ARTICLE

ACS chemical biology

Chemical Basis of Nitrogen Recovery through the Ureide Pathway: Formation and Hydrolysis of *S*-Ureidoglycine in Plants and Bacteria

Fabio Serventi[†,§], Ileana Ramazzina[†], Ilaria Lamberto[†], Vincenzo Puggioni[†], Rita Gatti[‡], and Riccardo Percudani[†,*]

[†]Dipartimento di Biochimica e Biologia Molecolare and [‡]Dipartimento di Medicina Sperimentale, Sezione di Istologia, Università di Parma, 43100 Parma, Italy, [§]Present address: Institute of Microbiology, Eidgenössische Technische Hochschule, 8093 Zürich, Switzerland.

ABSTRACT While some organisms, including humans, eliminate oxidized purines to get rid of excess nitrogen, for many others the recovery of the purine ring nitrogen is vital. In the so-called ureide pathway, nitrogen is released as ammonia from allantoate through a series of reactions starting with allantoate amidohydrolase (AAH), a manganese-dependent enzyme found in plants and bacteria. We report NMR evidence that the true product of the AAH reaction is *S*-ureidoglycine, a nonstandard α-amino acid that spontaneously releases ammonia *in vitro*. Using gene proximity and logical genome analysis, we identified a candidate gene (*ylbA*) for *S*-ureidoglycine metabolism. The proteins encoded by *Escherichia coli* and *Arabidopsis thaliana* genes catalyze the manganese-dependent release of ammonia through hydrolysis of *S*-ureidoglycine. Hydrolysis then inverts the configuration and yields *S*-ureidoglycolate. *S*-Ureidoglycine aminohydrolase (UGHY) is cytosolic in bacteria, whereas in plants it is localized, like allantoate amidohydrolase, in the endoplasmic reticulum. These findings strengthen the basis for the known sensitivity of the ureide pathway to Mn availability and suggest a further rationale for the active transport of Mn in the endoplasmic reticulum of plant cells.

Animals degrade the nitrogenous bases present in DNA and RNA to eliminate excess nitrogen. However, in various terrestrial and aquatic environments the availability of nitrogen limits biological production (*1, 2*), and many organisms depend on efficient use of nitrogen for growth. Plants, fungi, and bacteria have metabolic pathways for the reuse of the nitrogen present in pyrimidine and purine bases (*3, 4*). Purine catabolism is particularly relevant in tropical legumes that move the nitrogen fixed into the nodules to the aerial portion of the plant primarily in the form of ureides, allantoin, and allantoate (*5*). In these plants, the nitrogen fixed as a result of the symbiotic association with nodule-forming bacteria is combined with carbon through *de novo* purine synthesis. Through a series of enzymatic steps, purines are oxidized to allantoin and allantoate, compounds that have a favorable N:C ratio (see Figure 1) and serve for nitrogen transport and storage. In some important crops, accumulations of ureides appear related to water deficit sensitivity, and manganese supplementation has been shown to modify relative susceptibility to this process in a cultivar-dependent manner (*5, 6*).

In plants, the first step of ureide metabolism occurs in the cytoplasm and leads to the production of oxopurines, such as hypoxanthine, xanthine, and uric acid. The opening of the uric acid ring and elimination of carbon dioxide yields *S*-allantoin through a three-step peroxisomal pathway (*7*). Hydrolysis of *S*-allantoin by allantoinase in the endoplasmic reticulum yields allantoate (*8*), and the breakdown of allantoate produces usable nitrogen for subsequent anabolic reactions. In general,

*Corresponding author, riccardo.percudani@unipr.it.

Received for review October 10, 2009 and accepted December 29, 2009.

Published online December 29, 2009

10.1021/cb900248n

allantoate can be acted upon by two different enzymes to release ammonia (upon allantoate amidohydrolase) or urea (upon allantoicase). Allantoicase is typically found in animals (though not necessarily involved in purine degradation (*9*)), fungi and several bacteria, while some other bacteria and all plants use allantoate amidohydrolase (*10, 11*). Allantoate amidohydrolase (AAH; EC number 3.5.3.9) has been extensively studied. The gene (*allC*) encoding the protein (*12*) and the crystal structure of the protein in complex with the substrate (*13*) are known, dependency on manganese for catalysis has been established (*11, 14*), and the localization of the plant enzyme in the endoplasmic reticulum has been determined (*11*). Paradoxically, however, a fundamental question about this catalyst has never been fully answered: what reaction does this protein catalyze?

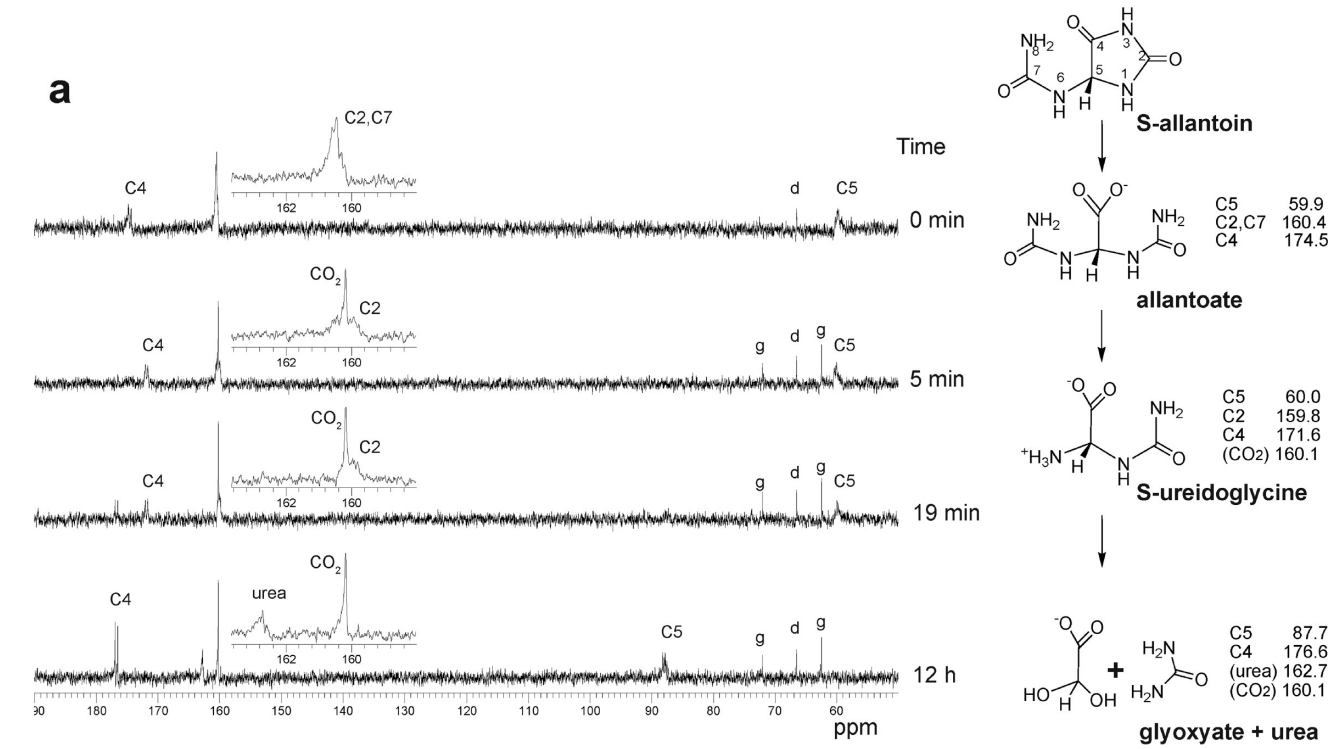
Early studies have established that the AAH reaction produces 2 mol of ammonia, 1 mol of carbon dioxide, and 1 mol of ureidoglycolate. Accumulation of an intermediate during the reaction was assumed because of the formation of glycine through a side transamination reaction in the presence of glyoxylate (*14*). The identification of ureidoglycine as being this intermediate is based on this observation and on similarities with other amidohydrolases rather than on direct evidence. Moreover, it is unclear whether the ureidoglycine intermediate is converted to ureidoglycolate by the action of AAH (*14, 15*) or by the action of a distinct, as yet unidentified, enzyme (*13, 16, 17*). These two possibilities are alternatively accounted for by the major databases of biological reactions. For example, in the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg), EC3.5.3.9 is responsible for hydrolysis of both allantoate and ureidoglycine, whereas in the Encyclopedia of Metabolic Pathways (http://metacyc.org) EC3.5.3.9 and an unidentified enzyme (EC3.5.3.−) are responsible for hydrolysis of allantoate and ureidoglycine, respectively.

