YfiT from *Bacillus subtilis* Is a Probable Metal-Dependent Hydrolase with an Unusual Four-Helix Bundle Topology[†,‡]

Shyamala S. Rajan,[§] Xiaojing Yang,[§] Ludmilla Shuvalova,[§] Frank Collart,[||] and Wayne F. Anderson*[,§]

Molecular Pharmacology and Biological Chemistry, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, and Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439

Received June 25, 2004; Revised Manuscript Received September 16, 2004

ABSTRACT: YfiT, a 19-kDa polypeptide from *Bacillus subtilis*, belongs to a small sequence family with members predominantly from Gram positive bacteria. We have determined the crystal structure of YfiT in complex with Ni^{2+} to a resolution of 1.7 Å. YfiT exists as a dimer and binds Ni^{2+} in a 1:1 stoichiometry. The protein has an unusual four-helix bundle topology and coordinates Ni^{2+} in an octahedral geometry with three conserved histidines and three waters. Although there is no similarity in their overall structures, the coordination geometry of the metal and the residues that constitute the putative active site in YfiT are similar to those of metalloproteases such as thermolysin. Our structural analyses suggest that YfiT might function as a metal-dependent hydrolase.

Assignment of functions to novel, uncharacterized gene products based on amino acid sequence and three-dimensional structure is often a challenge. This is a particularly difficult problem because function is a multifaceted concept and also because small changes in protein sequence can be associated with significant changes in function (*1*). As such, many individual proteins of known sequence and structure present challenges to the understanding of their function. In cases where purely sequence-based methods fail, three-dimensional structure and identification of structural homologues can provide clues to protein function because evolution retains the folding pattern long after sequence similarity becomes undetectable. Structural homology methods often rely solely on identifying proteins with similar overall fold/topology. The complex relationship between protein fold and function necessitates looking beyond the global fold of a protein to the specific arrangement of functional sites (*2*). Closer inspection of putative active sites, ligand-binding sites, and their architecture is particularly important because even structurally homologous proteins can have different functions. Conservation patterns in primary structures among members of a functionally uncharacterized family for which sequences and structures are known are also indicators of functionally important sites. In the case of proteins from completely sequenced genomes, comparative genomics information such as the evolutionary conservation of the encoding genes, local gene context, and phylogenetic profiles allow new approaches to function prediction. Such methods use protein−protein interaction patterns and correlations between occurrences of related proteins in different organisms as indicators of functional properties (*1*). Even when it is possible to propose a function for a gene product, there still remains the possibility that the protein has multiple functions. Here, we present the structural characterization of YfiT, a case study for some of these problems and approaches in identifying function.

We have determined the crystal structure of YfiT, a ~20-kDa polypeptide from *Bacillus subtilis*, at 1.7 Å resolution. The core of the protein is a four-helix bundle, and the protein exists as a dimer in both crystals and solution. YfiT binds Ni^{2+} in a 1:1 stoichiometry. The three histidines involved in coordinating the metal are invariant within its sequence family. The topology of YfiT is somewhat similar to apolipophorin, a lipid-binding protein that is a five-helix bundle. In an attempt to understand the function of YfiT, we have analyzed its structure, sequence, and genetic neighborhood along with those of its structural and sequence homologues. While these analyses did not help assign a biochemical function to YfiT, its (putative) active-site architecture being surprisingly similar to that of metalloproteases such as thermolysin suggests an enzymatic function for YfiT.

MATERIALS AND METHODS

The cloning, expression, and purification of recombinant YfiT in *Escherichia coli* were carried out essentially as described (*3*) with some modification. The 23-residue N-terminal polyhistidine tag (MHHHHHHSSGVDLGTEN-LYFQSNA) was cleaved between Q-S off the purified, selenomethionine-labeled YfiT protein using TEV protease (*4*), which carries a noncleavable polyhistidine tag. After its separation from the protease and the tag via nickel-affinity chromatography, the cleaved YfiT protein in 10 mM Tris-HCl at pH 8.3, 0.5 M NaCl, 5 mM β-mercaptoethanol, and

[†] This work was funded by an NIH Grant (GM062414) to the Midwest Center for Structural Genomics.

[‡] The coordinates of YfiT have been deposited in the Protein Data Bank under accession code 1RXQ.

* To whom correspondence should be addressed: Department of Molecular Pharmacology and Biological Chemistry, Ward 8-264, Feinberg School of Medicine, Northwestern University, 303 E. Chicago Ave., Chicago, IL 60611. Telephone: (312) 503-1697. Fax: (312) 503-5349. E-mail: wf-anderson@northwestern.edu.

[§] Northwestern University.

[||] Argonne National Laboratory.

10.1021/bi048665r CCC: $27.50
Published on Web 11/18/2004

5% glycerol was used at a concentration of ~10 mg/mL for crystallization in hanging drops (1:1 protein/reservoir). Optimization of initial Hampton Research screen conditions yielded diffraction-quality crystals in 0.1 M sodium cacodylate at pH 6.5 and 23% PEG 8K at 22 °C. Single-wavelength anomalous dispersion (SAD) datasets were collected on a MAR-CCD detector at the 5ID-B beamline at the Advanced Photon Source at the peak wavelength of Se (0.979 Å) with 1° oscillation and an exposure of 5 s per image. A total of 360° of data were collected to a resolution of 1.7 Å with an average of 6-fold redundancy. Data were integrated and scaled using MOSFLM (*5*) and SCALA (*6*). The data were then used in a SAD routine implemented in the SOLVE/RESOLVE program suite (version 2.2; *7*). SOLVE was used to locate the selenium sites, and RESOLVE was used for density modification and interpretation of the electron density maps. These programs found 12 of the 16 possible selenium sites and interpreted 94% of the 712 residues in the 4 molecules. Starting with this initial model, XtalView (*8*), CNS (version 1.1; *9*), and REFMAC (*10*) were used for further model building and refinement. Missing or misinterpreted residues were manually built and refined in subsequent cycles to complete the model. Anomalous difference electron density maps showed a strong peak (at 18−22σ relative to 29σ of the Se peaks) in each of the 4 monomers. On the basis of the size and shape of the peak and its surrounding residues and waters, this density was interpreted to be that of a metal ion. We modeled Ni^{2+} into the density after X-ray fluorescence scans of the crystal confirmed the presence of nickel. Also, one of the monomers showed additional contiguous density starting at the metal ion. We modeled the 5-residue peptide ESATS into this density. The final model contains 468 H_2O, 4 Ni^{2+}, and the peptide ligand in addition to 686 of the expected 712 residues of the protein. All of the missing residues are at the N termini.

