

Logical Identification of an Allantoinase Analog (*puuE*) Recruited from Polysaccharide Deacetylases^{*[5]}

Received for publication, February 13, 2008, and in revised form, May 7, 2008. Published, JBC Papers in Press, June 12, 2008, DOI 10.1074/jbc.M801195200

Ileana Ramazzina^{†1}, Laura Cendron[§], Claudia Folli[‡], Rodolfo Berni[‡], Daniela Monteverdi[¶], Giuseppe Zanotti[§], and Riccardo Percudani^{‡2}

From the Departments of [†]Biochemistry and Molecular Biology and [¶]Mathematics, University of Parma, 43100, Parma, Italy and the [§]Department of Chemistry, University of Padua, 35131 Padua, Italy

The hydrolytic cleavage of the hydantoin ring of allantoin, catalyzed by allantoinase, is required for the utilization of the nitrogen present in purine-derived compounds. The allantoinase gene (*DAL1*), however, is missing in many completely sequenced organisms able to use allantoin as a nitrogen source. Here we show that an alternative allantoinase gene (*puuE*) can be precisely identified by analyzing its logic relationship with three other genes of the pathway. The novel allantoinase is annotated in structure and sequence data bases as polysaccharide deacetylase for its homology with enzymes that catalyze hydrolytic reactions on chitin or peptidoglycan substrates. The recombinant PuuE protein from *Pseudomonas fluorescens* exhibits metal-independent allantoinase activity and stereospecificity for the *S* enantiomer of allantoin. The crystal structures of the protein and of protein-inhibitor complexes reveal an overall similarity with the polysaccharide deacetylase β/α barrel and remarkable differences in oligomeric assembly and active site geometry. The conserved Asp-His-His metal-binding triad is replaced by Glu-His-Trp, a configuration that is distinctive of PuuE proteins within the protein family. An extra domain at the top of the barrel offers a scaffold for protein tetramerization and forms a small substrate-binding cleft by hiding the large binding groove of polysaccharide deacetylases. Substrate positioning at the active site suggests an acid/base mechanism of catalysis in which only one member of the catalytic pair of polysaccharide deacetylases has been conserved. These data provide a structural rationale for the shifting of substrate specificity that occurred during evolution.

Allantoin ((2,5-dioximidazolidin-4-yl)urea) is one of the most nitrogen-rich organic compounds. Some plants economize carbon by using this compound as a nitrogen carrier

instead of glutamine and glutamate that have less favorable nitrogen-carbon ratios (1). Conversely, soil animals eliminate the excess of nitrogen present in purine bases mainly in the form of uric acid and allantoin. In turn, microorganisms can utilize these compounds as nitrogen sources by degrading them to carbon dioxide, glyoxylate, and ammonia (2).

The metabolic routes to and from allantoin are straightforward; there is only one known path for its formation and degradation (see Fig. 1A). Allantoin is formed from uric acid, in a three-step enzymatic pathway that has been recently elucidated (3). The last step of the pathway is the decarboxylation of 2-oxo-4-hydroxy-4-carboxy-ureido-imidazoline (OHCU)³; OHCU is unstable and spontaneously decomposes to racemic allantoin (4). However, because OHCU decarboxylase is strictly stereoselective, only the *S* enantiomer of allantoin is formed in nature (5, 6). Allantoin, either deriving from endogenous purine breakdown or imported through specific permeases (7), is hydrolyzed by allantoinase (EC 3.5.2.5) to allantoate. In turn, allantoate can be acted upon by two different enzymes to release ammonia (upon allantoate amidohydrolase) or urea (upon allantoicase). Given that the cleavage of the hydantoin ring is required for the utilization of the nitrogen present in the allantoin molecule, no organism is expected to be able to use allantoin for growth without possessing allantoinase activity. This activity is indeed especially common in microorganisms, and it has been documented in various bacterial and yeast species (2).

A gene coding for allantoinase (*DAL1*) was first identified in *Saccharomyces cerevisiae* (8). The corresponding protein revealed a clear sequence similarity with other enzymes acting on cyclic amide rings, such as hydantoinases and dihydroorotases (9). By structural comparisons, these proteins were then included in a large superfamily (comprising urease, adenine deaminase, and other hydrolases) characterized by a common structural core consisting of an ellipsoidal (β/α)₈ barrel and a conserved metal-binding site (10). Genes coding for proteins with allantoinase activity have been documented in amphibians, insects, bacteria, and plants (11–14); they are all homologous to *DAL1* and constitute a monophyletic group in the amidohydrolase superfamily, thus making identification of allantoinase genes in sequence data bases straightforward. Surprisingly, however, no *DAL1* sequences are found in organisms

^{*} This work was supported by the Universities of Parma and Padua. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

^[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Table S1.

The nucleotide sequence(s) reported in this paper has been submitted to the DDBJ/GenBankTM/EBI Data Bank with accession number(s) EU293536.

The atomic coordinates and structure factors (codes 3CL6, 3CL7, and 3CL8) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ Supported by a fellowship from SARCE SpA.

² To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, Parco area delle Scienze, University of Parma, 43100. Tel.: 39-0521-905140; Fax: 39-0521-905151; E-mail: riccardo.percudani@unipr.it.

³ The abbreviations used are: OHCU, 2-oxo-4-hydroxy-4-carboxy-ureido-imidazoline; HYN, hydantoin; ACA, 5-amino-4-imidazole carboxamide hydrochloride.

Logical Identification of *puuE*

that are known to use allantoin as a nitrogen source. It is known, for example, that *Pseudomonas aeruginosa* possesses allantoinase activity (15), and a strain with a defective allantoinase gene, named *puuE* (purine utilization E), has been described (16). Similarly, allantoinase activity and a mutant strain that cannot grow on media containing allantoin as a sole nitrogen source have been identified in *Schizosaccharomyces pombe* (17). Yet genes coding for allantoinase do not appear to be present in the genomes of these organisms.

To find the genes responsible for the allantoinase activity in these and other organisms, we have utilized genome comparison looking for a gene whose presence or absence in whole genomes followed a particular logic relation (18) with three other genes of the purine degradation pathways. In accordance with the biochemical pathway, the presence of this candidate gene in a given genome would imply the absence of *DAL1* and the presence of either allantoinase or allantoate amidohydrolase (see Fig. 1A). We show that a previously uncharacterized gene exhibiting this logic relationship, whereas related to polysaccharide deacetylases and lacking sequence similarity with the known allantoinase, encodes a protein that specifically catalyzes the hydrolysis of (S)-allantoin into allantoic acid. The determination of the three-dimensional coordinates of the protein in complex with substrate analogs allowed us to relate the functional properties of the protein to its structure.

EXPERIMENTAL PROCEDURES

Bioinformatics—We first selected genes known or presumed through chromosomal proximity to be involved in purine degradation. Hidden Markov models of the corresponding protein families were downloaded from Pfam and used to search protein sets deriving from complete genomes with the pfam_scan.pl script (19). Protein sets of 138 complete genomes were downloaded from the NCBI ftp site. Proteins obtained with the pfam_scan.pl search were divided in orthologous groups by using OrthoMCL (20) at an inflation index of 2.0. Proteins belonging to orthologous groups were encoded as present “1” or absent “0” in a binary vector describing their genomic distribution.

Vectors were then combined according to the logic relationship $D \leftrightarrow \neg A \wedge (B \vee C)$, where A, B, and C are *DAL1* allantoinase, allantoate amidohydrolase, and allantoinase, respectively, and D is an unknown protein. A dedicated Perl script was used to evaluate the agreement of different vector combinations to the logic relationship, as measured by the “uncertainty coefficient” (18). Significance of the logic relationship was calculated by applying the right-sided Fisher’s exact test to a 2×2 contingency table with the observed outcomes of the logic expressions. The Ngram Statistic Package of Perl was used for calculations.