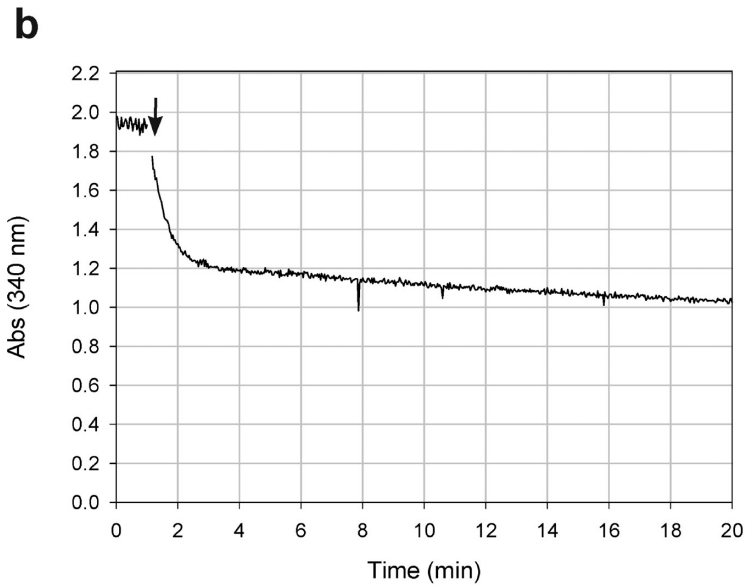
A major hurdle in understanding the pathway initiated by allantoate amidohydrolase has been the use of analytical methods that cannot distinguish between the true product of the AAH reaction and compounds downstream in the pathway, such as ureidoglycolate and glyoxylate (*11, 12, 15*). Here we combine detailed spectroscopic characterization of reactions catalyzed by recombinant proteins with bioinformatic analysis to discover an enzyme acting on *S*-ureidoglycine, a compound which we show to be a true product of the AAH reaction. We report on the catalysis and stereochemistry of *S*-ureidoglycine aminohydrolase (UGHY), we describe the structure−function relationship of the bacterial protein, and we determine the localization of the protein in plant cells.

RESULTS AND DISCUSSION

Identification of the True Product of the Allantoate Amidohydrolase Reaction. Recombinant, histidine-tagged allantoate amidohydrolase from *E. coli* (AllC) was produced using an ASKA library clone (*18*) and purified until near-homogeneity was acquired, using affinity chromatography. The reaction catalyzed by AllC was studied by NMR spectroscopy using ^{13}C- labeled allantoate as substrate. The time course of the reaction (Figure 1, panel a) shows rapid formation of the primary hydrolysis product and its subsequent decay. Formation of the primary hydrolysis product is accompanied by release of carbon dioxide (uncoupled peak at 160.1 ppm). On the basis of chemical shift values, the primary products of the reaction can be identified as ureidoglycine. The ureidic carbon of ureidoglycine was assigned to an ill-defined peak around 160.0 ppm (Figure 1, panel a, expansion). A well-defined peak, however, was obtained with allantoate specifically labeled at C2 (see below). The NMR data is consistent with a hydrolytic attack of a ureidic carbonyl of allantoate yielding ureidoglycine and carbamic acid. This latter compound is known to decay immediately to carbon dioxide and ammonia, while ureidoglycine could undergo slower hydrolyses to yield stable reaction products that were identified as urea and the *gem*-diol form of glyoxylate (*19*) on the basis of comparison with chemical standards.

NMR time courses of the reaction suggest that a first ammonia molecule is released by allantoate amidohydrolase, while a second molecule is released along with the hydrolysis of ureidoglycine. To observe the release of ammonia we used a coupled assay with glutamate dehydrogenase, by monitoring the decrease of absorbance of the NADPH cofactor at 340 nm (*14, 17*). With this assay, we observed a rapid decline of absorbance to a value corresponding with the release of 1 mol of ammonia per mole of substrate, followed by a slow decline in absorbance to a value corresponding to the release of a second mole of ammonia (Figure 1, panel b). The two reaction steps were clearly distinct, with only the velocity of the first step apparently proportional to





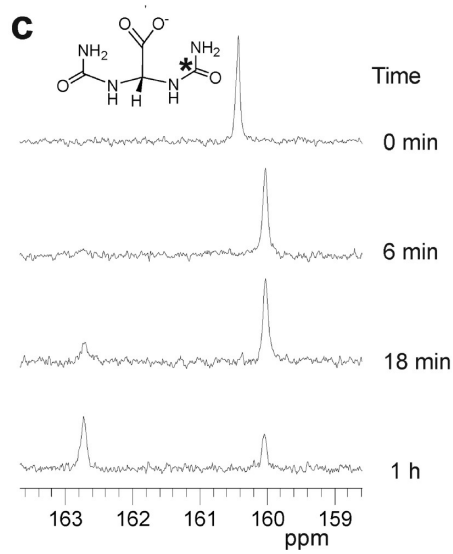


Figure 1. *S*-Ureidoglycine is the true product of the AllC reaction. a) Time course of the AllC-catalyzed hydrolysis of uniformly labeled [^{15}N,^{13}C] allantoate. ^{13}C spectra were collected at 25 °C in 0.1 M potassium phosphate, 80% D_2O, pD 7.6; signals from glycerol and dioxane (66.5 ppm, reference) are marked by lowercase letters. Chemical shifts assignments are indicated using the carbon numbering scheme of *S*-allantoin; stereochemistry is based on NMR analysis of specifically labeled allantoate (see below). b) Time course of ammonia release by the AllC reaction monitored spectrophotometrically using a coupled assay with glutamate dehydrogenase. The reaction was initiated by the addition of 0.15 mM allantoate (arrow) to a solution containing 0.13 μM AllC. c) Time course of the AllC-catalyzed hydrolysis of specifically labeled [2-^{13}C]allantoate.

the enzyme concentration. To rule out the possibility that the hydrolysis of ureidoglycine was a slow reaction catalyzed by allantoate amidohydrolase (*14, 15*), the enzyme was removed by ultrafiltration after completion of the first reaction step. The deproteinized solution released ammonia at the same rate as the unfiltered solution, thus demonstrating spontaneous hydrolysis of ureidoglycine. The rate constant of ureidoglycine hydrolysis measured in the Tris-HCl buffer at pH 7.2 was 4×10^{-4} s^{-1}.

Allantoate is a prochiral molecule in which each carbamoyl group is attached to a nonstandard amino acid (ureidoglycine). The pro-*R* carbamoyl is attached to an ureidoglycine moiety in an *S*-configuration, while the pro-*S* carbamoyl is attached to an ureidoglycine moiety in the *R*-configuration. Hydrolysis of a carbamoyl group is thus expected to produce optically active ureidoglycine, and a transient optically active compound with a spectrum similar to that of L-amino acids was observed by monitoring the reaction with CD spectroscopy (Supplementary Figure 1). To understand the stereochemistry of the reaction we conducted NMR experiments with allantoic acid specifically labeled at the pro-*S* carbon. The time course of the reaction (Figure 1, panel c) demonstrates that hydrolysis involves the pro-*R* carbon, thus yielding the *S* enantiomer of ureidoglycine as a reaction product.

Identification of *ylbA* as a Candidate *S*-Ureidoglycine Aminohydrolase. Even if the hydrolysis of ureidoglycine is spontaneous *in vitro*, in nature there should be an enzyme to catalyze this reaction. First, the spontaneous release of ammonia from ureidoglycine appears to be too slow to sustain an efficient flux of nitrogen through the ureide pathway; second, in living cells, the reaction is known to produce a single enantiomer of ureidoglycolate, namely, *S*-ureidoglycolate (*20*). By monitoring the AllC reaction by CD spectroscopy one does not observe formation of optically active ureidoglycolate as a final product (data not shown).

