Characterization of an Archaeal PIN Domain as an Exonuclease*

Vickery L. Arcus‡§¶, Kristina Bäckbro‡§¶, Annette Roos‡¶, Emma L. Daniel‡¶**, and Edward N. Baker‡**

From the ‡School of Biological Sciences, §GResearch Structural Biology Laboratory and **Centre of Molecular Biodiscovery, University of Auckland, Private Bag 92019, Auckland, New Zealand

Received for publication, December 18, 2003, and in revised form, January 15, 2004
Published, JBC Papers in Press, January 20, 2004, DOI 10.1074/jbc.M313833200

Genome sequencing projects have focused attention on the problem of discovering the functions of protein domains that are widely distributed throughout living species but which are, as yet, largely uncharacterized. One such example is the PIN domain, found in eukaryotes, bacteria, and Archaea, and with suggested roles in signaling, RNase editing, and/or nucleotide binding. The first reported crystal structure of a PIN domain (open reading frame PAE2754, derived from the crenarchaeon, Pyrobaculum aerophilum) has been determined to 2.5 Å resolution and is presented here. Mapping conserved residues from a multiple sequence alignment onto the structure identifies a putative active site. The discovery of distant structural homology with several exonucleases, including T4 phage RNase H and flap endonuclease (FEN1), further suggests a likely function for PIN domains as Mg2+-dependent exonucleases, a hypothesis that we have confirmed in vitro. The tetrameric structure of PAE2754, with the active sites inside a tunnel, suggests a mechanism for selective cleavage of single-stranded overhangs or flap structures. These results indicate likely DNA or RNA editing roles for prokaryotic PIN domains, which are strikingly numerous in thermophiles, and in organisms such as Mycobacterium tuberculosis. They also support previous hypotheses that eukaryotic PIN domains participate in RNA, and nonsense-mediated RNA degradation.

The explosive growth of whole genome sequencing efforts, and the discovery that a large proportion of the asserted gene products are of unknown or poorly understood function, has focused attention on new approaches to assigning function. In the absence of sufficient sequence similarity to clearly infer homology with already characterized proteins, a variety of bioinformatic approaches have been used to obtain functional clues. These include, for example, analyses of genome location (seeking potential operons), phylogenetic profiling, and observations of gene fusions in different species (1, 2). An alternative, complementary, approach is to use analyses of protein three-dimensional structure to derive functional insights, because three-dimensional structure is conserved in evolution much more strongly than sequence. This provides a rationale for a number of structural genomics initiatives (3–5).

A part of a pilot structural genomics project aimed at the discovery of biological function, we have focused on gene products from the hyperthermophilic crenarchaeon Pyrobaculum aerophilum, an organism whose complete genome sequence was published recently (7). A whole-genome comparison of P. aerophilum and Mycobacterium tuberculosis, two organisms with very different, and in a sense extreme, lifestyles, led us to identify a set of 250 pairs of orthologous genes that are both widely distributed in nature and are shared by these two organisms. Among these were a set of four genes from P. aerophilum (PAE0151, PAE0285, PAE0337, and PAE2754) and four from M. tuberculosis (Rv0065, Rv0549, Rv0960, and Rv1720) that have since been clustered at NCBI as part of COG4113, with members drawn from Archaea, cyanobacteria, actinobacteria and α-proteobacteria (Table I). These are now annotated in Pfam as PIN domains.

PIN domains, named for their homology with the N-terminal domain of the pili biogenesis protein (PIN, Pfam N terminus) (8), comprise a very large family of proteins with representatives in all three kingdoms of life. They are classified in the Pfam database (www.sanger.ac.uk/Software/Pfam) as PF01850, currently with more than 340 members. Functional annotation of the PIN domains is equivocal. They were initially thought to function in signaling (9), but a recent bioinformatic analysis of nearly 100 PIN domain sequences identified a set of five conserved acidic residues and a sixth conserved position where there is either a serine or threonine residue. This has led to a suggested exonuclease function (10). In eukaryotes, for example, the Caenorhabditis elegans PIN domain, smg-5, and the yeast PIN domain, NMD4p, are postulated to be ribonucleases that bind to helicases as part of the machinery for RNA degradation via the RNA, and nonsense-mediated RNA degradation (NMD) pathways (10).

