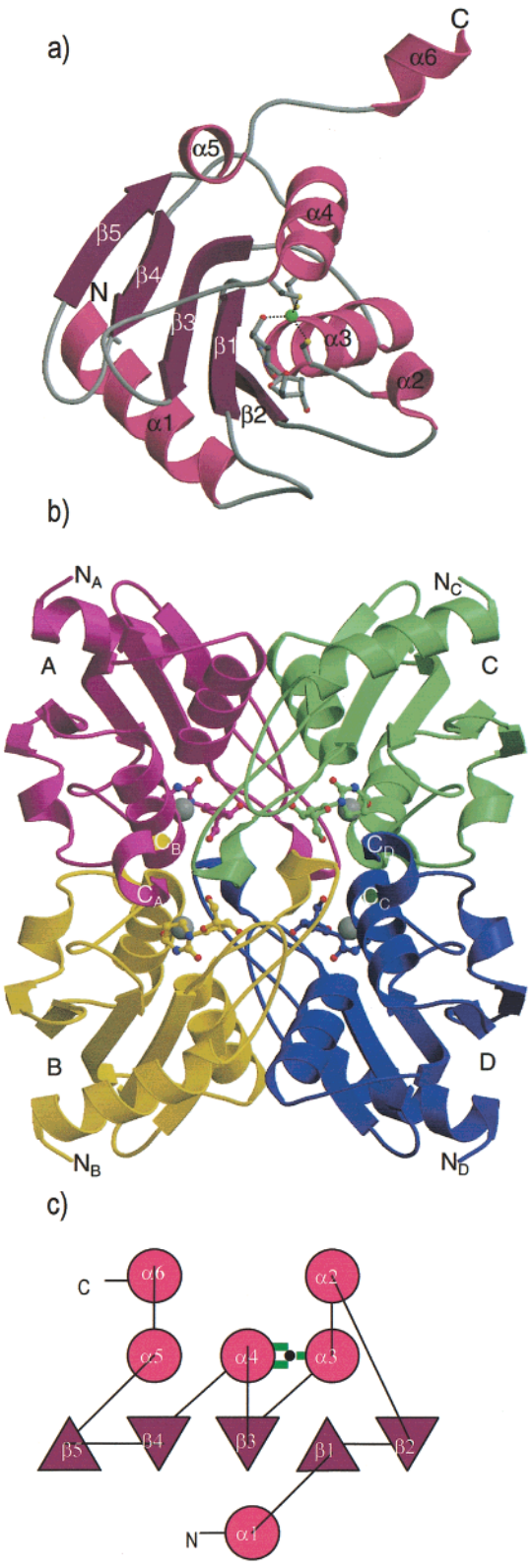

FIGURE 3: Fold of T-CDA from *B. subtilis*. Ribbon views of (a) the subunit and (b) the tetramer. The A, B, C, and D subunits are shown in magenta, yellow, green, and blue, respectively. The bound inhibitor and the zinc ligands are shown in ball-and-stick representation, and the zinc ions are shown as gray spheres. (c) Topology diagram of the T-CDA subunit. Green bars represent the zinc ligands.

