⌘ *Author's Choice*

Characterization of the Deoxynucleotide Triphosphate Triphosphohydrolase (dNTPase) Activity of the EF1143 Protein from *Enterococcus faecalis* and Crystal Structure of the Activator-Substrate Complex*[S]

Ivan I. Vorontsov[‡1], George Minasov[‡], Olga Kiryukhina[‡], Joseph S. Brunzelle[§], Ludmilla Shuvalova[‡], and Wayne F. Anderson[‡2]

From the ‡Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611 and the §Life Science Collaborative Access Team at the Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois 60439

The EF1143 protein from *Enterococcus faecalis* is a distant homolog of deoxynucleotide triphosphate triphosphohydrolases (dNTPases) from *Escherichia coli* and *Thermus thermophilus*. These dNTPases are important components in the regulation of the dNTP pool in bacteria. Biochemical assays of the EF1143 dNTPase activity demonstrated nonspecific hydrolysis of all canonical dNTPs in the presence of Mn^{2+}. In contrast, with Mg^{2+} hydrolysis required the presence of dGTP as an effector, activating the degradation of dATP and dCTP with dGTP also being consumed in the reaction with dATP. The crystal structure of EF1143 and dynamic light scattering measurements in solution revealed a tetrameric oligomer as the most probable biologically active unit. The tetramer contains four dGTP specific allosteric regulatory sites and four active sites. Examination of the active site with the dATP substrate suggests an in-line nucleophilic attack on the α-phosphate center as a possible mechanism of the hydrolysis and two highly conserved residues, His-129 and Glu-122, as an acid-base catalytic dyad. Structural differences between EF1143 *apo* and *holo* forms revealed mobility of the α3 helix that can regulate the size of the active site binding pocket and could be stabilized in the open conformation upon formation of the tetramer and dGTP effector binding.

Deoxynucleotide triphosphate triphosphohydrolases (dNTPases),[3] which hydrolyze deoxynucleotide triphosphates (dNTPs) into nucleosides and tripolyphosphate (PPP_i), may play a critically important role in cell survival through regulation of the pool of dNTPs. A striking example of the significance of such an activity came from a study of the productive infection of an *Escherichia coli* optA1 mutant by bacteriophage T7 carrying an additional phage gene 1.2 protein. This protein is able to inhibit the activity of the *Ec*-dGTPase enzyme (1). The inhibition is followed by a 200-fold increase of the dGTP concentration and resulted in host chromosome degradation. Here, we focus on a very distant homolog of the *Ec*-dGTPase, the EF1143 protein from *Enterococcus faecalis*. EF1143 demonstrates a 27% sequence identity over only 91 residues to the *Ec*-dGTPase, just a fifth of the full protein length, but belongs to the large, yet functionally unspecified, cluster #1 according to the Aravind-Koonin classification of HD-domain proteins (2). The HD/PDEase protein superfamily clan is defined by a highly conserved histidine and aspartyl reach H…HD…D motif participating in divalent cation binding (3). Although the activity and biochemical properties of *Ec*-dGTPase have been intensively studied since its discovery in 1958, no structure of the enzyme has been determined thus far (4, 5). In this article, we provide details of the characterization of EF1143 dNTPase activity and present the first crystal structures of the protein in *apo* form as well as complexed with dGTP activator and dATP substrate.

The structure of EF1143 with the dATP substrate provides important structural insights into a possible mechanism of the hydrolysis that is very unusual compared with the activities of other phosphohydrolases where one phosphate group or pyrophosphate is typically released in the course of reaction. Conformational differences between the EF1143 *apo* form and the structure with bound effectors suggests a possible mechanism for the dNTPase allosteric regulation. The first allosterically regulated dNTPase activity was only recently discovered in *Tt*-dNTPase from *Thermus thermophilus* (6). In contrast to the nonallosteric nature of *Ec*-dGTPase activity, it has been shown that for efficient hydrolysis, *Tt*-dNTPase requires the presence of three different dNTP substrates including dATP and dTTP as effectors. However, despite the extensive biochemical characterization, only the *apo* structure of the enzyme is known, and the location and existence of the special regulatory sites in its hexameric biologically active unit remain obscure (7). The

TABLE 1

Data and refinement statistics for the EF1143 protein from *E. faecalis*

	Native	Derivative (Hg^{2+})	Complex w/Ca^{2+}/dGTP/dATP
PDB ID	2O6I		3IRH
Space group	P3$_2$21	P3$_2$21	P2$_1$2$_1$2
a, b, c, Å	109.91, 109.91, 182.41	110.19, 110.19, 182.38	149.25, 188.66, 67.82
Unique reflection	38552	19765	71324
Test set	2049		3783
Resolution/last shell, Å	30.0-2.55/2.62-2.55	30.0-3.30/3.42-3.30	30.0-2.40/2.46-2.40
I/σ(I)	18.0/2.4	12.8/5.9	18.0/1.9
R_{merge}, %^{a}	8.0/50.0	10.1/33.1	8.0/46.0
Completeness, %	96.4/76.0	99.6/99.9	99.2/95.3
Overall redundancy	6.1/4.3	4.5/4.5	3.4/1.9
Refinement:			
R_{work}^{b}	24.9		18.5
R_{free}^{c}	31.7		24.4
B, Å^2	68.7		40.7
Rmsd bond length, Å^2	0.005		0.009
Rmsd bond angle, deg	1.08		1.21
Ramachandran distribution:			
Most favored	702		1492
Allowed	64		112
Outside allowed	2		4
Water molecules	297		563

^{a} $R_{\text{merge}} = \Sigma_h\Sigma_i|I_{h,I} - \langle I_h\rangle|/\Sigma_h\Sigma_i I_{h,i}$, where $I_{h,i}$ is the ith observation of the reflection h, whereas $\langle I_h\rangle$ is the mean intensity of reflections h.

^{b} $R_{\text{work}} = \Sigma||F_o| - |F_c||/|F_o|$.

^{c} R_{free} was calculated with a small fraction (5%) of randomly selected reflections.

EF1143 protein differs from the *Tt*-dNTPase in that the biological assembly is a tetramer. The structure of another tetrameric homolog of EF1143, the O67745 protein from *Aquifex aeolicus* (PDB ID 2HEK), is also known (8), but no functional annotation of this protein has been reported yet. This work describes biochemical, biophysical, and structural characterizations that were all performed on the same protein, thus, revealing the structural basis of the dNTPase activity.