In Archaea and thermophilic bacteria, PIN domains have been associated with a possible role in DNA repair. A recent analysis of conserved gene context across fully sequenced prokaryotic genomes revealed, in most Archaea and some thermophilic bacteria, a previously unrecognized cluster of genes containing DNA polymerases, helicases, nucleases, and many conserved hypothetical open reading frames, one of which is a PIN domain clustering in COG1848 (11). This suggested a new
DNA repair system in these organisms. DNA repair, and particularly mismatch repair, in thermophiles is a vexing question. The absence of key mismatch repair enzymes such as MutS and MutL, which are highly conserved in mesophiles from *Escherichia coli* to humans, has been suggested to result in “mutator” lifestyles for some thermophiles (7), in which adaptive mutations enable the organism to adapt to stress or extreme environments (12). Also absent from many of the fully sequenced thermophilic genomes are several well conserved nucleotide excision repair enzymes. The discovery of a new DNA repair operon in Archaea addresses this question and by implicating PIN domains as part of this operon adds another piece of functional evidence for this large protein family.

Here we present the first crystal structure of a PIN domain, from the crenarchaeon *P. aerophilum*. The protein, the gene product of open reading frame PAE2754, proves to be a distant structural homologue of T4 RNase H and other exonucleases, despite insignificant sequence identity. Strict conservation of the active site residues suggests that this PIN domain is indeed an exonuclease. We have confirmed this functional hypothesis in *vitro*. This has important implications both for archaeal DNA editing and for the role of PIN domains in eukaryotic RNA editing. It is also an illustration of the power of structural genomics whereby deep phylogenetic lineages are apparent at the structural level and lead directly to functional characterization of proteins of previously unknown function or with equivocal and/or general functional annotation.

**EXPERIMENTAL PROCEDURES**

**Protein Expression, Purification, and Crystallization**—The predicted open reading frame PAE2754 was amplified from genomic DNA by PCR, subcloned into the expression vector pPROEx (Invitrogen), transformed into *E. coli* BL21(DE3) cells, and expressed as an N-terminal His$_6$-tagged protein. Purification involved a heat step, in which incubation for 40 min at 80 °C denatured a large fraction of the *E. coli* proteins, followed by Ni$^{2+}$ affinity chromatography and size exclusion chromatography, as described (13).

Two site-specific mutations, L65M and L80M, were designed and introduced to facilitate structure determination by multiancestral anomalous diffraction (MAD) methods using the selenomethionine (SeMet)-substituted protein. The single mutants were individually made and tested for expression and crystallization, followed by the double mutant L65M/L80M. Mutagenesis was performed with the QuikChange site-directed mutagenesis kit (Stratagene). The double mutant was stable at 80 °C suggesting that the structure was not significantly destabilized by these mutations. The plasmid encoding the double mutant was transformed into the methionine auxotroph *E. coli* strain DL41(DE3) and grown in LeMaster medium with SeMet as the only methionine source. The SeMet-substituted double mutant protein (SeMet-PAE2754.MM) was then purified as above.

Both native PAE2754 and SeMet-PAE2754.MM were crystallized as described (13) and flash-cooled for data collection by soaking in cryo-protectant (mother liquor plus 10% glycerol) immediately prior to placement in a stream of cold N$_2$ gas at 110 K.