Active Site. The catalytic zinc ion is coordinated by three cysteine residues at the N-terminal end of α-helices α3 (Cys53) and α4 (Cys86 and Cys89) (Figure 3), their negative charge being partly neutralized by the helix dipoles. The zinc ion is tetrahedrally coordinated with the bound inhibitor THU as the fourth ligand (Figure 5). The coordination distances range from 2.2 to 2.4 Å where the lowest value is the distance to the THU 4-hydroxyl group. The inhibitor THU is bound in an anti conformation and is completely buried in the interior of the enzyme in a pocket totally devoid of solvent. At 2.0 Å resolution, a well-ordered solvent molecule would be visible in a difference electron density map, especially since the surrounding atoms have low *B*-factors, implying that this part of the structure is well-ordered (Table 2). The absence of water in the T-CDA active site is in accordance with the crystal structure of *E. coli* D-CDA in complex with zebularine-3,4-hydrate (*28*). Zebularine-3,4-hydrate and THU differ at two points: the 2′-hydroxyl group of zebularine-3,4-hydrate and the C5−C6 bond which is unsaturated in zebularine-3,4-hydrate. Apparently, these two CDA inhibitors do not make room for a water molecule that was observed adjacent to C4 in the crystal structures of *E. coli* D-CDA in complexes with 3,4-dihydrozebularine (*28*), 3-deazacytidine (*29*), or uridine (*30*). In T-CDA, THU is bound to the polypeptide backbone and residues that are conserved to different degrees in the sequence alignment in Figure 2. The THU bound to subunit A also interacts with residues from subunits B and D; i.e., each THU molecule in the tetramer is in contact with three of the four subunits. The side chains of residues that make specific interactions (shorter than 3.9 Å) with THU are Phe24(A), Val26(A), Asn42(A), Glu44(A), and Phe125-(B). The main chain NH group of residue Tyr48(D) is hydrogen-bonded to the 5′-hydroxyl group of THU. Tyr48 is replaced by a Phe in some of the T-CDA amino acid sequences, while this residue is not conserved in the D-CDA amino acid sequences. The 3′-hydroxyl group of ribose in THU is hydrogen-bonded to the highly conserved Asn42 and Glu44; the former is invariant in the alignment of both T-CDA and D-CDA sequences in Figure 2. There are no obvious interaction partners in T-CDA for the 2′-hydroxyl group in cytidine, one of the substrates of CDA. This could explain why cytidine and deoxycytidine are equally good substrates and why the enzyme does not discriminate between ribose and deoxyribose. The other residues that interact with THU (Phe24, Val26, Glu44, and Phe125) are either conserved or changed to residues with side chains of the same size and properties in the other T-CDA and D-CDA amino acid sequences. The only sequence that does not conform to the pattern is the one from *Streptomyces coelicolor*. However, the protein corresponding to this sequence has not been studied, and its CDA activity may be questioned.

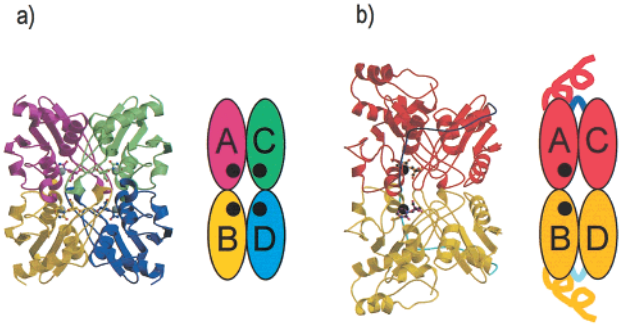

FIGURE 4: Cartoons of (a) T-CDA and (b) D-CDA that display the structures in equivalent views.

