BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Gene Context Analysis Reveals Functional Divergence between Hypothetically Equivalent Enzymes of the Purine−Ureide Pathway

Vincenzo Puggioni,[†] Ambra Dondi,[†] Claudia Folli,[‡] Inchul Shin,[§] Sangkee Rhee,[§] and Riccardo Percudani*[,†]

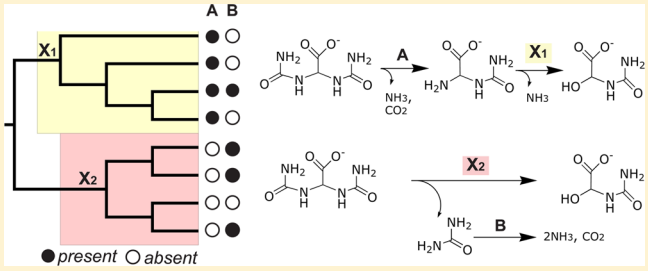
[†]Laboratory of Biochemistry, Molecular Biology, and Bioinformatics, Department of Life Sciences, University of Parma, Italy

[‡]Department of Food Science, University of Parma, Italy

[§]Department of Agricultural Biotechnology, Seoul National University, Seoul, Korea

Supporting Information

ABSTRACT: A major problem of genome annotation is the assignment of a function to a large number of genes of known sequences through comparison with a relatively small number of experimentally characterized genes. Because functional divergence is a widespread phenomenon in gene evolution, the transfer of a function to homologous genes is not a trivial exercise. Here, we show that a family of homologous genes which are found in purine catabolism clusters and have hypothetically equivalent functions can be divided into two distinct groups based on the genomic distribution of functionally related genes. One group (UGLYAH) encodes proteins that are able to release ammonia from (*S*)-ureidoglycine, the enzymatic product of allantoate amidohydrolase (AAH), but are unable to degrade allantoate. The presence of a gene encoding UGLYAH implies the presence of AAH in the same genome. The other group (UGLYAH2) encodes proteins that are able to release ammonia from (*S*)-ureidoglycine as well as urea from allantoate. The presence of a gene encoding UGLYAH2 implies the absence of AAH in the same genome. Because (*S*)-ureidoglycine is an unstable compound that is only formed by the AAH reaction, the *in vivo* function of this group of enzymes must be the release of urea from allantoate (allantoicase activity), while ammonia release from (*S*)-ureidoglycine is an accessory activity that evolved as a specialized function in a group of genes in which the coexistence with AAH was established. Insights on the active site modifications leading to a change in the enzyme activity were provided by comparison of three-dimensional structures of proteins belonging to the two different groups and by site-directed mutagenesis. Our results indicate that when the neighborhood of uncharacterized genes suggests a role in the same process or pathway of a characterized homologue, a detailed analysis of the gene context is required for the transfer of functional annotations.



INTRODUCTION

As the output of sequencing projects has largely outpaced the experimental determination of gene functions, a biological role for a great number of genes and proteins is assigned solely on the basis of bioinformatic analysis.[1−4] In the very common homology-based sequence annotation process, an experimentally verified function, usually established for a particular gene or protein in model organisms, is used as a reference to annotate uncharacterized homologous sequences that, hypothetically, have the same activity as the characterized one. However, this extrapolation procedure ought to be implemented with caution, as common gene ancestry (i.e., homology) does not guarantee conservation of the functional properties.[5−7]

Various types of analyses have been developed to increase the confidence of transfer of functional annotation or to identify possible cases of functional divergence. Even though a correlation exists between the degree of global sequence identity in pairwise comparisons and function conservation, there is no general similarity threshold that can be used to assess if two genes have the same or a different function.[8,9] Thus, a common strategy is to infer the type of homologous relationship on the grounds that in genes related by speciation (i.e., orthologous) the same function is often preserved.[10] This type of inference is made by comparing gene trees and species trees[11,12] or using approximate methods based on blast bidirectional hits and partition algorithms.[13−15] However, orthology-based approaches suffer some limitations. On the one hand, it is possible for orthologues to diverge functionally, while on the other hand a function can be preserved in genes deriving from duplication (i.e., paralogous) or lateral gene transfer (i.e., xenologous).

An alternative approach is to use an operational definition of functionally equivalent genes, which does not imply orthology.