To identify a candidate ureidoglycine aminohydrolase we searched for uncharacterized genes present in bacterial genomic clusters containing allantoate amidohydrolase. As a further condition, we required that orthologous genes be present in organisms containing allantoate amidohydrolase (both prokaryotes and eukaryotes) and absent in organisms containing allantoicase (these organisms are not expected to form ureidoglycine). The search identified the *E. coli ylbA* gene (*21*) as a candidate ureidoglycine aminohydrolase (Figure 2). The *ylbA* gene is immediately downstream from the *allC* gene in the *E. coli* genome and is present in several purine degradation clusters along with *allC* (Figure 2, panel a). Homology searches and phylogenetic analysis (Figure 2, panel b) indicate that genes closely related to *ylbA* are present in plants and green algae (which possess allantoate amidohydrolase) but not in metazoa and the majority of fungi (which possess allantoicase). A more distant clade of *ylbA*-related genes is found in organisms (bacteria and some fungi) devoid of both allantoicase and allantoate amidohydrolase. Three-dimensional structures are available for the *E. coli* YlbA proteins (PDB code 1RC6) and for other three bacterial related proteins (Figure 2, panel b); all of these structures have been solved *via* structural genomics and correspond to uncharacterized proteins.

The *ylbA* gene of *E. coli* encodes a protein of 261 amino acids. The homologous gene of *Arabidopsis thaliana* ($p < 10^{-28}$) encodes a protein of 298 amino acids with 28% amino acid identity. Difference in protein length is determined by an extra N-terminal segment that is present only in plant sequences, and it is predicted that it contains a cleavable signal peptide (Figure 2, panel c).

Recombinant Expression of *ylbA* from Bacteria and Plants. The full-length *E. coli* YlbA protein was produced using an ASKA library clone (*18*). The plant homologous protein was produced in *E. coli* by cloning and expression of an *Arabidopsis thaliana* coding sequence isolated from whole plant cDNA. Both the full-length protein and a truncated variant starting at Pro51, lacking a putative signal peptide and a short plant-specific sequence (see Figure 2, panel c), were produced. Recombinant proteins had an N-terminal histidine-tag and were purified near to homogeneity by affinity chromatography. The protein corresponding to the full-length plant sequence was found to be largely insoluble, whereas the bacterial protein and the plant truncated variant were found to be soluble and were used for biochemical characterization.

YlbA Encodes a Mn-Dependent *S*-Ureidoglycine Aminohydrolase. By supplementing the AllC reaction mixture with the recombinant YlbA protein from plant or bacteria, we observed the rapid release of 2 mol of ammonia per mole of allantoic acid; there were no release of ammonia or urea, however, if AllC was omitted from the reaction mixture (data not shown), suggesting

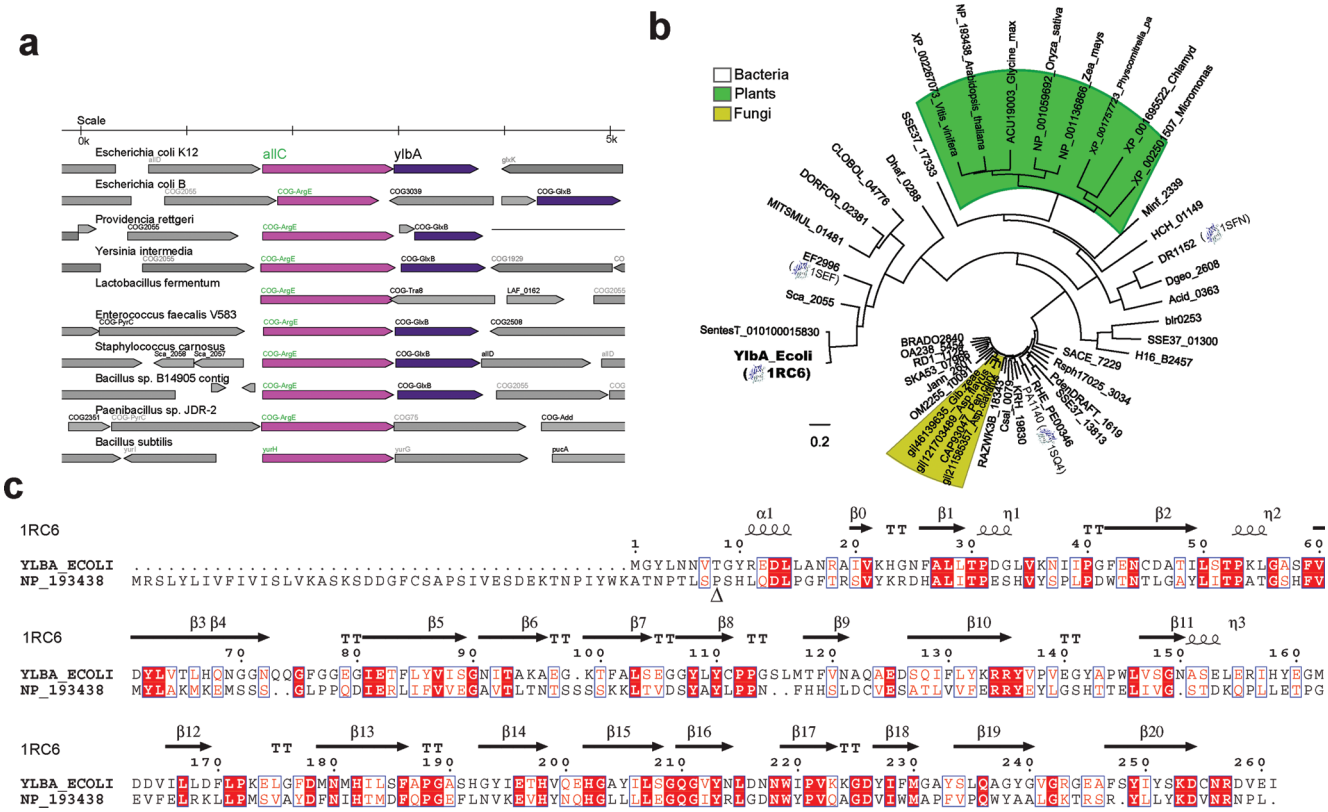


Figure 2. Identification of *ylbA* as a candidate *S*-ureidoglycine hydrolase. a) Comparison of the genetic context of the *allC* gene (magenta) in *E. coli* and other bacteria showing the frequent association with the *ylbA* gene (blue). b) Phylogenetic tree of the YlbA protein family. Protein sequences are identified by Microbesonline or Genbank accession numbers; PDB codes of the available 3D structures are indicated in parentheses. c) Alignments of *E. coli* YlbA with the homologous protein from *A. thaliana*. Conserved positions are shaded according with the similarity criteria of Espript (*36*); secondary structure elements derived from the 3D coordinates of *E. coli* YlbA (1RC6) are drawn above the alignment. The artificially truncated variant of the *A. thaliana* protein is indicated by a triangle.

that the YlbA uses the AllC reaction product, but not allantoic acid, as substrate. In fact when the YlbA protein from plant or bacteria was added after completing the first step of the AllC reaction, it caused the rapid release of a second mole of ammonia (Figure 3, panel a). The reaction product of allantoate in the presence of AllC and YlbA was found to be optically active, with a CD spectrum (Figure 3, panel b) matching a spectrum previously attributed to *S*-ureidoglycolate (*20*). We concluded from these experiments that the YlbA protein is able to catalyze the enantioselective hydrolysis of *S*-ureidoglycine to *S*-ureidoglycolate, with inversion of configuration at the chiral carbon atom.

Allantoate amidohydrolase from various sources is reported to be a Mn-dependent enzyme. We confirmed this observation with the recombinant AllC protein, which was activated by Mn^{2+} and, to a lesser extent by Co^{2+} and Ni^{2+} (data not shown). To obtain maximum activity, the manganese ion was typically included in the AllC reaction mixture. However, freshly purified AllC preparations were catalytically active (albeit at a lower rate) even without the addition of metals, allowing the metal dependency of YlbA to be determined. In the absence of added metals, YlbA from plant and bacteria was found to be inactive. The activity could be completely restored by the addition of Mn^{2+} and partially restored by the addition of Co^{2+}; all of the other divalent metals examined (Zn^{2+}, Cu^{2+}, Ca^{2+}, Ni^{2+}, Mg^{2+}) were ineffective (Figure 3, panel c).