**Structure Determination and Refinement**—Native PAE2754 x-ray diffraction data to 2.5-Å resolution were collected at the National Synchrotron Light Source, Brookhaven, on beamline X8C (λ = 1.000 Å). MAD data at two wavelengths were collected for SeMet-PAE2754.MM at the Stanford Synchrotron Radiation Laboratory, beamline 9-1. The data were processed using DENZO and SCALEPACK (14). Data collection and refinement statistics are given in Table II. The structure of PAE2754.MM was determined by single anomalous diffraction using a single SeMet data set at λ = 0.9794 Å. This is not where the anomalou differences are maximized for selenium, but this was necessary because the “remote” wavelength data set proved to be of poor quality due to crystal decay (data were initially collected at two wavelengths in accordance with Ref. 15). Using SOLVE (16, 17), a total of 17 out of a possible 24 selenium sites (from the 8 monomers in the crystal asymmetric unit) were located, based on anomalous differences. These gave initial phases to 2.8 Å with a figure of merit of 0.18 and a Z-score of 20.6. The phases were improved using maximum likelihood density modification via RESOLVE (16). Successful improvement of the phases was dependent on user-defined non-crystallographic symmetry elements based on the initial selenium positions.

**RESULTS**

**Crystal Structure of PAE2754**—The crystal structure of the protein encoded by the open reading frame PAE2754 from *P. aerophilum* was solved at 2.8-Å resolution, as a selenomethionine (SeMet)-labeled double mutant, incorporating two Leu → Met mutations, that was constructed to facilitate phasing by MAD methods. This derivative structure was then refined to give a final R-factor of 0.226 (R$_{free}$ = 0.279). The native structure was then solved by molecular replacement and refined at 2.5 Å resolution to a final R-factor of 0.250 (R$_{free}$ = 0.305). The resulting model has good stereochemistry (Table II) with 92% of residues in the most favored region of the Ramachandran plot.

The PAE2754 monomer forms a single domain in which the 133-residue polypeptide is folded as an αβ/α stack, with a central twisted parallel β-sheet of five short strands (Fig. 1). The strand order is 32145, and the twist of the sheet is such that the outer strands, which involve just three residues each, are oriented at 160° with respect to each other. Between strand 2 and strand 3, helices α2 and α3 pack in an antiparallel manner to form a long protrusion that extends orthogonally from the αβ/α stack. Hydrophobic cores above and below the central β-sheet stabilize the αβ/α stack, and a third hydrophobic mini-core is formed below the stack by the orthogonal pack-
ing of helices α4 and α5. These hydrophobic cores are highly populated by Ala, Leu, and Val residues, which compose no less than 40% of the PAE2754 sequence.

Dynamic light scattering measurements show that PAE2754 forms a tetramer in solution, and accordingly the 12 monomers in the crystal asymmetric unit are organized as three tetramers (eight monomers in the asymmetric unit are organized as two tetramers for PAE2754_MM). The tetramer is best described as a dimer of dimers (Fig. 2). An extensive dimer interface is formed by the 2-fold related packing of a nearly continuous region of sequence between residues 32 and 94, spanning helices 2, 3, and 4. This interface buries 1440 Å² of surface area (19% of the total monomer surface) and is dominated by hydrophobic interactions, marked by a striking interdigitation of many large hydrophobic side chains. There are just six hydrogen bonds between the two protein chains, all centered around the stacked histidine aromatic rings at the center of the interface.

The tetramer is formed by the association of two dimers via a relatively small interface, in which the C terminus of helix α2 from one monomer contacts the N terminus of helix α6 from another. Stabilizing interactions at this interface are modest and include a salt bridge between Arg-48 from one monomer and Asp-110 and Glu-112 from another. The backbone carbonyl group from Arg-48 also hydrogen-bonds across the interface to the amide group of Arg-111. Although only 450 Å² of surface area per monomer is buried on formation of this interface, the cooperative association of two dimers buries in total 450 Å². There is also a diagonally stabilizing interaction across the tetramer whereby a chloride ion linearly coordinates two arginine residues. The chloride lies at the center of a sphere of charged side chains that include Arg-48 (chain A), Glu-112 and Lys-116 (chain B), Glu-112 and Lys-116 (chain C), and Arg-48 (chain D).