Cell dimensions and space group assignment of YfiT crystals along with details of diffraction data collection and processing are shown in Table 1.

Table 1: Summary of Crystallographic Data: Data Collection Parameters and Processing and Refinement Statistics for the YfiT Crystal Structure

data collection	
space group	*P*1
unit cell dimensions	
a, *b*, *c* (Å)	46.4, 50.4, 89.4
α, β, γ (deg)	104.4, 90.6, 112.1
protein molecules/unit cell	4
Se-Met/unit cell	16
wavelength (Å)	0.979
data processing	
resolution range (Å)	30.0−1.7 (1.74−1.70)[a]
total reflections	458 604
unique reflections	74 337
completeness (%)	93.6 (88.4)[a]
anom. compl. (%)	92.8
$I/\sigma I$	16.3 (3.1)[a]
R_{merge}[b] (%)	7.2 (21.0)[a]
R_{anom} (%)	4.1
FOM (before den. mod.)	0.33
FOM (after den. mod.)	0.56
refinement	
resolution range (Å)	30.0−1.7
reflections	70 594
R_{work} (%)	16.65 (22.1)[a]
R_{free} (%)	19.65 (26.4)[a]
rmsd of	bond lengths = 0.01 Å, bond angles = 0.8°
protein residues	686 of 712
other molecules	468 waters, 4 Ni^{2+}, 1 peptide ligand
mean *B* factor	21.1

[a] Highest resolution shell. [b] $R_{merge} = \sum|I(k) - [I]|/\sum I(k)$, where $I(k)$ is the value of the kth measurement of the intensity of a reflection, $[I]$ is the mean value of the intensity of that reflection, and $\sum$ is of all of the measurements. $R_{factor} = \sum|F_{obs} - F_{calc}|/\sum F_{obs}|$.

RESULTS

YfiT is 1 of a family of 8 bacterial proteins that are related by sequence. Homologues of this protein are found in several Gram positive species including the bacilliform *B. cereus*, *B. anthracis*, *B. halodurans*, and *Oceanobacillus iheyensis*, as well in *Deinococcus radiodurans*. These proteins show ~50% identity to each other (Figure 1A). Of this sequence family, *D. radiodurans* and *O. iheyensis* each have two proteins. These two pairs of proteins show differences in certain key residues while maintaining overall similarity. These proteins are likely products of paralogous genes related by duplication that have since followed divergent evolution and probably perform somewhat different functions. The paralogs from *D. radiodurans* show ~50% identity and those of *O. iheyensis* are ~40% identical to each other. All of these homologous proteins from different species have ~180 residues except for one paralog of *D. radiodurans*, which is longer by 50 residues.

The location of a gene in the genome sometimes provides clues to the function of its product. In the *B. subtilis* genome, *yfiT* is flanked by transport proteins *yfiS* and *yfiU* (Figure 1B), which code for a putative macrolide efflux pump and a putative multidrug resistance protein, respectively. Both these genes are transcribed from the opposite DNA strand relative to *yfiT*. In *B. halodurans*, the *yfiT* homologue is flanked by a very small open-reading frame and a gene that encodes a putative bacitracin-efflux pump. In both *B. cereus* and *B. anthracis*, the *yfiT* homologues are flanked by genes that likely encode a (polyphosphate/acetyl) kinase and a ribosomal modification protein (alanine acetyl transferase), in what seem to be tightly linked three-gene operons. Both of the *yfiT* paralogs in *D. radiodurans* also seem to be in tightly linked three-gene operons. However, the products of the other genes in these two operons do not bear any similarity in sequence to any known proteins. In *O. iheyensis* and *Cytophaga hutchinsonii*, the *yfiT* homologues are well-separated from their nearest neighbors in their respective genomes and might be single-gene operons.

In *B. subtilis*, *yfiT* is separated from *yfiS* by ~120 bp (Figure 1B). *yfiS* and *yfiR* that encode a putative transcriptional regulator similar to QacR from the Tet repressor family are tightly linked in a presumably two-gene operon.

Overall Structure. A striking feature of the YfiT structure is the presence of four very long (~20 residues) helices, which are connected by long loops. These helices form a closed bundle that shows a left-handed superhelical twist. Helix 1 is more twisted than others and is antiparallel to helices 2 and 3 and parallel to helix 4. This four-helix bundle is interrupted by an extended crossover loop between helices 2 and 3 and by a twisted two-stranded antiparallel β sheet