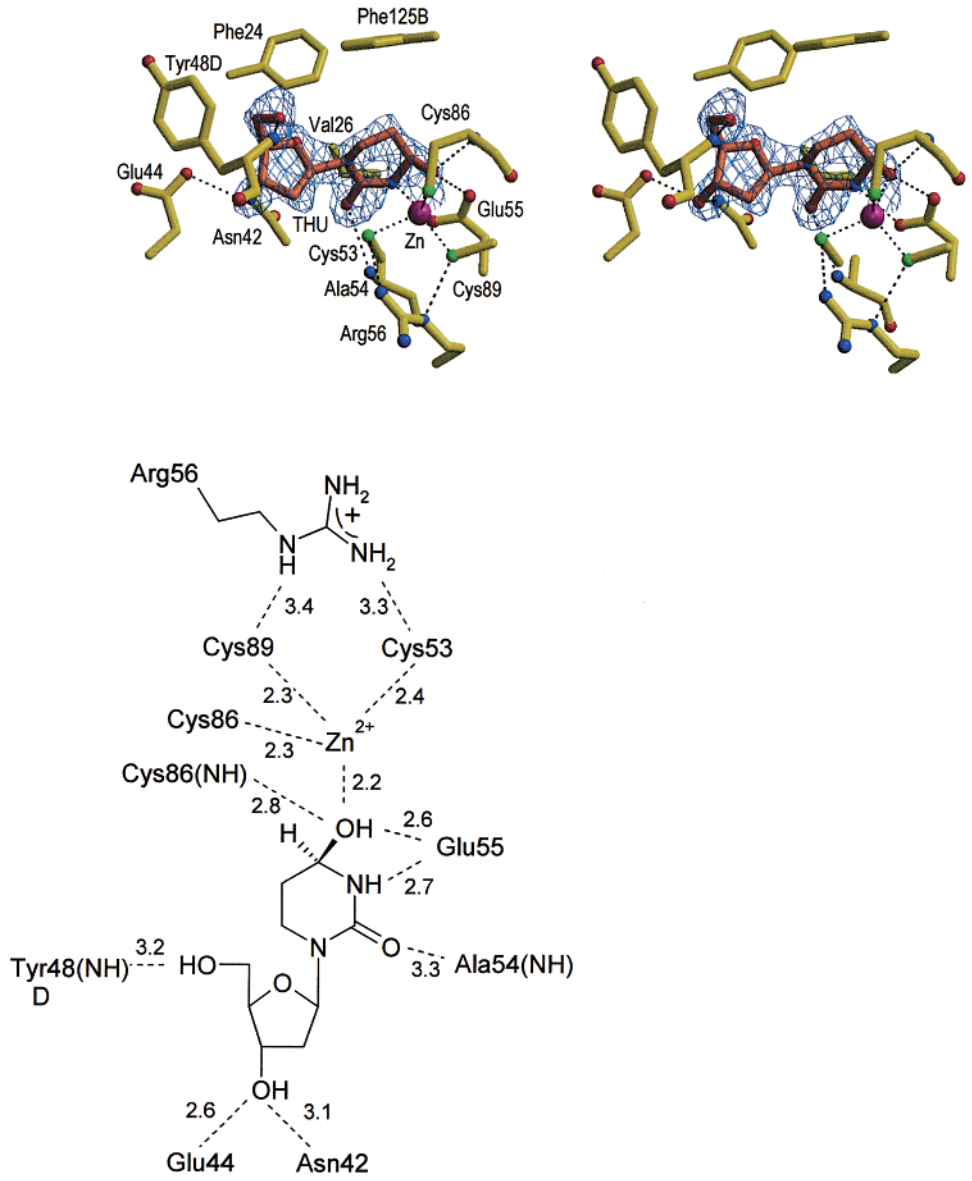

FIGURE 5: Active site of T-CDA from *B. subtilis* with the bound inhibitor THU. Hydrogen bonds and coordination bonds are shown as dotted lines. (a) Stereoview. The F_o-F_c omit electron density map of THU is contoured at 1.2σ. (b) Schematic view. The bond lengths are given in angstroms.

Despite the construction of the active site with residues from three subunits, T-CDA from *B. subtilis* does not display any cooperativity between its four active sites (*7*). The entrance for the substrate into the active site pocket of subunit A appears to require an opening of the C-terminal helix (α6) from the adjacent subunit B. Inspection of the *B*-factors shows that this part of the structure (residues 125−131) has a higher mobility ($B \approx 20$ Å^2) than the average of the structure ($B = 13$ Å^2). The following five C-terminal residues (132−136) display even more flexibility as they could not be traced in the electron density. Phe125 stacks with the ribose ring of the inhibitor THU, and residues 126−131 only make internal hydrogen bonds, which implicate that these residues easily can swing out and let the substrate in or the product out.

Sequence Alignment of CDA Sequences and Conserved Residues. Almost all of the residues that are conserved among the T-CDAs and D-CDAs in the amino acid sequence alignment of 58 potential CDA sequences in Figure 2 are involved in interactions that are important for the quaternary structure, substrate and zinc ion binding. Residues Cys53/His102, Cys86/Cys129, and Cys89/Cys132 from T-CDA from *B. subtilis*/D-CDA from *E. coli* serve as ligands for the zinc ion, and Glu55/Glu105 hydrogen-bonded to THU functions as proton donor and acceptor in catalysis. The conserved residues Phe24, Asn42, Glu44, Ala54, and Phe125 mediate substrate binding while Ser22, Arg90, Gln91, Glu95, and Leu121, which also are conserved, are engaged in tetramer interactions for T-CDA and in dimer or domain−domain interactions in the case of D-CDA. The conserved glycine residues (Gly27, Gly35, Gly40, Gly87) may be important for the geometry of the surrounding residues, such as the Gly87 between the zinc coordinating residues Cys86 and Cys89. Pro20 is present where the protein chain makes a turn, and Pro122 is at the end of α-helix α5. The two conserved alanine residues (Ala13 and Ala27) are both placed in the middle of secondary structural elements (α1 and β1)

