Structural and biochemical characterization of a novel Mn^{2+}-dependent phosphodiesterase encoded by the *yfcE* gene

DARCIE J. MILLER,[1,2] LUDMILLA SHUVALOVA,[1] ELENA EVDOKIMOVA,[3] ALEXEI SAVCHENKO,[3] ALEXANDER F. YAKUNIN,[3] AND WAYNE F. ANDERSON[1,2]

[1]Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Feinberg School of Medicine, Chicago, Illinois 60611, USA
[2]Drug Discovery Program, Northwestern University, Feinberg School of Medicine, Chicago, Illinois 60611, USA
[3]C. H. Best Institute, University of Toronto, Toronto, Ontario M5G 1L6, Canada

(RECEIVED January 15, 2007; FINAL REVISION April 1, 2007; ACCEPTED April 3, 2007)

Abstract

Escherichia coli YfcE belongs to a conserved protein family within the calcineurin-like phosphoesterase superfamily (Pfam00149) that is widely distributed in bacteria and archaea. Superfamily members are metallophosphatases that include monoesterases and diesterases involved in a variety of cellular functions. YfcE exhibited catalytic activity against bis-*p*-nitrophenyl phosphate, a general substrate for phosphodiesterases, and had an absolute requirement for Mn^{2+}. However, no activity was observed with phosphodiesters and over 50 naturally occurring phosphomonoesters. The crystal structure of the YfcE phosphodiesterase has been determined to 2.25 Å resolution. YfcE has a β-sandwich architecture similar to metallophosphatases of common ancestral origin. Unlike its more complex homologs that have added structural elements for regulation and substrate recognition, the relatively small 184-amino-acid protein has retained its ancestral simplicity. The tetrameric protein carries two zinc ions per active site from the *E. coli* extract that reflect the conserved di-Mn^{2+} active site geometry. A cocrystallized sulfate inhibitor mimics the binding of phosphate moeities in known ligand/phosphatase complexes. Thus, YfcE has a similar active site and biochemical mechanism as well-characterized superfamily members, while the YfcE phosphodiester-containing substrate is unique.

Keywords: enzymes; metalloproteins; crystallography; genomics-structural; phosphodiesterase

In any sequenced genome, 30%–50% of genes encode proteins with unknown biochemical or cellular function. The major challenge in the post-genomic era is to define both biochemical and physiological functions of all proteins as rapidly as possible.

Sequence analysis of *Escherichia coli* YfcE (P76495, GI 2,501,607) revealed the presence of a ''calcineurin-like'' motif (DxH/24x/GD/21–34x/GN/53–60/HxH), suggesting that it can have metallophosphoesterase activity (PEDANT database) (Frishman et al. 2003). The calcineurin-like superfamily (PfamPF00149) consists of enzymes with diverse functions believed to have evolved from a common ancestor (Koonin 1994). Members of this superfamily hydrolyze a wide variety of protein and nucleotide substrates, including protein phosphoserine phosphatases, nucleotidases, sphingomyelin phosphodiesterases, and 2′,3′-cAMP phosphodiesterases, as well as nucleases such as bacterial SbcD or yeast Mre11 (Koonin 1994; Aravind and Koonin 1998). Many of these proteins have been characterized biochemically. In addition, representative structures are known for several members of this superfamily, allowing for comparison of the overall structure and the location of the active-site residues. These

Reprint requests to: Wayne F. Anderson, Northwestern University, Feinberg School of Medicine, 303 E. Chicago Ave., Ward Bldg. Room 8-264, Chicago, IL 60611, USA; e-mail: wf-anderson@northwestern.edu; fax: (312) 503-5349.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.072764907.

proteins share a common β-sandwich architecture, which presumably has been retained to allow productive metal coordination. Many individual sequence families have acquired additional regulatory domains and secondary structure elements responsible for substrate recognition and catalysis. The relatively small 184-amino-acid YfcE protein is a prototype of a small subset of phosphatases within the superfamily that has maintained its ancestral simplicity. These YfcE sequence family homologs from bacteria and archaea are >30% identical and are expected to share the same biological function.

Results

Enzymatic studies of YfcE

Purified YfcE was screened for catalytic activity using several general enzymatic assays (Materials and Methods). In the presence of a $Mg^{2+}$and Mn^{2+} mixture, the protein showed significant activity toward bis-*p*-nitrophenyl phosphate (bis-*p*NPP), a general substrate for phosphodiesterases and nucleases. This protein also hydrolyzed two other artificial phosphodiesterase substrates, thymidine 5′-monophosphate *p*-nitrophenyl ester (*p*NP-TMP) and *p*-nitrophenylphosphorylcholine (*p*NPPC). Additional assays were performed to determine the pH and metal requirements for activity. With bis-*p*NPP, YfcE demonstrated maximal activity at pH 8.5–9.8 and had an absolute requirement for Mn^{2+} (Fig. 1). With Co^{2+}, the phosphodiesterase activity

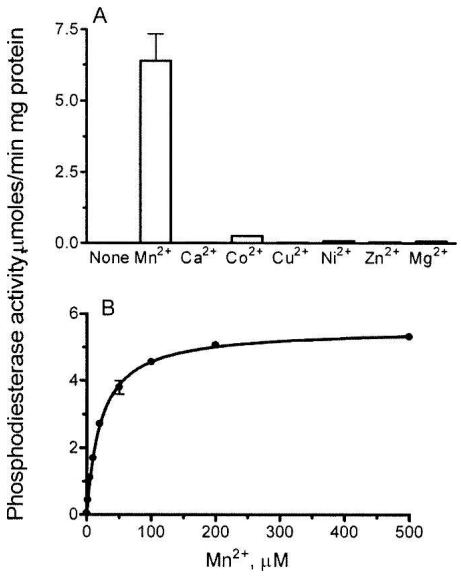

Figure 1. Metal dependence of *E. coli* YfcE phosphodiesterase activity. (*A*) Phosphodiesterase activity of *E. coli* YfcE in the presence of divalent cations. (*B*) Phosphodiesterase activity of YfcE as a function of Mn^{2+} concentration. Enzymatic assays were performed as described in the Materials and Methods and contained 0.66 nM of bis-*p*NPP and 1.75 μg of YfcE.

was 25 times lower than with that of Mn^{2+}, whereas other metals did not support the activity. When purified in the absence of divalent metal cations, the protein showed no activity without metal (Mn^{2+}) addition. After overnight dialysis against 50 mM HEPES-K buffer (pH 7.5) containing 1 mM Mn^{2+} with subsequent removal of Mn^{2+} by chromatography on Q-Sepharose, the phosphodiesterase activity of YfcE was no longer metal dependent and the protein showed the same high activity without Mn^{2+} addition. This indicates that YfcE has high affinity for Mn^{2+} and retains this metal after incubation. The apparent dissociation constant for Mn^{2+}, calculated from the dependence of phosphodiesterase activity of YfcE on Mn^{2+} concentration (Fig. 1B), was determined to be $K_D = 5.56 \pm 0.06$ μM (Table 1).