Multiple sequence alignments were carried out with the profile mode of Clustalw (21) using the superposition of the polysaccharide deacetylase structures as a guide; structural superpositions were obtained with Vector Alignment Search Tool at the NCBI. Protein alignments decorated with structural elements were visualized using the Esript server. The size of the *PuuE* tetramer was estimated from the three-dimensional coordinates with Moleman2. Interactions between *PuuE* chains were analyzed with PDBsum. Pymol was used to study models

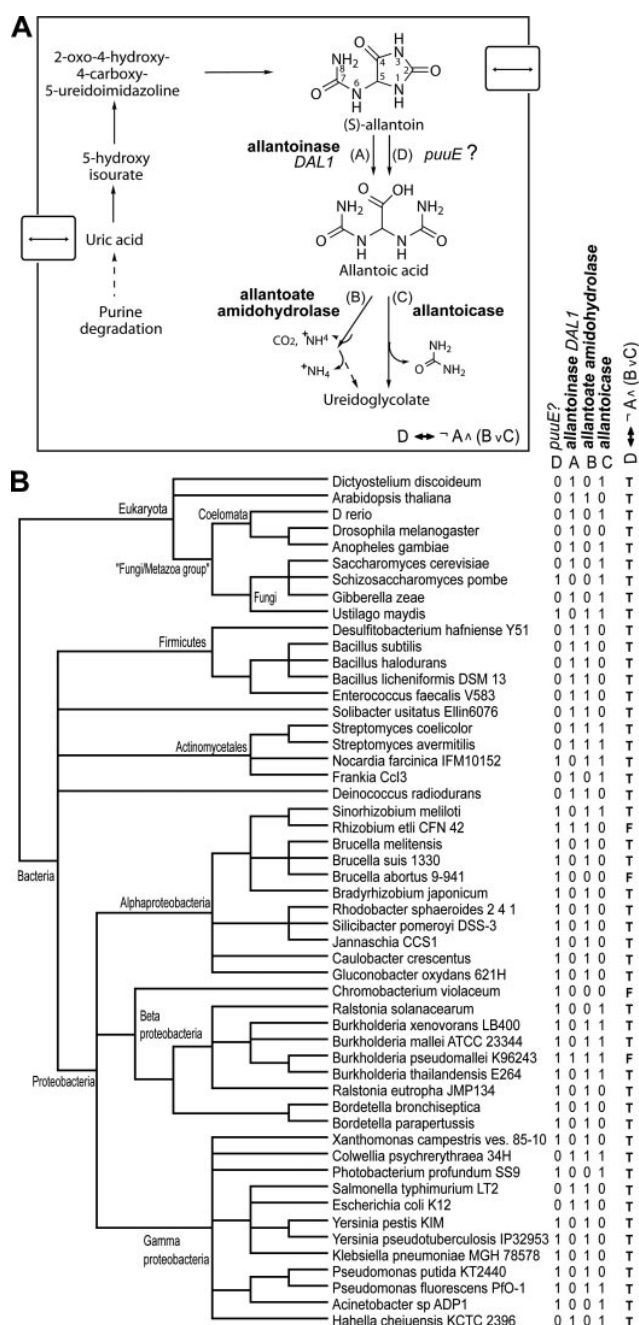


FIGURE 1. Logical identification of *puuE* allantoinase. A, pathway for the biosynthesis and utilization of allantoin as deduced from biochemical and genetic evidence. B, patterns of presence (marked 1) and absence (marked 0) of four logically related genes in fully sequenced genomes. Shown are 52 genomes possessing the *DAL1* gene or the putative *puuE* gene, of a preselected set of 138 genomes (see supplemental Table S1). The result of the logic expression $D \leftrightarrow \neg A \wedge (B \vee C)$ (where \neg = not; \wedge = and; \vee = or) derived from the metabolic pathway and applied to the presence/absence of genes, is shown as a truth table (T, true; F, false). Phylogenetic relationships among species are represented according to the NCBI taxonomy data base.

and generate figures. The topological representation of *PuuE* was constructed with TopDraw (22).

Chemicals—All of the reagents were from Sigma unless otherwise indicated. *Pseudomonas fluorescens* type strain was

obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSM 50090). (S)-Allantoin used in the CD assays was produced through the degradation of uric acid (0.2 mM) in the presence of 0.16 units of urate oxidase from *Candida utilis*, 0.12 μ g of zebrafish 5-hydroxyisourate hydrolase (23), and 6 μ g of zebrafish OHC decarboxylase (6). The reactions were conducted in 1 ml of 0.1 M potassium phosphate, pH 7.6. ^{15}N , ^{13}C -Labeled (S)-allantoin used in NMR assays was produced through the degradation of ^{15}N , ^{13}C -labeled uric acid (3) (6.9 mM) in the presence of 20 units of urate oxidase from *C. utilis*, 28 units of catalase from *Corynebacterium glutamicum* (Fluka), 3.7 μ g of zebrafish HIU hydrolase, and 25 μ g of zebrafish OHC decarboxylase. The reactions were conducted in 1.1 ml of 0.1 M potassium phosphate, 80% D_2O , pD 7.6.

Cloning, Gene Expression, and Protein Purification—Genomic DNA was prepared from *P. fluorescens* using described procedures (24). The region encoding the *P. fluorescens* PuuE protein (GenBankTM accession number EU293536) was PCR-amplified using *Go Taq* DNA polymerase (Promega) and two sequence specific primers: an NdeI-tailed upstream primer (5'-CATATGAGCGCTGACTACCC-3') and a downstream primer (5'-TTCGTAGATGTCGGCGAAG-3').

The amplification product was inserted into the pGEM vector (Promega) to generate the intermediate vector pGEM-PuuE. The restriction fragment obtained from NdeI digestion of plasmid pGEM-PuuE was then ligated into the dephosphorylated NdeI site of the expression vector pET11b (Novagen). The resulting plasmid was sequenced on both strands and electroporated into *Escherichia coli* BL21 codon plus cells (DE3).

The expression of PuuE was induced at an optical density at 600 nm of 0.6 with 0.5 mM isopropyl-1-thio- β -D-galactopyranoside; after 72 h at 12 °C the cells were resuspended in 180 ml of lysis buffer (50 mM sodium phosphate, 0.3 M NaCl, 1 mM β -mercaptoethanol, 10% glycerol, 1 μ M pepstatin, 1 μ M leupeptin, 100 μ M phenylmethylsulfonyl fluoride, 2 mg/ml lysozyme, pH 8) and incubated on ice for 30 min. The cells were lysed by four 15-s bursts of sonication. PuuE was purified to homogeneity, as assessed by 12% SDS-PAGE analysis, by gel filtration (Sephadex-G100, Pharmacia) with a final yield of ~ 3.5 mg/l of cell culture. The $\epsilon_{280\text{ nm}}$ (extinction coefficient) of PuuE was estimated to be $63830\text{ M}^{-1}\text{ cm}^{-1}$, based on the amino acid sequence.

The molecular weight of purified PuuE was estimated by gel filtration on a Sephadex G-100 column (Pharmacia) (55×1.5 cm), using buffer (0.1 M potassium phosphate, 0.3 M NaCl, pH 7.6) for equilibration and elution. The markers used were human retinol-binding protein (21.3 kDa), human transthyretin (55 kDa), bovine serum albumin (66.4 kDa), and cystathionine β -lyase (163.9 kDa).