Received: July 26, 2013
Revised: January 11, 2014
Published: January 13, 2014

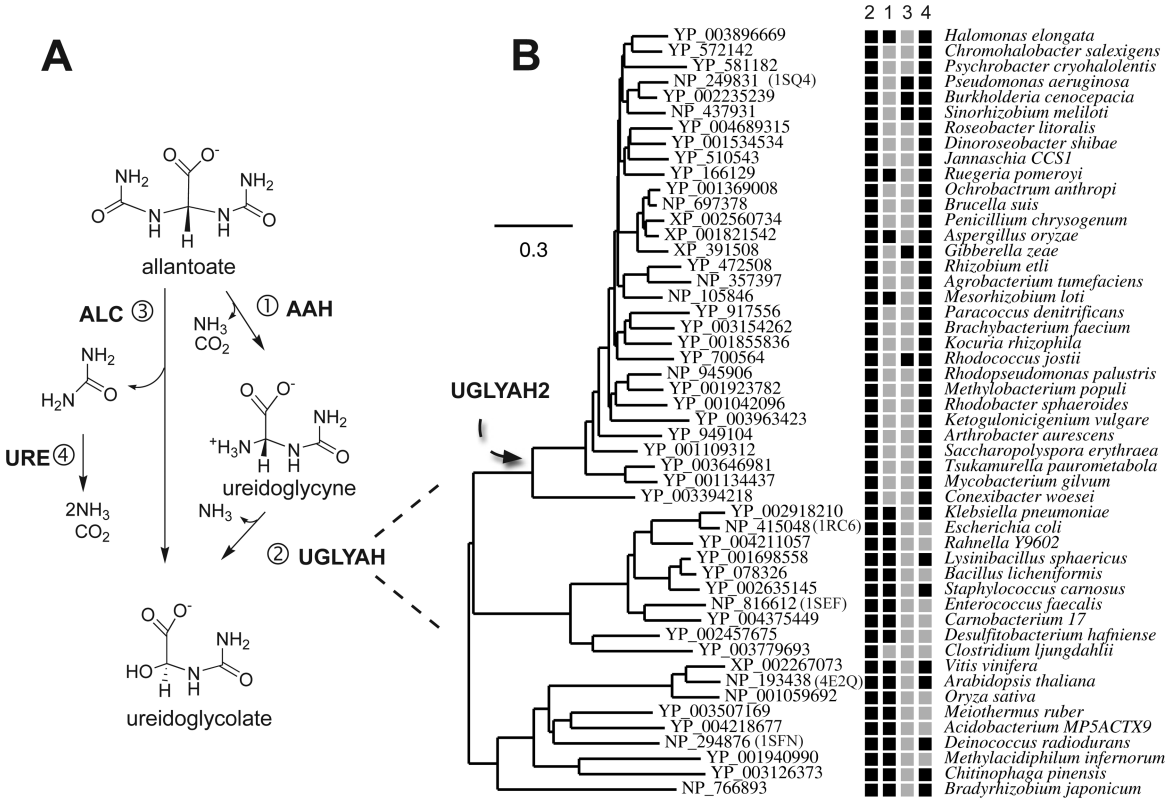


Figure 1. Metabolic network and gene co-occurrence of ureidoglycine hydrolase (UGLYAH). (A) Scheme of the reactions involving nitrogen mobilization from allantoate. (B) Distribution of the genes involved in the metabolic pathway mapped on the midpoint-rooted phylogenetic tree of UGLYAH proteins from complete genomes. Proteins with experimentally determined structures are indicated by PDB codes (in parentheses) next to the GenBank accession numbers. For each genome encoding UGLYAH genes, the presence (black square) or the absence (gray square) of the genes assigned to the metabolic pathway is indicated; genes are designated by numbers corresponding to reaction steps, as indicated in the metabolic scheme.

This approach is implemented, for example, by the protein families database TIGRFAMs, in which the term *equivalog* describes a relationship of conserved function among homologues,[16] as well as in specialized databases.[17,18] A family of functionally equivalent genes comprises one or more members with experimentally characterized functions and uncharacterized homologues that are thought to have the same molecular function. Different types of bioinformatic evidence can be used as inclusion or exclusion criteria. In prokaryotes, where gene clustering of functionally related genes is quite common, the conservation of the genomic context, i.e., the physical association with genes involved in the same process of pathway, provides additional evidence for the inclusion of homologous genes within a family of equivalogues.[19] On the contrary, the occurrence of homologous genes in a different genomic context can suggest exclusion of the gene from a family of equivalogues. For instance, proteins homologous to the pyrrolo-quinoline-quinone biosynthesis protein C are excluded from the TIGR02111 family and assigned to a different family (TIGR04305) on the grounds that the genomic context suggests a different function in folate biosynthesis.

The genes involved in the oxidative degradation of purines or their nitrogen-rich derivatives called ureides (allantoin and allantoate) are typically clustered in bacterial genomes and arguably represent the best understood case of evolution of a metabolic gene cluster in a model eukaryote (*Saccharomyces cerevisiae*).[20] As a primary metabolic pathway, the degradation of purine bases is characterized by an extreme variability across different organisms. The variability reflects the different purposes of this metabolic process in different species. In some organisms, typically in metazoa, the pathway is used to eliminate excess nucleobase and excess nitrogen, whereas in other organisms, as in plants, some bacteria, and fungi, the pathway enables the recovery of vital nitrogen from the nucleobases.[21,22] In the case of the latter, purines or purine derivatives subjected to degradation can derive from the endogenous metabolism, but they are also taken up from the environment. Depending on the preferred means of elimination or on the type of imported purine derivative, discrete parts of the pathway are found in certain organisms, while they are absent in others. Consequently, genes involved in consecutive steps of the pathway often exhibit a strong genomic association, that is, they either occur together, whether clustered or not, or are missed together in a given genome. The analyses of purine degradation clusters (in bacteria) and of gene co-occurrence (in both bacteria and eukaryotes) have enabled in the last few years the identification of several missing genes in purine catabolism.[23,24] For the presence of analogous enzymes and pathway branching points, an improved method of gene co-occurrence analysis[25] has been shown to have more power than the traditional analysis of "phylogenetic profiles"[26] for the prediction of protein activities in purine catabolism.[27,28]

Here, we show that the analysis of gene co-occurrence, in the frame of the phylogenetic tree of a protein family, can identify subtle functional divergence between homologous proteins involved in different reactions of the same metabolic pathway.

dx.doi.org/10.1021/bi4010107 | *Biochemistry* 2014, 53, 735−745

In particular, this analysis allowed us to identify a different activity for a group of bacterial and fungal proteins homologous to the (*S*)-ureidoglycine aminohydrolase (UGLYAH), an enzyme involved in purine degradation recently identified in *E. coli* and *A. thaliana*.[24,29] The reaction catalyzed by UGLYAH is part of a metabolic pathway that allows plants and bacteria to obtain mineral nitrogen from ureides. The release of ammonia from allantoate, catalyzed by allantoate amidohydrolase (AAH), produces (*S*)-ureidoglycine, which is acted upon by UGLYAH for the release of a second mole of ammonia and the formation of ureidoglycolate (see Figure 1). The genes encoding UGLYAH-like protein are typically found in purine degradation clusters in bacteria, and according to various criteria, they could be considered functionally equivalent to UGLYAH genes. They are, for instance, included in the same equivalogue family (TIGR03214) by TIGRFAMs. However, UGLYAH-like genes can be clearly distinguished from true UGLYAHs based on their different association profiles with other genes of the pathway. The analysis of the genomic distribution of functionally related genes allowed us to formulate defined predictions on the activity of the encoded protein, which has been verified through targeted experiments. We present biochemical and structural evidence that UGLYAH-like proteins, although able to release ammonia from (*S*)-ureidoglycine, act in the biochemical pathway in the release of urea from allantoate, thus catalyzing the formation of ureidoglycolate in a single reaction step.