The dependence of the velocity of the reaction on the substrate concentration (bis-*p*NPP, *p*NP-TMP, and *p*NPPC) in the presence of 100 μM Mn^{2+} is presented in Figure 2. YfcE had low affinity for these artificial substrates with bis-*p*NPP > *p*NP-TMP > *p*NPPC. Only for bis-*p*NPP were saturation kinetics observed, and the concentration dependence fitted the Michaelis-Menten equation (Fig. 2A). Specific activities with *p*NP-TMP and *p*NPPC were at least 10 times lower, and no saturation was observed at high substrate concentrations (40–80 mM). The *E. coli* YfcE had the same high K_m values for bis-*p*NPP (9.74 mM) (Table 1) as several eukaryotic phosphodiesterases (BRENDA database) (Schomburg et al. 2002).

In an attempt to define a physiologically relevant YfcE substrate, the biochemical activity of YfcE against naturally occurring phosphodiesters was also tested. These included candidates from the three main classes of naturally occurring phosphodiesters: nucleic acids, cyclic nucleotides, and phospholipids. The *E. coli* YfcE showed no nuclease activity against single-stranded M13 DNA, double-stranded λ DNA, *E. coli* genomic DNA, or total ribosomal or transfer RNA (data not shown). Moreover, no phosphohydrolase activity was found toward various 2′,3′- and 3′,5′-cyclic nucleotides or phosphatidylcholine. In addition, no YfcE phosphatase activity was observed using a panel of 57 natural phosphatase substrates (nucleotides, phosphorylated sugars, organic acids, and amino acids). Taken together, these results indicate that the *E. coli* YfcE is highly specific for its physiological substrate.

Structure determination and model quality

Since diffraction quality crystals of seleno-methionine YfcE crystals were not obtainable, a search of heavy atom derivatives was performed. Despite the detection of an anomalous signal for some of these data sets, a substructure solution initially remained elusive due to incorrect

Table 1. *Kinetic parameters for* E. coli *YfcE*

Variable substrate	K_m (mM)	V_{max} (μmoles/min mg)	k_{cat} (sec^{-1})	k_{cat}/K_m (M^{-1} sec^{-1})
Mn^{2+}	0.022 ± 0.001	5.56 ± 0.06	2.1 ± 0.02	9.63×10^4
bis-*p*NPP	9.74 ± 0.65	52.4 ± 1.47	19.8 ± 0.56	2.03×10^3
bis-*p*NPP (+ PO_4^{3-})[a]	18.3 ± 1.2	53.49 ± 1.83	20.3 ± 0.69	1.11×10^3

[a] Measured in the presence of 0.5 mM phosphate.

space group assignment. YfcE crystals belong to the space group $P3_1$ and include a tetramer per asymmetric unit. Initially, however, the crystals appeared to be of higher symmetry ($P3_121$, dimer per asymmetric unit) as judged by a self-rotation search and estimates of symmetry from merging statistics. For example, processing data from crystal 1 in $P3_1$ gives an R_{merge} of 5.7 and I/σ of 25, while processing in $P3_121$ gives an R_{merge} of 5.9 and I/σ 31.6. Self-rotation searches indicated twofold peaks perpendicular to the crystallographic threefold. The peak heights at kappa = 180° are 99% of the kappa = 120° origin peak. Although this suggested the crystals were $P3_121$, the twofolds represent near-perfect pseudosymmetry, not crystallographic symmetry. The rotational operator that relates dimers A-B and C-D is locally perfect, is parallel to a crystallographic axis, and lies in a plane perpendicular to the crystallographic 3_1 axis, but it does not intersect a 3_1, as required for the higher symmetry space group. The noncrystallographic twofold is offset by ~20 Å from the 3_1 axis at the origin.

A Pt derivative was successfully used for initial phasing to 2.8 Å resolution by single isomorphous replacement with anomalous scattering (SIRAS), with one Pt bound for each of the four polypeptides in the asymmetric unit. Following density modification, phase extension to 2.5 Å, and automated electron density interpretation, an 80% complete model was obtained. A better-quality native crystal, which was less isomorphous than the original Hg soaked ''native'' crystal with the Pt crystal, was used for structure refinement after rigid body refinement of the initial model and phase extension.

The final 2.25 Å resolution model contains 732 protein residues, 362 waters, and six SO_4^{2-} from the 0.2 M ammonium sulfate crystallization buffer. In addition, eight Zn^{2+} were modeled, two bound per active site along with one sulfate. Zinc was not added to the crystallization conditions and presumably bound the protein from the *E. coli* cell extract and was carried through the purification. The predominance of zinc over other divalent cations was judged by the overwhelming zinc fluorescence signal detected from crystals used for data measurement (data not shown). The electron density in the anomalous difference Fourier is also consistent with zinc binding. Nickel from the metal-chelating affinity purification protocol was not able to replace the zinc. The inability of manganese, the biochemically relevant divalent cation, to displace zinc during cell growth may simply reflect the relative abundance of zinc (Finney and O'Halloran 2003).

Due to poor electron density, the N-terminal His-tag was not modeled, as well as the first and last residues of chains C and D, and the side chain of D183. According to PROCHECK (Laskowski et al. 1993), 84.5% of residues were in the most favored Ramachandran plot region, 14.2% were in the additionally allowed region, 0.3% were in the generously allowed region, and residues His127 (A-D) and Asn163 (A,B) were in the disallowed region. As will

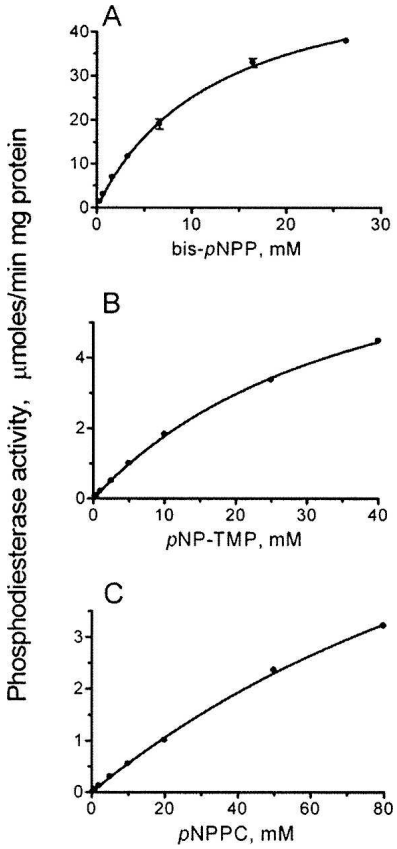

Figure 2. Phosphodiesterase activity of *E. coli* YfcE as a function of substrate concentration. (*A*) bis-*p*NPP; (*B*) *p*NP-TMP; and (*C*) *p*NPPC. Reactions were performed as described under the Methods and Materials in the presence of 0.5 mM Mn^{2+} and contained 0.07 μg YfcE (for bis-*p*NPP) or 1.75 μg YfcE (for *p*NP-TMP and *p*NPPC).

be discussed below, His127 is in the metal coordinated active site, and the geometry observed presumably reflects that required for direct zinc binding. Asparagines A163 and B163 are located in a tight β-turn. The deviations of their geometry may result from the close proximity of symmetry-related molecules. Asn163 of molecules C and D are exposed to solvent and do not have unfavorable main-chain torsion angles. The differing protein environments also have subtle effects on the overall similarity of subunits. Molecule A is more similar to B than to C or D. The RMSD between molecule A and B is <0.1 Å in Cα positions. In contrast, the RMSD between molecule A and C is ~1.0 Å in Cα positions. The density for molecule A is generally better defined and thus will be the chain described unless otherwise stated. The final R_{work} and R_{free} were 21.3% and 25.2%, respectively. A summary of data collection and refinement statistics is given in Table 2.