Protein Crystallization, Data Collection, and Structure Determination—The crystals were obtained with the isothermal vapor diffusion technique in the hanging drop setting at 4 °C, after mixing equal volumes of 10 mg/ml recombinant PuuE in 0.1 M NaCl, 0.03 M $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$, pH 7.6, and a precipitant solution containing 0.2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M sodium cacodylate, pH 6.5, 30% polyethylene glycol 8000. They belong to the P4 tetragonal space group, with cell parameters $a = b = 98.13\text{ \AA}$ and $c = 62.10\text{ \AA}$. The Matthews coefficient (V_m) value

of about 2.05 corresponds to the presence of two monomers/asymmetric unit and a solvent content of about 39% of the cell volume. Analogous crystals were grown for complexes with 0.7 mM hydantoin (HYN) and 0.7 mM 5-amino-4-imidazole carboxamide hydrochloride (ACA).

Diffraction data were measured at the ID14-eh1 and ID29 beamlines of the European Synchrotron Radiation Facility in Grenoble, France. Three data sets were collected without adding any cryoprotectant for the uncomplexed enzyme (1.58 Å resolution) and the two complexes with HYN (1.8 Å resolution) and ACA (2.25 Å resolution). X-ray data collections were processed with the software MOSFLM (25) and merged and scaled with SCALA (26). The structure of the native enzyme was solved using the molecular replacement method with the program Phaser (27), starting from the crystal structure of a putative polysaccharide deacetylase from *P. aeruginosa* (PAO1) as a template (Protein Data Bank code 1Z7A). The refinement was carried out including two monomers in the asymmetric unit, by the software Refmac (28) and SHELXL (29), alternated by manual adjustments. The examination of the $F_o - F_c$ electron density map for the complex with HYN showed a clear difference density in the putative active site cavity of both the monomers present in the asymmetric unit, and this volunteer electron density had the expected molecular features of this small planar molecule. Analogous results were obtained for the complex with ACA, even if the electron density was less well defined. Coordinates and topology of HYN and ACA compounds were generated by the Dundee PRODRG2 Server (30). The crystallographic R factors for the uncomplexed enzyme and for the complexes with HYN and ACA were reduced to final values of 0.208 ($R_{\text{free}} = 0.232$), 0.199 ($R_{\text{free}} = 0.225$), and 0.207 ($R_{\text{free}} = 0.230$), respectively.

Biochemical Assays—Circular dichroism measurements were carried out at 25 °C in a 10-mm path length cuvette with a Jasco J715 spectropolarimeter. PuuE activity was determined using (S)-allantoin (0.2 mM) or (R,S)-allantoin (0.4 mM) as substrates in the presence of PuuE (0.37 μ M) in 1 ml of 0.1 M potassium phosphate, pH 7.6. CD spectra were recorded in the 200–350-nm range for 10 min, one spectrum every min. PuuE activity was also determined using a Varian Cary 1E UV-visible spectrophotometer, by the addition of (R,S)-allantoin (0.63 mM) in 1 ml of 0.1 M potassium phosphate, pH 7.6.

To evaluate the velocity of the enzyme-catalyzed reaction at different substrate concentrations, CD signals were monitored at a fixed wavelength (215 nm). The reactions were initiated by the addition of PuuE (0.19 μ M) to 1 ml of 0.1 M potassium phosphate, pH 7.6, containing different concentrations of (R,S)-allantoin or allantoic acid as substrates. The substrate concentrations varied between 0.015 and 0.86 mM for allantoin and between 0.028 and 7.81 mM for allantoic acid. The reaction rate was obtained for each substrate concentration by linear fitting of data points collected over the first 15 s of the reaction. The kinetics experiments were carried out in triplicate.

The equilibrium constants for the conversion of allantoin to allantoic acid were calculated by measuring the direction of CD signals ($\lambda = 215\text{ nm}$, 25 °C) upon enzyme addition at different product/substrate ratios. Allantoic acid/(R,S)-allantoin ratios varied between 70/30 and 99/1.

Logical Identification of *puuE*

The metal content of active *PuuE* preparations was determined spectrophotometrically using the chelator 4-(2-pyridylazo)resorcinol. The 4-(2-pyridylazo)resorcinol solution (5 mM) was prepared in Milli-Q water, pH 9. The enzyme (10 μ M) was denatured in 1 ml of 0.1 M potassium phosphate, 8 M urea, pH 8, and supplemented with 0.1 mM 4-(2-pyridylazo)resorcinol. The mixture was incubated at 25 °C for 10 min, and the absorbance in the 200–800-nm range was recorded using a Varian Cary 1E spectrophotometer. A calibration curve was obtained using ZnCl_2 (0.0–6.0 μ M) by monitoring the absorbance at 500 nm. The effect of the addition of metal ions on the enzyme activity was determined by preincubating the enzyme with buffers containing from 2- to 500-fold excess metal ions (Zn^{2+} , Co^{2+} , Cu^{2+} , Ca^{2+} , Mn^{2+} , and Mg^{2+}). The protein activity was monitored by UV-visible spectroscopy. The experiments were also carried out by preincubating the enzyme with EDTA (2.5–250 mM).

To obtain ^{13}C NMR spectra of the product of the *PuuE*-catalyzed reaction, a 1.1-ml solution of 0.1 M potassium phosphate, 80% D_2O , pD 7.6, containing ^{15}N , ^{13}C -labeled (*S*)-allantoin (6.9 mM) was supplemented with *PuuE* (0.54 μ M); the solution was gently stirred for 1 min and then transferred in a 5-mm NMR tube to collect spectra at different times. The ^{13}C NMR spectra were proton decoupled and were collected at 25 °C with a VARIAN Inova 600 instrument.

RESULTS

Logical Identification of *puuE*—Given three genes of the purine degradation pathway, *DAL1* allantoinase (*A*), allantoate amidohydrolase (*B*), and allantoicase (*C*), we want to find a fourth gene (*D*) whose presence in complete genomes is consistent with the logic relationship $D \Leftrightarrow \neg A \wedge (B \vee C)$. In other words, *D* is present if and only if *A* is absent and *B* or *C* are present. Genes of purine degradation are often clustered in prokaryotes, and in some cases even in eukaryotes (31). The search was therefore restricted to genes that are found, at least in some instances, in the proximity of genes involved in purine degradation.

After careful distinction between orthologous and paralogous families (see methods for details), the search in a set of 138 genomes identified a single strong candidate (Fig. 1). The logic expression relating *D* to the other three genes is false in four of the 52 genomes where *A* or *D* are found. In the remaining 86 genomes, 13 contain genes that were classified as allantoicase or allantoate amidohydrolase. Exceptions to the logical rule deduced by the metabolic pathway are not unexpected. They can arise by limitations of the gene search and classification procedure or represent true biological exceptions. An example of the latter case is the presence in mammals of an allantoicase ortholog that has a different function (32). Overall, the logic expression relating *D* to *DAL1* allantoinase, allantoate amidohydrolase, and allantoicase is true in 121 of 138 genomes, with an uncertainty coefficient (*U*) (18) of 0.43 ($p < 10^{-14}$).