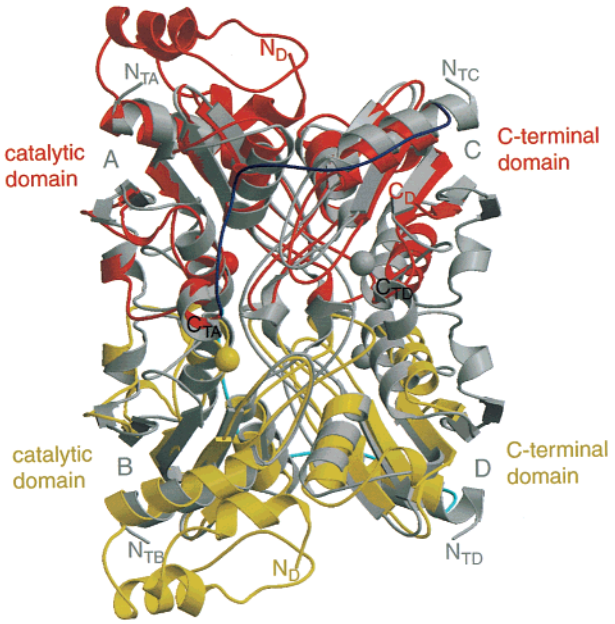

FIGURE 6: Ribbon diagram of the superimposed structures of *B. subtilis* T-CDA in gray and *E. coli* D-CDA (PDB code 1CTT) with its two subunits colored in red and yellow, respectively. The zinc ions of *E. coli* D-CDA are shown as red and yellow spheres, while the additional two zinc ions that are present in *B. subtilis* T-CDA are shown as gray spheres. The linker connecting the catalytic and the C-terminal domain of D-CDA is colored dark blue in the red subunit and cyan in the yellow subunit.

with their side chains pointed into the hydrophobic core of the protein.

Comparison of T-CDA and D-CDA. The overall structure of T-CDA from *B. subtilis* displays significant similarity to D-CDA from *E. coli*. A superposition of the entire *B. subtilis* homotetramer and the *E. coli* homodimer with the program O (*13*) using default parameters gives an rmsd of 1.7 Å for 422 Cα atoms. The two structures superimposed are depicted in Figure 6.

The D-CDA dimer is constructed from two identical subunits, composed of three domains. In the crystal structure, they are related by a crystallographic 2-fold axis perpendicular to the plane of the paper in Figure 6 (*3*). The additional 2-fold pseudo-symmetry evoked by the structural similarity of the catalytic and the C-terminal domain of D-CDA gives rise to an approximate overall *222* symmetry, similar to the symmetry of T-CDA. All secondary structural elements of *B. subtilis* T-CDA have corresponding secondary structural elements in the catalytic domain of *E. coli* D-CDA. A structural superposition of the Cα atoms in the catalytic domain of *E. coli* D-CDA with the A and B subunits of *B. subtilis* T-CDA using default parameters in the program LSQMAN (*31*) gives rmsd values of 1.3 and 1.4 Å with 116 and 119 Cα atoms superimposed, respectively. A sequence alignment based on the structural alignment shows 28% identity between the *E. coli* D-CDA catalytic domain and the *B. subtilis* T-CDA subunit. The polypeptide chain of the C-terminal domain of *E. coli* D-CDA is 29 residues shorter than the one of the catalytic domain, and as a consequence, this domain lacks the β-strand corresponding to β5 and the α-helices corresponding to α5 and α6. In this case, the sequence identity is 26% based on a structural alignment. A structural alignment of the C-terminal domain of *E. coli* D-CDA with either the A or the B subunit of T-CDA using default parameters in the program LSQMAN (*31*) gives 87 aligned Cα atoms with an rmsd value of 1.3 Å. However, there are some major differences between the D-CDA and the T-CDA structures. The C-terminal domain of D-CDA does not contain the zinc coordinating residues, and hence no zinc ion. As a consequence, there are in total two active sites per D-CDA as compared to the four active sites in T-CDA (Figure 6). Another significant difference is the N-terminal domain of the D-CDA subunit that is missing in T-CDA. The function of this domain is not resolved, and the structure of T-CDA has not thrown any new light on this aspect.