Effect of sulfate and phosphate on YfcE activity

The presence of sulfate in the YfcE active site suggested sulfate or phosphate may mimic binding of a natural ligand. Therefore, the effect of these ions on hydrolysis was tested. Sulfate had a significant inhibiting effect on the hydrolysis of bis-*p*NPP by YfcE (K_i = 25.3 mM). However, 5%–10% residual YfcE activity was observed in the presence of saturating sulfate concentrations. Phosphate inhibited YfcE activity at least 20 times more strongly (K_i = 0.57 mM) (Fig. 3). The inhibition by phosphate had competitive character and resulted in a significant increase of K_m for bis-*p*NPP (Table 1).

Topology

A ribbon diagram and Cα trace of the YfcE A chain is shown in Figure 4, A and B, respectively. The protein consists of 11 β-strands and three α-helices. The first 82 residues make up a typical Rossmann fold (β1-α1-β2-α2-β3-α3; parallel β-strands). Residues 91–143 form a β-sheet with order β4-β5-β6-β8-β7. C-terminal β-strands 9–11 have an antiparallel arrangement and make up an extended β-sheet with strands 1–3 (order 3–2-1–9-10–11). This six-stranded β-sheet is situated parallel to the five-stranded β-sheet.

Sequence conservation

YfcE shares up to 28% sequence identity with diadenosine tetraphosphatases (cluster of orthologous groups COG0639) (Tatusov et al. 2001) and 20%–27% identity with a variety of other phosphatase families. This high degree of sequence similarity is due to the strong conservation of metal binding residues and suggests certain biochemical parallels may be drawn between YfcE and other phosphatase families. However, our biochemical studies indicate the unknown target of YfcE phosphoesterase activity is novel. In order to find conserved amino acids with potential importance for substrate selectivity, a search for ≥30% YfcE sequence homologs was performed using PSI-BLAST (Altschul et al. 1997). There are fewer than 20 sequences with ≥30% sequence identity to YfcE, representing 10 species. The result of the alignment generated using Multalin (Corpet 1988) is shown in Figure 5. All but one of these proteins are from a fully sequenced genome and belong to COG0622.

Table 2. *Data measurement and refinement statistics*

Crystal details	C1	C2 (Pt)	C3 (native)
Space group $P3_1$ (no. 144)			
Cell: α,β = 90.0°, γ = 120.0°	a,b = 71.05, c = 173.84 Å	a,b = 70.84, c = 174.01 Å	a,b = 71.02, c = 173.68 Å
Processing	C1	C2	C3
Wavelength (Å)	0.9787	0.9785	0.9794
Resolution limit (Å)	2.8	2.8	2.2
Reflections (unique)	320,628 (24,041)	388,999 (24,050)	691,292 (48,352)
% Complete (last shell)	99.9 (99.3)	99.9 (99.7)	100.0 (100)
I/σ (last shell)	25 (15.0)	15 (10.0)	5.9 (2.5)
% R_{merge} (last shell)	5.7 (12.4)	6.4 (10.9)	9.5 (29.4)
Phasing power (last shell)		isomorphous: 0.956 (0.380); anomalous: 0.45 (0.13)	
Refinement			
Resolution (Å)		20–2.25	
No. reflections		46,449	
%R_{work} (% R_{free})		21.3 (25.2)	
RMSD		bonds 0.007 Å, angles 1.5°	
No. protein residues		732 of 828 (missing residues: 4 from YfcE and 92 from His-tag)	
No. other molecules		362 waters, 6 sulfates, 8Zn^{2+}	
Mean B factor (Å^2)		37.1	

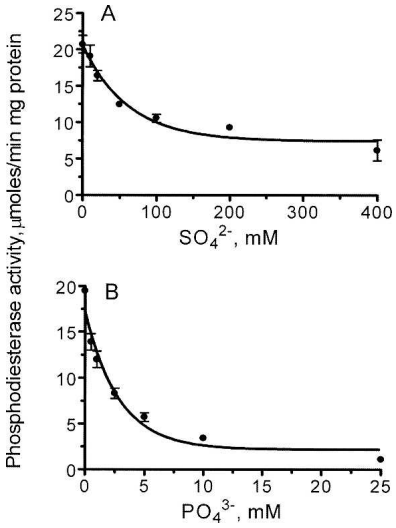

Figure 3. Inhibition of phosphodiesterase activity of *E. coli* YfcE by sulfate (*A*) or phosphate (*B*). Reaction mixtures contained 0.5 mM Mn^{2+}, 6.6 mM of bis-*p*NPP, and 0.14 μg of YfcE, and assays were performed as described in the Materials and Methods.

From the sequence alignment and YfcE structure, seven sequence motifs were identified as having functional significance. Several motifs contain residues directly involved in metal-coordination and reside in loops following β-strands. These include motif I after β1 (^{8}SDxHG; uppercase = identical, x = any), motif II after β2 (^{35}&GDxLN/YHGpRN; & = hydrophobic, lowercase = conserved), motif III after β3 (^{71}rGNCDseVDqm&&), motif IV after β5 (^{105}HGh), and motif V before β7 (^{126}GHTH&P). Motif VI (^{142}&NPGS) presumably has an indirect affect on metal binding via stabilization of motif V residues. The consensus motifs II and III are similar to many phosphatases (&&x&GDx&& and GNH[ED], respectively) (Koonin 1994). In addition to metal binding, motifs II and III contain oligomerization and predicted phosphate binding residues. Motif VII (^{51}YxPxxV/LAe) from α2 also helps stabilize the tetramer. A more detailed discussion of tetramer formation and phosphate binding follows. Refer to Figure 5 for location of these motifs within the YfcE sequence and secondary structure.