The pattern of presence and absence of the four genes across genomes is largely independent from species phylogeny. While maintaining the logic relation, the assortment of the four genes has been modified several times during evolution by events of gene loss or gain (Fig. 1*B*). As pointed out in previous studies,

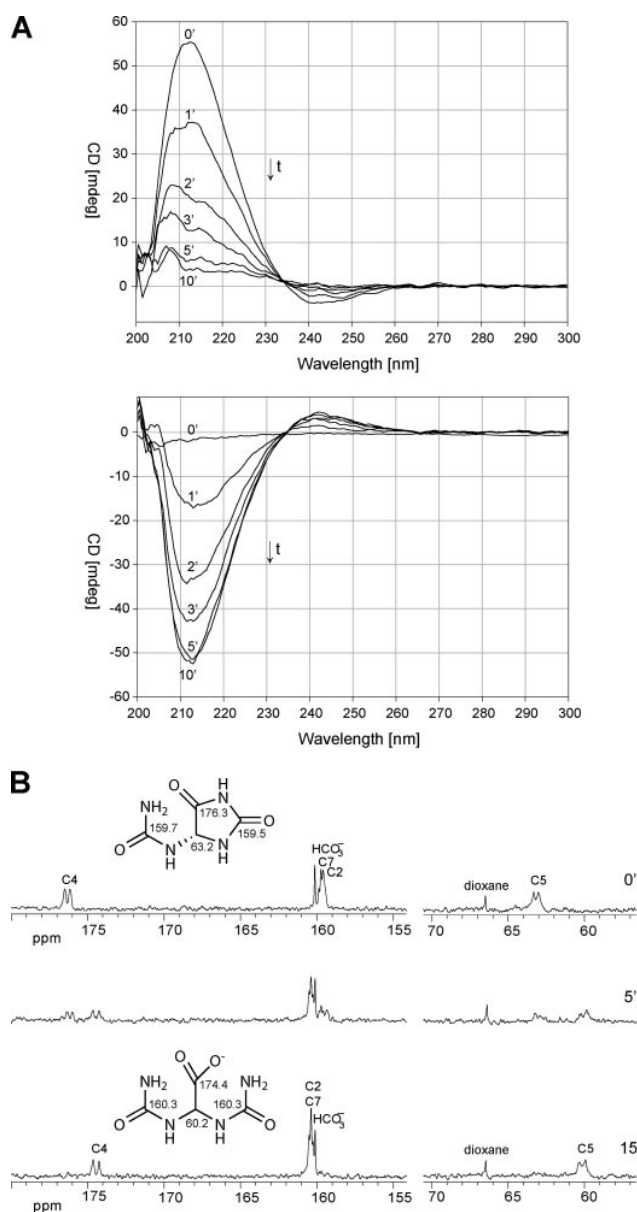


FIGURE 2. Biochemical activity of the *PuuE* protein. *A*, time-resolved CD spectra of (*S*)-allantoin (0.2 mM, upper panel) and (*R,S*)-allantoin (0.4 mM, lower panel) in the presence of *PuuE*. *B*, ^{13}C NMR spectra of ^{15}N , ^{13}C -labeled (*S*)-allantoin in the presence of *PuuE*. Carbon assignment is reported above peaks; the atoms are numbered as in Fig. 1. Dioxane (66.5 ppm) was used as chemical shift reference; HCO_3^- derives from decarboxylation of a labeled (*S*)-allantoin precursor (OHCU) (3).

multiple correlated events of gene loss or gain make a compelling case for the existence of a functional link (33).

Because of its relations with other genes of the purine degradation pathway, we considered the gene *D* responsible for the allantoinase activity found in *Pseudomonas*, *S. pombe*, and other organisms and named it *puuE* after the allantoinase mutant described in *P. aeruginosa* (16). The protein encoded by *puuE* has sequence similarity with enzymes that remove *N*-linked or *O*-linked acetyl groups of cell wall polysaccharides. Although acting on quite different substrates, polysaccharide

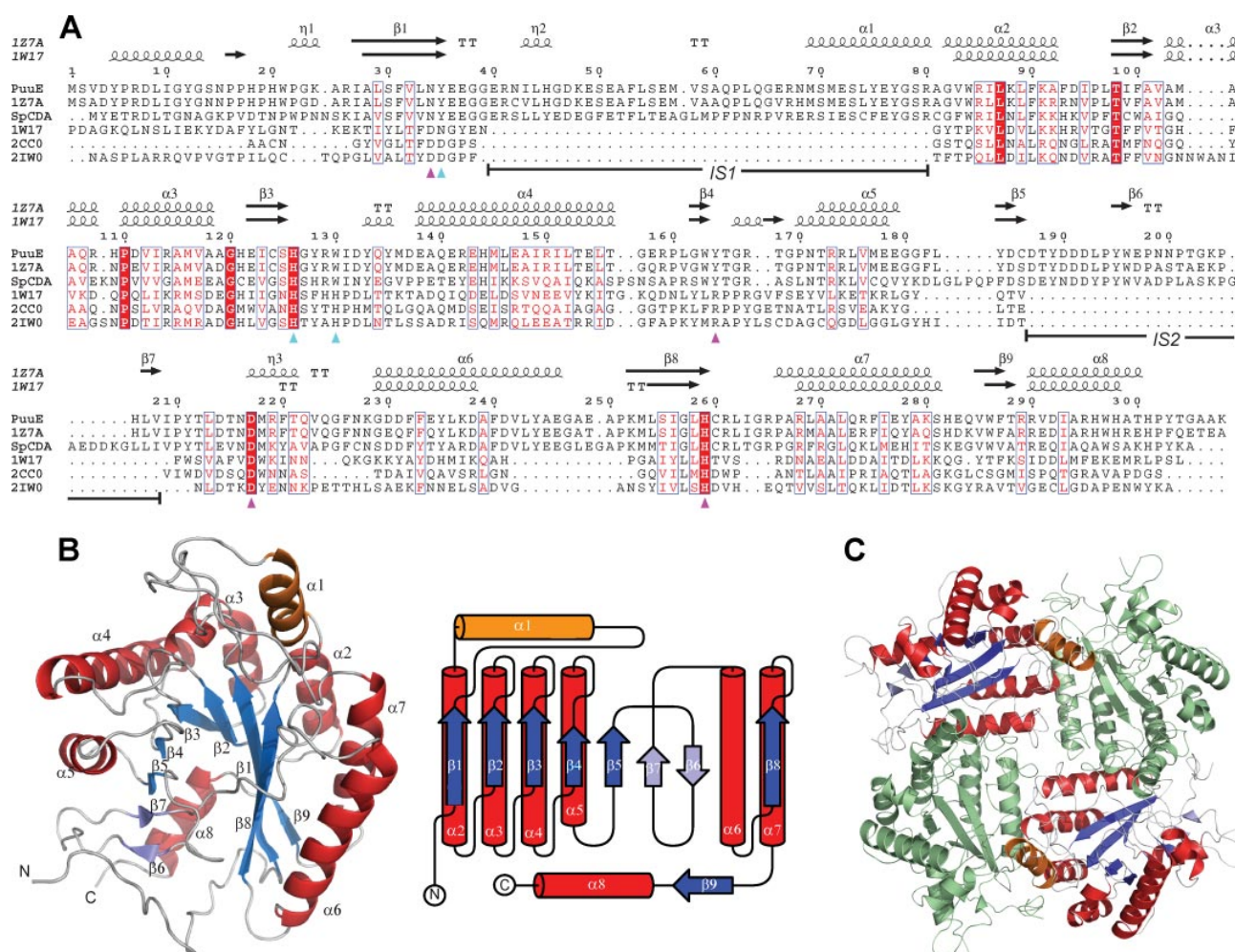


FIGURE 3. Sequence and structure of PuuE. A, the PuuE sequence is aligned with a hypothetical polysaccharide deacetylase from *P. aeruginosa* (Protein Data Bank code 1Z7A), a putative chitin deacetylase from *S. pombe* (SpCDA), a peptidoglycan deacetylase from *B. subtilis* (Protein Data Bank code 1W17), acetyl xylan esterases from *S. lividans* (Protein Data Bank code 2CC0), and a chitin deacetylase from *C. lindemuthianum* (Protein Data Bank code 2IW0). Secondary structure elements deriving from Protein Data Bank coordinates are drawn over the alignment. Residues involved in metal-binding or in catalysis (38) are denoted by cyan (metal-binding) and magenta (catalysis) arrowheads. B, cartoon drawing of the structure of the PuuE monomer alongside its topological representation. The helices are colored red, and the strands are blue, except for helix $\alpha 1$ (orange) and strands $\beta 6$ and $\beta 7$ (light blue), which do not fit the canonical PsDA fold. C, cartoon drawing of the PuuE tetramer. Subunits A and C are colored as in B; subunits B and D are colored green.

deacetylases catalyze a hydrolytic reaction similar to that catalyzed by allantoinase.