MATERIALS AND METHODS

Gene Co-occurrence Analysis. Protein sequence collections from complete genomes containing putative UGLYAH genes, as determined by homology searches in the RefSeq database, were downloaded from the NCBI database. The set was filtered by excluding all but one species per genus, and the resulting collection of 50 complete proteomes was used to identify functionally related genes according to the metabolic scheme reported in Figure 1. To minimize false identification (i.e., homologous proteins with different function), we used a dedicated procedure based on homology searches with a reference set of proteins *bona fide* involved in the pathway ("in-pathway" set) and a reference set of homologous proteins *bona fide* not involved in the pathway ("out-pathway" set). A gene was considered to be present if the best score found by Blastp searches with the in-pathway set was significant ($E < 10^{-6}$) and higher than the best out-pathway score. Accession numbers of the proteins used in the reference sets (Table S1, Supporting Information) and the proteins identified in complete genomes (Table S2, Supporting Information) are reported in the Supporting Information. Gene presence/absence data were mapped on the UGLYAH phylogenetic tree using a Perl procedure and the Scripttree program (http://lamarck.lirmm.fr/scriptree). The statistical significance of the associations of different phylogenetic groups of *UGLYAH* genes with *AAH* genes was calculated by applying Fisher's Exact test to the 2 × 2 contingency tables obtained with the logic expressions UGLYAH ↔ AAH, UGLYAH2 ↔ !AAH.

Sequence and Structure Analysis. The classification of the *A. tumefaciens* UGLYAH2 protein with respect to experimentally determined ureidoglycine aminohydrolase was inspected in the databases TIGRFAMs (http://www.jcvi.org/cgi-bin/tigrfams/index.cgi), Roundup (http://roundup.hms.harvard.edu), PhylomeDB (http://phylomedb.org), OrthoMCL (http://orthomcl.org), and OrthologueDB (http://www.pathogenomics.sfu.ca/ortholugedb). The analysis of genetic clusters containing UGLYAH genes was conducted using the Microbesonline Web server (http://microbesonline.org). Sequence alignments were carried out with Clustalw[30] with manual adjustments suggested by structural superimpositions. The phylogenetic tree of UGLYAH proteins was constructed with the neighbor-joining algorithm[31] of ClustalW applying the Kimura correction for multiple substitutions. The tree was visualized and rooted using the FigTree program (http://tree.bio.ed.ac.uk/software/figtree). Sequence similarity plots were constructed with the Plotcon program of the Emboss package, as implemented by the Mobyle@Pasteur server (http://mobyle.pasteur.fr). Structures were analyzed with the Pymol software (http://www.pymol.org), and structural superimpositions were carried out with the Fatcat server (http://fatcat.burnham.org).

Materials. All reagents were from Sigma unless specified. Potassium allantoate was obtained through basic hydrolysis of allantoin.[32] Briefly, 2.8 g of commercial allantoin was dissolved into 20 mL of a solution 1 M KOH; the solution was stirred for 30 min at 75 °C, cooled in ice, and allowed to precipitate overnight at 4 °C by adding 200 mL of 95% ethanol. The precipitate was washed with a small amount of warm water and recrystallized overnight at 4 °C with 95% ethanol. The precipitate was dried at 100 °C and stored at room temperature for later use. *E. coli* allantoate amidohydrolase and *A. thaliana* ureidoglycine aminohydrolase were obtained by recombinant expression in *E. coli* as previously described.[29]

Gene Cloning and Protein Purification. To clone the UGLYAH2 coding sequence, DNA was prepared from *Agrobacterium tumefaciens* GV3101 (a laboratory strain of the C58 background) using described procedures.[33] As the gene in the sequence database (Atu3205) corresponded to a protein apparently shortened at the N terminus while an in-frame ATG codon located 30 bp upstream produced a protein with N-terminal sequence similar to other UGLYAH2s, the upstream ATG was selected as the starting codon. Amplicons encoding the full-length protein (274 aa) were obtained by PCR amplification of genomic DNA with primers forward 5′-ATGGCTGAAATGAAGAGATATTATTC and reverse 5′-TTACCAGAGCTTCACGTGGCG-3′. Amplicons were cloned directly into the expression vector pET28-SnaBI (Bolchi, A., unpublished work) using a one-step cloning procedure.[34] The plasmid was then transformed into BL21 (DE3)-RIL *E. coli* cells, and the insert was verified through sequencing. Transformed cells were grown at 37 °C in LB medium. Gene expression was induced at an optical density of 0.6 using 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG); after 3 h at 30 °C, the cells were resuspended in 1/10 volumes of the initial grown medium with sonication buffer (50 mM sodium phosphate, 0.3 M NaCl, 10% glycerol, and 1 mg/mL lysozyme, pH 7.6) and incubated on ice for 30 min. Cells were lysed by ten 15-s bursts of sonication. The protein was incubated with Talon resin (Clontech) for 1 h and purified by affinity chromatography as assessed by SDS−PAGE analysis. Loaded columns were washed with buffer (50 mM sodium phosphate, 0.3 M NaCl, 10% glycerol, and 5 mM imidazole), and protein was eluted with 100 mM imidazole. The protein was stored in elution buffer at −20 °C. For the CD experiment, the protein was dialyzed with 50 mM sodium phosphate and 0.3 M NaCl buffer at pH 7.6 and stored at 4 °C.

Site-Directed Mutagenesis. The site-directed mutants V196 M and L242M were prepared by PCR using the plasmid

pET28b-SnaBI UGLYAH2 as template, a high-fidelity thermostable DNA polymerase (Pfu Ultra II Fusion HS DNA polymerase; Stratagene, La Jolla, CA, USA), and mutagenic primers complementary to opposite strands (FW V196M, 5′-cgatatgcatctcaacatcAtGaccttcgaaccggg-3′, and Rev V196M, 5′-cccggttcgaaggtCaTgatgttgagatgcatatcg-3′; Fw L242M, 5′-ggtgatttcatgtggAtgcgtgcctattgcc-3′, and Rev L242M, 5′-ggcaataggcacgcaTccacatgaaatcacc-3′). For each mutation, the product of reaction was treated with DpnI (New England Biolabs, Beverly, MA, USA) to digest the parental DNA template. This procedure allowed us to select the newly synthesized and potentially mutated plasmids. The products of each digestion were used to transform *E. coli* XL1 Blue cells. Single clones were then sequenced to confirm the occurrence of the desired mutation.

Biochemical Assays. The allantoate urea release activity of UGLYAH2 or UGLYAH was monitored spectrophotometrically with a continuous assay coupled with glutamate dehydrogenase (EDH). The typical incubation mixture (1 mL) contained 0.1 M KP buffer at pH 7.6, 0.28 mM NADH, 2.5 mM α-ketoglutarate, 19.36 U of bovine liver EDH, 0.085 mM allantoic acid, and 3 U of jack bean urease type c-3. The reaction was initiated by the addition of UGLYAH or UGLYAH2 (6 μg), and the decrease in absorbance at 340 nm due to the oxidation of NADH was recorded. Ammonia release from allantoate was evaluated by excluding urease from the reaction mixture. The ureidoglycine activity was monitored using the same reaction mixture at pH 8.5 and with the addition of recombinant AAH (34 μg) to generate *in situ* the unstable substrate; UGLYAH2 or UGLYAH (8.4 μg) was added after the first phase of the AAH reaction.