Structure–function Relationship of *E. coli* YlbA. The three-dimensional structure of the YlbA protein has been solved at 2.6 Å resolution in the framework of

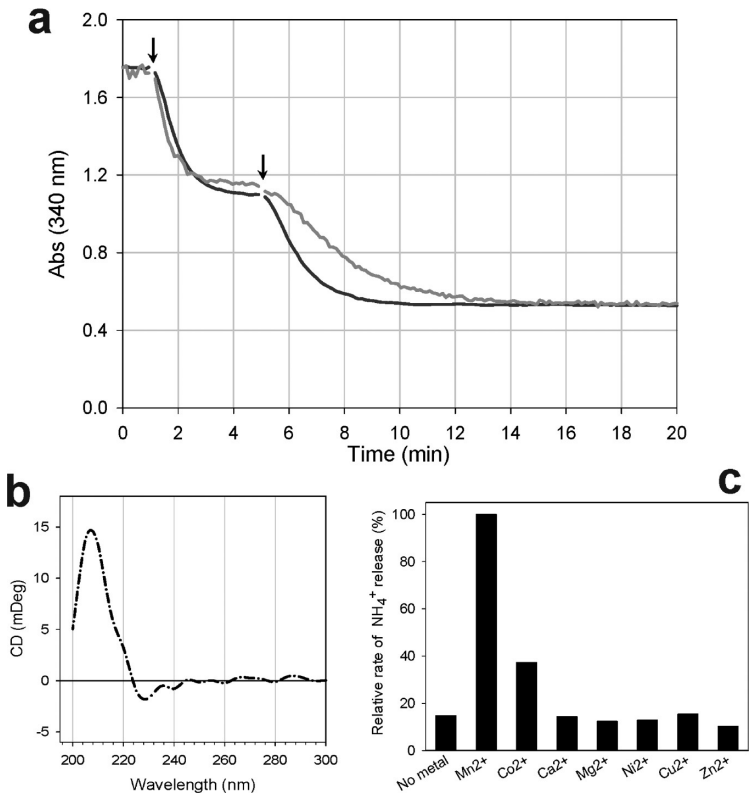


Figure 3. Biochemical characterization of recombinant YlbA from bacteria and plants. a) Time courses of ammonia release from allantoate, catalyzed by the successive additions of AllC and *E. coli* YlbA (gray line) or AllC and *A. thaliana* YlbA-related protein (black line) monitored spectrophotometrically using a coupled assay with glutamate dehydrogenase. b) CD spectrum of the optically active product obtained from allantoate in the presence of both AllC and YlbA. c) Relative rates of ammonia release by the YlbA-catalyzed reaction in the presence of various divalent metals.

structural genomics (*22*) and was deposited in the Protein Data Bank in 2003 (PDB code 1RC6). The structure of the YlbA monomer shows two cupin domains (Figure 4, panel a). Each cupin domain is made up of a small barrel (*cupa* in Latin) featuring two antiparallel β-sheets (A and B in domain I, A′, and B′ in domain II), with a topology characteristic of the cupin fold (Figure 4, panel b). The six-stranded β-sheets A (β2, β3, β5, β8, β10, β11) and A′ (β1, β12, β13, β15, β18, β20) have a 236451′ topology, where β-strand 1′ is contributed by the neighboring domain by domain-swapping. The four-stranded β-sheets B (β4, β6, β7, β9) and B′ (β14, β16, β17, β19) have a 1423 topology.

The 3D structure reveals strong structural similarities between the N- and C-terminal halves of the protein, indicating that the bidomain structure of YlbA originated from an ancestral duplication of a single cupin domain. This bicupin organization is also present in other proteins with a structure known to belong to the YlbA family (see also Figure 2, panel b). Related proteins from *Enterococcus faecalis* (1SEF), *Deinococcus radiodurans* (1SFN), and *Pseudomonas aeruginosa* (1SQ4) have structures similar to that of YlbA (rmsd of 1.7, 2.1, and 2.2 Å, respectively), although some differences are found in the quaternary organization. On the basis of crystal data, the biological units were determined to be monomeric for 1RC6 and 1SEF, dimeric for 1SFN, and octameric for 1SQ4. No metals were included in the solution structures of these proteins. However, this does not exclude manganese binding, since identification of metals at binding sites is not trivial even when 3D data is available from X-ray crystallography (*23*).

The presence of a cupin domain is found in several proteins classified in the cupin superfamily, functionally one of the most diverse seen thus far (*24*). It comprises 20 families with members that perform diverse functions ranging from enzymatic to nonenzymatic functions. YlbA is the prototypic member of a family that encompasses proteins whose function is unknown (*25*). Structural similarities, however, reveal an evolutionary relationship between YlbA and proteins whose function is known. Interestingly, among the proteins of the cupin superfamily there are several Mn-dependent enzymes, such as ureidoglycolate hydrolase (*26*) (another enzyme of the purine catabolic pathway), oxalate oxidase, and oxalate decarboxylase.

In most bicupins, an enzymatic activity is associated with only one of the two domains. Active domains usually have a metal-binding site (usually for Mn, Fe, or Zn) localized at the mouth of the barrel. Thus, structural alignment with known metal-binding cupins can identify residues involved in metal coordination (*25*). The comparison of N- and C-terminal domains of YlbA proteins with the Mn-binding domain of oxalate decarboxylase (Figure 4, panel c) suggests the existence of a metal-binding site in domain II of YlbA. Unlike domain I, in which residues involved in metal coordination are replaced by not conserved hydrophobic residues, domain II shows the presence of potential metal-binding residues that are strictly conserved in the YlbA family. Although the configuration of potential metal-binding residues in the domain II of YlbA (His, Glu, His, Gln) differs from the canonical configuration of cupins (His, His, Glu,