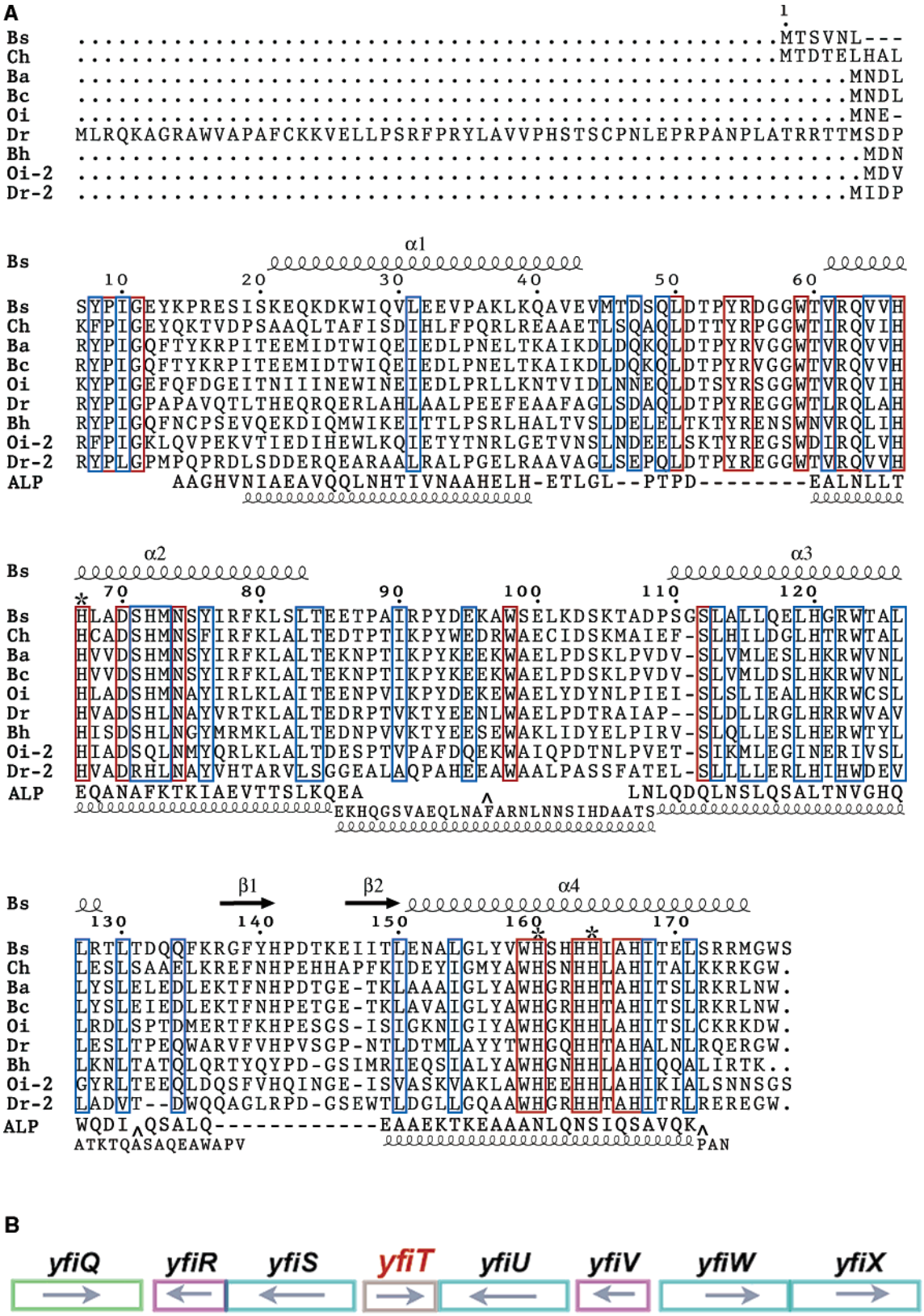

FIGURE 1: (A) Alignment of YfiT with sequence homologues. The sequence of YfiT (Bs: NP_388720) is aligned with homologues from *B. anthracis* (Ba: NP_656581), *B. cereus* (Bc: AAP09662), *B. halodurans* (Bh: BAB03996), *O. iheyensis* (Oi: BAC12738; Oi-2: BAC12369), *D. radiodurans* (Dr: AAF10418; Dr-2: AAF10598), and *C. hutchinsonii* (Ch: ZP_00117077). Structure-based sequence alignment of apolipophorin (ALP: 1AEP) is also shown. Broken lines indicate deletions. Invariant residues are shown in red boxes, while some positions with conservative replacements are shown in blue boxes. Asterisks mark residues involved in coordinating Ni^{2+}. Secondary structural elements shown at the top of the sequences are those of YfiT, and those at the bottom are of ALP. Insertion notations indicate the regions of ALP that do not align well with YfiT. Sequences were aligned using MultAlin/ESPript (*27*, *28*) programs and modified. (B) Genome context of YfiT. The genetic neighborhood of *yfiT* is shown along with the direction of transcription of constituent genes. Putative transporter genes and those of drug pumps are colored cyan, and putative transcription regulators are colored purple. *yfiQ* encodes a putative intercellular adhesion protein.

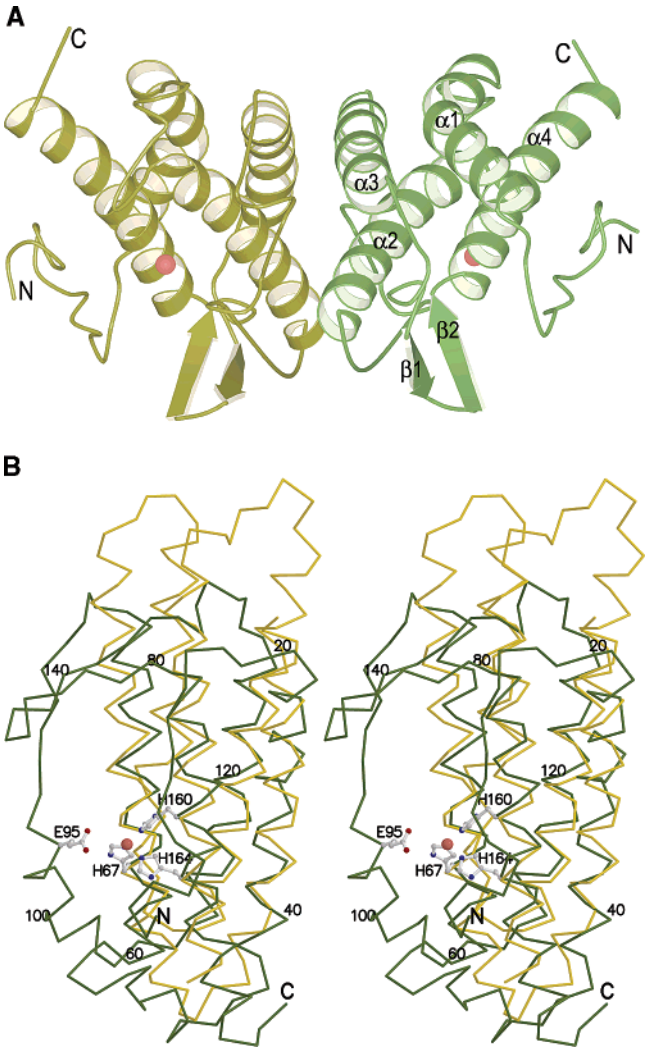

FIGURE 2: Structure of YfiT. (A) Ribbon diagram. The two monomers in the dimer are distinguished by the different intensities of colors. Ni^{2+} ions are presented in cpk format and colored salmon. The figures in A and D were prepared with MOLSCRIPT (*29*) and rendered in Raster3D (*30*). (B) Structural alignment with ApoLp-III. Cα traces of YfiT and ApoLp-III are presented in dark green and gold, respectively. The metal ion of YfiT is depicted in cpk format and colored salmon. The two structures were fit by their peptide backbones using CCP4 (*6*) programs. The figure was generated with MOLSCRIPT and rendered in Raster3D.

located between helices 3 and 4 (Figure 2A). This sheet is outside the core of the bundle and projects into the solvent.