The effect on the UGLYAH2 activity of different chelating compounds was evaluated by incubating the enzyme overnight with 1−10 mM CDTA (*trans*-1,2-diaminocyclohexane-*N*,*N*,*N*′,*N*′-tetraacetic acid monohydrate), DTPA (diethylenetriaminepentaacetic acid), EDTA (ethylenediaminetetraacetic acid), or HQSA (8-hydroxyquinoline-5-sulfonic acid). The rescue of the activity by metal ions was tested after 30 min of incubation of the enzyme treated with 1 mM HQSA in a 5 mM solution of different metal ions ($Zn^{2+,}$ Co^{2+}, Cu^{2+}, Ca^{2+}, $Ni2^{+}$, Mn^{2+}, and Mg^{2+}).

The formation of optically active compounds was monitored by circular dichroism (CD) analysis carried out in a 10 mm path length cuvette with a JascoJ-715 spectropolarimeter. The formation of (*S*)-ureidoglycolate was observed after 60 min of incubation of allantoate (2.8 mM) with dialyzed UGLYAH2 (24 μg) in 1 mL of 20 mM KP at pH 7.6. The enzyme was removed by ultrafiltration using a vivaspin column with a 3 KDa cutoff, and the low-molecular mass fraction was diluted 1:2 with the reaction buffer for CD analysis. The formation of (*R*)-ureidoglycolate was observed after 60 min of incubation of glyoxylate (1.25 M) and urea (1.25 M) with dialyzed UGLYAH2 or UGLYAH (60 mg). The ultrafiltrate obtained using a vivaspin column with a 10 KDa cutoff was diluted 1:500 for CD analysis.

To monitor the UGLYAH2 reaction through NMR spectroscopy, 0.6 mL of a solution of 90% D_2O/10% H_2O containing 50 mM KP and 20 mM allantoate was transferred to a 5-mm NMR tube, and the 1H NMR spectra was recorded in the absence of the enzyme; the solution was mixed with the enzyme (1 μM) in a small plastic tube, gently stirred for 1 min, and retransferred to the NMR tube to collect spectra at different times. 1H NMR spectra were recorded at 25 °C with a 600 MHz Varian Inova instrument equipped with a triple resonance probe. Double PFG spin echo (DPFGS) sequence was used for water suppression with 64 scans, a sweep window of 5000 Hz, and 16k points. The spectra were processed and analyzed with MestReNova. 8.1 software.

RESULTS AND DISCUSSION

Gene Context Analysis Suggests Functional Divergence among *UGLYAH* Genes. Genes encoding (*S*)-ureidoglycine aminohydrolase (UGLYAH) are often found in gene clusters together with genes encoding allantoate amidohydrolase (AAH). This association, which is found, for example, in *Escherichia coli* and related enterobacteria, has been key to the identification[24,29] of the *E. coli* and *A. thaliana* proteins as the enzymes catalyzing the hydrolysis of (*S*)-ureidoglycine, the product of the AAH reaction, into (*S*)-ureidoglycolate (Figure 1A).

However, an association with *AAH* genes is not observed for all *UGLYAH* genes.[24] In fact, in a number of bacteria *bona fide* UGLYAH-encoding genes are found in purine degradation clusters lacking AAH-encoding genes (Figure S1, Supporting Information); in most of these cases, homology searches fail to identify AAH-encoding genes at other genomic locations. Similarly, no candidate *AAH* genes are found in the fungal genomes possessing putative *UGLYAH* genes. These observations call into question the catalytic activity of the corresponding gene products, given that (*S*)-ureidoglycine is an unstable compound that should only be produced by the AAH reaction.

Homologous UGLYAH proteins that are found in species lacking AAH genes do not possess additional domains and do not have particularly divergent sequences with respect to affirmed UGLYAHs. The experimentally validated *E. coli* and *A. thaliana* UGLYAHs share a 36% sequence similarity. The putative UGLYAH protein from *A. tumefaciens* (a species lacking AAH genes) shares 47% and 37% sequence similarity with the *A. thaliana* and *E. coli* proteins, respectively. Moreover, in most databases, UGLYAH proteins from species lacking AAH are enclosed in the same orthologous group with affirmed UGLYAHs. For instance, those proteins are included in the same group as UGLYAH by TIGRFAMs,[35] Roundup,[36] PylomeDB,[37] and OrthoMCL,[38] while they are not considered orthologous for OrthologeDB.[39]

On the basis of phylogenetic analysis (Figure 1B), UGLYAH proteins can be divided into three groups that do not reflect species relationships: two groups comprise the proteins that have been characterized as (*S*)-ureidoglycine aminohydrolase in plants and bacteria, while the third group (henceforth UGLYAH2) comprises all uncharacterized proteins from fungi and other bacteria. Interestingly, this phylogenetic subdivision corresponds to a different co-occurrence relationship[25] with AAH genes. The presence of a UGLYAH gene implies the presence of AAH in the genome, whereas the presence of a UGLYAH2 gene implies the absence of AAH in the genome ($P < 10^{-8}$). However, the presence of either UGLYAH or UGLYAH genes implies the absence of ALC, the enzyme catalyzing the direct conversion of allantoate into (*S*)-ureidoglycolate through the release of urea (see Figure 1A). This suggests that in the biochemical pathway the activity of ALC can be surrogated either by the combination of the AAH and UGLYAH proteins or by the UGLYAH2 protein alone.

Ureidoglycine Aminohydrolase and Allantoicase Activity of UGLYAH2 Protein. The gene co-occurrence

analysis described above provides evidence that proteins belonging to the UGLYAH2 group are involved in a different reaction of the same metabolic pathway (purine degradation). To assign a function to this group of proteins, we selected the gene from *Agrobacterium tumefaciens* (gene ID Atu3205) based on the availability of its genomic DNA in our laboratory. Histidine-tagged recombinant UGLYAH2 was cloned from *A. tumefaciens*, and the corresponding protein was overproduced in *E. coli* and purified by affinity chromatography to near homogeneity as determined by gel electrophoresis and mass spectrometry (Figure S2, Supporting Information).