Oligomeric state

According to gel-filtration experiments, the native molecular mass of YfcE was 79.4 ± 2 kDa, indicating that the protein exists as a tetramer in solution. This is consistent with the crystal structure. A Cα trace of the tetramer colored according to the level of sequence conservation among ≥30% sequence homologs is shown in Figure 6. The A-C (or B-D) interface contains residues from oligomerization motifs II, III, and VII. The interactions are mostly van der Waals, with a few important exceptions. Invariant residues His41 and Tyr51 hydrogen bond using side-chain heteroatoms. The backbone NH of invariant Asn73 makes a hydrogen bond to the backbone carbonyl of invariant Val78. As will be discussed later, this interaction suggests maintenance of the A-C interface is uniquely coordinated to proper metal and phosphate binding, as Asn73 binds the metal and sulfate inhibitor directly, while Val78 is in van der Waals contact with the sulfate. Semiconserved residues Arg71 and Glu77 make hydrogen bonds between side chains. The buried surface area is 1632 Å^2 (B-D = 1581 Å^2). The A-D (or B-C) interface does not contain strictly conserved residues. It is held together by close hydrogen bonds between

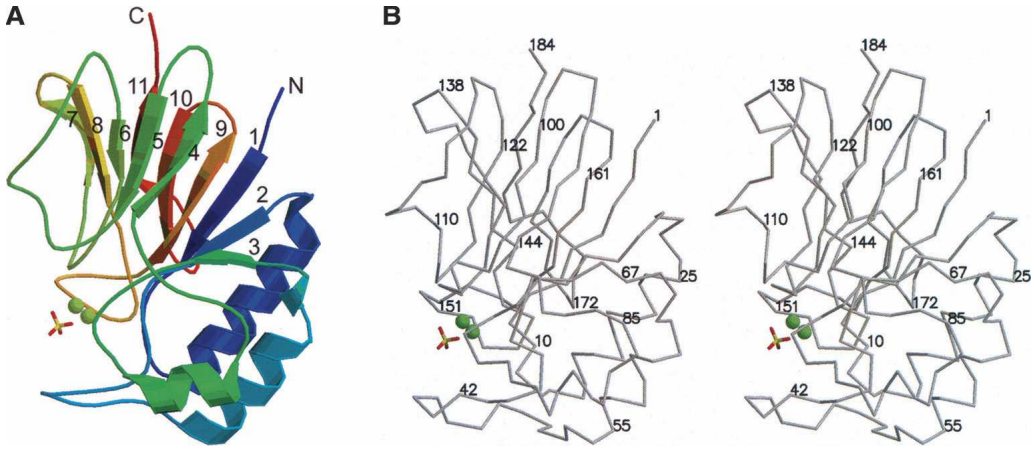

Figure 4. Structure of YfcE A chain. (*A*) Ribbon drawing of YfcE. The protein is colored blue to red from N to C terminus. β-Strands are numbered according to the order in which they appear in the primary sequence. The active site is defined by two bound zincs and a sulfate, which are represented as green cpk atoms and ball-and-stick atoms, respectively. (*B*) Stereogram of YfcE Cα trace.

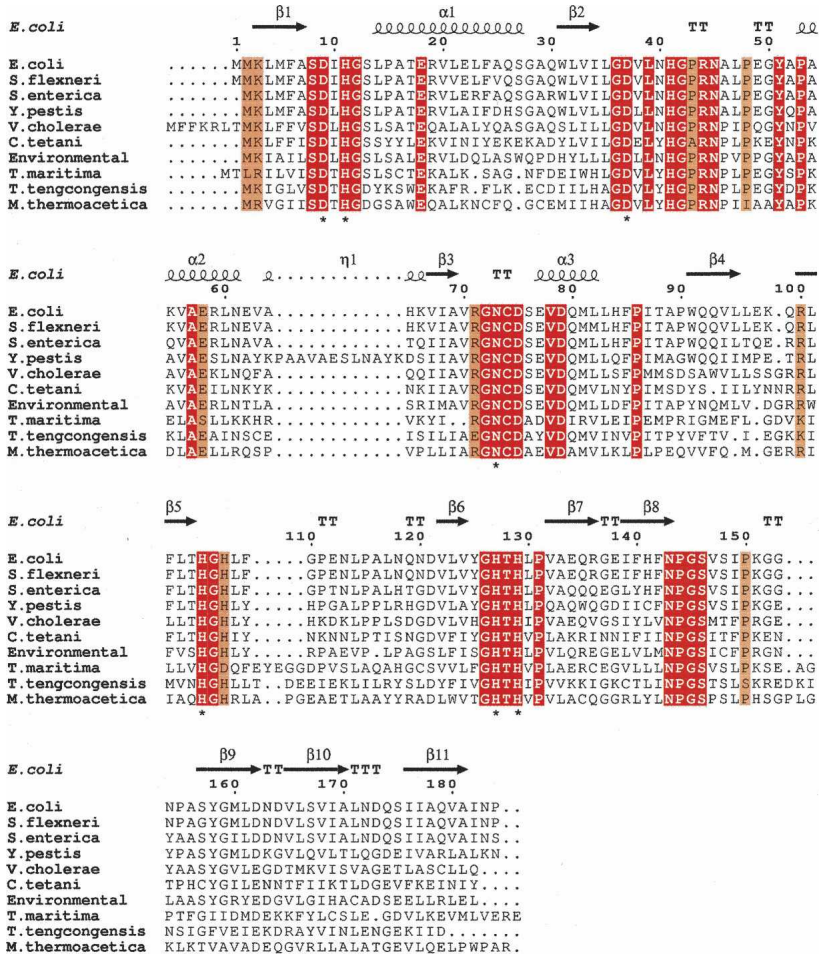

Figure 5. YfcE sequence alignment. The sequences shown share ≥30% identity to YfcE. The sequence numbering corresponds to that of YfcE. Secondary structure elements are labeled *above* the residues in which they appear. Invariant residues are highlighted in solid red background. A subset of 90% identical residues is indicated with a solid orange background. Asterisks located *below* the aligned sequences denote residues that directly bind zinc.

backbone atoms of β-strand 4, creating a buried surface area of 2087 Å^2 (B-C = 2099 Å^2). This contact surface does not have strict sequence preference but does contain residues with similar hydrophobic properties.

Discussion

E. coli YfcE, predicted by sequence homology to be a calcineurin-like phosphatase (Koonin 1994; Frishman et al. 2003), demonstrates metallophosphatase activity in vitro. The ability of recombinant YfcE to hydrolyze a number of artificial and naturally occurring phosphate compounds was tested. The best substrate identified was bis-*p*NPP, a commonly used artificial substrate to detect phosphodiesterase activity. The affinity of YfcE for this compound (K_m = 9.74 mM) is similar to that of other eukaryotic phosphodiesterases (BRENDA database) (Schomburg et al. 2002). YfcE requires Mn^{2+} for bis-*p*NPP hydrolysis ($K_D = 5.56 \pm 0.06$ μM) and is optimally active at pH 8.5–9.8. As no activity was detected against monoesters and a wide variety of naturally occurring phosphodiesters, YfcE is expected to have narrow substrate specificity. Additionally, YfcE may require a cellular binding partner for substrate recruitment or activation, roles normally performed by multi-domain phosphatases.

Since an atomic resolution structure was not available for YfcE or any of its close sequence family homologs, the X-ray crystal structure of YfcE was determined. While biochemical studies establish YfcE as a Mn^{2+}-dependent phosphodiesterase, analysis of the YfcE/sulfate/Zn^{2+} crystal structure complex sheds light on key details of metal binding. In addition, the 2.25 Å resolution structure helps define critical residues involved in substrate recognition and tetramer formation.