The study of mechanisms for balancing dNTP concentration in the cell has an important practical side. The process of dNTP degradation opposes the process of their synthesis by ribonucleotide reductases (RNRs) (9, 10). The allosteric mechanism of RNR regulation has been shown to be very important for regulation of the DNA replication process and for cell proliferation, especially in cancer (11, 12). It could be speculated that allosterically regulated dNTPases might present potential targets for anti-microbial therapeutics since the sequential, structural and functional homologs of EF1143 are widely represented in bacteria and *E. faecalis* is itself a life-threatening pathogen resistant to many common antibiotics (13, 14).

EXPERIMENTAL PROCEDURES

Expression and Purification—The cloning, expression, and purification of the *E. faecalis* HD-domain protein EF1143 were carried out as described for other protein targets at the Midwest Center for Structural Genomics (15–18).

Protein Crystallization and X-ray Data Collection—For crystallization experiments the protein was used at a concentration of 8 mg/ml. Crystal screening was done in 96-well Corning crystallization plates. Diffraction quality crystals belonging to a trigonal space group P3$_2$21 were grown in 0.2 M di-sodium tartrate di-hydrate (pH 8.3), and 20% w/v polyethylene glycol 3350 in a 2-μl sitting drop at 22 °C. A mercury derivative was prepared by a 30 min soak of the native crystal at 22 °C in the well solution containing 2.5 mM HgCl$_2$. The protein stock solution for co-crystallization trials contained 5 mg/ml protein, 0.36 mM CaCl$_2$, 50 mM dGTP, 50 mM dATP, and 0.25 M NaCl.

Orthorhombic crystals with the space group P2$_1$2$_1$2 were obtained in 0.2 M sodium bromide, 0.1 M bis-Tris propane pH 7.5 and 20% w/v polyethylene glycol 3350 in a 2-μl sitting drop at 22 °C. Crystals were cryo-cooled in liquid nitrogen. Diffraction data were collected at 100 K at the SER-CAT (22-ID), SBC-CAT (19-ID), and LS-CAT (21-ID-F) beamlines at the Advanced Photon Source, Argonne National Laboratory. All the data were processed with HKL2000 (19).

Structure Determination—The *apo* structure was determined using the SIRAS method (20). Four mercury sites were found from anomalous maps using Phenix-HYSS (21). Positions, occupancies, and B-factors of these sites were refined and initial phases were calculated with FOM 0.26 (0.28 for centric reflections) at 3.5 Å with SHARP (22). The phases were modified by SOLOMON with assumption of 60% solvent content in the crystal and extended to 2.7 Å (23). An initial model was autotraced with ARP/wARP (24) and completed using iterative cycles of model building with TURBO-FRODO (25) and refinement with the REFMAC program (26). A structure of the EF1143 complex with dATP and dGTP was solved by molecular replacement using Phaser (27) and the chain A from the *apo* EF1143 structure as a search model. The structure was completed using Coot (28) and REFMAC. Crystallographic data and the refinement statistics are in Table 1.

Sequence and structure homology analyses were performed using VAST, BLAST, and ClustalX 1.8 (29–31). Graphical presentations were done with ALSCRIPT and PyMOL (32, 33).

NTPase and dNTPase Activity—All assays were carried out at room temperature. The reactions were initialized and later terminated by additions of the enzyme and 0.5 M EDTA to 5 mM of final concentration, respectively. A final volume of 500 μl was used for all reactions. Reaction mixture incubation times were 30 min unless otherwise stated.

A preliminary assay for nucleotide triphosphate triphosphohydrolase (NTPase) activity was performed on the mixture of four canonical ribonucleotide triphosphates: ATP, GTP, TTP,

and CTP (NTPs). The reaction mixture contained 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 1 mM MgCl$_2$, 500 μM MnCl$_2$, 50 μM ZnCl$_2$, 100 μM NTPs each, and 10 μM enzyme. A preliminary assay for the dNTPase activity was performed the same way except that a mixture of canonical dNTPs was used. Control mixtures did not contain protein. Reactions were incubated for 1 h. The activity was detected by monitoring the consumption of the substrates. Analysis and separation of substrates and products in the reaction and control mixtures were performed by FPLC anion-exchange chromatography using a MonoQ HR 5/5 column on an ÄKTAprime plus system (34). The column was equilibrated with 10 mM ammonium phosphate buffer (pH 7.5) and eluted at 0.7 ml/min flow rate with a linear gradient of ammonium phosphate in the concentration range 10–500 mM. Absorbance at 254 nm was used to register eluted compounds in all cases.

Assays with mixtures and single dNTPs were performed in a buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 1 mM MgCl$_2$, 500 μM MnCl$_2$, and 50 μM ZnCl$_2$. Reaction mixtures contained 5 μM enzyme and three, two, or one dNTP with 200 μM concentration each. After incubation the mixtures were injected to a MonoQ HR 5/5 column on an ÄKTAprime plus system. The column was equilibrated and eluted at 1 ml/min flow rate with a stepwise profile using 10 mM and 1 M ammonium phosphate buffers (pH 7.5), respectively.

Dependence of dNTPase activity on different divalent cations was analyzed in the presence of single metal ion species: Mn^{2+}, Mg^{2+}, Zn^{2+}, or Ca^{2+}. Reaction mixtures contained 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 200 μM of dATP, 200 μM of dGTP, 5 μM enzyme and one of the metal ions in the following concentrations: 500 μM of MnCl$_2$, MgCl$_2$, CaCl$_2$, or 50 μM ZnCl$_2$. Similar assays with combinations of two ion species (500 μM MnCl$_2$, 500 μM MgCl$_2$, or 50 μM ZnCl$_2$) were additionally performed. AcroSep Q HyperD F anion exchange column on an ÄKTAprime plus system and the stepwise elution scheme were used in these and all subsequent assays.

The effect of Mn^{2+} and Mg^{2+} on dNTPase activity with single dNTPs was analyzed in reaction mixtures containing 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 400 μM of single dNTP, 5 μM enzyme and 500 μM of MnCl$_2$ or MgCl$_2$. dNTPase activity with mixtures of dNTPs in the presence of only Mg^{2+} was analyzed in the buffer containing 50 mM Tris-HCl (pH 7.5), 100 mM KCl, and 500 μM MgCl$_2$. Reaction mixtures contained 5 μM enzyme and three or two dNTP with 200 μM concentration each.

Dependence of dNTPase activity on different deoxyguanosine nucleotide activators in the presence of Mg^{2+} was analyzed in reaction mixtures containing 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 200 μM dATP, 500 μM MgCl$_2$, 5 μM enzyme, and 200 μM either dGTP, dGDP, or dGMP.