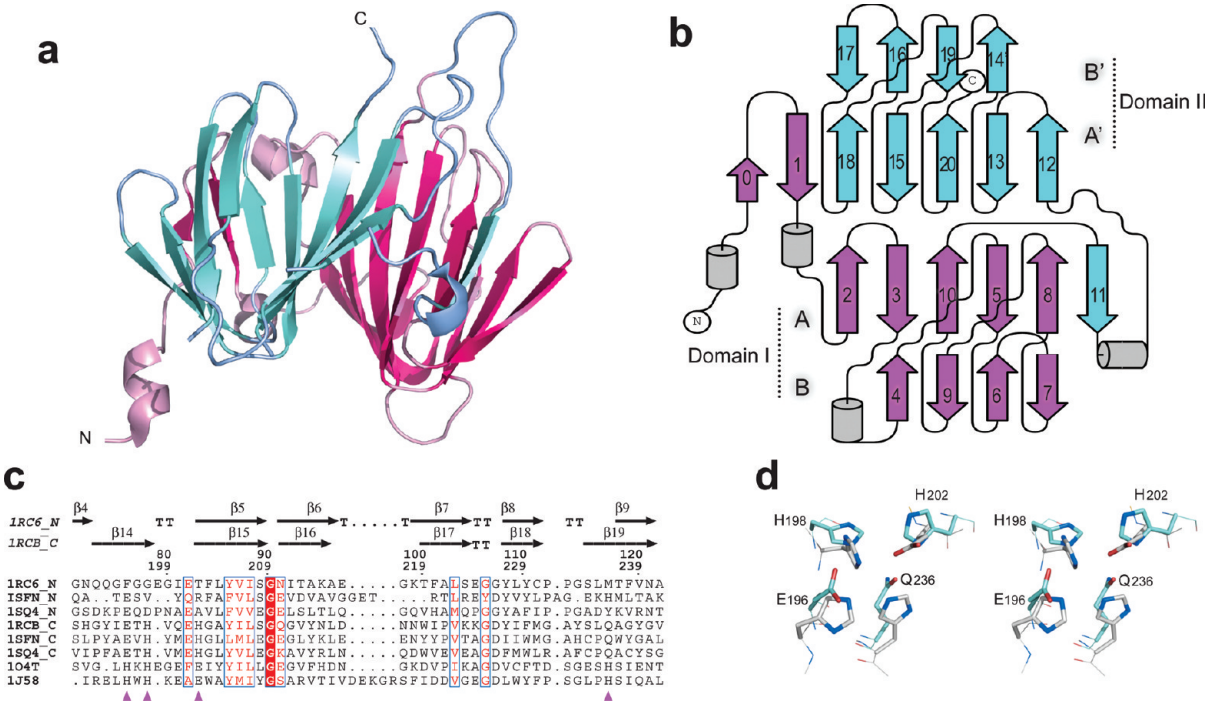


Figure 4. Structure of YlbA protein. a) Cartoon sketch of the YlbA structure, highlighting the secondary structural elements and the bidomain structure, with the N-terminal half (aa 1—135) in magenta and the C-terminal half (aa 136—261) in cyan. b) Topology diagram of YlbA, showing the β-sheet organization of domains I and II. c) Structure-based sequence alignment of N- and C-terminal domains of YlbA and related proteins with oxalate decarboxylase (Oxd) from *B. subtilis* (1J58) and *T. maritima* (1O4T). Amino acid residues that coordinate a manganese ion in the Oxd structures are indicated by triangles. d) Stereoview comparison of the metal binding site of Oxd (1O4T, gray) with the corresponding region of the C-terminal domain of YlbA (1SEF, cyan); residue numbering is according *E. coli* YlbA.

His), the 3D superposition shows conservation of the geometry of the metal binding site (Figure 4, panel d). Interestingly, examination of the electron density map of 1SEF, the best resolved YlbA structure (2.0 Å), reveals the presence of an uninterpreted electron density in a region corresponding to the Mn atom of oxalate decarboxylases.

Endoplasmic Localization of Plant *S*-Ureidoglycine Aminohydrolase. In plants, allantoate amidohydrolase is localized in the endoplasmic reticulum (ER) (*11*). The presence of *S*-ureidoglycine aminohydrolase in the same compartment where the unstable ureidoglycine is formed would make biological sense. Plant YlbA sequences have an extra N-terminal segment of ~40 aa that has been predicted to contains a signal for the co-translational insertion of the protein in the ER (see Figure 2, panel c). The *Arabidopsis* as well as other plant sequences do not contain the known signature for ER retention (K/HDEL). However, the presence of such a signal is not an absolute requirement for ER proteins (*27*).

To demonstrate the subcellular localization of *S*-ureidoglycine aminohydrolase, we constructed C-terminal fusions of full length *Arabidopsis* YlbA sequence with the yellow fluorescent protein (YlbA-YFP). The fusion proteins were transiently expressed in *Arabidopsis* protoplast or also co-expressed with a green fluorescent protein with a signal peptide and an ER retention signal (GFP-HDEL), used as a marker for the ER. Upon transient expression, YlbA-YFP localized to cellular structures with the typical shape of the ER (Supplementary Figure 2). The co-localization of YlbA-YFP and GFP-HDEL fusion proteins further supports an ER localization for the plant enzyme (Figure 5). We concluded from confocal microscopy studies that in the plant cell, *S*-ureidoglycine aminohydrolase is localized in the same

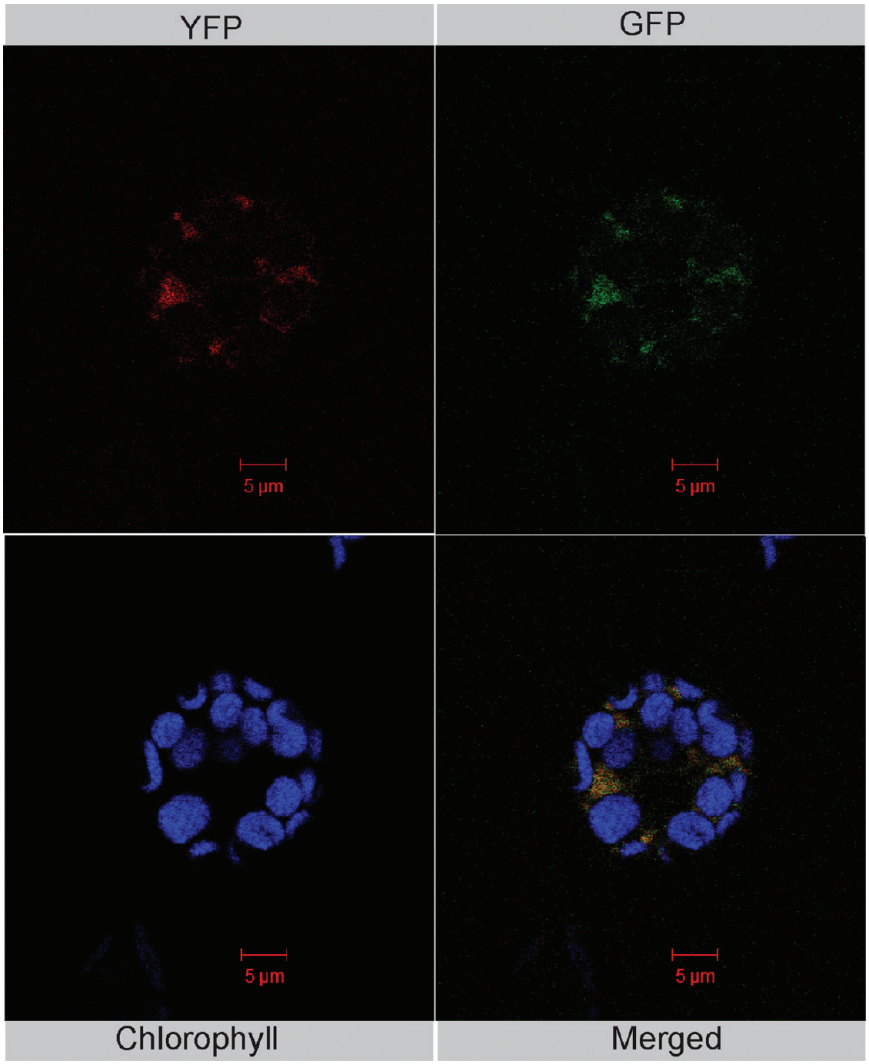


Figure 5. Localization of *S*-ureidoglycine hydrolase in plant cells. The co-localization of YlbA-YFP and GFP-HDEL proteins in double transformants was observed in *Arabidopsis* protoplasts using confocal microscopy. Fluorescence was monitored in the range of 499—510 nm (GFP), nm 531—596 (YFP), and 649—767 nm (chlorophylls); detection channels and three-channel merged images are indicated by column labels.

subcellular structure where *S*-ureidoglycine is formed. Retention of the protein in the ER does not depend on the presence of a canonical K/HDEL signal.

Conclusions and Outlook. The true product of the AllC reaction is *S*-ureidoglycine. The distribution of the *allC* gene indicates *S*-ureidoglycine is formed in several bacteria and in all Viridiplantae (plants and green algae). NMR evidence presented here on the AllC reaction product is in agreement with the evolutionary classification of the enzyme. By sequence homology, AllC belongs to *N*-carbamoyl-L-amino acid amidohydrolases. These proteins differ to *N*-carbamoyl-D-amino acid amidohydrolases, which have a different fold and metal-independent catalysis. Before the identification of the *allC* gene, bacterial extracts with allantoate aminohydrolase activity were found to be able to hydrolyze *N*-carbamoyl-L-asparagine but not *N*-carbamoyl-D-asparagine (*20*). These observations are consistent with hydrolysis of the pro-*R* carbamoyl group of allantoate and formation of *S*-ureidoglycine by the AllC-catalyzed reaction.