D-CDA may have arisen by a gene duplication of T-CDA, and pronounced structural similarity of the catalytic and the C-terminal domains of D-CDA strengthens this hypothesis. So does the presence of a 20 amino acid residue long linker that lacks secondary structure between the catalytic and the C-terminal domains of D-CDA (Figure 6). The magnitudes of the buried surfaces between subunits and domains in T-CDA and D-CDA are given in Table 3 and Figure 4. Not surprisingly do the covalently connected domains in D-CDA display a much more extended interaction surface than the dimer−dimer interaction. Several residues that constitute the interaction surfaces are highly conserved in the T-CDA and D-CDA amino acid sequences (Figure 2). This applies to residues Phe24, Arg90, Gln91, Leu121, and Phe125 in the A and B subunit interface of T-CDA and also to residues Tyr21, Ser22, Glu44, and Arg56 in the A and D subunit interface of T-CDA with two exceptions: Tyr21 and Arg56 are only conserved among T-CDA amino acid sequences. Despite the large interaction surface, there are fewer conserved residues between the A and C subunits in T-CDA. The conserved residues that are part of this surface comprise Tyr18, Tyr21, Asn45, and Glu95, and among these, only Glu95 is conserved in both T-CDA and D-CDA. As a remark, the residue corresponding to Tyr21 (Tyr207 in *E. coli* D-CDA numbering) is conserved in the C-terminal domain of D-CDA (sequence alignment not shown).

The zinc ion does not have the same ligands in D-CDA and T-CDA. In *E. coli* D-CDA, the ligands are one histidine (His102) and two cysteine residues (Cys129 and Cys132), while in *B. subtilis* T-CDA all three residues are cysteines (Cys53, Cys86, and Cys89). The function of the catalytic zinc ion in CDA is to activate a water molecule to form a hydroxide ion that in turn makes a nucleophilic attack on the C4 carbon in the pyrimidine ring of cytidine (*3*). Glu55 is a key residue in the proposed model for catalysis. The carboxylic group of this residue functions both as a proton acceptor and as a proton donor in the uridine-forming reaction, where the amino group of cytidine is replaced with an oxo group stemming from the water initially coordinated to the zinc ion.

The activity of an enzyme containing a zinc ion coordinated by three negatively charged cysteine residues as in T-CDA would be expected to be lower than one where the zinc ligands are two cysteines and a histidine residue. This is supported by the investigations of Bertini and co-workers in a series of ab initio Hartree−Fock calculations for zinc in complexes with different simple ligands, e.g., a hydroxide ion or a water molecule coordinated to the zinc ion together with three ammonia molecules and the corresponding

complexes where ammonia is substituted with CH_3S^- groups (*32*). These calculations show that the pK_a value of the bound water molecule increases with the number of thiol ligands and the formation of a hydroxide ion becomes less likely. It has been suggested that the third cysteine residue in T-CDA is positioned so it becomes a poor ligand for zinc (*7*). This is not seen in the *B. subtilis* T-CDA structure where the zinc−sulfur distances are virtually identical and the zinc ion and its ligands form an almost perfect tetrahedron. Another proposition, based on a model of T-CDA from D-CDA, was that Tyr48 should form a hydrogen bond to Cys53 and decrease its negative charge (*7*). The structure of *B. subtilis* T-CDA does not support this theory either. However, an arginine residue (Arg56) is hydrogen-bonded to two of the zinc ligands, Cys53 and Cys89. This important residue and the position of the cysteine residues at the positive end of the dipole of the α-helices α3 and α4 compensate part of the negative charges of the cysteine side chains and thereby lower the pK_a value of the coordinated water molecule.

CONCLUSIONS

The structure of tetrameric T-CDA from *B. subtilis* has confirmed the hypothesis that the overall structure of T-CDA is similar to D-CDA. The T-CDA lacks the N-terminal domain of D-CDA, and the tetramer subunit displays the largest degree of similarity to the catalytic domain of D-CDA. T-CDA contains four active sites in the homotetramer, while D-CDA has only two. Despite having very similar structural cores, the C-terminal domain of D-CDA does not contain the zinc coordinating residues and hence no zinc ion.

The zinc-liganding residues differ between the *B. subtilis* T-CDA and the *E. coli* D-CDA, being three cysteine residues in the first case and two cysteine residues and one histidine in the second. How the zinc ion in T-CDA is able to activate a water molecule and create the hydroxide ion used in catalysis has been an enigma due to the negative charges of the three coordinating cysteine residues. The three-dimensional structure of *B. subtilis* T-CDA revealed an arginine (Arg56) positioned at hydrogen bonding distances to two of the cysteine ligands (Cys53 and Cys89). The excess negative charge in the environment of the zinc ion is compensated by this residue and the positive charge from the dipoles of α-helices α3 and α4. Together this explains how the zinc ion is able to fulfill its catalytic role.