Dynamic Light Scattering (DLS)—Experiments were done on a Zetasizer Nano ZEN1600 (Malvern Instruments) DLS photometer. Preliminary measurements were performed at 25 °C with samples containing 0.25–1.00 mg/ml protein, 50 ml Tris-HCl (pH 7.5) and 250 mM NaCl. Temperature dependence of the stability of the oligomeric unit in the presence of dGTP and Mg^{2+} was studied in the temperature range of 25–55 °C with 5 °C steps using a solution containing 0.50 mg/ml protein, 50 ml Tris-HCl (pH 7.5), 250 mM NaCl, 100 μM MgCl$_2$ and 200 μM

dGTP. Molecular weight of oligomers was estimated using Mark-Houwink-Kuhn-Sakurada (MHKS) relations for globular proteins (35).

RESULTS

Phosphohydrolase Activities—The presence of the H…HD…D divalent metal binding pattern in the sequence suggested that EF1143 might possess phosphohydrolase activity. Indeed, a generic phosphodiesterase assay with bis-pNPP confirmed that EF1143 is a phosphohydrolase. Because the HD-domain protein family is closely related to the well characterized 3′,5′-cyclic PDEase family we checked the activity of EF1143 with cAMP, a representative substrate for PDEases (supplemental Fig. S1*a*) (36). However, no significant activity was detected. Using the experimentally determined crystal structure (see below), a VAST search using only the extended C-terminal subdomain of EF1143 (residues 330–456) revealed its close similarity to the DNA-binding fragment of Bovine Papillomavirus-1 E2 protein (supplemental Fig. S1*b*) (37). This suggested a possible nuclease activity for EF1143 and this was tested. Its activity toward single- and double-stranded DNAs (0.7 × 10^{-3} and 0.4 × 10^{-3} ΔA$_{260}$/min/(mg/ml), respectively), was found to be six orders of magnitude less than the activity of the "true" nuclease from *Staphylococcus aureus* (1.3 × 10^3 ΔA$_{260}$/min/(mg/ml)) (38). Finally, iterative sequence analysis revealed a very distant similarity of EF1143 and the dGTPase from *E. coli* (39). A set of quantitative assays with canonical nucleotide and deoxynucleotide triphosphates was developed to test the corresponding activities. Significant triphosphohydrolase activity of EF1143 was detected only toward a mixture of dNTPs (supplemental Fig. S2). After 1 h of incubation of the EF1143 protein with a mixture of four canonical dNTPs almost all of them were hydrolyzed into triphosphate and nucleosides, which were eluted in the flow-through without binding to the anion exchange column; no mono- or diphosphate products were detected in this or any following reactions.

Substrate Specificity and Allosteric Regulation—To reveal a specificity of the dNTPase activity of EF1143 we analyzed its reactions with single canonical triphosphate deoxynucleotides and their mixtures. Reactions were performed with a mixture of three di-cations, Mg, Mn, and Zn, with relative concentrations of 20:10:1. The first surprising finding was that EF1143 does not hydrolyze single dNTPs (supplemental Fig. S3). Reactions with mixtures of dNTPs revealed that dGTP and at least one of two other nucleotides, dATP or dCTP, has to be present for efficient hydrolysis. Thus, dGTP plays a role as an activator. At the same time, hydrolysis of dTTP was insignificant even in the presence of dGTP (supplemental Fig. S4). The second and the most puzzling observation was that in the reaction with dATP and dGTP, the latter was also very efficiently consumed in the reaction meaning that dGTP is also a substrate. This was especially surprising because in the reaction with only dGTP it does not activate its own hydrolysis. Finally, in the presence of the Mg^{2+}/Mn^{2+}/Zn^{2+} mixture, the EF1143 protein efficiently hydrolyzes three deoxynucleotide triphosphates, dATP, dCTP, and dGTP, if dGTP and either dATP or dCTP are present.

Effect of Divalent Cations on Activity—To reveal the metal dependence of hydrolysis, in the next set of assays we used a

TABLE 2

Essential aspects of the dNTPase activity of EF1143 from *E. faecalis*

Efficiency of the hydrolytic activity roughly designated as follows: (+ + +) - about 90%; (+ +) - about 60–80%; (+) - less then 20% of the substrate was consumed in 30 min under the conditions tested (see "Experimental Procedures" and supplemental Fig. S3–S9).

		Substrate			
Metal	Activator	dGTP	dATP	dCTP	dTTP
Ca^{2+}/Zn^{2+}		$-^a$	−	−	−
Mn^{2+}		+ + +	+ +	+ +	+
Mg^{2+}		−	−	−	−
Mg^{2+}	dATP	+ + +	−	−	−
Mg^{2+}	dGTP	−	+ +	+ +	−
Mg^{2+}	dGDP	n/a^b	+	n/a	n/a
Mg^{2+}	dGMP	n/a	$+/-^c$	n/a	n/a

[a] −, activity was not detected.

[b] n/a, activity was not tested.

[c] +/−, measurement result is not conclusive.

mixture of EF1143 with dATP and dGTP. We checked the activity in the presence of Mg^{2+}, Mn^{2+}, Zn^{2+}, and Ca^{2+} ions (supplemental Fig. S5). Two reactions were used as controls: a reaction in the absence of metals and enzyme, and a reaction with the mixture of three metal ions, Mg, Mn, and Zn. No activity was detected in the reaction mixtures containing Zn^{2+} or Ca^{2+} (Table 2). Competition assays with Mn^{2+}/Zn^{2+} and Mg^{2+}/Zn^{2+} pairs additionally confirmed the inhibitory effect of Zn^{2+} (supplemental Fig. S6). In contrast, in the mixture with Mn^{2+} almost all substrates were consumed in 30 min (~90% based on the height reduction of the peak, corresponding to the mixture of dATP and dGTP). In the reaction with Mg^{2+} approximately a half of the substrates was hydrolyzed in the same period of time. In the control reaction with three ions the amount of dNTPs was reduced by ~80%. In conclusion, the enzymatic activity of the EF1143 protein efficiently proceeds in the presence of Mg^{2+} or Mn^{2+}, while Zn^{2+} or Ca^{2+} does not support hydrolysis.