Residues that are conserved across COG4113 (Figs. 2 and 3) are clustered in a pocket formed at the C-terminal end of the β-sheet and the N termini of helices α2 and α6. This arrangement brings together four conserved acidic residues (Asp-8,
Glu-38, Asp-92, and Asp-110 in PAE2754) that point into the pocket and create a highly negatively charged hole. Two other conserved residues on either side of Asp-110, Thr-108, and Leu-112, also flank the acidic pocket with Thr-108 being hydrogen-bonded to Asp-92. In the PAE2754 dimer (Fig. 2), the two acidic pockets are ~20 Å apart and are separated by an intriguing structure formed by the one remaining fully conserved residue, Tyr-91; the two Tyr-91 side chains lie adjacent at the dimer interface, with their aromatic rings parallel and 6 Å apart. Upon formation of the tetramer, the four active site pockets lie in the interior of a tunnel with restricted access via two openings on opposite sides of the tetramer (Fig. 2). Adjacent lysine and glutamic acid residues (Lys-45 and Glu-46) from each monomer flank the entrances to the tunnel.

Structural Comparisons—A DALI search (22) using the monomer structure as a search model gave no significant matches to structures in the Protein Data Bank. The best structural match was to a porcine D-amino acid oxidase (DAO) (23) with a Z-score of 3.3, a root mean square difference (r.m.s.d.) in atomic positions of 3.9 Å for 97 Ca positions, and 8% sequence identity. The topological match between PAE2754 and DAO in the overlaid region was relatively good, with matches for all the major elements of secondary structure in PAE2754 except α4, although DAO does have two large insertions of 90 amino acids (between α2 and α3) and 110 amino acids (between β4 and β5). DAO has an FAD cofactor whose nucleotide component binds into the hole where the active site is hypothesized for PAE2754. What was striking, however, was the observation that residues that are conserved across COG4113, including the acidic residues at the putative active site, aligned structurally with similar residues in T4 RNase H that are involved in Mg²⁺ binding and catalysis (Fig. 4). Furthermore, these residues are also conserved across a large family of related prokaryotic exonucleases. This led us to test Mg²⁺ binding and DNase activity in vitro.

Thus, although the fold of PAE2754 is most closely related to domains that bind ADP or FAD, suggesting nucleotide binding as part of its core function, sequence conservation at the active site predicts that PAE2754 (along with the many other PIN domains) belongs to the T4 RNase H family of exonucleases (26). Furthermore, once the superposition of T4 RNase H on to the PAE2754 structure could be established, it became clear that other members of the exonuclease family, including the exonuclease domain of TaqDNA polymerase (27, 28) and the Flap endonuclease-1 (FEN-1) (29, 30), which has both endo- and exonuclease activities, are also structurally related to PAE2754 and, by extension, to other PIN domains.

In Vitro Tests for Exonuclease Activity—Exonuclease assays were carried out using synthetic DNA primers designed to give a long 5′–3′ single-stranded overhang. A time course incubation clearly shows PAE2754 has a Mg²⁺-dependent exonuclease activity (see Fig. 5). This experiment was repeated with Mn²⁺ in place of Mg²⁺ with equivalent results (data not shown), but no activity was seen in the absence of a suitable divalent cation. These initial tests show that the cleavage of single-stranded DNA by PAE2754 is slow and requires equimo-
lar amounts of DNA, Mg\(^{2+}\), or Mn\(^{2+}\) and protein to provide catalysis. The sluggish reaction may be the result of the non-optimal substrate and/or the non-optimal temperature of the assay; we presume that the optimal temperature for this enzyme is 95–100 °C. Assays to determine substrate specificity and the optimal temperature are the subject of ongoing work.