Function of the PuuE Protein—Recombinant PuuE (308 amino acids) was cloned from *P. fluorescens* and overproduced in *E. coli*. The purified protein eluted from a calibrated gel filtration column with an estimated molecular mass of ~ 140 kDa, consistent with a homotetrameric assembly.

When PuuE was assayed by CD spectroscopy using (S)-allantoin as substrate, we observed the disappearance of the (S)-allantoin spectrum (Fig. 2A, upper panel). When PuuE was assayed with the racemic mixture of (S)- and (R)-allantoin as substrate, we observed the formation of the (R)-allantoin spectrum (Fig. 2A, lower panel); once it had reached its maximum value, the spectrum decayed at the same rate as the spontaneous racemization of allantoin (34) (not shown), indicating that the enzyme has an absolute stereospecificity for the S enantiomer of allantoin. The formation of allantoic acid in the reaction catalyzed by PuuE could be unambiguously assessed by

NMR spectroscopy using ^{13}C -labeled (S)-allantoin (Fig. 2B). The reaction appears to be reversible; when PuuE was assayed by CD spectroscopy using high concentrations of allantoic acid as substrate, we observed the formation of the (S)-allantoin spectrum (not shown). An equilibrium ratio of allantoic acid to (S)-allantoin of 68:1 was calculated by measuring the direction of shift upon enzyme addition at different product/substrate ratios.

The velocity of the enzyme-catalyzed hydrolysis of (S)-allantoin at different substrate concentrations shows a hyperbolic relationship, with a K_m of 0.26 mM and a K_{cat} of 12.6 s^{-1} . The reverse reaction, the stereoselective condensation of allantoic acid to form (S)-allantoin, is catalyzed by the enzyme with a K_m of 2.4 mM and a K_{cat} of 1.6 s^{-1} . The enzyme activity was also assayed with compounds similar to allantoin: hydantoin, hydantoin-5-acetic acid, 2-imidazolidone, 4(5)-imidazole-carboxaldehyde, and 5-amino-4-imidazole-carboxamide hydrochloride. None of these compounds proved to be a sub-

Logical Identification of *puuE*

TABLE 1
Statistics on data collection and refinement

	PuuE	PuuE-HYN	PuuE-ACA
X-ray data processing^a			
X-ray source	ID29	ID14-eh1	ID14-eh1
Wavelength (Å)	0.9700	0.9340	0.9340
Cell parameters, <i>a</i> , <i>b</i> , <i>c</i> (Å)	98.30, 62.39	98.14, 62.10	97.72, 61.32
Resolution (Å)	52.71–1.58 (1.67–1.58)	53.00–1.80 (1.90–1.80)	69.20–2.25 (2.37–2.25)
Unique reflections	81,449 (11,879)	54,454 (7,551)	24,175 (3,346)
Multiplicity	3.9 (3.9)	4.3 (3.0)	3.6 (3.3)
Completeness (%)	99.8 (100)	99.2 (95.2)	87.2 (83.6)
[<i>I</i> / <i>s</i> (<i>I</i>)]	8.4 (1.7)	14.2 (3.0)	9.0 (2.8)
<i>R</i> _{merge}	0.072 (0.441)	0.084 (0.410)	0.151 (0.357)
Refinement			
Number of reflection used/resolution shell	77,053 (10.00–1.58)	51,686 (98.06–1.80)	22,962 (97.59–2.25)
<i>R</i> _{cryst} / <i>R</i> _{free} (%)	20.8/23.2	19.9/22.5	20.7/23.0
Total number of atoms, including ligands and solvent	5,244 (292 water molecules)	5,228 (252 water molecules)	5,090 (136 water molecules)
Mean B value (Å ²)	14.9	14.4	13.6
Ramachandran plot (%)			
Most favored	87.8	86.1	83.4
Additionally allowed	12.2	13.9	16.6
Generously allowed	0	0	0
Disallowed	0	0	0
Root mean square on bonds length (Å), angles (°)	0.009/1.6	0.012/1.4	0.011/1.6

^a The crystals belong to the P4 space group. They were frozen at 100 K under a nitrogen gas cold stream, without any cryoprotectant. Rotations of 0.5° were performed.

strate of the enzyme. However, hydantoin and 5-amino-4-imidazole-carboxamide hydrochloride were found to inhibit allantoin hydrolysis.

Because all of the known PuuE homologs as well as DAL1 allantoinases have been reported to depend on bivalent metals for catalysis (35–37), we carried out experiments to determine whether the catalytic activity of PuuE depends on metals. Metal chelators did not influence the enzyme activity, and the enzyme was not activated by incubation with Zn²⁺, Cu²⁺, Ca²⁺, Co²⁺, Mn²⁺, or Mg²⁺. Furthermore, in 4-(2-pyridylazo)resorcinol-based assays, enzymatically active PuuE preparations revealed a very low (<0.1) metal/protein ratio.

The PuuE Family—PuuE allantoinase belongs to a large family named polysaccharide deacetylase that comprises in sequence data bases more than 2000 proteins from about 700 species. The structures of several members of the family are known from structural genomics and dedicated studies (36–40).

The family encompasses only proteins with activities on polysaccharide substrates and involved in cell wall modifications: peptidoglycan *N*-acetylmuramic acid and peptidoglycan *N*-acetyl glucosamine deacetylases (41, 42), chito-oligosaccharide and chitin deacetylases (43, 44), and xylan esterases (45). Common features of polysaccharide deacetylases are metal ion dependence (36) and the recognition of multimeric substrates. Chitin deacetylases, for examples, need at least a chito-oligosaccharide trimer for activity but prefer longer multimers (46, 47). This marks a clear difference between polysaccharide deacetylases and PuuE, which appears to be metal-independent and acts on a small substrate molecule.

Sequence alignment with polysaccharide deacetylases shows that PuuE proteins have some distinguishing features (Fig. 3A): an inserted segment between positions 40 and 80 (IS1), which does not have a counterpart in polysaccharide deacetylases and is conserved in all PuuE proteins, and a smaller segment between positions 189 and 207 (IS2), which is similarly present only in PuuE proteins. Only the first histidine of the polysaccharide deacetylases Asp-His-His metal-binding triad is

retained in PuuE sequences (H126). Amino acid substitutions are also observed for residues that have been implicated in catalysis, with the exception of an aspartate residue (Asp²¹⁷) and a histidine residue (His²⁵⁹) that are thought to play a major role in catalysis (36).

Sequence comparison reveals high similarity with the structural genomics deposition 1Z7A, corresponding to a selenomethionine-substituted structure of a probable polysaccharide deacetylase from *P. aeruginosa* (Fig. 3). The *P. aeruginosa* gene encoding the protein is orthologous and syntenic to the *P. fluorescens* gene encoding PuuE. The three-dimensional coordinates of 1Z7A, deposited in 2005 by the Midwest Center for Structural Genomics, thus represent the first known three-dimensional structure of an allantoinase. Comparison of the 1Z7A structure with polysaccharide deacetylases revealed the presence of structural elements that could account for the different substrate specificity exhibited by PuuE.

Overall Structure of PuuE—To gain further insight into the evolution of the peculiar substrate specificity of PuuE proteins, the crystal structure of the *P. fluorescens* protein was solved for the uncomplexed enzyme and for complexes of the enzyme with the inhibitors HYN and ACA (Table 1).