Next, we tested the activity of EF1143 toward single dNTPs in the reaction mixtures with either Mg^{2+} or Mn^{2+} (Table 2). Rather surprisingly, in the presence of Mn^{2+} the EF1143 protein efficiently hydrolyzes all canonical dNTPs except dTTP (supplemental Fig. S7). After 30 min of reaction we observed ~70, 75, and 90% consumption of the initial amounts of dCTP, dATP, and dGTP, respectively. Hydrolysis of dTTP was detectable but the initial substrate reduction was less than 10%. Thus, in the presence of only Mn^{2+} EF1143 reveals nonspecific dNTPase activity without activator. In contrast, no hydrolytic activity with any single dNTP was detected in the presence of Mg^{2+}. Then, we repeated reactions with two dNTPs in the presence of only Mg di-cation (supplemental Fig. S8). In the reaction mixture with dATP and dGTP both substrates were hydrolyzed as was observed in the presence of the Mn^{2+}/Mg^{2+} mixture. In the reaction with dCTP and dGTP, the latter again activated hydrolysis of dCTP but the consumption of dGTP itself was not detected. No hydrolysis was observed in the reaction with dTTP/dGTP or in the mixture containing dATP, dCTP, and dTTP.

The features of EF1143 dNTPase activity in the presence of the Mn^{2+}/Mg^{2+} mixture could be summarized in the following way (Table 2). (i) dNTPase activity of EF1143 has two distinctive modes depending on the di-cation species. (ii) The EF1143 protein nonspecifically hydrolyzes single dNTPs in the presence of Mn^{2+} with the efficiency order dGTP>dATP≥dCTP≫dTTP. (iii) In the presence of Mg^{2+} EF1143 reveals dGTP-activated hydrolysis of two other dNTPs with the efficiency dATP>dCTP; in a mixture with dATP, dGTP is also getting consumed.

Effect of Activators on Hydrolysis—To test the importance of the triphosphate chain of dGTP for the activation, we assayed dATP hydrolysis in the presence of dGDP and dGMP (supplemental Fig. S9). We found that dGDP is able to initiate some hydrolysis though with significantly less efficiency compared with dGTP. Activation of the reaction by dGMP was insignificant (Table 2).

Overall, biochemical characterization of the dNTPase activity of the EF1143 protein suggests the existence of at least two binding sites in the biologically active unit: one site, for the reaction and another, for the protein activation. One of the possibilities to acquire multiple binding sites for EF1143 is to form an oligomer.

Oligomerization State in Solution—DLS experiments on 0.25–1.0 mg/ml solutions of the EF1143 protein at 25 °C revealed a single peak in the particle size distribution by volume with a hydrodynamic diameter of 11.8 and width of 3.0 nm. The molecular weight of the oligomer was estimated (using MHKS relations for globular proteins), to be 214.2 kDa, corresponding mainly to the EF1143 tetramer in agreement with the theoretically calculated value of 212.4 kDa. The presence of Mg^{2+} shifts the denaturation temperature from the interval of 34–41 °C to 43–46 °C. An addition of dGTP further stabilized the protein, increasing the denaturation temperature to the range of 49–52 °C. Thus, in solution EF1143 exists predominantly in a tetrameric oligomerization state, which can be additionally stabilized by Mg^{2+} and dGTP.

Overall Structure—The overall shape of the monomer, presented on Fig. 1*a*, could be described as a "bowl-with-handle" with an average diameter ~65 Å and a depth of ~30 Å (8). A central binding pocket is formed by helices α3, α4, α5, α6, α11, α13, and α14. Helix α4 (residue His-66), a loop between helices α5 and α6 (His-101 and Asp-111), and helix α11 (Asp-183) bring together a metal coordinating H…HD…D motif. The tertiary structure of EF1143 also includes four antiparallel β-sheets. The largest β-sheet consists of three β-strands (β6, β7, and β10) and separates the "bowl" from the "handle," represented by two helices α19 and α21. The role of the two-strand β-sheets consisting of β8 and β9 and β4-turn-β5 sheet is, most probably, to affix a long polypeptide stretch (residues 366–416)
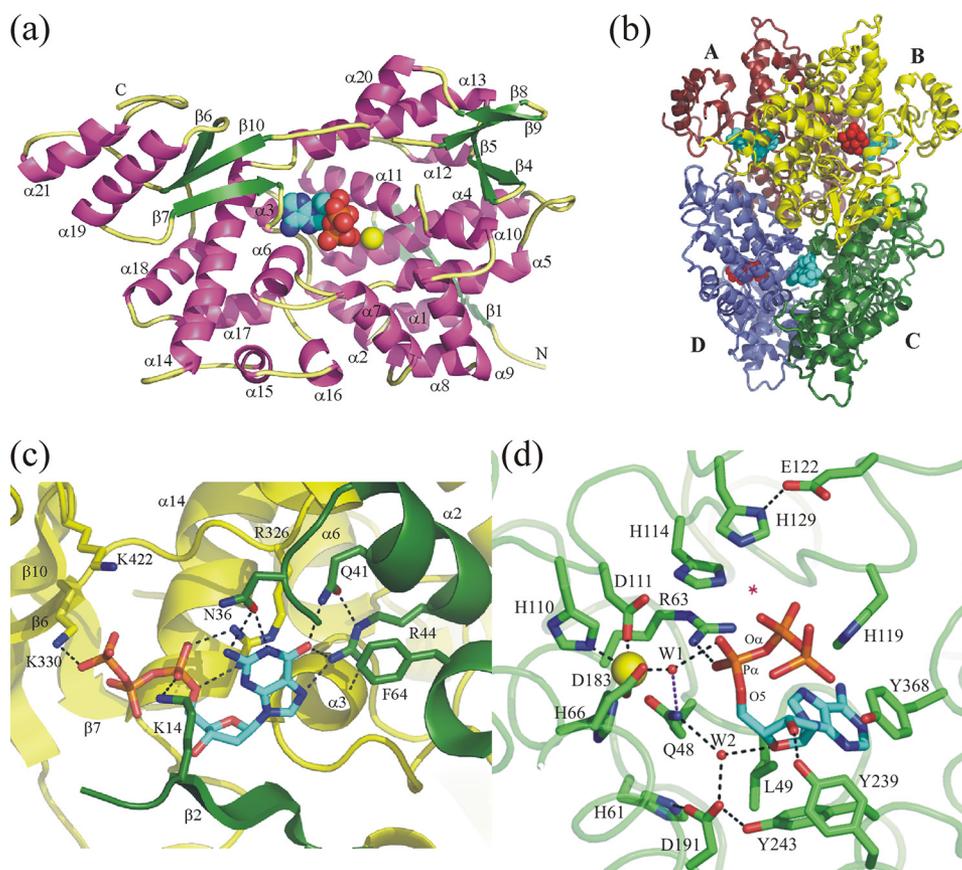
FIGURE 1. *a*, general view of the EF1143 polypeptide chain B from the *holo* (3IRH) crystal structure. The dATP substrate and Ca^{2+} ion are shown as *spheres*. *b*, EF1143 homotetramer with approximate 222 symmetry. dATP substrates and dGTP effectors are shown as *red* and *cyan spheres*, respectively. *c*, detailed view of the effector binding site in the interface of two polypeptide chains A (*green*) and B (*yellow*). One side of the binding pocket is formed by an antiparallel β-sheet (strands β2 and β3) and helices α1, α2, and α4 of one monomer, another, by the secondary structure elements of the adjacent polypeptide chain: helices α3 and α14, and loops α18→β6 and β10→α21. *d*, active site in chain B(3IRH). A *magenta star* designates the place where a water molecule could be activated and localized to perform an in-line nucleophilic attack on Pα-O5 bond. His-129–Glu-122 is a putative catalytic dyad. His-114 could assist in correct positioning of the attacking water. A possible function of highly conserved residues Gln-48, His-61, Asp-191, and Tyr-243 would be to provide H-bond network for the structural W2 water molecule which, together with Tyr-239, Arg-63, and W1, appears to play an important role in recognition and proper positioning of the deoxyribose unit of the substrate.