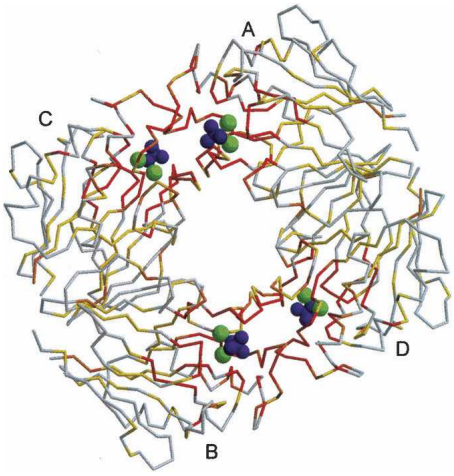

Figure 6. Conserved residue location in the YfcE tetram**er.** Individual chains are labeled A, B, C, or D. The Cα trace is colored according to the level of identity within YfcE family members shown in Figure 5. Identical amino acids are colored red. Residues with 90% identity or semiconserved are colored orange. Residues with 80% identity or sharing the same residue type (i.e., hydrophobic) are colored yellow. Zincs (green) and sulfate (purple) are shown as cpk atoms.

Active site-metal coordination and inhibitor binding

YfcE has a two-metal binding site shown in Figure 7. As mentioned previously, the predominant metal species bound in the crystals is Zn^{2+} carried along from the *E. coli* extract. Zinc 1 is coordinated by invariant residues Asp9, His11, His129, and Asp37. Zinc 2 is coordinated by invariant residues Asp37, Asn73, His105, and His127. The two zinc ions are separated by only 3.45 Å. A sulfate from the crystallization solution binds both metals. Hydrogen bonds are formed with the metal-coordinating residues Asn73 and His127, as well as the invariant residue Arg44 and a water molecule. Invariant residue Val 78 from neighboring subunit C and Cys74 are in van der Waals proximity. The sulfate may mimic phosphate in the natural ligand, since it is able to inhibit binding of substrates in vitro. Sulfate had significant inhibiting effect on the hydrolysis of bis-*p*NPP by YfcE (K_i = 25.3 mM) (Fig. 3). Not surprisingly, phosphate is a better phosphatase inhibitor, binding with higher affinity and producing at least 20 times stronger inhibition of YfcE activity. Unlike sulfate, phosphate acts as a competitive inhibitor.

Structural comparison with Mre11 nuclease

The *E. coli* YfcE β-sandwich architecture is common among phosphatases and has been proposed to represent a common ancestral origin. YfcE has retained this simple fold, whereas most other phosphatases have acquired additional domains or secondary structure elements responsible for substrate specificity and regulation. A DALI (Holm and Sander 1993) structure search identifies a wide variety of phosphatases as having significant similarity to YfcE. Superposition of YfcE with proteins having a Z-score better than four illustrates the similarity of their β-sandwich architecture, although the relative orientation of the β-sheets to one another is variable. *Pyrococcus furiosus* Mre11 nuclease is the closest YfcE structural homolog among extensively characterized proteins and is the focus of the following comparative analysis (PDB ID 1II7, Dali Z-score 13.1, RMSD 3.0).

Briefly, Mre11 is involved in many critical DNA metabolic processes and is found in all kingdoms. It contains Mn^{2+}-dependent nuclease activity and a single-stranded DNA endonuclease/3′-5′ exonuclease mechanism (Furuse et al. 1998; Paull and Gellert 1998; Trujillo et al. 1998). The N-terminal domain is responsible for biochemical activity, while a DNA binding domain aids in recruiting the substrate for catalysis. The exonuclease activity of Mre11 is regulated by heterotetrameric association with Rad50, an ATPase. In eukaryotes, the protein Nbs1 binds Mre11/Rad50 to regulate processes such as repairing DNA damage, checkpoint signaling, and telomere silencing (Carney et al. 1998; Haber 1998; Varon et al. 1998).

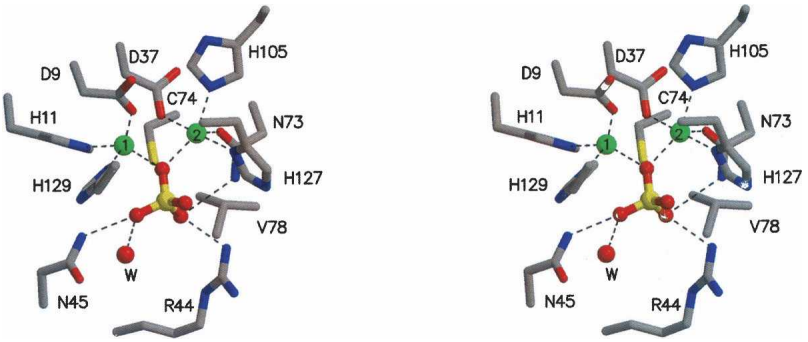

Figure 7. Stereo representation of the YfcE active site. Interactions with zinc ions (green) and sulfate are indicated with broken lines.

Structural alignment of YfcE with Mre11 (Hopfner et al. 2001) using metal binding residues is shown in Figure 8A. The active site based superposition is virtually indistinguishable from that based on polypeptide fragments identified by DALI. The core fold is conserved among the proteins, although β4 of YfcE runs antiparallel to its Mre11 counterpart β7, and C-terminal β11 of YfcE does not have a structural equivalent. Instead, Mre11 continues in the opposite direction to complete domain II, a DNA substrate recognition domain. In addition to domain II, Mre11 contains three extra β-strands flanking β4 of YfcE. Interestingly, two of these strands are equivalent to β4 and β5 of YfcE from the neighboring subunit (A-D interface), thereby extending the β-sheet surface similarly to Mre11.

In addition to topology, the active site of the YfcE/sulfate/Zn^{2+} complex bears striking similarity to the Mre11/dAMP/Mn^{2+} crystal complex. The active sites are shown in Figure 8B. Since both of these proteins display manganese-dependent activity, it is not surprising they contain identical metal chelating residues. Equivalent residues are Asp9-Asp8, His11-His10, Asp37-Asp49, Asn73-Asn84, His105-His173, His127-His206, and His129-His208. The position of the divalent cations is also preserved, although the coordination of the metals differs, reflecting the preferences for tetrahedral versus octahedral geometry when zinc is bound versus manganese. However, it is possible that when manganese is bound to YfcE, Asp37, like Asp49 of Mre11, binds both metals, not just metal 2. Additional waters, like the metal bridging water of Mre11, may also be present. The location of the sulfate inhibitor in the YfcE crystal complex mimics the location of α-phosphate in the Mre11-dAMP complex. In YfcE, the sulfate is bound by Asn73, Arg4, Asn45, and a water molecule. As mentioned above, Asn73 is equivalent to Asn84 in Mre11 and plays a dual role in binding both metal 2 and the substrate for both enzymes.