We first tested the activity of the protein on (*S*)-ureidoglycine, the enzymatic product of AAH (Figure S3, Supporting Information). The unstable (*S*)-ureidoglycine substrate was generated *in situ* by the AAH reaction,[29] as monitored by the release of ammonia from the allantoate molecule in a coupled assay with glutamate dehydrogenase. In the absence of other enzymatic activities, (*S*)-ureidoglycine spontaneously releases ammonia and urea[28] at a rate constant of 4×10^{-4} s^{-1}. The addition of UGLYAH2 caused the rapid release of a second mole of ammonia in a manner similar to that observed for *A. thaliana* UGLYAH (Figure S3a, Supporting Information). Next, UGLYAH2 activity was tested directly on allantoate, a substrate that is not attacked by UGLYAH proteins. No ammonia release was observed under the standard assay conditions (not shown). However, the rapid formation of 2 mol of ammonia was observed in the presence of urease in the reaction mixture. Conversely, UGLYAH was confirmed to be unable to release ammonia or urea from allantoate (Figure S3b, Supporting Information).

From these experiments and the comparison of the kinetic constants for the two substrates (Table 1 and Figure S4, Supporting Information), we concluded that the UGLYAH2 protein is able to catalyze both the hydrolysis of the amino group of (*S*)-ureidoglycine and, at variance with UGLYAH, the hydrolysis of the ureido group of allantoate. This latter enzymatic activity, known as allantoicase activity (EC 3.5.3.4), has been described in enzymes of various sources[40] and later experimentally assigned to the product of the *ALC* gene in fungi and metazoa.[41−43] *Bona fide ALC* genes can be identified by homology in purine degradation clusters of various bacteria, such as *Pseudomonas*, *Vibrio*, and *Burkholderia*. In spite of the similarity in their catalytic activity, UGLYAH2 and ALC do not have a significant sequence or structure similarity.

Table 1. Kinetic Constants Obtained from Spectrophotometric Coupled Enzyme Assays

	allantoate[a]		ureidoglycine[b]
enzyme	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	k_{cat} (s^{-1})
UGLYAH	<0.004	<3	≥3
UGLYAH2	2.4	4.9×10^3	≥0.6
V196M	0.5	6.1×10^2	≥0.15
L242M	1.6	1.4×10^3	≥0.6

[a]The assay mixtures (0.2 mL and 100 mM KP, pH 7.6) contained allantoate (0.04 to 1.7 mM), 11.5 U urease from *Canavalia ensiformis*, 2.5 mM α-ketoglutarate, 4 U glutamate dehydrogenase from bovine liver, and 0.24 mM NADH. The decrease in absorbance at 340 nm was quantitated after the addition of the enzymes (1 to 20 μg) preincubated for 30′ with 2.5 mM $MnCl_2$. [b]The assay mixtures (0.2 mL and 100 mM KP, pH 8.5) contained allantoate (0.04 to 0.17 mM), 6 μg of recombinant AAH from *E. coli*, 2.5 mM α-ketoglutarate, 4 U glutamate dehydrogenase from bovine liver, and 0.24 to 0.35 mM NADH. The formation *in situ* of ureidoglycine was monitored by following the change in absorbance at 340 nm. The decrease in absorbance at 340 nm was then quantitated after the addition of the enzymes (1 to 3 μg) preincubated for 30′ with 2.5 mM $MnCl_2$. Because of the instability of the ureidoglycine substrate, only the lower limits for the k_{cat} values were determined.

Metal Dependence and Stereospecificity of UGLYAH2. The (*S*)-ureidoglycine aminohydrolase activity of the bacterial and fungal UGLYAH has been previously reported to be dependent on Mn^{2+}.[29] Purified UGLYAH2 was catalytically active both with (*S*)-ureidoglycine and allantoate substrates in the absence of added metals. However, the enzyme was found to be sensitive to metal chelators. When tested with allantoate as a substrate, the activity was only partially reduced by incubation with CDTA, DTPA, and EDTA (data not shown) and completely abolished by incubation with HQSA. The activity could be completely restored with the addition of Mn^{2+}, while a higher activity (120%) was observed with Co^{2+}; the enzyme was partially restored by the addition of Ni^{2+}, while the other divalent metals examined (Zn^{2+}, Cu^{2+}, Ca^{2+}, and Mg^{2+}) were ineffective (Figure 2A).

The expected product of the allantoicase and ureidoglycine aminohydrolase reactions is the S enantiomer of ureidoglycolate. In addition, allantoicase from various sources has been reported to have the peculiar ability to attack the opposite enantiomer of their reaction product, (*R*)-ureidoglycolate, to form glyoxylate and urea.[40] By analyzing through CD spectroscopy the product of the UGLYAH2 reaction with allantoate as substrate (Figure 2B), we observed an optically active compound with a spectrum previously attributed to (*S*)-ureidoglycolate.[44] To test the enzyme activity on (*R*)-ureidoglycolate, we exploited the reversibility of the reaction[21] and analyzed the product obtained with glyoxylate and urea as substrates. In the presence of UGLYAH2, we observed an optically active compound with a CD spectrum corresponding to the mirror image of the (*S*)-ureidoglycolate spectrum (Figure 2C), thus demonstrating that the ability to degrade (*R*)-ureidoglycolate is a property of UGLYAH2 enzymes.

The capacity of UGLYAH to catalyze the hydrolysis of the R enantiomer of ureidoglycolate opens up the possibility that (*S*)-ureidoglycolate is formed not as the primary reaction product but through the enzymatic formation of racemic ureidoglycolate followed by the enzymatic hydrolysis of the R enantiomer. However, by monitoring the reaction with 1H NMR, the rapid decay of the allantoate proton signal at 5.27 ppm is observed with the simultaneous formation of a novel signal at 5.25 ppm that we assigned to ureidoglycolate; the formation of glyoxylate, revealed by a a peak at 5.09 ppm,[45] is observed only after the completion of the enzymatic reaction (Figure S5, Supporting Information) and can be ascribed to the spontaneous decay of ureidoglycolate.[44]