A previously uncharacterized gene, *ylbA*, encodes an enzyme catalyzing the hydrolysis of *S*-ureidoglycine into *S*-ureidoglycolate. This enzyme accelerates spontaneous hydrolysis of ureidoglycine and, importantly, selects the correct enantiomer of ureidoglycolate for the subsequent reaction of the pathway, which in most organisms is performed by a stereospecific *S*-ureidoglycolate hydrolase (*4*). An additional advantage of the enzyme can be the avoidance of undesired side-effects promoted by the unstable ureidoglycine. In fact, the AllC reaction has been reported to yield different end-products *in vitro* depending on the composition of the reaction mixture (notably in the presence of pyridoxal phosphate) (*14*). In eukaryotes, this role is further supported by compartmentalization. Ureidoglycine is only produced in the endoplasmic reticulum in plants, which possess an endoplasmic allantoate amidohydrolase (*11*), and as shown here, in plant cells the enzyme acting on ureidoglycine is also localized in the endoplasmic reticulum. Compartmentalization in the ER is also consistent with the Mn dependency of both AllC and YlbA. Manganese, an essential metal nutrient for plants, is actively sequestered in the ER of these organisms (*28*). At variance with other organelles, where the transport of Mn has been related to its participation in enzymatic reactions (a cluster of Mn atoms is required as the catalytic center of photosystem II in chloroplasts, and Mn-depedent superoxide dismutases are

present in peroxisomes and mitochondria), transport of Mn into the ER has been so far explained in terms of detoxification and storage (*28, 29*). The requirement of manganese for two consecutive enzymes in the ureide pathway indicates that this metal also plays an important metabolic role in plant ER.

The *ylbA* gene name derives from a temporary nomenclature adopted for *E. coli* genes of unknown function (*21*). Once a y-gene is functionally characterized, it is suggested that a mnemonic gene name be assigned to replace the y-name. We suggest renaming the gene encoding *S*-ureidoglycine aminohydrolase (UGHY) as *ughy*.

An interesting question concerning the *in vivo* role of this enzyme can be approached through the phenotypic analysis of gene deletion mutants. Early genetic evidence for the existence of ureidoglycine hydrolase was obtained by the isolation of mutants able to use allantoic acid as a source of nitrogen but not as a carbon source (*16*). More recently, AAH from *A. thaliana* was used to complement a yeast strain deleted in the gene encoding allantoicase. This experiment, which can be considered equivalent to the deletion of ureidoglycine hydrolase (*S*-ureidoglycolate is formed by allantoicase or by the consecutive action of AAH and UGHY), showed partial recovery of the ability to use allantoate as a nitrogen source and the accumulation of a metabolite, indicated by a yellow-orange pigmentation of the transformed colonies (*10*). We used available strains from the Keio collection of *E. coli* knockout mutants (*30*) in the attempt to investigate the phenotype of *ughy* mutants grown on oxidized purines as a source of nitrogen. *E. coli* is known to be able to utilize certain purine derivatives as a source of nitrogen under anaerobic conditions, although strain to strain differences were observed (*4, 12*). In our experiments we could not observe growth on allantoin or allantoate for the Keio *ylbA* mutant and for the wild-type strain BW25113. Lack of growth of the wild-type strain depended on gene regulation rather than on a defect in enzymes of the catabolic pathway, since a BW25113 mutant deleted in the repressor of purine degradation (*allR*) could grow on allantoin (but not on allantoate) with glucose or xylose as carbon sources (Supplementary Figure 3).

The present study provides an example of how the combination of a detailed characterization of the chemistry of "known" biological reactions and *in silico* genome analysis can lead to the identification of novel gene functions. This methodology has been inspired by recent findings in the same metabolic pathway. A detailed spectroscopic characterization of the urate oxidase reaction suggested the presence of missing enzymatic activity in the urate metabolism (*31*), leading to the identification of two additional enzymes acting on unstable intermediates and involved in the conversion of urate into *S*-allantoin (*7*). To date, seven enzymatic steps have been identified in the ureide pathway for the complete conversion of xanthine to glyoxylate, carbon dioxide, and ammonia. Three out of these seven steps are performed by enzymes acting on unstable compounds and have only been identified recently. This suggests that the very low number of known enzymes acting on short-lived chemical species is mainly due to the fact that these catalysts easily escape identification by classical biochemical or genetic approaches. Other examples of such catalysts are expected to be revealed in the future thanks to the availability of complete genomes and advanced analytical tools.

While this article was under revision, another study was published that reported the independent identification of ureidoglycine hydrolase through comparative genomics (*32*). The evidence presented by Werner *et al.* on the function of the *ylbA* gene, in the context of a comprehensive analysis of the ureide pathway in plants and bacteria, are consistent with the results of the present study.

METHODS

Synthesis of Labeled Compounds and NMR Spectroscopy. Uniformly labeled [^{15}N,^{13}C]allantoate was obtained through hydrolysis of [^{15}N,^{13}C]S-allantoin, which was synthesized from [^{15}N,^{13}C]*R*-adenosine (Spectra Stable Isotope) using six enzymatic steps (*33*). [^{15}N,^{13}C]*S*-Allantoin hydrolysis was performed in 1.1 mL of 0.1 M potassium phosphate, 80% D_2O, pD 7.6 using 28.8 μg of recombinant PuuE allantoinase (*33*). After the completion of the reaction, the solution was ultrafiltrated to remove proteins and degassed using a syringe to remove labeled CO_2 from previous reactions. Synthesis and hydrolysis of *S*-allantoin were performed consecutively to avoid spontaneous racemization (*34*). Specifically labeled [2-^{13}C]allantoate was obtained enzymatically from [8-^{13}C]urate, which was synthesized by condensing 5,6-diaminouracil with [^{13}C]urea (Sigma) according to a described protocol (*31*). [8-^{13}C]Uric acid was converted into [2-^{13}C]*S*-allantoin using the same procedure described for uniformly labeled uric acid (*33*), and [2-^{13}C]*S*-allantoin was hydrolyzed to [2-^{13}C]allantoate using the same procedure described above.

^{13}C NMR spectra of the product of the AllC reaction were obtained using a 0.6 mL solution of 0.1 M potassium phosphate, 80% D_2O, pD 7.6, containing uniformly labeled [^{15}N,^{13}C]-allantoate (6.9 mM) or specifically labeled [2-^{13}C]allantoate (9.5 mM). The solutions, supplemented with 20 μg of AllC plus 80 μM $MnCl_2$ or with 50 μg of AllC plus 45 μM $MnCl_2$, respectively, were gently stirred for 1 min and then transferred into 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.

Bioinformatics. Search and comparison of genetic clusters containing allantoate amidohydrolase were conducted using the Microbesonline web server (http://microbesonline.org). Sequences of the YlbA family were retrieved from Microbesonline or from GenBank through homology searches. Sequence alignments were carried out with Clustalw (*35*) and visualized with Espript (*36*). Phylogenetic analysis was performed using the neighbor-joining method (*37*) implemented in Clustalw, and the resulting tree was visualized as a polar layout with FigTree (http://tree.bio.ed.ac.uk). Sequences of the YlbA family with known structure were identify by sequence homology in the PDB database. Prediction of the signal peptide in the plant sequence was performed with the SignalP server (http://www.cbs.dtu.dk/services/SignalP).

Identification of structural similarity and pairwise structural alignment were performed with Fatcat (*38*). Pymol (http://pymol.sourceforge.net) was used to study and visualize protein models, and Coot (*39*) was used to study electron density maps.