YfcE mechanistic predictions

Details of metal coordination, the position of the cleaved phosphate oxygens, and location of structural or catalytic waters may differ in the presence of Mn^{2+} and/or substrate. However, there are sufficient structural and mechanistic similarities among metallophosphatases to warrant a general discussion of the basic YfcE catalytic mechanism. Briefly, metal-dependent phosphatases are believed to hydrolyze the scissile bond via nucleophilic attack by a hydroxide ion in an in-line transfer mechanism. The pentacoordinated intermediate is stabilized by conserved residues, which directly bind the negatively charged phosphate. Product release occurs following protonation of the leaving group (Hopfner et al. 2001; Knöfel and Sträter 2001; Arthur et al. 2004). The location of the nucleophile with respect to the substrate and the base assisting product release may differ and are debated even for well-characterized metallophosphatases. In most cases, investigators speculate that the nucleophile comes directly from a metal-bound water that is converted to a hydroxide ion due to the metal-Lewis-acidity of a metal ion. The pH-dependent activity of YfcE

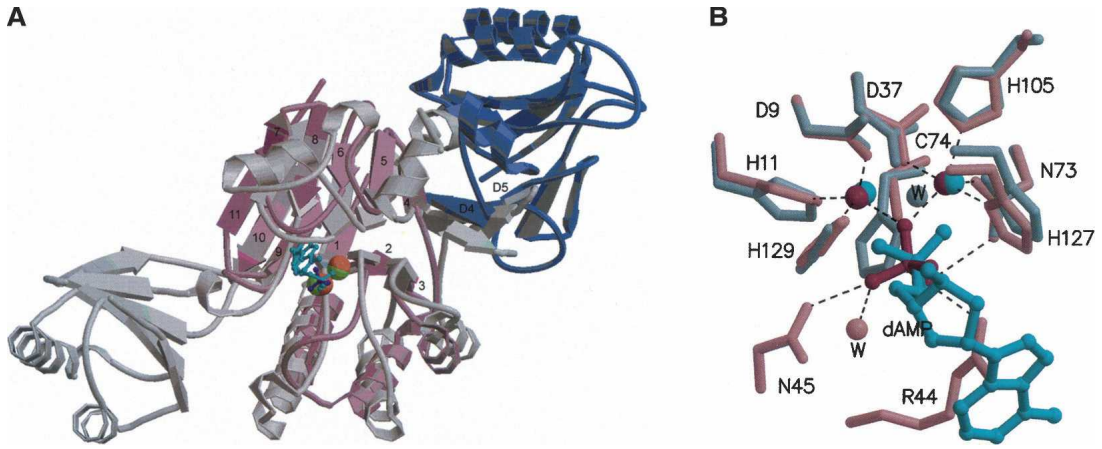

Figure 8. Superposition of *E. coli* YfcE and *P. furiosus* Mre11. Metal chelating residues were used to create the superposition. (*A*) A ribbon diagram of YfcE and Mre11 crystal complexes. The YfcE A chain is colored plum; the D chain, blue-violet. β-Strands are numbered for YfcE only. A ball-and-stick representation of sulfate bound to YfcE is colored dark purple, and zinc ions are shown as green cpk atoms. Mre11 is colored light cyan, while a bound dAMP molecule is colored cyan. Bound Mn^{2+} ions are shown as salmon-colored cpk atoms. (*B*) The same color scheme applies, except zinc and manganese ions are colored in purple and cyan, respectively. The metal bridging water of Mre11 is shown in light cyan. Amino acid labels are shown for YfcE only. The orientation of YfcE is the same as for Figure 7.

supports this hypothesis. As was mentioned earlier, the optimal pH range of YfcE is 8.5–9.8. As in other metallophosphatases, YfcE contains conserved residues that presumably bind the cleaved phosphate and could stabilize a negatively charged pentaphosphate intermediate.

One difference between YfcE and a majority of superfamily members with possible implications for phosphate binding and catalysis is the substitution of a cysteine for histidine at position 74. His85 of Mre11 and His117 of *E. coli* 5′-nucleotidase, for example, contact the cleaved phosphate directly. Based on biochemical and structural observations, this interaction is believed to be important in transition state stabilization, and is required for catalysis (Hopfner et al. 2001; Knöfel and Sträter 2001; McMillan et al. 2003; Arthur et al. 2004). However, in our structure, Cys74 is 3.6 Å or 3.9 Å away from sulfate oxygens. It is possible that Cys74, which is invariant among the close sequence homologs shown in Figure 5, exerts its effects on substrate binding and transition state stabilization via a water molecule that is absent from our inhibitor complex. Alternatively, the backbone of Cys74 may move closer to the substrate, allowing the side chain of Cys74 to bind oxygen of the cleaved phosphate. However, this latter scenario seems unlikely given the constraints imposed by neighboring residue Asn73 on proper active site geometry and tetramer formation. Any other YfcE insights, such as the likely location of the nucleophilic water upon substrate binding and the directionality of product release, would be highly speculative.

Substrate recognition

Combined sequence and structure analysis of YfcE suggests the most likely substrate-binding region is located in the shallow cleft between molecules A and C (or B and D). Most of the conserved and surface exposed residues are located in this region, which includes residues from both subunits located in proximity to the predicted phosphate-binding site (Arg44, Asn45, Asn73, Cys74, Val77, Met81, and Lys151). This feature is apparent from Figure 9A, a space-filling depiction of the YfcE tetramer colored according to the level of sequence conservation found among ≥30% sequence homologs. Interestingly, the conserved binding cleft coincides with a positively charged patch on the surface of YfcE (Fig. 9B). Since YfcE is an acidic protein (theoretical pI ~5.7), the positively charged patch on the otherwise evenly mixed electrostatic surface may aid in recruitment of a negatively charged substrate to the catalytic site. The electrostatic profile of YfcE, along with its high degree of structural similarity to Mre11, suggests YfcE may even function as part of a DNA processing complex.