ACKNOWLEDGMENT

We thank Lisbeth Stauning and Flemming Hansen for excellent technical assistance and Dr. Silvia Vincenzetti for making her unpublished data available to us. We acknowledge access to synchrotron radiation at beamline 711, MAX-laboratory, Lund, and support from the EC Access to Research Infrastructure (ARI) program.

REFERENCES

1. Neuhard, J. (1983) in *Metabolism of nucelotides, nucleosides and nucleobases in microorganisms* (Munch-Pedersen, A., Ed.) pp 95−148, Academic Press, London.
2. Song, B. H., and Neuhard, J. (1989) *Mol. Gen. Genet. 216*, 462−468.
3. Betts, L., Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1994) *J. Mol. Biol. 235*, 635−656.
4. Carlow, D. C., Smith, A. A., Yang, C. C., Short, S. A., and Wolfenden, R. (1995) *Biochemistry 34*, 4220−4224.
5. Lipscomb, W. N., and Sträter, N. S. (1996) *Chem. Rev. 96*, 2375−2433.
6. Vallee, B. L., and Falchuk, K. H. (1993) *Physiol. Rev. 73*, 79−118.
7. Carlow, D. C., Carter, C. W., Jr., Mejlhede, N., Neuhard, J., and Wolfenden, R. (1999) *Biochemistry 38*, 12258−12265.
8. Mejlhede, N., Atkins, J. F., and Neuhard, J. (1999) *J. Bacteriol. 181*, 2930−2937.
9. Jancarik, J., and Kim, S. H. (1991) *J. Appl. Crystallogr. 29*, 584−587.
10. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol. 276*, 307−326.
11. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) *Acta Crystallogr., Sect. D D55*, 484−491.
12. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pann, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr., Sect. D D54*, 905−921.
13. Jones, T. A., Zou, J., Cowan, S., and Kjeldgaard, M. (1991) *Acta Crystallogr., Sect. A A47*, 110−119.
14. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr., Sect. D D50*, 760−763
15. Kleywegt, G. J., and Jones, T. A. (1996) *Acta Crystallogr., Sect. D D52*, 829−832.
16. Lascowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 282−291.
17. Cambi, A., Vincenzetti, S., DeSanctis, G., Neuhard, J., Natalini, P., and Vita, A. (2001) *Protein Eng.* (in press).
18. Tham, T. N., Ferris, S., Kovacic, R., Montagnier, L., and Blanchard, A. (1993) *J. Bacteriol. 175*, 5281−5285.
19. Kuhn, K., Bertling, W. M., and Emmrich, F. (1992) *Biochem. Biophys. Res. Commun. 190*, 1−7.
20. Kurtz, J. E., Exinger, F., Erbs, P., and Jund R. (1999) *Curr. Genet. 36*, 130−136.
21. Yang, C., Carlow, D., Wolfenden, R., and Short, S. A. (1992) *Biochemistry 31*, 4168−4174.
22. Vincenzetti, S., Cambi, A., Neuhard, J., Schnorr, K., Grelloni, M., and Vita, A. (1999) *Protein Expression Purif. 15*, 8−15.
23. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) *Trends Biochem. Sci. 23*, 403−405.
24. Madsen, D., and Kleywegt, G. J. (2001) http://alpha2.bmc.uu.se/dennis/manual/.
25. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
26. Merrit, E. A., and Murphy, M. E. P. (1994) *Acta Crystallogr., Sect. D D50*, 869−873.
27. Esnouf, R. M. (1997) *J. Mol. Graphics 15*, 133−138.
28. Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1995) *Biochemistry 34*, 4516−4523.
29. Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1996) *Biochemistry 35*, 1335−1341.
30. Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1997) *Biochemistry 36*, 4768−4774.
31. Kleywegt, G. J., and Jones, T. A. (1997) *Methods Enzymol. 277*, 525−545.
32. Bertini, I., Luchinat, C., Rosi, M., Sgamellotti, A., and Tarantelli, F. (1990) *Inorg. Chem. 29*, 1460−1463.

BI011849A