In the *P*1 space group of the YfiT crystals, there are 4 molecules in the unit cell. Buried surface area calculations suggest that the crystallographic units are actually 2 dimers, and the biological entity is likely to be a dimer as seen in solution (dynamic light-scattering data not shown). The noncrystallographic 2-fold axis of symmetry that relates the two monomers is close to helices 2 and 3 and almost perpendicular to the helix bundle. The total buried surface area at the interface of the putative homodimer is ~2954 $Å^2$, and the largest buried surface area between any two dimers is ~634 $Å^2$. Each of the monomers can be superimposed on the others by its peptide backbone with a root-mean-square deviation (rmsd) of ~0.45 Å as is the case with the dimers. The subunits within a dimer are not identical as indicated by the alternate conformations of several residues at the interface. The residues that form the dimer interface belong to helices 2 and 3 as well as the long surface loop that connects them. Lys, Arg, Glu, Asp, Ser, Thr, Asn, Gln, His, and Tyr constitute ~56% of the residues that make up the dimer interface and are involved in polar interactions. Leu and Ala make up most of the remaining residues. Residues of the interface that are part of helix 2 and the following loop are mostly invariant or conservatively replaced in the family, suggesting that dimer formation is conserved. The only exceptions are in the paralogs occurring in *D. radiodurans* and *O. iheyensis*. These differences may be associated with modified functions.

Structural Homology. A search for structural homologues of YfiT by DALI (*11*) revealed that the highest similarity is to an insect protein that binds lipids, apolipophorin III (ApoLp-III, 1AEP; *12*). ApoLp-III is a 161-residue protein arranged in an up and down five-helix bundle. Similarity of ApoLp-III with YfiT extends to 4 helices (117 residues) that can be aligned via the peptide main-chain atoms with an rmsd of 2.9 Å (Figures 1A and 2B). Some of the other structurally similar proteins that DALI identified included the prothrombin fragment (1NU7; *13*), interferon τ (1AVO; *14*), and interferon β (1IFA; *15*). These are up and down helix bundles and can be aligned with YfiT via their main-chain atoms with an rmsd of 3.4, 12.0, and 13.7 Å respectively. Their structural homology with YfiT is over only 3 of its 4 helices.

Putative Active Site. The YfiT structure revealed the presence of two ligands that presumably co-purified with the protein. In all 4 molecules, a metal ion was present, which was determined to be predominantly nickel by X-ray fluorescence (*16*). Most of the residues that are spatially close to the metal-binding site are invariant within the family. In 3 of the 4 monomers, the Ni^{2+} is bound by 3 conserved histidines and 3 waters in an N/O (nitrogen/oxygen atoms) octahedral coordination (Figure 3A). In the fourth monomer, 2 of the 3 waters in the coordination complex appear to be partially replaced by a ligand whose electron density seems to be consistent with that of a peptide (Figure 3B). We modeled the sequence ESATS into this density. Given the relatively poor density in this region and the suboptimal occupancy of the peptide, we cannot be completely certain of the identity of this ligand.

A search among the nickel-binding proteins in the Protein Data Bank (PDB) for similarity to the (putative) active site of YfiT did not yield any positive results. However, a zinc-binding protein, thermolysin (2TMN; *17*), was surprisingly similar in the architecture of its metal-binding region to that of YfiT, even though the two structures differ considerably in their overall topology. Thermolysin is a heat-resistant, two-domain α/β protein from *B. thermoproteolyticus* (*17*). It is a Zn^{2+}-dependent neutral protease that is specific for peptide bonds on the imino side of hydrophobic residues such as Ile, Leu, and Phe (*17*). Alignment of thermolysin with monomer D of YfiT by their metals and the coordinating groups shows their active-site architecture to be very similar (Figure 3B and parts A and B of Figure 4). Conserved residues Glu95 and His163 of YfiT are in the same position as the catalytic groups Glu143 and His231 in thermolysin. Their distances from the metal and the peptide ligand are similar to those in thermolysin, where the phosphoramidate inhibitor is in the same relative position as the peptide in monomer D.