on the edge of the "bowl". The remaining β-sheet, β2-turn-β3, is located on the outer surface of the "bowl" and together with several helices plays a role in formation of the biological assembly.

The asymmetric unit of the *holo* (3IRH) crystal structure contains a EF1143 tetramer, consisting of chains A, B, C, and D (Fig. 1*b*). The tetramer has internal approximate 222 symmetry. Three helices, α2, α17, and α18 form a contact interface between chains A and B, and similarly between chains C and D. Also, there are two binding pockets on each interface A/B and C/D. In the *holo* structure all four "interface" binding pockets are fully (100%) occupied by dGTP molecules. The crystal structure of the EF1143 *apo* form (PDB ID 2O6I) has only two crystallographically independent polypeptide chains, which form a dimer equivalent to the A/D (B/C) pairs in the tetrameric biological unit. The *apo* crystal structure (2O6I) is built of EF1143 dimers.

A possible role of the "handle" (helices α19 and α21) remains unclear from the analysis of the EF1143 monomer and its biological assembly. The "handles" stick out of the tetramer surface and do not participate in formation of interfaces between monomers. Nevertheless, based on the weak nuclease activity

and high structural similarity of the EF1143 C-terminal domain to the DNA-binding domain of the Papilloma virus E2 protein (Fig. S1b) we could speculate that EF1143 may possess DNA-binding capability. More experiments are needed to test this hypothesis.

Identification of Binding Sites—Full occupation of all four "interface" binding sites was obvious at an early stage of the refinement (supplemental Fig. S10). The binding pocket is formed by residues on strand β2 and helices α2 and α4 of one chain together with the secondary structure elements β6, β7, α3, α14, and a loop α18→β6 of the neighboring chain (Fig. 1*c*). Residues Lys-14, Asn-36, Gln-41, Arg-44, Phe-64, Arg-326, and Lys330 are in direct contact with the ligand and make this site highly specific to dGTP. All these residues except Asn-36 are highly conserved (supplemental Fig. S11). Residue Arg-326 forms a stacking-like interaction with the guanine base and donates two hydrogen bonds to the α-phosphate group. The Arg-44 residue is involved in a network of H-bonds with the dGTP base, oxygen atoms of the α4 helix main chain and Gln-41. Positively charged residues Lys-14, Lys-330, and, very probably, Lys-422 participate in interactions with the triphosphate group. The high sequence conservation of Lys-330, which

forms an H-bond with the third phosphate group and is well ordered in structures of all four dGTPs, demonstrates that the "interface" binding sites are specifically designed to be occupied by the triphosphate nucleotide. This fact provides structural evidence that the triphosphate but not di- or monophosphate deoxyguanosine is the best activator of the dNTPase activity.

In both *apo* (2O6I) and *holo* (3IRH) forms of EF1143 the reaction binding sites were at least partially occupied by divalent cations. Metal species, Zn^{2+} in the *apo* and Ca^{2+} in the *holo* structures, were primarily assigned based on the coordination distances (supplemental Fig. S10). Importantly, in the *holo* structure (3IRH) only Ca^{2+} was added to the crystallization mixture to prevent cleavage of the substrate. In both cases coordination of the metal ions is close to square-based pyramidal (T_5) where H...HD...D motif occupies three basal apices (residues His-66, Asp-111 and Asp-183) and the apical apex (His-110). No water molecules which could complete this T_5 or similar octahedral (T_6) coordination were observed in the *apo* structure. In chains B and D of the *holo* structure we observed a water molecule W1 at the average Ca-O distance of 2.28 Å which is located at the second (opposite to the Nϵ of His110) apical position of the Ca^{2+} coordination sphere, in chain A two water molecules complete the metal coordination to distorted octahedron.

Only two out of four active sites in the EF1143 tetrameric unit in the *holo* structure 3IRH are occupied by substrates. Whether this observation is of functional significance remains to be determined. In the *holo* crystal structure the closest distances from the active sites to the crystallographic neighbors were found to be systematically smaller, but only by about 0.5 Å for the unoccupied active sites A and C. If this crystal environment difference is the source of the occupancy difference is not clear. In chains A and C continuous electron density in close proximity to the active site was assigned to the $\beta7{\rightarrow}\beta8$ loop (residues 366–385), not visible in the *apo* form 2O6I. Continuous electron densities in the active sites B and D were successfully fitted by dATP molecules with partial occupancies of ~2/3 (supplemental Fig. S10). The partial occupancy of dATP at the active site suggests that its binding is weaker than that of dGTP at the interface site. This is not surprising since the site has to accommodate different dNTPs. Additionally, if the binding of dATP at the apparently unoccupied A and C active sites was weaker by a factor of 2 or more compared with the B and D sites, the density there would be too weak to be interpretable. In contrast with the "interface" dGTP binding site, we detected only two direct H-bonds of the dATP ligand with the protein: between Tyr-239 and the hydroxyl group of the deoxyribose and between an oxygen of the α-phosphate group and Arg-63 (Fig. 1*d*). Instead, the nucleoside fragment forms several van-der-Waals contacts with residues of the secondary structure elements $\alpha3$, $\alpha6$, $\alpha11$, loops $\beta7{\rightarrow}\beta8$, $\alpha3{\rightarrow}\alpha4$, and $\alpha13{\rightarrow}\alpha14$, and, in particular, with residues Leu-49, His-119, Tyr-187, Tyr-243, and Tyr-368. Another oxygen of the α-phosphate group interacts with water molecule W1 which provides a bridge to the Ca ion (Ca...P$_\alpha$ distance is 5.4 Å). Another interesting feature of the binding in the active site is an *anti*-conformation of the dATP glycosidic bond, $\chi = 138.9°$, while in the "interface" site the dGTP nucleotide is bound in its *syn*-conformation, $\chi = $

65.5°. In conclusion, the binding pocket in the active center of EF1143 does not show clear preference to one particular base, in agreement with the enzyme ability to hydrolyze different canonical deoxynucleotides.