Comparison of UGLYAH and UGLYAH2 Structures. Among the UGLYAH family members with known structures, three microbial proteins from *Escherichia coli* (1RC6), *Deinococcus radiodurans* (1SFN), and *Enterococcus faecalis* (1SEF), and the eukaryotic protein from *Arabidopsis thaliana* (4E2Q) belong to the UGLYAH group, whereas the protein from *Pseudomonas aeruginosa* (1SQ4) belongs to the UGLYAH2 group (see Figure 1). Because of the high sequence similarity (64% identity) and the conservation of critical residues, the *P. aeruginosa* protein is a *bona fide* isofunctional homologue of the *A. tumefaciens* protein characterized here. Similar to that in *Agrobacterium*, the corresponding gene in

dx.doi.org/10.1021/bi4010107 | *Biochemistry* 2014, 53, 735−745

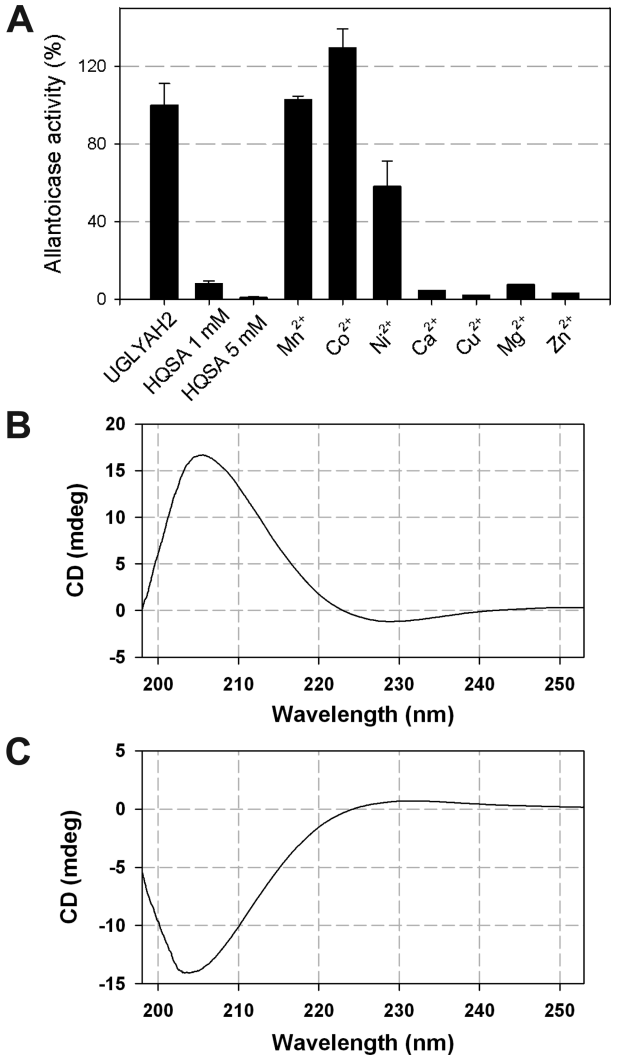


Figure 2. Metal dependence and stereospecificity of UGLYAH2. (A) Relative activity of UGLYAH2 in the presence of the HSQA metal chelator and upon incubation with various metals. The values are the mean and standard deviations of three independent experiments. (B) CD spectrum of the reaction product in the presence of allantoate; (C) CD spectrum of the reaction in the presence of glyoxylate and urea.

Pseudomonas (see Figure S1, Supporting Information) is not located in the main purine degradation cluster, which contains a typical allantoicase gene, suggesting that the *UGLYAH* locus may be involved in the utilization of a purine-derived metabolite (i.e., allantoate or ureidoglycolate) taken up from the environment.

The three-dimensional structure of microbial UGLYAH proteins has been known from structural genomics before the determination of the function of the UGLYAH family. Proteins belonging to this family are made up of two structurally similar cupin domains located at the N and C termini. Analysis of sequence conservation suggested that only the C-terminal cupin domain is endowed with metal-binding and catalytic activity.[29] This has recently been confirmed with the determination of the structure of the protein from *A. thaliana*, both in ligand-free and substrate-bound forms, allowing the precise localization of the metal- and substrate-binding site within the C-domain.[46] Sequence and structure comparison (Figure 3) show that UGLYAH2 proteins are similarly organized in two cupin domains, with only the C-domains showing a conservation profile compatible with metal binding and catalytic activity (Figure 3A and Figure S6, Supporting Information). As in UGLYAH proteins, the residue conservation in the multiple alignments of UGLYAH2 proteins is stronger in the C-domain than in the N-domain; however, for both domains the average sequence similarity is higher in UGLYAH2 proteins than in UGLYAH proteins (Figure S7, Supporting Information).

The X-ray structure of the *P. aeruginosa* protein has been solved at 2.7 Å resolution by the Northeast Structural Genomics Consortium[47] and deposited in the PDB in 2004 as a protein of unknown function. In both the author-provided and software-determined[48] biological assemblies, 1SQ4 forms an octameric structure consisting of two layers of tetramers (Figure 3B). The four monomers within a layer are related by a 4-fold symmetry axis running through the center of the octamer. The same quaternary organization has been reported for *A. thaliana* UGLYAH based on size-exclusion chromatography and crystallographic evidence.[46]

The structure of the 1SQ4 monomer from *P. aeruginosa* can be superimposed to the 4E2Q monomer from *A. thaliana* with an RMSD of 2.6 Å over 195 residues (Figure S8a, Supporting Information). The residues involved in metal coordination in 4E2Q are strictly conserved in 1SQ4 and in related proteins (see Figure 3A). However, the catalytic loop located between stands β15 and β16 and the metal-binding residues E208 and H210 are slightly displaced in 1SQ4 (Figure 3C), possibly due to the presence of a thiocyanate ion[49] deriving from the crystallization buffer bound at the active site (Figure S8b, Supporting Information).

Structural Rationale of the Different Substrate Specificities of UGLYAH and UGLYAH2. Proteins belonging to the UGLYAH and UGLYAH2 groups share the ability to catalyze ureidoglycine hydrolysis with the release of ammonia, but only UGLYAH2 proteins are able to catalyze the hydrolysis of allantoate with the release of urea. It is noticeable that the reactions are mechanistically similar and could be mediated by the same catalytic mechanism (Figure S9, Supporting Information) previously proposed for ureidoglycine aminohydrolase.[46] In keeping with this observation, residues involved in catalysis are invariants in the two groups (see Figure 3A). The different activities of the proteins could be explained by different affinities for the substrates. In particular, the active site of UGLYAH2 proteins could accommodate the two carbamoyl groups of allantoate as well as the single carbamoyl group of ureidoglycine (Figure S9a,b, Supporting Information), whereas the UGLYAH active site could only accommodate the smaller ureidoglycine molecule.