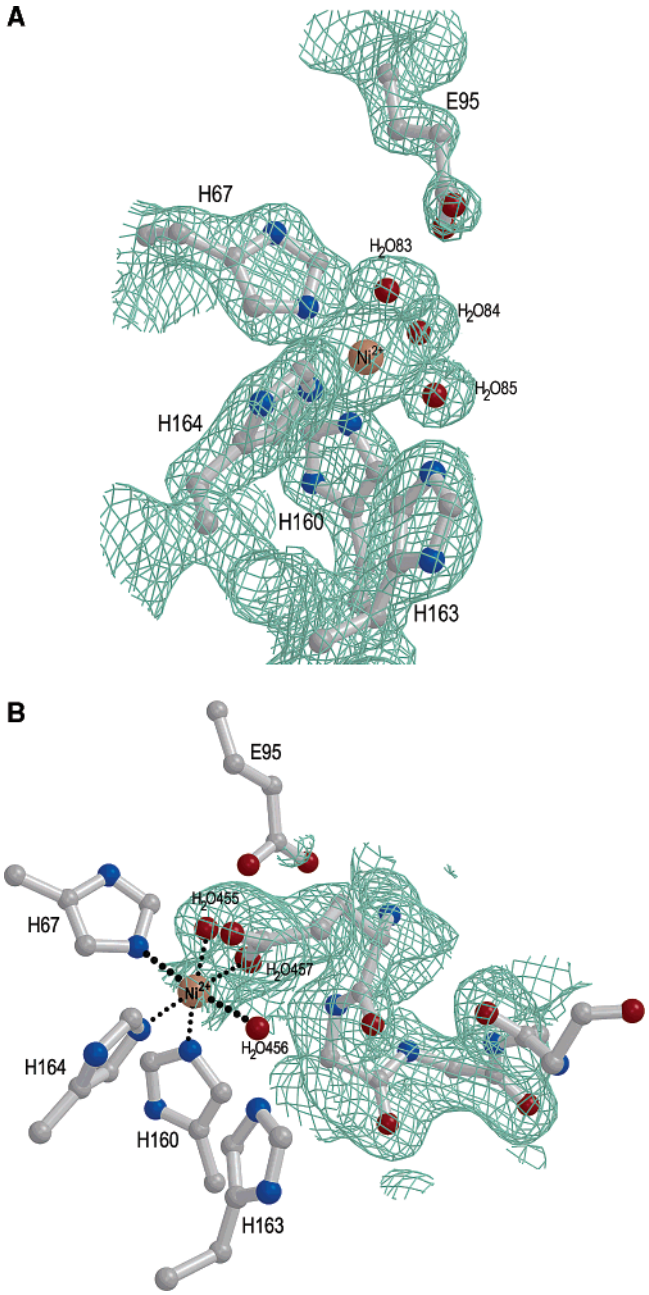

FIGURE 3: (Putative) active-site architecture of YfiT. (A) Monomer B. The 6 ligands involved in coordinating Ni^{2+} include 3 conserved histidines and 3 waters. A total of 2 of the histidines and 2 of the waters are bound to Ni^{2+} in a single plane, while the remaining groups are bound to the metal in opposite directions perpendicular to this plane. Some of the residues close to the metal binding site that are invariant in the family are also shown. The $2f_o - f_c$ electron density map is contoured at 1.5σ. This figure was made with BOBSCRIPT (*31*) and rendered in Raster3D. (B) Monomer D. A total of 1 of the 4 monomers in the unit cell has a ligand bound to the metal ion. Shown here is the omit map (1σ contour) of the ligand into which the peptide ESATS was modeled. For clarity, some of the side chains of the peptide are not shown. The rest of this region is similar to that shown from another monomer in Figure 3A.

DISCUSSION

YfiT is a prototype of a small family of proteins that bear considerable sequence similarity. The size of the family, which consists of only 8 proteins from 6 microorganisms (Figure 1A), and the availability of genome sequence data from partial or entire genomes of these bacteria allowed us to conduct a comprehensive analysis of these proteins in the context of their sequence and genome location. Identification of likely operons, search for homologues of the constituent genes and their products, and the probable regulatory sequences of these operons (Figure 1B) did not provide convincing clues regarding the biochemical function of these proteins. Lack of any biochemical or genetic data on any member of this family compounded the problem of assigning a functional role to YfiT.

The structure of YfiT has some unusual features. While the helix-bundle topologies among proteins are quite common, the YfiT structure is interesting in that the long helices in the four-helix bundle do not have the common up and down topology and are connected by long loops. The loop that connects helices 2 and 3 goes up and around helix 2 and then comes down the length of this helix to connect with helix 3 (parts A and B of Figure 2). This is different from the other up and down helical connections within this bundle and those from other proteins, some of which were identified by DALI as being structurally similar to YfiT. As an example, the four-helix bundle of hemerythrin (2CCY) has a similar overall structure with a 2-stranded β sheet that connects 2 of the helices. However, all of the helices in this protein have simple up and down connections as seen in apolipophorin (Figure 2B). The overall topology of YfiT and its structural homologues did not furnish any clues to the function of YfiT.

The most revealing feature of the YfiT structure in terms of its possible function is the presence of the metal, its coordination, and its location (Figures 2 and 3). The fact that this is the active site of the protein is indicated by the architecture of the metal-binding site, the evolutionary conservation of most of the residues that localize to this region, and the fortuitous presence of a ligand bound to the metal in one subunit. These observations also implicate the metal ion in not just a structural but a functional role as well. Another finding that may be of relevance in its function is that YfiT exists as a dimer in the crystal. While dimer formation may be important for function, the putative active sites are not part of the dimer interface. They are situated at opposite sides of the dimer facing the bulk solvent away from this interface (parts A and B of Figure 2).

The identification of a putative active site did not lend itself to easy identification of the biochemical function(s) of YfiT. We compared the structurally significant sequence motifs of YfiT, along with an analysis of its metal-coordination geometry and the coordinating groups, as well as the architecture of the metal-binding site with the available structures in the PDB.

A search among the known structures for the two conserved sequence motifs (WXVRQVVHH and WHXX-HHIAHI) that contain the metal-coordinating histidines, revealed their individual presence in seemingly unrelated, histidine-rich proteins, some of which bind metals. However, the histidines from these sequences are not involved in coordinating metal in any of these cases, unlike the situation in YfiT. As an example, the first sequence is found in a peptide deformylase (1IX1) from *Pseudomonas aeruginosa.* The protein binds Zn^{2+} via 2 His and a Cys in a pyramidal coordination. But the histidines of the deformylase that bind the metal are not part of our conserved sequence motif, which occurs at the C terminus of the protein exposed to the solvent.