**Putative Active Site**—The four conserved acidic residues in each monomer, Asp-8, Glu-38, Asp-92, and Asp-110, are clustered together in a surface pocket, facing into the tunnel through the center of the tetramer. Two of these residues, Asp-8 and Asp-110, are at the N termini of helices α1 and α6, respectively, and their carboxylate groups are fixed in place by typical helix N-cap hydrogen bonds (with Ala-11 NH and Tyr-91 carboxylate groups). This residue is almost fully conserved as Lys or Arg. The distance between the two Mg\(^{2+}\) ions is 8 Å, too small for double-stranded DNA or RNA, but consistent with the protein threading overhanging single-stranded nucleic acids or flap structures, as is the case for flap endonucleases. Aromatic residues have been found previously (33, 34) to perform such a function in, for example, single-stranded RNA and DNA binding domains. Second, side chains of Lys-45 from each monomer project into the tunnel and could be involved in binding to nucleic acid phosphate groups. This residue is almost fully conserved as Lys or Arg.

The active site is only accessible from inside the tunnel through the tetramer, implying that nucleic acid substrates must thread through it. The diameter of the opening to the tunnel is an oval with dimensions of about 10 × 14 Å, too small for double-stranded DNA or RNA, but consistent with the protein cleaving overhanging single-stranded nucleic acids or flap structures, as is the case for flap endonucleases.

**DISCUSSION**

PIN domains, of which Pae2754 is the first to be structurally characterized, appear to be highly abundant and to have a deep lineage through all three kingdoms of life. Thus, Clissold and Ponting (10) identified 95 PIN domain sequences from a 90% non-redundant data base using HMM sequence searching methods. Makarova and Koonin (35) find 122 PIN sequences, of which PAE2754 is the first to be structurally characterized. No fewer than 10 related COGs bear the annotation "predicted nucleic acid-binding protein, conserved PIN domain." Thus, the invariable threonine residue, Thr-108, is a candidate for involvement in catalysis but is fairly well buried, hydrogen-bonded to Asp-8 Oδ2 and Asp-110 NH, and it is difficult to see how it can play any direct role. Two other hydroxyl-containing residues, Ser-10 and Thr-89, which are almost fully conserved, are also adjacent to the metal site where they are fully exposed in the central tunnel and could play a role in catalysis or binding. Two other features of the active site region seem likely to be important. First, the pairwise, parallel, stacking (Fig. 2) of the aromatic rings of the conserved Tyr-91 residues on the inner surface of the central tunnel suggests that these residues could participate in substrate binding by stacking on either side of a nucleotide base. Aromatic residues have been found previously (33, 34) to perform such a function in, for example, single-stranded RNA and DNA binding domains. Second, side chains of Lys-45 from each monomer project into the tunnel and could be involved in binding to nucleic acid phosphate groups. This residue is almost fully conserved as Lys or Arg.

The active site is only accessible from inside the tunnel through the tetramer, implying that nucleic acid substrates must thread through it. The diameter of the opening to the tunnel is an oval with dimensions of about 10 × 14 Å, too small for double-stranded DNA or RNA, but consistent with the protein cleaving overhanging single-stranded nucleic acids or flap structures, as is the case for flap endonucleases.
able to cluster the PIN domains with exonucleases and hypothesize their role in RNAi and nonsense-mediated RNA degradation when just 5 of 135 sequence positions are conserved across the domain.

Our structure of PAE2754 places these conserved residues close together in three-dimensional space. Together with our experimental evidence for nuclease activity, and the demonstrated structural homology with known exonucleases such as T4 ribonuclease H and the flap exonuclease, it further provides compelling evidence that PIN domains do indeed play a role in DNA and/or RNA editing processes through Mg$^{2+}$-dependent exonuclease activity.