Sequence and structure comparisons provide insight on the molecular basis of this different specificity of the two proteins. Mapping amino acid substitutions at the active site of the two types of proteins highlights two positions in which conserved residues in UGLYAH2 correspond to bulkier residues in UGLYAH (see Figure 3A). In the *A. tumefaciens* protein, these residues are valine 196, which is substituted by leucine or methionine in UGLYAHs, and leucine 242, which is substituted by methionine in UGLYAHs. The examination of the *A. thaliana* UGLYAH structure in complex with the substrate (Figure 3C) suggests a possible role of the residue at position 196 in determining substrate specificity. Under the assumption that allantoate and ureidoglycine have the same binding mode, a bulk residue at position 196 would cause steric hindrance with

dx.doi.org/10.1021/bi4010107 | *Biochemistry* 2014, 53, 735−745

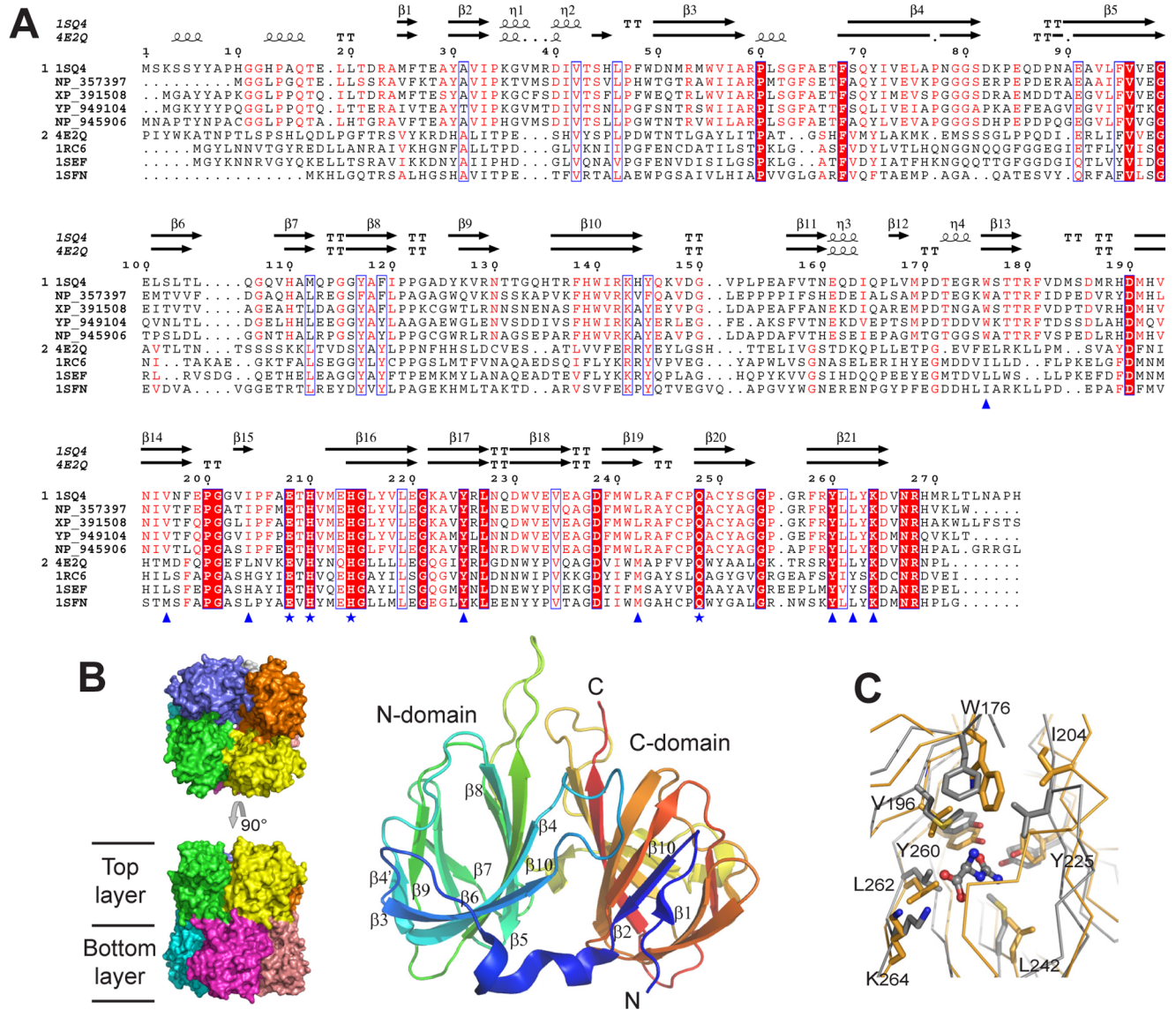


Figure 3. Sequence and structure comparison of UGLYAH proteins. (A) Multiple alignment of UGLYAH2 (group 1) and UGLYAH (group 2) proteins. Residues that are conserved across groups are boxed; invariant residues are colored white on a red background. Residues that are conserved only within groups are colored red. Secondary structure elements derived from the X-ray coordinates of group 1 (1SQ4) and group 2 (4E2Q) proteins are drawn over the alignment. Blue asterisks and triangles indicate residues involved in the metal coordination shell and substrate binding, respectively, as observed in the binary complex of the *A. thaliana* protein.[46] (B) Surface representation of the 1SQ4 octamer shown in top view (upper) and side view (lower) alongside the cartoon representation of the 1SQ4 monomer in rainbow colors with β-strands numbered as in panel A. (C) Close up comparison of the active sites of 1SQ4 (orange carbons) and 4E2Q (gray carbons) showing the overall conservation of substrate-binding residues. Residues relevant for metal (black sphere) and ureidoglycine (yellow carbons) binding in the *A. thaliana* binary complex and equivalent residues in 1SQ4 are represented as sticks; numbering is according to the 1SQ4 sequence shown in panel A.

the pro-S carbamoyl group of allantoate. The binding of allantoate would instead be permitted by the small valine residue present at this position in UGLYAH2 proteins (see Figure 3A).