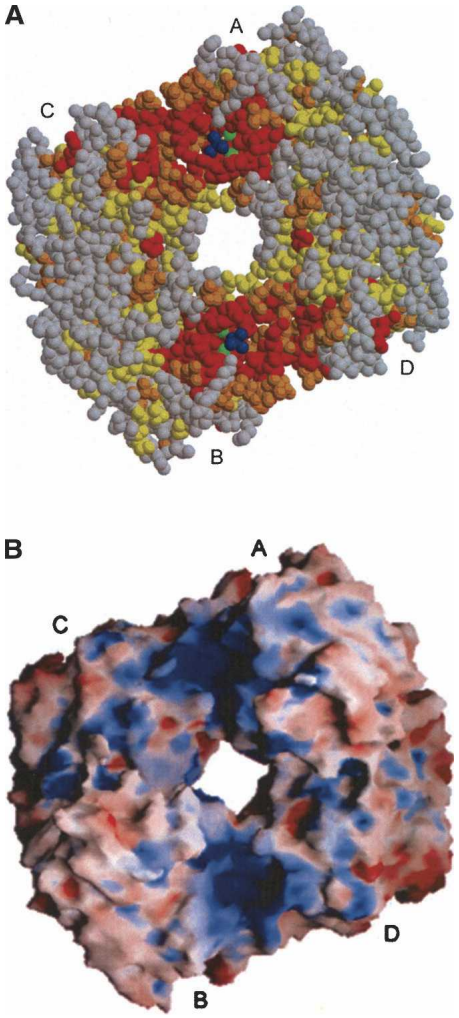

Figure 9. Postulated docking site on the surface of YfcE. (*A*) Conserved residue location in the YfcE tetramer. Individual chains are labeled A, B, C, or D. Residues, drawn as cpk objects, are colored according to the level of identity within YfcE family members shown in Figure 6. Identical amino acids are colored red. Residues with 90% identity or semi-conserved are colored orange. Residues with 80% identity or sharing the same residue type (i.e., hydrophobic) are colored yellow. Zincs (green) and sulfate (dark blue) are shown as cpk atoms. (*B*) Electrostatic surface of the YfcE tetramer. Protein and zinc ions are shown. Pullman charges were assigned to the protein using SYBYL (SYBYL 2002), and a formal charge of +2 was assigned to zinc. The electrostatic potential surface was generated using GRASP (Nicholls et al. 1991). Negative charges are indicated by red, positive charges by blue, and neutral charges by white. The scale ranges from –12 to +12 k_bT (k_b = Boltzmann constant, T = temperature).

Conclusion

YfcE demonstrates Mn^{2+}-dependent phosphodiesterase activity in vitro. In addition, the overall topology and conserved di-metal active site of YfcE are reminiscent of other known metal-dependent phosphatases. The YfcE substrate is not yet known but is expected to be highly

specific for YfcE activated hydrolysis due to the inability of carefully chosen ligands to serve as YfcE targets. The YfcE substrate is expected to dock in the conserved and positively charged cleft between dimers A and C (or B and D) of the tetramer. Future studies will probe the identity of the YfcE substrate and address whether YfcE functions as part of a larger protein complex containing additional substrate recognition elements or regulatory domains.

Materials and Methods

Cloning, cell-growth, expression, and purification

Cloning, cell-growth, expression, and purification of *E. coli* YfcE were performed essentially as described by Korolev et al. (2001). The expression construct coded for a protein with an N-terminal hexa-histidine tag for purification by metal-chelating affinity chromatography.

Determination of YfcE oligomeric state using size exclusion chromatography

HPLC size exclusion chromatography was performed as described by Zhang et al. (2002), with few exceptions.

Enzymatic screening

The *E. coli* YfcE was screened for a variety of catalytic activities using general enzymatic assays as described previously (Yakunin et al. 2004). Phosphatase activity was measured in a reaction mixture containing 50 mM HEPES-K buffer (pH 7.5), 5 mM $MgCl_2$, 0.1 mM $MnCl_2$, and 40 mM *p*-nitrophenyl phosphate (*p*NPP) by following the absorption increase at 410 nm. Phosphodiesterase and nuclease activities were measured using 50 mM Tricine buffer (pH 8.5), 5 mM $MgCl_2$, 0.1 mM $MnCl_2$, and 0.83 mM bis-*p*NPP with detection at 410 nm. Screening with natural phosphatase substrates was performed in 96-well plates using 160 μL reaction mixtures containing 50 mM HEPES-K (pH 7.5), 0.1 mM substrate (various nucleotides and phosphorylated sugars, amino acids, and organic acids from Sigma), 6.25 mM $MgCl_2$, 0.6 mM $MnCl_2$, 0.6 mM $NiCl_2$, and 2–5 μg of protein. After 30- to 60-min incubation at 37°C, the reaction was terminated by the addition of 40 μL of malachite green reagent (Baykov et al. 1988), and the production of Pi was measured at 630 nm.

Enzymatic assays

Phosphodiesterase activity of the *E. coli* YfcE against chromogenic substrates (bis-*p*NPP, thymidine 5′-monophosphate *p*-nitrophenyl ester, *p*-nitrophenylphosphorylcholine) was measured in a reaction mixture containing 50 mM Tricine buffer (pH 8.5), 0.66 mM substrate (or as indicated), and 0.1 mM Mn^{2+} with detection at 410 nm. Phosphodiesterase activity with various 2′,3′- and 3′,5′-cyclic nucleoside monophosphates (from Sigma) was measured using a quantitative assay based on a measurement of the alkaline phosphatase-sensitive nucleotide product. All incubations (0.8 mL) contained 50 mM of Tricine buffer, 1–2 mM of substrate, and 0.1 mM $MnCl_2$. After 20- to 40-min incubation at 37°C with enzyme (1–2 μg), 1 unit of alkaline phosphatase was added, and the reaction was incubated for 10 min at 37°C. Liberated Pi was assayed with malachite green reagent (Baykov et al. 1988). Hydrolysis of single-stranded M13 DNA, double-stranded λ DNA, *E. coli* genomic DNA, total ribosomal RNA, and total transfer RNA was assayed by analyzing reaction products by agarose gel electrophoresis. Kinetic parameters (K_m and V_{max}) were determined from Lineweaver-Burk plots by nonlinear curve fitting using GraphPad Prism program (version 4.00 for Windows, GraphPad Software; http://www.graphpad.com), and k_{cat} was determined by $k_{cat} = V_{max}/[\text{enzyme}]$.

Crystallization, heavy-atom soaking, and data measurement

Crystals were grown at 22°C in hanging drops by the vapor diffusion method. Upon setup, the drop contained 0.05 M of sodium acetate (NaAc at pH 6.5), 5 mM HEPES (pH 7.5), 10% poly-ethylene glycol (PEG) 4K, 0.1 M of ammonium sulfate (AmS), 0.25 M NaCl, 1 mM of β-mercaptoethanol, and 5 mg/mL of protein. The reservoir contained 0.1 M NaAc (pH 6.5), 20% PEG 4K, and 0.2 M AmS. Crystals belonging to the trigonal space group $P3_1$ appeared in 1–3 mo. Crystal 1 was soaked with 3 mM CH_3HgCl for 15 min. The crystal was then soaked in mother-liquor containing 10% glycerol for cryoprotection prior to flash cooling in liquid nitrogen. Crystal 2 was soaked in 6 mM K_2PtCl_4 for 1 h prior to cryoprotection and mounting. Heavy atom derivative data sets were measured at the 32-ID beamline of the Advanced Photon Source (APS), Argonne National Laboratory (ANL), operated by the Life Sciences Collaborative Access Team. Data were integrated and scaled using HKL (Otwinowski and Minor 1997). Native data (crystal 3) were measured at the Dupont-Northwestern-Dow 5ID beamline of the APS. Data were integrated and scaled with MOSFLM and SCALA, respectively (Evans 1993; Leslie 1997).