Multiple sequence alignments of PIN domains based on COGs or Pfam show the following three principal features: (i) three conserved aspartic acid residues, one near the N terminus of the protein and two clustered near the C terminus; (ii) a conserved threonine or serine adjacent to the last conserved aspartic acid, forming a (T/S)XD motif in which the threonine or serine possibly plays a catalytic role; (iii) a well conserved acidic residue (either Asp or Glu) in the center of the sequence. The acidic residues are clustered in such a way that they could support the coordination of either one or two Mg$^{2+}$ ions. If two Mg$^{2+}$ ions are bound, as is commonly the case in polymerases and nucleases (25), and has been proposed as essential for catalysis (31), it is probable that one will be bound to Asp-8, Asp-92, and Asp-110, which are in close proximity, and the second to Glu-38. The former site may be of higher affinity given that the side chains of Asp-8 and Asp-110 are fixed at the

![Fig. 4](image_url)

**Fig. 4. Structural comparison of the PAE2754 PIN domain with T4 RNase H.** A, stereo view of the structures of PAE2754 (gray) and T4 RNase H (blue) aligned according to the superposition matrix from DALI. This produces a close alignment of the central β-sheet of the two proteins, although other secondary structural elements match less well. PAE2754 helices are labeled according to Fig. 2, and the position of the two Mg$^{2+}$ ions in RNase H are shown as blue spheres. B, an expanded view, in stereo, of the active site showing just the secondary structure elements for PAE2754 and the matching residues from each of the two structures along with the two Mg$^{2+}$ ions from the T4 RNase H structure. Putative active site residues for PAE2754 are shown in gray, and active site residues for T4 RNase H are shown in blue. For clarity, just three of the five conserved positions are labeled (PAE2754 sequence and numbering). The remaining unlabeled residues are Asp-92 and Thr-108.

![Fig. 5](image_url)

**Fig. 5. In vitro exonuclease activity of PAE2754.** A polyacrylamide/urea denaturing gel showing DNA stained by ethidium bromide (see text). Lane A shows the 54-bp oligonucleotide alone. Lanes B–G show 1–5- and 19-h incubations, respectively, of annealed oligonucleotides (54 + 18 bp) with PAE2754 and MgCl$_2$ at 37 °C. Lane H shows a 19-h incubation of annealed oligonucleotides (54 + 18 bp) with PAE2754 at 37 °C in the absence of MgCl$_2$.
N termini of α-helices. On the other hand, a direct and essential role for Thr-108 in catalysis seems less likely, given its relatively buried location.

The ways in which PIN domains associate into oligomers, or are combined with other domains or other proteins, are likely to determine the types of editing in which they are involved and the types of substrates on which they act. Thus, the PAE2754 structure is tetrameric, both in solution (as shown by dynamic light scattering and gel filtration) and in the crystal, and the tunnel through the center of the tetramer provides quite restricted access to the active site. Our preliminary modeling suggested that only single-stranded nucleic acids, or single-stranded overhangs from duplex structures, could serve as substrates. This hypothesis is supported by our in vitro assays, using double-stranded DNA with a single-stranded overhang as a template, although we have not demonstrated any specificity in either substrate or the direction of DNA cleavage, and further functional studies clearly need to be carried out. We also note that the FEN1 endonuclease from *Pyrococcus horikoshii* forms a dimer that is topologically similar to the PAE2754 tetramer, with the two active sites inside the dimer and access via a hole to the exterior.

In contrast, a second example of what is clearly a PIN domain structure has recently been deposited in the Protein Data Bank (ID code 1O4W, deposited by the Joint Centre for Structural Genomics (San Diego)). This protein shares 23% sequence identity with PAE2754 and forms a very similar monomer in which 101 residues match with an r.m.s.d. of 3.0 Å. This *Archaeoglobus fulgidus* PIN domain forms a dimer in which the monomers are tethered by a stretch of 10 amino acids at the C terminus of the protein, thus separating the monomer active sites by 42 Å and creating a very different molecular environment from that of the four active sites in the PAE2754 tetramer.

An intriguing feature of the phylogenetic distribution of PIN domains is that they seem to be amplified in a number of species, sometimes to a remarkable extent. These include *A. fulgidus*, *P. horikoshii*, and *Methanococcus jannaschii*, all of which are thermophilic euryarchaeota. Among mesophilic bacteria, PIN domains also seem to be extraordinarily amplified in *M. tuberculosis*. COG1848, proteins from which have been predicted to be part of a new DNA repair system (11), includes no fewer than 14 PIN domain proteins from *M. tuberculosis*. Additionally, three other COGs are found to include a further 12 *M. tuberculosis* PIN domain proteins between them. This presents the intriguing question as to why *M. tuberculosis* would have 26 PIN domains as part of its DNA or RNA editing machinery.