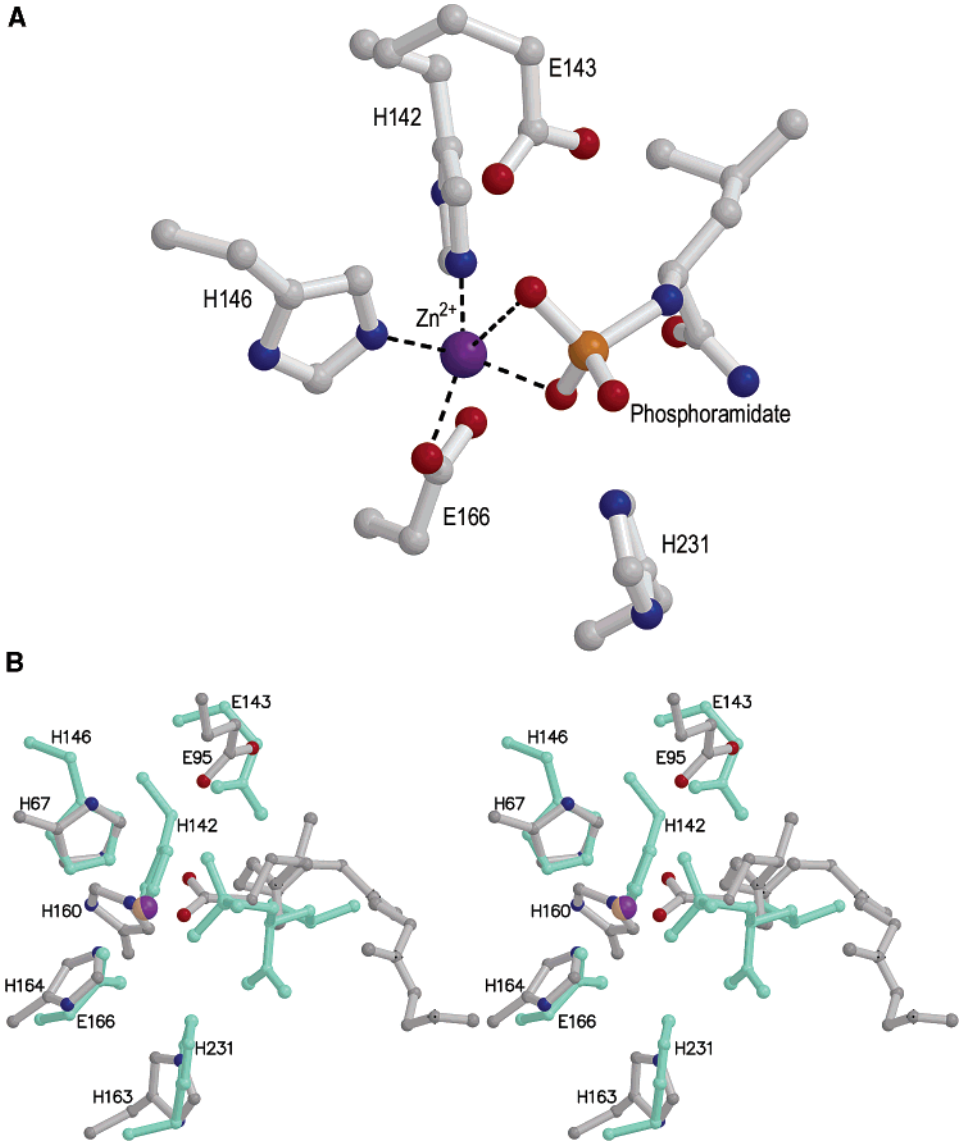

FIGURE 4: Structural comparison of YfiT with thermolysin. (A) Active site of thermolysin. The Zn^{2+} in thermolysin is bound by 2 histidines and a glutamate. Catalytic residues (Glu143 and His231) and a reaction inhibitor (phosphoramidate) are also shown. The metal ion is in cpk format and is colored purple. (B) Superposition of the active sites in stereo. Active sites of YfiT monomer D and thermolysin were aligned via their metal ions and coordinating groups using XtalView (*8*). The protein residues of thermolysin and its inhibitor phosphoramidate are colored aquamarine, and the zinc ion is colored purple. The Ni^{2+} of YfiT is colored plum. Figures in both A and B were prepared using MOLSCRIPT and rendered in Raster3D.

Similarly, the second sequence is found in leucine carboxy methyltransferase (1RJD) and palmitoyl protein thioesterase-2 (1PJA; *18*). Neither of these structures has metal ions bound, and the sequence of interest is exposed to the solvent. In the former, it is again located at the C terminus. Interestingly, the sequence WXVRQVVHH is also found in EmrB, a drug efflux protein occurring in *E. coli* and *Hemophilus influenza*. YfiU, the protein flanking YfiT, is a homologue of the EmrB/QacA multidrug resistance transporter family. The significance, if any, of this observation is not yet clear.

The Ni^{2+} is encircled by a number of aromatic residues within a 10 Å radius. Surprisingly, they seem to occur in pairs (Tyr8 and Tyr93, Tyr54 and Tyr157, and Trp59 and Trp159) that are nearly symmetric about the metal ion. In addition, a pair of histidines (His72 and His167) not involved in coordinating the metal are symmetrically placed in relation to the Ni^{2+} at ~7 Å. All of these residues are either invariant or are conservatively replaced in the family. The only exceptions to this observation are at two positions (Tyr93 and Tyr157) in the paralogs of *D. radiodurans* and *O. iheyensis*, which may have diverged in function as well as in sequence. The role of these aromatic residues near the active site is not yet understood.

Transition metals such as nickel are just as important in the enzymatic regulation of several cellular processes as they are toxic to the cell by interfering with the functioning of DNA, lipids, and proteins. The few nickel-binding proteins in the cell include those involved in the recognition, binding, and transport of the metal into the cell, others that bind and sequester it following its entry into the cell, still others that are required to insert it into nickel enzymes, and finally the nickel enzymes themselves. To date, there are only 7 known nickel-binding enzymes (*19*) including urease, hydrogenase, carbon monoxide dehydrogenase, methyl-coenzyme M reductase, Ni-superoxide dismutase (Ni-SOD), glyoxylase I, and cis−trans isomerase. Most of these enzymes have complex quaternary structures with complicated active-site architectures often involving metal clusters (*19*). These

proteins are clearly not similar to YfiT in their metal coordination or their active-site architectures.

We also explored the possibility that within the cell, YfiT might actually bind some other transition metal, such as zinc, which during protein purification by nickel-affinity chromatography was displaced by nickel. It is possible that the 3 histidines could coordinate a Zn^{2+} ion with 1 water molecule being the fourth tetrahedral ligand. X-ray fluorescence spectra (data not shown) of YfiT protein that has had Zn^{2+} added followed by dialysis to remove excess metal indicates that the protein will bind Zn^{2+}.

Of the 4 molecules (2 dimers) in the unit cell, only 1 shows a stretch of contiguous positive density in difference electron density maps that starts at 2 of the waters involved in Ni^{2+} coordination (Figure 3B). The peptide that we modeled into this density could be that of a reaction product, a nonproductive reaction complex, or an inhibitor. This monomer also shows movement of several conserved residues in the neighborhood of the metal-coordination site. While most zinc-binding proteins have either a pyramidal or tetrahedral coordination (*20*), the higher coordination of the metal seen here could reflect an intermediate as seen in some metalloenzymes. Alternatively, this could indeed be a nickel-binding protein with multiple binding sites for water in an octahedral coordination favored by nickel. Because the metal bound to the peptide ligand occurs only at this one site of the four, it accounts for only 25% of the metal content in the unit cell. Also, because the peptide ligand is present at suboptimal occupancy, the possibility exists that the metal bound to the peptide ligand is something other than the predominant Ni^{2+}.