Cloning, Production of Recombinant Proteins and Cell Growth. Clones encoding full-length histidine-tagged AllC and YlbA proteins from *E. coli* K12 were obtained from the ASKA library (*18*). Clones were verified by sequencing using primers based on the pCA24N vector. Amplicons encoding the full-length YlbA protein from *A. thaliana* (AtYlbA) were obtained by PCR amplification of total plant cDNA with primers forward 5′-ATCGGTCCGGCAAAAATGCGATCACTTTAC-3′ and reverse 5′-ATCGGACCGCTTTGATCACAATGGATTTCG-3′. Amplicons encoding a protein lacking the signal peptide and a plant-specific N-terminal sequence, starting from amino acid 51 of the *A. thaliana* YlbA (ΔAtYlbA), were obtained using a different forward primer (5′-ATCGGTCGGATGCCTTCACACCTCCAAGACTTG-3′). AtYlbA amplicons were cloned into pGEM-T Easy vector (Promega) and subcloned into the expression vector pET28-CpoI (A. Bolchi, unpublished). ΔAtYlbA amplicons were cloned directly into the expression vector pET28-SnaBI (A. Bolchi, unpublished) using a one-step cloning procedure (*40*). The ligated plasmid was then transformed into BL21 (DE3)-RIL *E. coli* cells (Stratagene), and the inserts were sequence-verified. Transformed cells were grown at 37 °C in LB medium. Gene expression was induced at an optical density at 600 nm of 0.3 (AllC, YlbA) or 0.6 (AtYlbA, ΔAtYlbA) using 0.5 (AllC, YlbA) or 1 mM (AtYlbA, ΔAtYlbA) isopropyl-1-thio-β-D-galactopyranoside (IPTG); after 48 h at 15 °C (AllC, YlbA) or 24 h at 22 °C (AtYlbA, ΔAtYlbA) the cells were resuspended in 120 mL of sonocation buffer (50 mM sodium phosphate, 0.3 M NaCl, 10% glicerol, 1 μM pepstatin, 1 μM leupeptin, 100 μM PMSF, 2 mL mL^{-1} lysozyme, pH 8) with (AllC, YlbA) or without (AtYlbA, ΔAtYlbA) 1 mM β-mercaptoethanol and incubated on ice for 30 min. Cells were lysed by four 15-s bursts of sonication, and the supernatant obtained after centrifugation of the crude extract was concentrated by ultrafiltration in an Amicon cell (YM-10 membrane, Millipore). Proteins were purified by metal affinity chromatography with Talon resin (Clontech) to a near-homogeneity, as assessed by SDS−PAGE analysis. Loaded columns were washed with buffer 50 mM Tris-HCl, 5% glycerol, 0.2 M NaCl, pH 7.0, and proteins were eluted with 50 mM (AllC) or 100 mM (YlbA) imidazole. AllC was conserved at 4 °C in the presence of 1 mM EDTA. This condition ensured a better, though not complete, stability of the enzymatic activity. YlbA was conserved at —20 °C in the elution buffer. In these conditions, the enzymatic activity was stable for several months.

Experiments with the Keio knockout mutants (*30*) were conducted with cells grown at 37 °C in a minimal medium consisting of 34 mM NaH_2PO_4, 64 mM K_2HPO_4, 1 μM $FeSO_4$, 0.1 mM $MgSO_4$, 10 μM $CaCl_2$, 10 μM $MnCl_2$, pH 7.1 in the presence of 15 mM allantoin or 30 mM $(NH_4)_2SO_4$. Glucose or xylose (120 mM) was utilized as carbon sources. Anaerobic growth was carried out in 2-mL vials completely filled with medium and maintained under constant rotation for 72 h.

Biochemical Assays. Ammonia release by the AllC reaction was determined spectrophotometrically using a coupled assay with glutamate dehydrogenase (GDH) (*14, 17*). The typical incubation mixture consisted of 1 mL of 0.2 M Tris-HCl buffer, pH 8.5, with 0.3 mM NADPH, 2.5 mM α-ketoglutarate, 4.12 units of GDH from *Proteus* sp. (Sigma), 100 μM $MnCl_2$, and 6 μg of AllC. The reaction was initiated by the addition of allantoic acid (0.15 mM), and the decrease in absorbance at 340 nm due to the oxidation of NADPH was recorded. Enzyme activity was monitored in a pH range of 7.2−9.5. The effect of metal ions on the enzyme activity was determined by preincubating the enzyme with 10- to 1000-fold excess metal ions (Zn^{2+}, Co^{2+}, Cu^{2+}, Ca^{2+}, Ni^{2+}, Mn^{2+}, Mg^{2+}) in 0.2 M Tris-HCl, 1 mM EDTA, pH 8.5. Maximum activation was obtained with a ratio of Mn^{2+}/enzyme of 60:1. Ammonia release by the YlbA reaction was determined using the same coupled assay described above, in the absence or in the presence of AllC. Recombinant YlbA (6 μg) from bacterial or plant source was added to the initial reaction mixture or after the fast phase of the AllC reaction. The effect of metal ions on the activity of YlbA was determined by excluding Mn^{2+} from the initial reaction mixture and by adding YlbA plus 100 μM metal ions (Zn^{2+}, Co^{2+}, Cu^{2+}, Ca^{2+}, Ni^{2+}, Mn^{2+}, Mg^{2+}) after the first phase of the AllC reaction.

Formation of optically active compounds in the reactions catalyzed by AllC and YlbA was monitored by CD measurements carried out in a 10 mm path length cuvette with a Jasco J-715 spectropolarimeter. The degradation of allantoic acid (700 μM) in 1 mL of 20 mM potassium phosphate, 100 μM $MnCl_2$, pH 7.6, was monitored in the 200—300 nm range in presence of *E. coli* AllC (20 μg) and in the presence or in the absence of *A. thaliana* YlbA (30 μg). Ureidoglycolate has a CD signal in a region of the spectrum where absorption by other compounds of the reaction mixture interferes with the measurements. CD signals could be obtained with appropriate reactant and buffer concentrations and by lowering the imidazole content of protein solutions.

***In Vivo* Localization of Fluorescent Proteins.** To generate chimeric fusion construct of YlbA with the Yellow Fluorescent Protein (YFP), a pGEM plasmid containing the full-length plant AtYlbA sequence was amplified using primers forward 5′-ATCGGTCCGGCAAAAATGCGATCACTTTAC-3′, and reverse 5′-TGCTCACCATCAATGGATTTCGATTCACAT-3′, introducing a plant ribosome binding site at the 5′ of the gene and a 5′-YFP sequence (10 nt) at the 3′ of the YLBA sequence. The reverse primer eliminates the TAA stop codon from YlbA, allowing YFP fusion. YFP cDNA (Clontech) was amplified using primers forward 5′-TCGAAATCCATTGATGGTGAGCAAGGGCGAG-3′ and YFP reverse 5′-ATTCTAGATTACTTGTACAGCTCGTCCATG-3′ introducing a 3′ YlbA sequence (13 nt) at the 5′ of YFP and a 3′ *Xba1* site. A third PCR with primers YLBA forward and YFP reverse, using as template the YLBA and YFP amplicons, gave the fusion construct YlbA-YFP, which was cloned in pBluescript II KS/SK (+) following a published protocol (*40*). The ligated plasmid was then transformed into XL1B *E. coli* cells (Stratagene), and the insert was sequence verified. The plasmid was subsequently treated with *EcoRI* and

XbaI and subcloned into the plant expression vector pART7. The construct GFP-Hdel-pVKH18en6 (*41*) was a kind gift from Janet Evins.

The fusion constructs were introduced by poly(ethylene glycol)-mediated transformation into protoplasts prepared from plant leaves (*42*). Standard *Arabidopsis thaliana* ecotype Columbia-0 (European *Arabidopsis* Stock Centre) plants were grown 3—4 weeks on soil in 16-h light/8-h dark cycle at 25 °C. Transformed protoplasts were incubated in darkness at 23 °C for 16—24 h before checking the fluorescence. Cells were mounted in custom-made chambers and observed by confocal microscopy (x100 objective lens, 488 nm excitation) using a LSM 510 Meta scan head equipped with an Axiovert 200 M inverted microscope (Carl Zeiss). Unlabeled samples were used to establish the levels and locations of the autofluorescence due to plastids, and single-label controls were used to assess bleed-through between fluorochromes, allowing the selection of emission filter sets giving negligible cross-talk between fluorochromes with partially overlapping emission spectra. For 3D reconstructions, stacks of digital images were processed with the Axiovision software (Carl Zeiss) by applying the "shadow" algorithm.

Accession Codes: The sequences of *E. coli* and *A. thaliana* YlbA have been submitted to GenBank with accession numbers GQ303359 and GQ303360.

Acknowledgment: We thank A. Peracchi for discussion, D. Acquotti and the University of Parma CIM for NMR measurements, A. Secchi for help with the synthesis of labeled compounds, B. Montanini and A. Ronda for technical assistance, and F. Caristi for the help given during protein expression. This work has been funded with an FIL grant from the University of Parma.