We did not observe any definite electron density corresponding to the water molecule which could be involved in the $O_{wat}{\rightarrow}P\alpha$-O5 in-line nucleophilic attack on the Pα center (Fig. 1*d*). However, there is just enough space between Pα and the closest residue in its apical direction, His-129, to secure a reactive solvent molecule in the "correct" place: Pα...Nϵ(His-129) distance is 6.0 and 5.7 Å in chains B and D, respectively. In this case, a His-129—Glu-122 couple could play the role of a catalytic dyad for the generation of the nucleophilic OH^- hydroxyanion, and a nearby His-114 residue could additionally assist in correct positioning of this attacking group: the averaged Pα...Nϵ(His-114) distance is 4.4 Å. In summary, the analysis of the reaction center in the *holo* structure revealed that EF1143 forms a deep hydrophobic binding pocket which can be occupied by a dATP nucleotide with the α-phosphate group positioned for possible in-line nucleophilic attack by a reactive water species activated by a putative His-129–Glu-122 catalytic dyad. These features are in agreement with the fact that EF1143 possesses dNTP triphosphohydrolase activity.

DISCUSSION

Structural and Sequential Comparative Analyses—Homologs of the EF1143 protein are well represented in bacteria. The closest homologs from the first twenty different species show more than 60% identity over the full-length protein, and all of them are annotated as putative HD-domain phosphohydrolases. At the same time, known biochemically characterized dNTPases including *Ec*-dGTPase, *Tt*-dNTPase and two *Pseudomonas aeruginosa* dNTPases (PA1124 and PA3043) have 16–34% homology to EF1143 over typically only 85–105 residues according to the BLAST search. Thus, EF1143 is the first biochemically characterized representative for a large group of its close (at least >60% identity) sequence homologs.

The characterized dNTPases mentioned above could be considered the functional analogs of EF1143. *Ec*-dGTPase is highly specific to dGTP but can also hydrolyze dTTP and noncanonical dITP, dUTP, and dXTP (40). Substrate specificity for PA1124 is somewhat similar to *Ec*-dGTPase and includes dGTP, dTTP, dITP, and 8-oxo-dGTP. PA3043 can hydrolyze all canonical dNTPs plus dUTP and dITP (41). None of these enzymes require any activator for the reaction to occur, except for the presence of Mg^{2+} or Mn^{2+} cations. In contrast, *Tt*-dNTPase not only shows different specificity for the substrates depending on the metal species (Mn^{2+} or Mg^{2+}), but has to be activated in the presence of Mg^{2+}. This enzyme exhibits no activity toward single dNTPs in the presence of Mg^{2+}, but in the mixture with dATP and dTTP the hydrolysis of all canonical dNTPs, as well as dITP, 8-oxo-dGTP and 8-oxo-dATP, can be induced (6). At the same time, with Mn^{2+} and Co^{2+} ions in the active site, any single dNTP can be degraded (42). The activity of EF1143 with canonical dNTPs is qualitatively similar to that of *Tt*-dNTPase: (*a*) degradation of all single dNTPs in the presence of Mn^{2+}, and (*b*) hydrolysis in the presence of Mg^{2+} requires activation. An interesting and puzzling peculiarity of
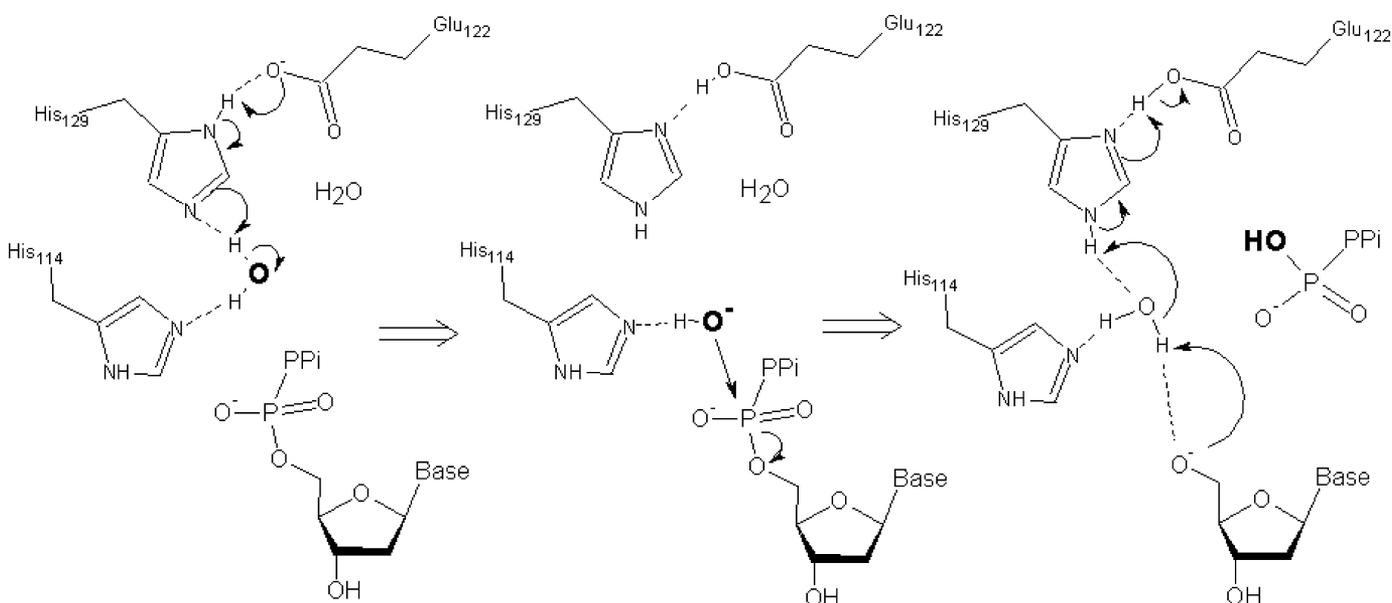
FIGURE 2. **Scheme of a possible mechanism of the hydrolysis.**

both enzymes is that in the latter case not only substrates but also effectors may be effectively consumed in the reaction. The differences, in particular, include, first, the order of specificity dGTP>dATP>dCTP in EF1143 *versus* dCTP>dGTP>dTTP> dATP in *Tt*-dNTPase, and, second, only one effector, dGTP, is needed in EF1143 *versus* two, dATP and dTTP, in *Tt*-dNTPase. At physiological conditions, where the concentration of Mg^{2+} typically exceeds that of Mn^{2+} by at least two orders of magnitude, both enzymes will probably work in the "allosteric regulation" mode (42). Thus, EF1143 presents a model for the allosteric regulation of the dNTPase activity.