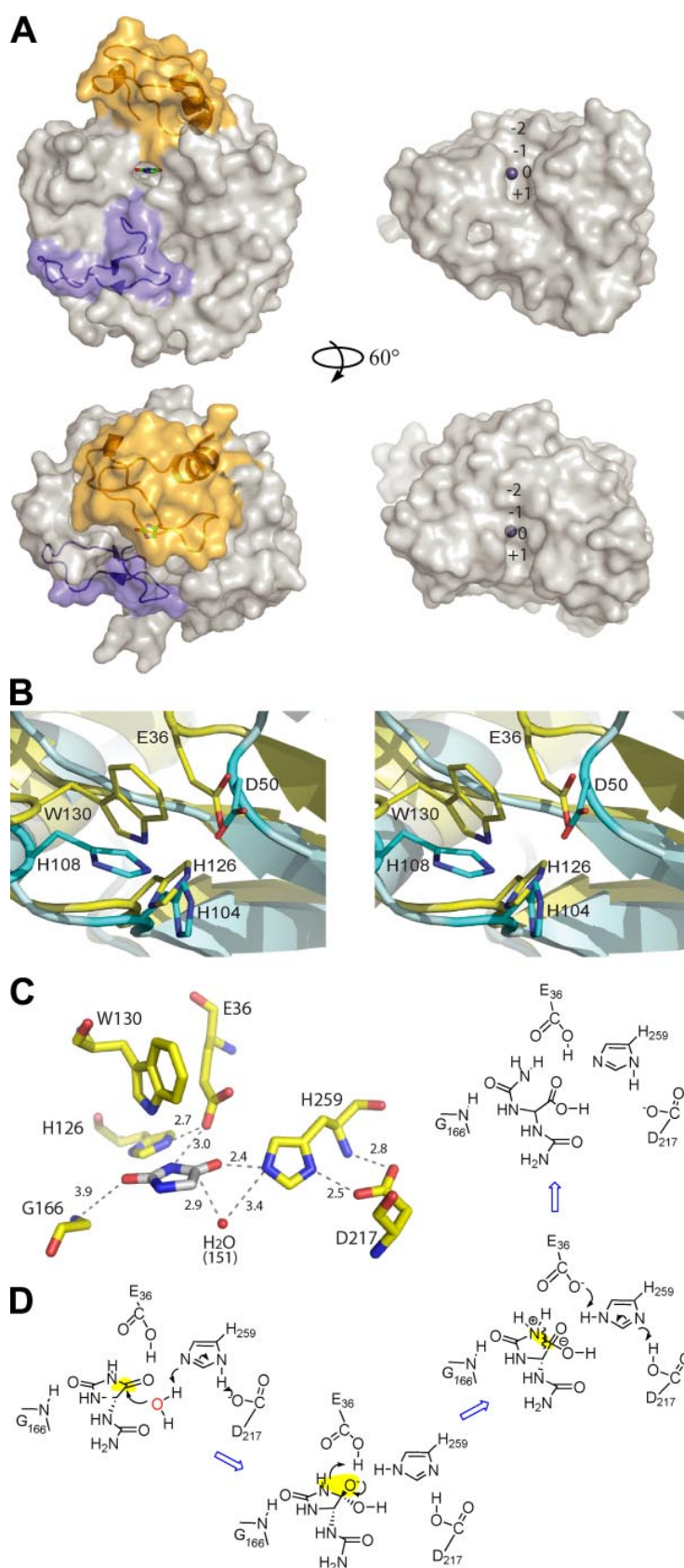
Like the other members of the family, the monomer structure of PuuE reveals a deformed (β/α) barrel with C-terminal β -strand (β_9) and α -helix (α_8) sealing the bottom of the barrel (Fig. 3B). Structural alignment with VAST (48) superimposes 188 residues of PuuE on the *Bacillus subtilis* BspDaA structure (Protein Data Bank code 1W17) with an root mean square deviation of 2.6 Å. Similar values are obtained with *Colletotrichum lindemuthianum* chitin deacetylase ClCDA (Protein Data Bank code 2IW0; 183 superimposed residues, root mean square deviation of 2.5 Å) and *Streptomyces lividans* xylan esterase (Protein Data Bank code 2CC0; 161 superimposed residues, root mean square deviation of 2.2 Å). The PuuE barrel has a small antiparallel β sheet interposed between β_5 and α_6 , corresponding to the sequence segment IS2 (Fig. 3, A and B); strand β_7 of this inserted element occupies in the structure the same position of strand β_2 of BspDaA and strand β_1 of ClCDA. The other PuuE

sequence segment that does not have a counterpart in polysaccharide deacetylases (IS1) folds as a separated domain at the top of the barrel, with a long loop and an α -helix that is connected to the first helix of the barrel (Fig. 3, A and B).

At variance with all known polysaccharide deacetylases, the PuuE crystal structure reveals a tetrameric organization (Fig. 3C). The results obtained with gel filtrations suggest that this is also the functional form of the enzyme. PuuE assembles as a planar homotetramer with an approximate size of $96 \times 97 \times 54 \text{ \AA}^3$. Very few contacts are established between diagonally opposed subunits, whereas several polar or hydrophobic interactions are established between contiguous subunits, which share an interface area of about 1900 \AA^2 . Residues involved in interchain contacts are clustered in the PuuE extra domain and particularly on the α helix $\alpha 1$. Six of twelve residues of helix $\alpha 1$ (Glu⁷², Ser⁷³, Glu⁷⁶, Tyr⁷⁷, Ser⁷⁹, and Arg⁸⁰) establish hydrogen bonds at subunit interfaces, mainly with residues present in the loop between strands $\beta 5$ and $\beta 6$ of a contiguous subunit (Tyr¹⁸⁹, Asp¹⁹⁰, Asp¹⁹¹, and Asp¹⁹²).

Reshaping of the Substrate-binding Site in PuuE Proteins—In polysaccharide deacetylases, the substrate binds to a large groove at the top of the barrel that contains at least four subsites for different sugar moieties (-2 , -1 , 0 , and $+1$ where subsite 0 is the site of catalysis) (38, 46). The structures of HYN and ACA complexes show that in PuuE the substrate binds to a small cavity corresponding to a part of the subsite 0 of polysaccharide deacetylase (Fig. 4A).

The bottom wall of the cavity is delimited by a residue (Tyr¹⁸⁹) that belongs to the inserted segment IS2 (Fig. 3) and by a residue (Leu²¹³) that closely follows IS2 and is displaced by more than 6 \AA compared with polysaccharide deacetylases. The top and back walls of the cavity are formed by the inserted segment IS1. In particular, the long loop that



Logical Identification of *puuE*

connects strand $\beta 1$ to helix $\alpha 1$ is positioned at the top of the barrel along the groove and completely covers the surface area of the sugar-binding subsites -2 and -1 of polysaccharide deacetylases (Fig. 4A, lower panels).

Loss of Metal Binding at the Active Site of *PuuE* Proteins—The electron density maps of *PuuE* did not show the presence of metal at the active site; this was confirmed by a wide spectrum x-ray fluorescence scan on the crystals (data not shown). A close comparison with the metal-binding site of chitin deacetylase 2IW0 illustrates that *PuuE* has a different residue configuration (Fig. 4B).

In the $C\alpha$ trace alignment, the position corresponding to the aspartate residue of the chitin deacetylase metal-binding triad (Asp⁵⁰) is occupied in *PuuE* by a tyrosine residue (Tyr³⁵; see Fig. 3A) whose side chain points in the opposite direction (not shown). The carboxyl group of a subsequent glutamate residue (Glu³⁶), however, occupies a position similar to the carboxyl group of Asp⁵⁰. The side chain of Glu³⁶ is tethered by a tryptophan residue (Trp¹³⁰) that replaces the canonical histidine of the polysaccharide deacetylase metal-binding triad. Glu³⁶ and Trp¹³⁰ are invariant in the *PuuE* alignment, allowing an easy distinction of *PuuE* sequences from polysaccharide deacetylases.