Supporting Information Available: This material is available free of charge *via* the Internet.

REFERENCES

1. Codispoti, L. A. (2007) An oceanic fixed nitrogen sink exceeding 400 Tg N a-1 vs the concept of homeostasis in the fixed-nitrogen inventory, *Biogeosciences 4*, 233–253.
2. Reich, P. B., Hobbie, S. E., Lee, T., Ellsworth, D. S., West, J. B., Tilman, D., Knops, J. M., Naeem, S., and Trost, J. (2006) Nitrogen limitation constrains sustainability of ecosystem response to CO_2, *Nature 440*, 922–925.
3. Tipton, P. A. (2006) Urate to allantoin, specifically (S)-allantoin, *Nat. Chem. Biol. 2*, 124–125.
4. Vogels, G. D., and Van der Drift, C. (1976) Degradation of purines and pyrimidines by microorganisms, *Bacteriol. Rev. 40*, 403–468.
5. Todd, C. D., Tipton, P. A., Blevins, D. G., Piedras, P., Pineda, M., and Polacco, J. C. (2006) Update on ureide degradation in legumes, *J. Exp. Bot. 57*, 5–12.
6. Serraj, R., Vadez, V. V., Denison, R. F., and Sinclair, T. R. (1999) Involvement of ureides in nitrogen fixation inhibition in soybean, *Plant Physiol. 119*, 289–296.
7. Ramazzina, I., Folli, C., Secchi, A., Berni, R., and Percudani, R. (2006) Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes, *Nat. Chem. Biol. 2*, 144–148.
8. Hanks, J. F., Tolbert, N. E., and Schubert, K. R. (1981) Localization of enzymes of ureide biosynthesis in peroxisomes and microsomes of nodules, *Plant Physiol 68*, 65–69.
9. Vigetti, D., Monetti, C., Prati, M., Gornati, R., and Bernardini, G. (2002) Genomic organization and chromosome localization of the murine and human allantoicase gene, *Gene 289*, 13–17.
10. Todd, C. D., and Polacco, J. C. (2006) AtAAH encodes a protein with allantoate amidohydrolase activity from *Arabidopsis thaliana*, *Planta 223*, 1108–1113.
11. Werner, A. K., Sparkes, I. A., Romeis, T., and Witte, C. P. (2008) Identification, biochemical characterization, and subcellular localization of allantoate amidohydrolases from *Arabidopsis* and soybean, *Plant Physiol 146*, 418–430.
12. Cusa, E., Obradors, N., Baldoma, L., Badia, J., and Aguilar, J. (1999) Genetic analysis of a chromosomal region containing genes required for assimilation of allantoin nitrogen and linked glyoxylate metabolism in *Escherichia coli*, *J. Bacteriol. 181*, 7479–7484.
13. Agarwal, R., Burley, S. K., and Swaminathan, S. (2007) Structural analysis of a ternary complex of allantoate amidohydrolase from *Escherichia coli* reveals its mechanics, *J. Mol. Biol. 368*, 450–463.
14. van der Drift, C., de Windt, F. E., and Vogels, G. D. (1970) Allantoate hydrolysis by allantoate amidohydrolase, *Arch. Biochem. Biophys. 136*, 273–279.
15. Xu, Z., de Windt, F. E., and van der Drift, C. (1995) Purification and characterization of allantoate amidohydrolase from *Bacillus fastidiosus*, *Arch. Biochem. Biophys. 324*, 99–104.
16. Wu, C. H., Eisenbraun, E. J., and Gaudy, E. T. (1970) Enzymatic degradation of ureidoglycine by *Pseudomonas acidovorans*, *Biochem. Biophys. Res. Commun. 39*, 976–982.
17. Muratsubaki, H., Satake, K., and Enomoto, K. (2006) Enzymatic assay of allantoin in serum using allantoinase and allantoate amidohydrolase, *Anal. Biochem. 359*, 161–166.
18. Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research, *DNA Res. 12*, 291–299.
19. Meany, J. E., and Pocker, Y. (2002) The dehydration of glyoxalate hydrate: general-acid, general-base, metal ion and enzymic catalysis, *J. Am. Chem. Soc. 113*, 6155–6161.
20. Gravenmade, E. J., Vogels, G. D., and Van der Drift, C. (1970) Hydrolysis, racemization and absolute configuration of ureidoglycolate, a substrate of allantoicase, *Biochim. Biophys. Acta 198*, 569–582.
21. Rudd, K. E. (1998) Linkage map of *Escherichia coli* K-12, edition 10: the physical map, *Microbiol. Mol. Biol. Rev. 62*, 985–1019.
22. Chance, M. R., Fiser, A., Sali, A., Pieper, U., Eswar, N., Xu, G., Fajardo, J. E., Radhakannan, T., and Marinkovic, N. (2004) High-throughput computational and experimental techniques in structural genomics, *Genome Res. 14*, 2145–2154.
23. Schymkowitz, J. W., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F., and Serrano, L. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field, *Proc. Natl. Acad. Sci. U.S.A. 102*, 10147–10152.
24. Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R., and Goodenough, P. W. (2001) Evolution of functional diversity in the cupin superfamily, *Trends Biochem. Sci. 26*, 740–746.
25. Agarwal, G., Rajavel, M., Gopal, B., and Srinivasan, N. (2009) Structure-based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold, *PLoS One 4*, e5736.
26. Raymond, S., Tocilj, A., Ajamian, E., Li, Y., Hung, M. N., Matte, A., and Cygler, M. (2005) Crystal structure of ureidoglycolate hydrolase (AllA) from *Escherichia coli* O157:H7, *Proteins 61*, 454–459.
27. Lewis, M. J., and Pelham, H. R. (1996) SNARE-mediated retrograde traffic from the Golgi complex to the endoplasmic reticulum, *Cell 85*, 205–215.
28. Pittman, J. K. (2005) Managing the manganese: molecular mechanisms of manganese transport and homeostasis, *New Phytol. 167*, 733–742.
29. Peiter, E., Montanini, B., Gobert, A., Pedas, P., Husted, S., Maathuis, F. J., Blaudez, D., Chalot, M., and Sanders, D. (2007) A secretory pathway-localized cation diffusion facilitator confers plant manganese tolerance, *Proc. Natl. Acad. Sci. U.S.A. 104*, 8532–8537.
30. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol. Syst. Biol.* 2006.0008.

31. Kahn, K., Serfozo, P., and Tipton, P. A. (1997) Identification of the true product of the urate oxidase reaction, *J. Am. Chem. Soc. 119*, 5435–5442.
32. Werner, A. K., Romeis, T., and Witte, C. P. (2009) Ureide catabolism in *Arabidopsis thaliana* and *Escherichia coli*, *Nat. Chem. Biol. 6*, 19–21.
33. Ramazzina, I., Cendron, L., Folli, C., Berni, R., Monteverdi, D., Zanotti, G., and Percudani, R. (2008) Logical identification of an allantoinase analog (puuE) recruited from polysaccharide deacetylases, *J. Biol. Chem. 283*, 23295–23304.
34. Kahn, K., and Tipton, P. A. (2000) Kinetics and mechanism of allantoin racemization, *Bioorg. Chem. 28*, 62–72.
35. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res. 22*, 4673–4680.
36. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics 15*, 305–308.
37. Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol. 4*, 406–425.
38. Ye, Y., and Godzik, A. (2004) FATCAT: a web server for flexible structure comparison and structure similarity searching, *Nucleic Acids Res. 32*, W582–585.
39. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126–2132.
40. Bolchi, A., Ottonello, S., and Petrucco, S. (2005) A general one-step method for the cloning of PCR products, *Biotechnol. Appl. Biochem. 42*, 205–209.
41. Saint-Jore, C. M., Evins, J., Batoko, H., Brandizzi, F., Moore, I., and Hawes, C. (2002) Redistribution of membrane proteins between the Golgi apparatus and endoplasmic reticulum in plants is reversible and not dependent on cytoskeletal networks, *Plant J. 29*, 661–678.
42. Abel, S., and Theologis, A. (1994) Transient transformation of *Arabidopsis* leaf protoplasts: a versatile experimental system to study gene expression, *Plant J. 5*, 421–427.