Results of the structural and biophysical characterization of EF1143 and other dNTPases provide a possible explanation of the origin of the allosteric regulation mechanism. Currently, all dNTPases could be divided into two groups depending on their biological assembly (supplemental Figs. S11 and S12). Biologically active oligomers of PA1124, PA3043, *Tt*-dNTPase (PDB ID 2DQB) and its structural homolog from *Pseudomonas syringae*, *Ps*-dGTPase (PDB ID 2PGS), are hexamers, while *Ec*-dGTPase, EF1143 and its biochemically uncharacterized structural homologs from *Bacteroides thetaiotaomicron* (PDB ID 2Q14, sequence homology 29%), and *Aquifex aeolicus*, (PDB ID 2HEK, sequence homology 38%) form homotetramers. The more complex construction of the hexameric unit, a double ring of trimers, carries six active sites and at least two different types of voids at the polypeptide interfaces, which potentially could be the allosteric binding sites (7). This might explain a more sophisticated mechanism of the activity in *Tt*-dNTPase, but no complexes for hexameric dNTPases have been reported so far to test this hypothesis. At the same time, in the tetrameric enzymes (2Q14 and 2HEK) all interface binding sites contain, respectively, ADP and GDP molecules, which were accidentally co-crystallized with the proteins. In contrast, our *holo* structure, 3IRH, provides the first experimental evidence that allosteric binding sites in the interfaces can be specifically occupied by the triphosphate deoxynucleotide effector, dGTP, in accord with the EF1143 biochemical characterization.

Possible Mechanisms of the Reaction—The structure of the EF1143 active site provides clear evidence that the cleavage of the Pα-O5 bond of the dNTP substrate may occur according to the traditional "in-line nucleophilic attack" mechanism (43). dNTPase catalyzes the following reaction: $dNTP+H_2O = dN+PPPi$, which requires the presence of the hydrolytic water molecule or hydroxyl ion. A magenta "star" on Fig. 1*d* designates the area where a water molecule could be activated and localized in the apical Pα position for the in-line attack from the side opposite to the Pα-O5 bond. A possible scenario could include (i) coordination of the water molecule by H-bonds with Nϵ of His-114 and His-129, (ii) transfer of an H^+ to His-129–Glu-122 catalytic dyad and generation of HO^- reactive hydroxide, (iii) approach of HO^- for the attack that can be assisted by a preserved H-bond to His-114, (iv) nucleophilic attack on Pα accompanied by dissociation of the Pα-O5 bond and inversion of the phosphorus center, (v) departure of the PPPi group from the active center and back-donation of the proton from His-129—Glu-122 dyad to O5, most probably mediated by another water molecule replacing the attacking water species in the space between His-129 and dN (Fig. 2). All features of this scenario are in agreement with quantum mechanical and/or structure-based experimental models of hydrolysis proposed for several other phosphohydrolases. The dual acid-base role of the Glu-His dyad in generation of the HO^- hydroxide ion and protonation of the reaction product was described for the cyclic nucleotide hydrolysis in the phosphodiesterase (44). Hydroxide ion as a reactive species was proposed for the 5'-deoxyribonucleotidase activity in the *Ec*-YfbR protein (45). Water mediated proton shuttle mechanisms were studied in detail for GTP hydrolysis in RAS and RAS-GAP complexes (46, 47). Most importantly, a structure-based mechanism of EF1143 activity described here provides a testable hypothesis for theoretical and experimental site-directed mutagenesis studies. As a first step, one could check the catalytic function of the His-129–Glu-122 dyad (His135-Glu120 in *Tt*-dNTPase), which is absolutely conserved in structurally characterized putative tetra-
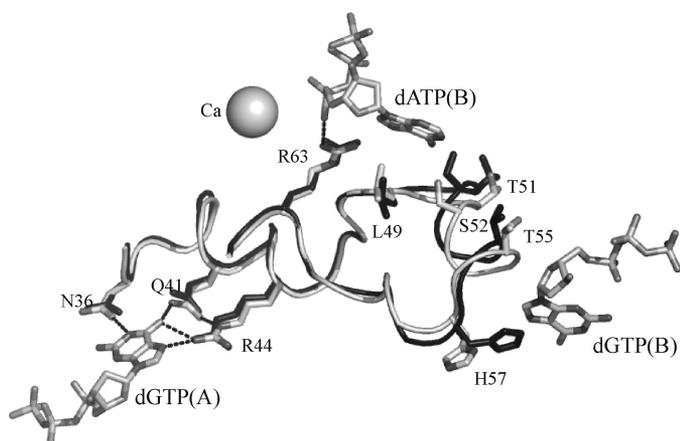
FIGURE 3. **Mechanism of the allosteric regulation of the EF1143 activity.** Superposition of the chain A(2O6I) (*black*) from the *apo* structure with the chain B(3IRH) (*gray*) from the *holo* form together with Ca²⁺, two dGTPs and dATP. Formation of the tetramer and binding of the dGTP(B) effector results in the stabilization of the α3 helix, residues 52–56, in the "open" conformation (*gray*). In the "closed" conformation (*black*) the Ser-52 residue protrudes its side chain into the interior of the binding pocket, thus, preventing docking of the substrate into the active site.

meric dNTPases and EF1143 close sequence homologs (supplemental Fig. S11).

In conclusion, the *holo* crystal structure, 3IRH, of the EF1143 protein proposes the first structure-based mechanism of dNTP hydrolysis, which can be used as a starting point for future more detailed studies of this enzymatic reaction.