A search of the PDB for Zn^{2+}-binding proteins, where 3 of the coordinating groups are histidines, yielded several eukaryotic proteins including matrix metalloproteinases (1BUD, *21*; 1A85, *22*), carbonic anhydrase (1AZM; *23*), and human interferon-β (1AU1; *24*). Except for carbonic anhydrase, the rest of the structures that we examined have the metal-binding sites exposed to the bulk solvent. In the dimeric form of interferon-β, the metal is at the interface of the dimer bound by both subunits, unlike YfiT. Architecture of the metal-binding region in these structures is similar to that of YfiT only to the extent of the coordinating ligands. These structures are not as histidine-rich nor do they have the ring of aromatic residues observed in YfiT.

The most unexpected and surprising discovery, however, came from another Zn^{2+}-binding protein, thermolysin (2TMN). The N-terminal domain of thermolysin consists of 9 β strands and 1 helix, while its C-terminal domain comprises 6 helices. Helices 1 and 3 are long, being more than 20 residues in length. Superposition of thermolysin and YfiT via their metals and the coordinating groups shows helix 3 of thermolysin to be parallel to helices 2 and 4 of YfiT, while helix 5 superimposes on helix 3 of YfiT. The metal ion in thermolysin is found in a groove between the β-sheet-rich N-terminal domain and the helical C-terminal domain. Although the protein folds are completely different, the (putative) active-site architecture of YfiT is strikingly similar to that of thermolysin (Figure 3B and parts A and B of Figure 4). In native thermolysin, Zn^{2+} is in a tetrahedral coordination with 1 water, a glutamate, and 2 histidines (*25*). When a peptide substrate binds, the carbonyl oxygen forms a fifth zinc ligand and displaces the water molecule toward Glu143, which acts as a general base. The water molecule activated by the combined influence of Glu143 and the metal attacks the carbonyl carbon of the scissile bond (*17*). The His231 hydrogen bonds to one of the oxygens bound to the zinc. It is not totally clear, but the proton appears to be donated by Glu143 to the peptide nitrogen, forming a tetrahedral intermediate in which 2 of the oxygens are liganded to the Zn^{2+}. This state is stabilized by both hydrogen bonds to His231, Tyr157, and other residues, as well as hydrophobic interactions. Finally, the C−N bond of the intermediate breaks to yield the products. Although other zinc-requiring proteases such as carboxypeptidase A and angiotensin-converting enzyme bear little similarity to thermolysin in sequence, it is now known that their active sites share common features (*17*). Thus, a similar proteolytic mechanism is observed in carboxypeptidase A (*17*), where Glu270 occupies the position of Glu143 and Tyr248 is analogous to His231 in thermolysin. In YfiT, we have very similar architecture, where Glu95 and His163 replace Glu143 and His231, respectively, of thermolysin (Figure 3B and parts A and B of Figure 4). In the other 3 monomers of YfiT (that do not contain the ligand), the side-chain oxygens of Glu95 hydrogen bond with 2 of the 3 waters that coordinate the metal. In thermolysin, the equivalent glutamate side chain hydrogen bonds to the water interacting with the active-site Zn^{2+} (*17*). In the ligand-bound monomer of YfiT, binding of the ligand changed the orientation of the Glu95 side chain so that it swings away from the metal and faces the ligand. In this new position of the Glu95, 1 of the side-chain oxygens hydrogen bonds with the main-chain N of Glu1 of the peptide ligand. The second oxygen bonds with a water that took the place of one of the Glu95 side-chain oxygens seen in other monomers. Other differences in the active site of the monomer with the ligand compared to the ones without include the movement of the side chains of Lys96 and Trp159.

Besides the putative catalytic groups, several other residues near the active site are similarly located in both thermolysin and YfiT. Leu155, Ser169, Trp115, Tyr76, and Asp226 near the active of thermolysin are replaced by the conserved residues Val65, Ser71, Trp98, Phe139, and His167 of YfiT. In the structural alignment, Tyr157 of thermolysin is found to be sandwiched between Tyr54 and Trp59 of YfiT, both of which are invariant in the YfiT family. The inhibitor *N*-phosphoryl-L-leucinamide in the thermolysin structure is in a similar position as the peptide that was modeled in monomer D of YfiT (Figures 3B and 4B).

The location of the metal-binding site exposed to the bulk solvent and the 3 replaceable coordinating waters and, above all, the finding of a peptide-like ligand bound to 1 of the monomers allow us to speculate that the possible substrates for YfiT activity might be proteins or peptides. It is conceivable that YfiT, like thermolysin, could function as a protease. The comparative analyses of its (putative) active-site components and its architecture have so far provided our only clues to the possible function of YfiT. Consistent with the suggestion that YfiT may have hydrolase activity, preliminary experiments show that the hydrolysis of *p*-nitrophenyl acetate is catalyzed by YfiT (data not shown).

Another interesting observation of YfiT is its crystal packing. It is puzzling that the 2 dimers crystallized in a triclinic cell, where they are not related by crystallographic

symmetry. One explanation for the lack of symmetry is the peptide ligand bound to the metal ion in 1 of the 4 monomers. Presumably, the crystal packing resulted in selection for the presence of the ligand in one position and against it in the other three positions. This microheterogeneity in the protein sample often makes it difficult to obtain crystals or results in poor crystal order (*26*). However, in this case, it was fortuitously associated with a nonrandom selection of structures at the four noncrystallographically related positions.

CONCLUSIONS

While its genome location, its protein sequence, and its overall topology did not provide definitive indications regarding the function of YfiT, the metal ion, its coordination, and the architecture of the metal-binding region yield promising clues. The unexpected similarity to the active site of thermolysin suggests that YfiT might likely function as a metal-dependent hydrolase. One could speculate about the possible substrates for such a function. The location of the active site exposed to the bulk solvent is readily accessible to a macromolecule such as a protein or peptide. The finding of a peptide bound to the metal in one of the subunits strengthens this argument. Alternatively, YfiT may be involved in hydrolyzing xenobiotics into products that can either be utilized by the cell or transported out of the cell by the pumps flanking the protein.