This augmentation of the exonuclease family and the accumulation of PIN domain proteins in certain species raise intriguing questions as to the cellular function of the PIN domains. In thermophilic Archaea, it is reasonable to expect that there would be additional modes of DNA and RNA editing and repair mechanisms due to the elevated levels of oligonucleotide modification and damage caused by the high temperatures. Indeed, a predicted new DNA repair operon encompassing ~20 genes in Archaea contains a PIN domain protein (11, 12). However, the augmentation of the exonucleases in mesophiles is unexpected. *M. tuberculosis* is the most extreme example of the currently sequenced organisms, with its 26 PIN domain proteins, all of which contain the acidic quartet of Mg$^{2+}$-binding residues and the adjacent serine or threonine. This suggests that a large retinue of exo- or endonuclease enzymes are present in this pathogenic bacterium.

There has been speculation about the functional relevance of the presence or absence of DNA repair genes in *M. tuberculosis*. It has been suggested that an absence of repair enzymes might be beneficial under conditions of stress or therapeutic treatment during long periods of stationary phase as is the case for *M. tuberculosis* (36, 37). On the other hand, the correlation between those species lacking the *mutS* gene and thus assumed to be mismatch repair-deficient and the expansion, in the same species, of the PIN domain proteins may not be coincidental.

We conclude that these PIN domain proteins are very likely substitutes for a number of exo- and endonucleases, with roles in DNA repair that are apparently missing from the *M. tuberculosis* genome and from the genomes of various extremophile species (12).

The second possible role for these nucleases is as a defense arsenal designed to neutralize phage via DNA and/or RNA degradation. It is the case that these functions have been adopted in eukaryotic PIN domains to degrade RNA via the RNA, and NMD pathways (10). Hence, in the eukaryotes it seems that editing and repair functions for PIN domains have likely been adopted to degradation pathways.

Acknowledgments—We thank Alexey Murzin for helpful discussions surrounding PIN domains and their functional and structural classification. We thank Li-Wei Hung for the data collection at the National Synchrotron Light Source, Brookhaven, under the auspices of the International *M. tuberculosis* Structural Genomics Consortium. We thank Clyde Smith for MAD data collection at the SSRL.