Site determination, model building, and refinement

Using crystal 1 as a reference ''native'' data set and crystal 2 for substructure determination, four Pt atoms were identified with SOLVE-2.06 (Terwilliger and Berendzen 1999), implementing the SIR (single isomorphous replacement) routine and a 3.2 Å cutoff. These sites were refined using the single isomorphous replacement with anomalous scattering (SIRAS) protocol from autoSHARP to 2.8 Å. After phase extension to 2.5 Å using DM (Cowtan 1994), electron density interpretation was performed using ARP/wARP-6.0 (Perrakis et al. 1999). The model was 80% complete and included four polypeptides in the asymmetric unit. Model completion was aided by use of polypeptide A as a reference model for the relatively incomplete polypeptides C and D. Subsequently, rigid body refinement in CNS-1.1 (Brünger et al. 1998) was performed to improve the agreement of the model to native crystal 3 observations. After phase extension to 2.25 Å and iterative rounds of model building in O (Jones et al. 1991) and refinement with CNS-1.1, the final R_{work} and R_{free} were 21.3% and 25.2%, respectively. A summary of data collection and refinement statistics is given in Table 2.

Figure preparation

Figures 4, 6, 7, 8, and 9A were generated by Molscript (Kraulis 1991) and Raster3D (Merritt and Bacon 1997).

Data deposition

The atomic coordinates and structure factors have been deposited in the RCSB PDB, with accession code 1SU1, RCSB021989.

Acknowledgments

This work was supported by the Midwest Center for Structural Genomics which is supported by the National Institutes of Health through Grant GM-62414 and the Ontario Research and Development Challenge Fund. Support for D.J.M. was partially provided by the Northwestern University Drug Discovery Program grant NIH-T32-AG00260. Portions of this work were performed at the DuPont-Northwestern-Dow Collaborative Access Team (DND-CAT) Synchrotron Research Center located at Sector 5 of the Advanced Photon Source (APS) and at the Life Sciences CAT located at Sector 32. DND-CAT was supported by E. I. Dupont de Nemours & Co., the Dow Chemical Co., the U.S. National Science Foundation, and the State of Illinois. Use of the APS was supported by the U.S. Department of Energy, Basic Energy Sciences, and DND-CAT.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Aravind, L. and Koonin, E.V. 1998. Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res.* **26:** 3746–3752.

Arthur, L.M., Gustausson, K., Hopfner, K.-P., Carson, C.T., Stracker, T.H., Karcher, A., Felton, D., Weitzman, M.D., Tainer, J., and Carnery, J.P. 2004. Structural and functional analysis of Mre11-3. *Nucleic Acids Res.* **32:** 1886–1893.

Baykov, A.A., Evtushenko, O.A., and Avaeva, S.M. 1988. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.* **171:** 266–270.

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., Pannu, N.S., et al. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54:** 905–921.

Carney, J.P., Maser, R.S., Olivares, H., Davis, E.M., Le Beau, M., Yates 3rd, J.R., Hays, L., Morgan, W.F., and Petrini, J.H. 1998. The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: Linkage of double-strand break repair to the cellular DNA damage response. *Cell* **93:** 477–486.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16:** 10881–10890.

Cowtan, K. 1994. An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsl. Protein Crystallog.* **31:** 34–38.

Evans, P.R. 1993. SCALA, version 3.1.9, Medical Research Council Laboratory of Molecular Biology, Cambridge, UK.

Finney, L.A. and O'Halloran, T.V. 2003. Transition metal speciation in the cell: Insights from the chemistry of metal ion receptors. *Science* **300:** 931–936.

Frishman, D.M., Kosykh, D., Kastenmuller, G., Kolesov, G., Zubrzycki, I., Gruber, C., Geier, B., Kaps, A., Albermann, K., Volz, A., et al. 2003. The PEDANT genome database. *Nucleic Acids Res.* **31:** 207–211.

Furuse, M., Nagase, Y., Tsubouchi, H., Murakami-Murofushi, K., Shibata, T., and Ohta, K. 1998. Distinct roles of two separable in vitro activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J.* **17:** 6412–6425.

Haber, J.E. 1998. The many interfaces of Mre11. *Cell* **95:** 583–586.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233:** 123–138.

Hopfner, K.-P., Karcher, A., Craig, L., Woo, T.T., Carney, J.P., and Tainer, J.A. 2001. Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase. *Cell* **105:** 473–485.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47:** 110–119.

Koonin, E.V. 1994. Conserved sequence pattern in a wide variety of phosphoesterases. *Protein Sci.* **3:** 356–358.

Korolev, S., Ikeguchi, Y., Skarina, T., Beasley, S., Edwards, A., Joachimiak, A., Pegg, A.E., and Savchenko, A. 2001. The crystal structure of spermidine synthase with a multisubstrate adduct inhibitor. *Nat. Struct. Biol.* **9:** 27–31.

Knöfel, T. and Sträter, N. 2001. Mechanism of hydrolysis of phosphate esters by the dimetal center of 5′-nucleotidase based on crystal structures. *J. Mol. Biol.* **309:** 239–254.

Kraulis, P.J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24:** 946–950.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26:** 283–291.

Leslie, A.G.W. 1997. MOSFLM, version 6.2.2, Medical Research Council Laboratory of Molecular Biology, Cambridge, UK.

McMillan, L., Beacham, I.R., and Burns, D.M. 2003. Cobalt activation of *Escherichia coli* 5′-nucleotidase is due to zinc ion displacement at only one of two metal-ion binding sites. *Biochem. J.* **372:** 625–630.

Merritt, E.A. and Bacon, D.J. 1997. Raster3D photorealistic molecular graphics. *Methods Enzymol.* **277:** 505–524.

Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11:** 281–296.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode, In *Macromolecular crystallography* (eds. C.W. Carter, Jr., and R.M. Sweet), pp. 307–326. Academic Press, New York.

Paull, T.T. and Gellert, M. 1998. The 3′ to 5′ exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol. Cell* **1:** 969–979.

Perrakis, A., Morris, R.J., and Lamzin, V.S. 1999. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6:** 458–463.

Schomburg, I., Chang, A., and Schomburg, D. 2002. BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* **30:** 47–49.

SYBYL 2002. Tripos associates, St. Louis, MO.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29:** 22–28.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr.* **D55:** 849–861.

Trujillo, K.M., Yuan, S.S., Lee, E.Y., and Sung, P. 1998. Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J. Biol. Chem.* **273:** 21447–21450.

Varon, R., Vissinga, C., Platzer, M., Cerosaletti, K.M., Chrzanowska, K.H., Saar, K., Beckmann, G., Seemanova, E., Cooper, P.R., Nowak, N.J., et al. 1998. Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93:** 467–476.

Yakunin, A.F., Proudfoot, M., Kuznetsova, E., Savchenko, A., Brown, G., Arrowsmith, C.H., and Edwards, A.M. 2004. The HD domain of the *Escherichia coli* tRNA nucleotidyltransferase has 2′,3′-cyclic phosphodiesterase, 2′-nucleotidase, and phosphatase activities. *J. Biol. Chem.* **279:** 36819–36827.

Zhang, R.-G., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C., Edwards, A., Joachimiak, A., and Savchenko, A. 2002. Crystal structure of *Thermotoga maritima* 0065, a member of the IclR transcriptional factor family. *J. Biol. Chem.* **277:** 19183–19190.