ACKNOWLEDGMENT

X-ray diffraction data were collected at the DuPont−Northwestern−Dow Collaborative Access Team (DND-CAT) Synchrotron Research Center (Sector 5) of the Advanced Photon Source. DND-CAT is supported by the E.I. DuPont de Nemours and Co., The Dow Chemical Company, NSF (Grant DMR-9304725), and the State of Illinois through the Department of Commerce and the Board of Higher Education (Grant IBHE HECA NWU 96). Use of the APS is supported by the U.S. DOE (Contract W-31-102-Eng-38). Support from the R. H. Lurie Cancer Center of Northwestern University to the Structural Biology Center is also acknowledged. We thank J. Lee for help with the purification of the protein.

REFERENCES

1. Whisstock, J. C., and Lesk, A. M. (2003) Prediction of protein function from protein sequence and structure, *Q. Rev. Biophys. 36*, 307−340.
2. Jones, S., and Thornton, J. M. (2004) Searching for functional sites in protein structures, *Curr. Opin. Chem. Biol. 8*, 3−7.
3. Rajan, S. S., Yang, X., Collart, F., Yip, V. Y., Withers, S. G., Varrot, A., Thompson, J., Davies, G. J., and Anderson, W. F. (2004) Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD^{+}/Mn^{2+}-dependent phospho-α-glucosidase from *Bacillus subtilis*, *Structure 12*, 1619−1629.
4. Stols, L, Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelly, M. I. (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site, *Protein Expression Purif. 25*, 8−15.
5. Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data, Joint CCP4 + ESF-EAMCB, *Newsletter on Protein Crystallography 26*.
6. Bailey, S. (1994) The CCP4 Suite: Programs for protein crystallography, *Acta Crystallogr., Sect. D 50*, 760−763.
7. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr., Sect. D 55*, 849−861.
8. McRee, D. E. (1999) XtalView/Xfit−A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol. 125*, 156−165.
9. Brunger, A. T. (1992) The Free *R* value: A novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472−474.
10. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr., Sect. D 53*, 240−255.
11. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123−138.
12. Breiter, D. R., Kanost, M. R., Benning, M. M., Wesenberg, G., Law, J. H., Wells, M. A., Rayment, I., and Holden, H. M. (1991) Molecular structure of an apolipoprotein determined at 2.5 Å resolution, *Biochemistry 30*, 603−608.
13. Friedrich, R., Panizzi, P., Fuentes-Prior, P., Richter, K., Verhamme, I., Anderson, P. J., Kawabata, S., Huber, R., Bode, W., and Bock, P. E. (2003) Staphylocoagulase is a prototype for the mechanism of cofactor-induced zymogen activation, *Nature 425*, 535−539.
14. Knowlton, J. R., Johnston, S. C, Whitby, F. G., Realini, C., Zhang, Z., Rechsteiner, M., and Hill, C. P. (1997) Structure of the proteasome activator REGα (PA28α), *Nature 390*, 639−643.
15. Senda, T., Shimazu, T., Matsuda, S., Kawano, G., Shimizu, H., Nakamura, K. T., and Mitsui, Y. (1992) Three-dimensional crystal structure of recombinant murine interferon-*β*, *EMBO J. 11*, 3193−3201.
16. Newville, M. (2003) Fundamentals of XAFS at http://cars9.uchicago.edu/xafs/.
17. Tronrud, D. E., Monzingo, A. F., and Matthews, B. W. (1986) Crystallographic structural analysis of phosphoramidates as inhibitors and transition-state analogs of thermolysin, *Eur. J. Biochem. 157*, 261−268.
18. Calero, G., Gupta, P., Nonato, M. C., Tandel, S., Biehl, E. R., Hofmann, S. L., and Clardy, J. (2003) The crystal structure of palmitoyl protein thioesterase-2 (PPT2) reveals the basis for divergent substrate specificities of the two lysosomal thioesterases, PPT1 and PPT2, *J. Biol. Chem. 278*, 37957−37964.
19. Watt, R. K., and Ludden, P. W. (1999) Nickel-binding proteins, *Cell. Mol. Life Sci. 56*, 604−625.
20. Harding, M. M. (2001) Geometry of metal−ligand interactions in proteins, *Acta Crystallogr., Sect. D 57*, 401−411.
21. Gong, W., Zhu, X., Liu, S., Teng, M., and Niu, L. (1998) Crystal structures of Acutolysin A, a three-disulfide hemorrhagic zinc metalloproteinase from the snake venom of *Agkistrodon acutus*, *J. Mol. Biol. 283*, 657−668.
22. Brandstetter, H., Engh, R. A., Von Roedern, E. G., Moroder, L., Huber, R., Bode, W. and Grams, F. (1998) Structure of malonic acid-based inhibitors bound to human neutrophil collagenase. A new binding mode explains apparently anomalous data, *Protein Sci. 7*, 1303−1309.
23. Chakravarty, S., Yadava, V. S., Kumar, V., and Kannan, K. K. (1985) Drug−protein interaction at the molecular level−A study of sulfonamide carbonic-anhydrase complexes, *J. Biosci. 8*, 491.
24. Karpusas, M., Nolte, M., Benton, C. B., Meier, W., Lipscomb, W. N., and Goelz, S. (1997) The crystal structure of human interferon *β* at 2.2 Å resolution, *Proc. Natl. Acad. Sci. U.S.A. 94*, 11813−11818.
25. Kester, W. R., and Matthews, B. W. (1977) Crystallographic study of the binding of dipeptide inhibitors to thermolysin: Implications for the mechanism of catalysis, *Biochemistry 16*, 2506−2516.
26. Anderson, W. F., Boodhoo, A., and Mol, C. D. (1988) The importance of purity in the crystallization of DNA binding immunoglobulin Fab fragments, *J. Cryst. Growth 90*, 153−159.
27. Corpet, F. (1998) Multiple sequence alignment with hierarchical clustering, *Nucleic Acids Res. 16*, 10881−10890.
28. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: Analysis of multiple sequence alignments in PostScript, *Bioinformatics 15*, 305−308.
29. Kraulis, P. J. (1991) MOLSCRIPT−A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr. 24*, 946−950.
30. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic molecular graphics, *Methods Enzymol. 277*, 505−524.
31. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes generally enhanced coloring capabilities, *J. Mol. Graphics Modell. 15*, 132−134.

BI048665R