Possible Mechanism of the Allosteric Regulation—Comparison of the *apo* form of EF1143 from the 2O6I structure with the *holo* form from 3IRH provides mechanistic insight into the active site alteration upon the tetramer formation and binding of the effectors. The asymmetric unit of the *apo* structure contains a dimer of two polypeptide chains which resembles the A/D part of the tetramer from the *holo* structure and, therefore, does not have interfaces with effector binding sites. In the *holo* structure all three nucleotides (two dGTP effectors and a dATP substrate) have contacts with the polypeptide fragment 36–48 which includes helices α2 and α3 (Fig. 3). A superposition of the *apo* chain A (2O6I) and *holo* chain B (3IRH) demonstrates that the influence of dGTP binding results in a shift of the α3 helix. One of the dGTP effectors opens the binding pocket by "drawing back" the α3 helix and its Ser-52 residue which otherwise would block the penetration of the dATP substrate into the site. This residue is well conserved among sequence and structural homologs with Thr and Ala as alternative amino acids (supplemental Fig. S11). In conclusion, formation of the tetrameric unit and binding of at least one effector change the size of the active center binding pocket and allows substrate docking.

REFERENCES

1. Huber, H. E., Beauchamp, B. B., and Richardson, C. C. (1988) *J. Biol. Chem.* **263,** 13549–13556
2. Aravind, L., and Koonin, E. V. (1998) *Trends Biochem. Sci.* **23,** 469–472
3. Finn, R. D., Tate, J., Mistry, J., Coggill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008) *Nucleic Acids Res.* **36,** D281–D288
4. Kornberg, S. R., Lehman, I. R., Bessman, M. J., Simms, E. S., and Kornberg, A. (1958) *J. Biol. Chem.* **233,** 159–162
5. Quirk, S., and Bessman, M. J. (1991) *J. Bacteriol.* **173,** 6665–6669
6. Kondo, N., Kuramitsu, S., and Masui, R. (2004) *J. Biochem.* **136,** 221–231
7. Kondo, N., Nakagawa, N., Ebihara, A., Chen, L., Liu, Z. J., Wang, B. C., Yokoyama, S., Kuramitsu, S., and Masui, R. (2007) *Acta Crystallogr. D Biol. Crystallogr.* **63,** 230–239
8. Oganesyan, V., Adams, P. D., Jancarik, J., Kim, R., and Kim, S. H. (2007) *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63,** 369–374
9. Jordan, A., and Reichard, P. (1998) *Annu. Rev. Biochem.* **67,** 71–98
10. Eklund, H., Uhlin, U., Färnegårdh, M., Logan, D. T., and Nordlund, P. (2001) *Prog. Biophys. Mol. Biol.* **77,** 177–268
11. Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R., and Thelander, L. (2003) *Cell* **112,** 391–401
12. Håkansson, P., Hofer, A., and Thelander, L. (2006) *J. Biol. Chem.* **281,** 7834–7841
13. Amyes, S. G. (2007) *Int. J. Antimicrob. Agents* **29,** S43–52
14. Letkiewicz, S., Miedzybrodzki, R., Fortuna, W., Weber-Dabrowska, B., and Górski, A. (2009) *Folia Microbiologica* **54,** 457–461
15. Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelly, M. I. (2002) *Protein Express. Purif.* **25,** 8–15
16. Dieckman, L., Gu, M., Stols, L., Donnelly, M. I., and Collart, F. R. (2002) *Protein Express. Purif.* **25,** 1–7
17. Rajan, S. S., Yang, X., Collart, F., Yip, V. L., Withers, S. G., Varrot, A., Thompson, J., Davies, G. J., and Anderson, W. F. (2004) *Structure* **12,** 1619–1629
18. Rajan, S. S., Yang, X., Shuvalova, L., Collart, F., and Anderson, W. F. (2004) *Biochemistry* **43,** 15472–15479
19. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
20. Xu, H., and Hauptman, H. A. (2003) *Acta Crystallogr. Sect. A* **59,** 60–65
21. Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., and Terwilliger, T. C. (2004) *J. Synchrotron Radiation* **11,** 53–55
22. Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494
23. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31,** 34–38
24. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
25. Roussel, A., and Cambillau, C. (1991) *TURBO-FRODO*, Silicon Graphics Applications Directory. Silicon Graphics, Mountain View, CA
26. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255
27. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40,** 658–674
28. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132
29. Gibrat, J. F., Madej, T., and Bryant, S. H. (1996) *Curr. Opin. Struct. Biol.* **6,** 377–385
30. Madden, T. L., Tatusov, R. L., and Zhang, J. (1996) *Methods Enzymol.* **266,** 131–141
31. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882
32. Barton, G. J. (1993) *Protein Eng.* **6,** 37–40
33. De Lano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA
34. Kremmer, T., Paulik, E., Boldizsár, M., and Holényi, I. (1989) *J. Chromatogr.* **493,** 45–52
35. Harding, S. E. (1997) *Prog. Biophys. Mol. Biol.* **68,** 207–262
36. Huai, Q., Colicelli, J., and Ke, H. (2003) *Biochemistry.* **42,** 13220–13226
37. Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) *Nature* **359,** 505–512
38. Cuatrecasas, P., Fuchs, S., and Anfinsen, C. B. (1967) *J. Biol. Chem.* **242,** 1541–1547
39. Beauchamp, B. B., and Richardson, C. C. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85,** 2563–2567
40. Seto, D., Bhatnagar, S. K., and Bessman, M. J. (1988) *J. Biol. Chem.* **263,** 1494–1499

41. Mega, R., Kondo, N., Nakagawa, N., Kuramitsu, S., and Masui, R. (2009) *Febs J.* **276,** 3211–3221

42. Kondo, N., Nishikubo, T., Wakamatsu, T., Ishikawa, H., Nakagawa, N., Kuramitsu, S., and Masui, R. (2008) *Extremophiles* **12,** 217–223

43. Frey, P. A. (1982) *Tetrahedron* **38,** 1541–1567

44. Salter, E. A., and Wierzbicki, A. (2007) *J. Phys. Chem. B* **111,** 4547–4552

45. Zimmerman, M. D., Proudfoot, M., Yakunin, A., and Minor, W. (2008) *J. Mol. Biol.* **378,** 215–226

46. Grigorenko, B. L., Nemukhin, A. V., Shadrina, M. S., Topol, I. A., and Burt, S. K. (2007) *Proteins* **66,** 456–466

47. Grigorenko, B. L., Nemukhin, A. V., Topol, I. A., Cachau, R. E., and Burt, S. K. (2005) *Proteins* **60,** 495–503