Common and Idiosyncratic Elements in the Acid/Base Reaction Mechanism of *PuuE*—Interestingly, polysaccharide deacetylases can maintain a residual activity in the absence of metals (39), and the major role in the reaction appears to be played by two protein residues that act as an acid/base catalytic pair (36). The binding of substrate analogs at the *PuuE* active site allows assessment of the interactions established between enzyme and its substrate. ACA and HYN bind in a very similar fashion, mainly through hydrogen bonds with almost the same residues, and occupy a site that overlaps with the metal coordination sphere of polysaccharide deacetylases.

The HYN complex, whose electron density is better defined, illustrates the specific interactions that occur near the scissile bond (N-3–C-4) of the hydantoin ring of allantoin (Fig. 4C). HYN enters the cavity by displacing two ordered water molecules, and its positioning is consistent with the binding of allantoin only in the *S* configuration. The carbonyl oxygen of the hydantoin ring C-4 is at H-bond distance from the Ne2 atom of His²⁵⁹, a member of the catalytic pair of polysaccharide deacetylases (36). His²⁵⁹ interacts, through its Nd1 atom, with the buried Asp²¹⁷ residue, which is also invariant in the family alignment. The other member of the catalytic pair of polysaccharide deacetylases, an Asp residue, is substituted in *PuuE*, by Asn (Asn³⁴; Fig. 3A). This particular substitution inactivates the enzyme when introduced in peptidoglycan deacetylase (36). Also not conserved is an Arg residue that interacts with the catalytic aspartate and that is substituted by Tyr in *PuuE* (Tyr¹⁶⁴; Fig. 3A). In the *PuuE* active site, however, a glutamate residues (Glu³⁶) is able to H-bond with both the N-3 atom and

the C-4 carbonyl oxygen of hydantoin (Fig. 4C). Glu³⁶, which is invariant in *PuuE* sequences and is tethered by two conserved residues (Trp¹³⁰ and His¹²⁶), most likely completes the acid/base pair of *PuuE*.

In the hydrolytic mechanism of polysaccharide deacetylases, His is proposed to act as a catalytic acid and Asp is proposed to act as a catalytic base (36). Here, the positioning of the substrate and of a water molecule close to the C-4 atom, favors a mechanism in which His acts as a base and Glu as an acid (Fig. 4D). His²⁵⁹ could abstract a proton from the water molecule, creating a nucleophile to attack the carbonyl carbon in the substrate. Glu³⁶ could stabilize the oxyanion intermediate through H-bond and donate a proton to the N-3 atom of the substrate, creating a good leaving group. In this way Glu³⁶ would play the same role of the catalytic glutamate of imidazolonepropionase (49), an enzyme unrelated to *PuuE* acting on a substrate similar to allantoin.

DISCUSSION

The coarse function of *PuuE* could be guessed by its frequent chromosomal proximity with genes involved in purine degradation, and a definite prediction of its biochemical activity could be made through the analysis of the logic relations (18) with other genes of the pathway. By contrast, no hints about the actual function of *PuuE* are provided by more than 100 genome annotations of this protein besides its homology with polysaccharide deacetylases. This emphasizes the need for a better exploitation of the evidence deriving from genome comparison for the prediction of the function of uncharacterized genes.

The analysis of complex logic relationships in the case described here has been facilitated by the frequent clustering of the genes of the pathway, which allows one to restrict the analysis to a small set of candidates. Nevertheless, the significance of the logic association obtained suggests that such gene associations can also emerge from exhaustive comparisons. Although the method was conceived for the analysis of protein triplets, it can be readily extended to the analysis of more complex relationships. Here a logic association was obtained by analyzing protein quartets. The binary relation of *PuuE* with DAL1 ($U = 0.02$; $p < 10^{-2}$), and the ternary relations that include separately allantoicase ($U = 0.15$; $p < 10^{-5}$) or allantoate amidohydrolase ($U = 0.33$; $p < 10^{-11}$), are less significant predictors of the *PuuE* function.

Three of the four protein families compared here (*PuuE*, DAL1, and allantoate amidohydrolase) belong to large families comprising proteins with many different functions. Quite obviously, no functional associations can be obtained when the whole families are compared. A critical issue of this analysis is thus the splitting of large families into groups containing genes with the same function. Although no computational methods guarantee the attainment of this goal, the distinction of orthologous groups among gene families is considered a good

FIGURE 4. **Substrate binding and reaction mechanism of *PuuE*.** A, surface drawing of hydantoin-bound *PuuE* (left panel) compared with a zinc-bound chitin deacetylase (Protein Data Bank code 2IW0; right panel). The inserted segments IS1 and IS2 of *PuuE* are colored orange and light blue, respectively. The positions of sugar-binding subsites in 2IW0 are indicated. B, stereo view comparison of the metal-binding site of 2IW0 (cyan) and the corresponding region of *PuuE* (yellow). Metal-binding residues of 2IW0 and superimposable residues of *PuuE* are shown as sticks. C, the *PuuE* active site in complex with hydantoin. Key residues and ligands are represented in stick mode. D, proposed catalytic mechanism for (*S*)-allantoin hydrolysis.

approximation. Here we have used an automatic method of orthologous clustering (20) and found it accurate enough to distinguish groups of proteins with the same function. The suitability of such methods for the automatic clustering of very large sets of proteins allows the application of this strategy on a genome-wide basis.

We note that a protein included here in the *PuuE* family, CDA1_SCHPO from the yeast *S. pombe* (SpCDA in Fig. 3A), has been reported to function in yeast as chitin deacetylase based on the presence of a polysaccharide deacetylase domain and impairment in spore formation of the deletion mutant (50). Several lines of evidence indicate that allantoin hydrolysis is the actual function of CDA1_SCHPO: (i) the *S. pombe* protein is similar (43% identity) and likely orthologous to the *P. fluorescens* allantoinase, whereas it is more distantly related to *S. cerevisiae* chitin deacetylases CDA1 (15% identity) and CDA2 (14% identity); (ii) *S. pombe* does not have the *DAL1* allantoinase gene and possesses an allantoicase gene (Fig. 1); (iii) the gene coding for CDA1_SCHPO is a neighbor of a gene coding for ureidoglycolase, a protein involved in purine degradation; (iv) all of the sequence features that correspond to distinguishing structural elements of the *PuuE* family are present in the *S. pombe* sequence (Fig. 3); and (v) residues that were found to interact with the substrate at the active site are remarkably conserved in the *S. pombe* sequence (Figs. 3 and 4). The phenotype of the deletion mutant observed in sporulation, as well as the increased expression of the gene after nitrogen starvation, can be easily explained by the role of purine degradation in nitrogen metabolism.

PuuE and *DAL1* allantoinase do not have significant similarity at the sequence level. Although both proteins have a (β/α) barrel catalytic domain, they belong to distinct superfamilies and do not have significant similarity at the structural level, as can be deduced from the comparison of *PuuE* with L-hydantoinase (Protein Data Bank code 1GKR), a close homolog of *DAL1* with known structure. *PuuE* and *DAL1* are thus analogous enzymes: nonhomologous proteins with the same catalytic function (51). It has been shown that in many cases analogous enzymes evolved by the recruitment of existing enzymes that take over new functions by virtue of changed substrate specificity (52). Compared with *DAL1*, *PuuE* has a narrower organism distribution, being present only in fungi and bacteria (even though, in bacteria, *PuuE* has gained a large diffusion, accounting for about the 65% of the allantoinase activity encoded in prokaryotic genomes). Moreover, the polysaccharide deacetylase domain of *PuuE*, with the exception of *PuuE* itself, is associated with proteins involved in specialized pathways related to cell wall components, whereas the amidohydrolase domain of *DAL1* is often found in proteins involved in the primary metabolism, including proteins catalyzing other reactions of the purine metabolism.