REFERENCES

1. Marcotte, E. M., Pellegrini, M., Thompson, M. J., Yeates, T. O., and Eisenberg, D. (1999) *Nature* 402, 83–86
2. von Mering, C., Huynen, M., Jaeger, D., Schmidt, S., Bork, P., and Snel, B. (2003) *Nucleic Acids Res.* 31, 258–261
3. Goulding, C. W., Apostol, M., Anderson, D. H., Gill, H. S., Smith, C. V., Kuo, M. R., Yang, J. K., Waldo, G. S., Suh, S. W., Chauhan, R., Hale, A., Bachhawat, N., Mande, S. C., Johnston, J. M., Lott, J. S., Baker, E. N., Arcus, V. L., Leys, D., McLean, K. J., Munro, A. W., Berendzen, J., Sharma, V., Park, M. S., Eisenberg, D., Sacchettini, J., Alber, T., Rupp, B., Jacobs, W. J., and Terwilliger, T. C. (2002) *Curr. Drug Targets Infect. Disord.* 2, 141–147
4. Stevens, R. C., Yokoyama, S., and Wilson, I. A. (2001) *Science* 294, 89–92
5. Adams, M. W., Dailey, H. A., DeLucia, L. J., Luo, M., Prestegard, J. H., Rose, J. P., and Wang, B. C. (2003) *Acc. Chem. Res.* 36, 191–198
6. Teichmann, S. A., Murzin, A. G., and Chothia, C. (2001) *Curr. Opin. Struct. Biol.* 11, 354–363
7. Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Steeter, K. O., Simon, M. I., and Miller, J. H. (2002) *Proc. Natl. Acad. Sci. U.S.A.* 99, 984–989
8. Wall, D., and Kaiser, D. (1999) *Mol. Microbiol.* 32, 1–10
9. Noguchi, E., Hayashi, N., Azuma, T., Seki, T., Nakamura, M., Nakashima, N., Yanagida, M., He, X., Muddler, U., Sazer, S. and Nishimoto, T. (2005) *EMBO J.* 15, 5955–5965
10. Clissold, P. M., and Ponting, C. P. (2000) *Curr. Biol.* 10, R888–R890
11. Makarova, K. S., Aravin, A. D., Grishin, N. V., Rognazoni, I. B., and Kosinov, R. V. (2002) *Nucleic Acids Res.* 30, 482–496
12. Makarova, K. S., Aravin, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I., and Kosinov, E. V. (1999) *Genome Res.* 9, 608–628
13. Arcus, V. L., Backbóro, R., Roos, A., and Baker, E. N. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.*
14. Owsianowski, Z., and Minzor, W. (1997) *Methods Enzymol.* 276, 307–328
15. Gonzalez, A. (2003) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 59, 315–322
16. Terwilliger, T. C. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 58, 1937–1940
17. Terwilliger, T. C. (2001) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 57, 1755–1762
18. Holton, T., Ierger, T. R., Christopher, J. A., and Sacchettini, J. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 56, 722–734
19. Ierger, T. R., and Sacchettini, J. C. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 58, 2043–2054
20. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* 47, 110–119
21. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Grosse-Kunstleve, R. W., Jiang, S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* 54, 905–921
22. Holm, L., and Sander, C. (1996) *Trends Biochem. Sci.* 20, 478–480
23. Murzin, A. G. (1999) *Curr. Opin. Struct. Biol.* 9, 681–690
24. Barber, M. J., Neame, P. J., Lim, I. W., White, S., and Matthews, F. S. (1992) *Nature* 357, 612–616
25. Eom, S. H., Wang, J., and Steitz, T. A. (1996) *Nature* 382, 278–281
26. Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., and Steitz, T. A. (1995) *Nature* 376, 612–616
29. Hosfield, D. J., Mol, C. D., Shen, B., and Tainer, J. A. (1998) Cell 95, 135–146
30. Hwang, K. Y., Baek, K., Kim, H. Y., and Cho, Y. (1998) Nat. Struct. Biol. 5, 707–713
31. Beese, L. S., and Steitz, T. A. (1991) EMBO J. 10, 25–33
32. Ceska, T. A., Sayers, J. R., Stier, G., and Suck, D. (1996) Nature 382, 90–93
33. Draper, D. E. (1999) J. Mol. Biol. 293, 255–270
34. Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V., and Wuttke, D. S. (2002) Science 296, 145–147
35. Makareva, K. S., and Koonin, E. V. (2003) Genome Biol. 4, 115.111–115.117
36. Karunakaran, P., and Davies, J. (2000) J. Bacteriol. 182, 3331–3335
37. Mizrahi, V., and Andersen, S. J. (1998) Mol. Microbiol. 29, 1331–1339
38. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA
39. Nicholls, A., Sharp, K., and Honig, B. (1991) Proteins Struct. Funct. Genet. 11, 281–296
40. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) Nucleic Acids Res. 24, 4876–4882
Distant Structural Homology Leads to the Functional Characterization of an Archaeal PIN Domain as an Exonuclease

Vickery L. Arcus, Kristina Bäckbro, Annette Roos, Emma L. Daniel and Edward N. Baker

J. Biol. Chem. 2004, 279:16471-16478.
doi: 10.1074/jbc.M313833200 originally published online January 20, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M313833200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 38 references, 6 of which can be accessed free at http://www.jbc.org/content/279/16/16471.full.html#ref-list-1