It is thus likely that *DAL1* is a more ancient allantoinase that has been replaced in some organisms by a protein, *PuuE*, recruited from enzymes with a different role through the shifting of substrate specificity. The comparison with polysaccharide deacetylases shows that this functional transition has been accompanied by relevant structural modifications. Insertion of two novel structural elements has changed the oligomerization

of the protein and deeply reshaped the substrate-binding site. Nonconservative amino acid substitutions at the active site have conferred metal independency to the enzyme and adaptation to a ligand, allantoin, that is remarkably different from the original polysaccharide substrate.

Acknowledgments—We thank A. Peracchi and S. Petrucco for discussion and G. L. Rossi for support. We thank D. Acquotti and the University of Parma CIM for help with NMR measurements and the staff of the ID14-eh1 diffraction beamline of the European Synchrotron Radiation Facility for technical assistance.

REFERENCES

1. Todd, C. D., Tipton, P. A., Blevins, D. G., Piedras, P., Pineda, M., and Polacco, J. C. (2006) *J. Exp. Bot.* **57**, 5–12
2. Vogels, G. D., and Van der Drift, C. (1976) *Bacteriol. Rev.* **40**, 403–468
3. Ramazzina, I., Folli, C., Secchi, A., Berni, R., and Percudani, R. (2006) *Nat. Chem. Biol.* **2**, 144–148
4. Kahn, K., and Tipton, P. A. (1998) *Biochemistry* **37**, 11651–11659
5. Kim, K., Park, J., and Rhee, S. (2007) *J. Biol. Chem.* **282**, 23457–23464
6. Cendron, L., Berni, R., Folli, C., Ramazzina, I., Percudani, R., and Zanotti, G. (2007) *J. Biol. Chem.* **282**, 18182–18189
7. Yoo, H. S., Cunningham, T. S., and Cooper, T. G. (1992) *Yeast* **8**, 997–1006
8. Buckholz, R. G., and Cooper, T. G. (1991) *Yeast* **7**, 913–923
9. LaPointe, G., Viau, S., LeBlanc, D., Robert, N., and Morin, A. (1994) *Appl. Environ. Microbiol.* **60**, 888–895
10. Holm, L., and Sander, C. (1997) *Proteins* **28**, 72–82
11. Gaines, P. J., Tang, L., and Wisniewski, N. (2004) *Insect. Biochem. Mol. Biol.* **34**, 203–214
12. Hayashi, S., Jain, S., Chu, R., Alvares, K., Xu, B., Erfurth, F., Usuda, N., Rao, M. S., Reddy, S. K., Noguchi, T., et al. (1994) *J. Biol. Chem.* **269**, 12269–12276
13. Kim, G. J., Lee, D. E., and Kim, H. S. (2000) *J. Bacteriol.* **182**, 7021–7028
14. Yang, J., and Han, K. H. (2004) *Plant Physiol.* **134**, 1039–1049
15. Janssen, D. B., Joosten, H. M., Herst, P. M., and van der Drift, C. (1982) *J. Bacteriol.* **151**, 1176–1183
16. Matsumoto, H., Ohta, S., Kobayashi, R., and Terawaki, Y. (1978) *Mol. Gen. Genet.* **167**, 165–176
17. Fluri, R., and Kinghorn, J. R. (1985) *J. Gen. Microbiol.* **131**, 527–532
18. Bowers, P. M., Cokus, S. J., Eisenberg, D., and Yeates, T. O. (2004) *Science* **306**, 2246–2249
19. Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006) *Nucleic Acids Res.* **34**, D247–251
20. Li, L., Stoeckert, C. J., Jr., and Roos, D. S. (2003) *Genome Res.* **13**, 2178–2189
21. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
22. Bond, C. S. (2003) *Bioinformatics* **19**, 311–312
23. Zanotti, G., Cendron, L., Ramazzina, I., Folli, C., Percudani, R., and Berni, R. (2006) *J. Mol. Biol.* **363**, 1–9
24. Kalia, A., Rattan, A., and Chopra, P. (1999) *Anal. Biochem.* **275**, 1–5
25. Leslie, A. G. (2006) *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 48–57
26. Evans, P. (2006) *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 72–82
27. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *Journal of Applied Crystallography* **40**, 658–674
28. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D. Biol. Crystallogr.* **53**, 240–255
29. Sheldrick, G. M., and Schneider, T. R. (1997) *Methods Enzymol.* **277**, 319–343
30. Schüttelkopf, A. W., and van Aalten, D. M. (2004) *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 1355–1363
31. Wong, S., and Wolfe, K. H. (2005) *Nat. Genet.* **37**, 777–782

Logical Identification of *puuE*

32. Vigetti, D., Monetti, C., Prati, M., Gornati, R., and Bernardini, G. (2002) *Gene* **289**, 13–17
33. Barker, D., and Pagel, M. (2005) *PLoS. Comput. Biol.* **1**, e3
34. Kahn, K., and Tipton, P. A. (2000) *Bioorganic Chemistry* **28**, 62–72
35. Mulrooney, S. B., and Hausinger, R. P. (2003) *J. Bacteriol.* **185**, 126–134
36. Blair, D. E., Schuttelkopf, A. W., MacRae, J. I., and van Aalten, D. M. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15429–15434
37. Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Vincent, F., Brzozowski, A. M., Dupont, C., Shareck, F., Centeno, M. S., Prates, J. A., Puchart, V., Ferreira, L. M., Fontes, C. M., Biely, P., and Davies, G. J. (2006) *J. Biol. Chem.* **281**, 10968–10975
38. Blair, D. E., Hekmat, O., Schuttelkopf, A. W., Shrestha, B., Tokuyasu, K., Withers, S. G., and van Aalten, D. M. (2006) *Biochemistry* **45**, 9416–9426
39. Blair, D. E., and van Aalten, D. M. (2004) *FEBS. Lett.* **570**, 13–19
40. Oberbarnscheidt, L., Taylor, E. J., Davies, G. J., and Gloster, T. M. (2007) *Proteins* **66**, 250–252
41. Fukushima, T., Yamamoto, H., Atrih, A., Foster, S. J., and Sekiguchi, J. (2002) *J. Bacteriol.* **184**, 6007–6015
42. Vollmer, W., and Tomasz, A. (2000) *J. Biol. Chem.* **275**, 20496–20501
43. Kafetzopoulos, D., Martinou, A., and Bouriotis, V. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 2564–2568
44. John, M., Rohrig, H., Schmidt, J., Wieneke, U., and Schell, J. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 625–629
45. Collins, T., Gerday, C., and Feller, G. (2005) *FEMS Microbiol. Rev.* **29**, 3–23
46. Tokuyasu, K., Mitsutomi, M., Yamaguchi, I., Hayashi, K., and Mori, Y. (2000) *Biochemistry* **39**, 8837–8843
47. Hekmat, O., Tokuyasu, K., and Withers, S. G. (2003) *Biochem. J.* **374**, 369–380
48. Madej, T., Gibrat, J. F., and Bryant, S. H. (1995) *Proteins* **23**, 356–369
49. Yu, Y., Liang, Y. H., Brostromer, E., Quan, J. M., Panjikar, S., Dong, Y. H., and Su, X. D. (2006) *J. Biol. Chem.* **281**, 36929–36936
50. Matsuo, Y., Tanaka, K., Matsuda, H., and Kawamukai, M. (2005) *FEBS Lett.* **579**, 2737–2743
51. Fitch, W. M. (1970) *Syst. Zool.* **19**, 99–113
52. Galperin, M. Y., Walker, D. R., and Koonin, E. V. (1998) *Genome Res.* **8**, 779–790