To test the hypothesis that the residues described above are determinants of the different substrate specificities, the residues found in UGLYAH2 proteins were replaced by site-directed mutagenesis with residues found at the same position in UGLYAH2 proteins, generating the artificial mutants V196M and L242M (Figure S10, Supporting Information). The analysis of the steady-state kinetic constants (see Table1) showed that the catalytic efficiency for allantoate was diminished in the L242M mutant, resulting from a slightly lower k_{cat} and a higher K_m. Conversely, a significant difference in the lower limit of k_{cat} measured for ureidoglycine was not observed in the range of substrate concentration permitted by the assay conditions. The catalytic efficiency for allantoate was more severely diminished in the V196 M mutant. Unexpectedly, however, this mutation affected the k_{cat} value rather than the K_m value. A similar decrease in the k_{cat} value was observed for ureidoglycine, indicating that this single substitution in the absence of compensatory mutations, as they are observed at nearby positions in the comparison of UGLYAH sequences and structures (see Figure 3), is generally detrimental to the enzyme catalysis. Overall, these data suggest that more than a single mutational event may be necessary for the functional conversion of the enzymatic activities of UGLYAHs.

dx.doi.org/10.1021/bi4010107 | *Biochemistry* 2014, 53, 735−745

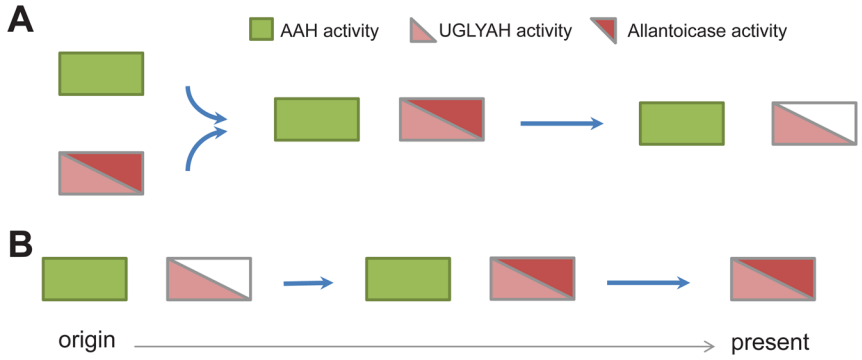


Figure 4. Alternative scenarios for the origin of the functional divergence and gene co-occurrence in UGLYAH and UGLYAH2 proteins. Boxes represent genes; colors represent enzymatic activities of the corresponding proteins. (A) Gene-gain function-loss hypothesis. (B) Function-gain gene-loss hypothesis.

Evolutionary Origin of UGLYAH functional divergence. Two alternative hypotheses could explain the evolutionary origin of the functional divergence of UGLYAH and UGLYAH2 proteins (Figure 4). In the gene-gain function-loss scenario (Figure 4A), urea release from allantoate was the ancestral activity of the protein, which also possessed a nonphysiological secondary activity on ureidoglycine. The appearance of an allantoate amidohydrolase activity forming ureidoglycine, made the exploitation of the ureidoglycine aminohydrolase activity favorable, allowing for urease-independent recycling of the purine nitrogen. The loss of selective pressure for the maintenance of urea-release activity ultimately caused the loss by mutations of this activity in a particular branch of the protein family and the establishment of the more specialized UGLYAH proteins. In the function-gain gene-loss scenario (Figure 4B), ureidoglycine hydrolysis was the ancestral activity of the protein, which later gained an additional allantoate hydrolase activity. This activity rendered the presence of allantoate amidohydrolase dispensable and caused the disappearance of the corresponding gene in the genomes possessing UGLYAH2. It is noteworthy that both scenarios (1) do not require gene duplication (i.e., paralogy) and (2) have a common intermediate state in which two alternative branches of the same pathway coexist. The evolutionary resolution of this coexistence may depend on the availability of urease: in the presence of urease, either the urea- or the ammonia-release pathways are viable, while in the absence of urease, the ammonia-release pathway is clearly preferable. Consistent with this scenario, in the different genomes the presence of UGLYAH2 implies the presence of urease, whereas urease genes are often absent in genomes possessing UGLYAH genes (see Figure 1B).

Because experimental data demonstrates that the ureidoglycine hydrolase activity is an accessory activity that UGLYAH2 proteins possess in the absence of enzymes producing the ureidoglycine substrate, the gene-gain function-loss scenario and the hypothesis of a more ancient origin of UGLYAH2 proteins seems more plausible. However, this is contrasted by other bioinformatics evidence indicating that UGLYAH (rather than UGLYAH2) are basal in the phylogenetic tree (see Figure 1B) and that sequence diversity is on average higher in UGLYAH than in UGLYAH2 proteins (see Figure S7, Supporting Information).

CONCLUDING REMARKS

Gene context analysis has been used to generate testable hypotheses about the divergent function of a group of proteins (UGLYAH2), which would have been considered isofunctional to ureidoglycine amino hydrolase (UGLYAH) according to various types of evidence. In the absence of indications from this analysis, the recognition of a functional divergence between UGLYAH and UGLYAH2 would have been particularly challenging because the two proteins appear to be the same enzyme when assayed with ureidoglycine as substrate (see Figure S3a, Supporting Information). However, the assay on allantoate, as suggested by bioinformatic evidence and biochemical reasoning, clearly distinguished the enzymatic activity of the two proteins (see Table 1 and Figure S3b, Supporting Information). Although the identified gene relationships, the presence of UGLYAH implies the presence of AAH and the presence of UGHLYAH2 implies the absence of AAH, hold in most organisms, exceptions are also observed (e.g., *C. ljungdahlii* in Figure 1B). Such counterexamples are not unexpected in this kind of analysis as they can derive from errors in the procedure of gene identification and classification or from true biological exceptions.

As illustrated by the case of UGLYAH2 proteins, even when the genomic context (i.e., the physical associations) of an uncharacterized gene suggests a role in the same process or pathway of a characterized homologue, a detailed analysis of the genomic distribution of functionally related genes can provide evidence for functional divergence. The occurrence of homologous protein catalyzing different reactions in the same pathway is not a rare phenomenon. It has been estimated that about one-third of the functionally divergent paralogues share the same pathway, although taking different metabolic steps.[50] Examples relevant to purine degradation are the deamination of adenine and adenosine, carried out by paralogous genes in fungi and metazoa,[51] and the hydrolysis of the amidic group of ureidoglycolate and allantoate, two reactions of the plant ureide pathway that are carried out by the closely related AAH and UAH paralogs.[24]

The evolution of a different gene function is expected to be accompanied by the establishment of novel associations with other genes. Upon functional divergence, for example, a protein establishing a physical interaction in multiprotein complexes will modify its interaction network, or an enzyme that is getting involved in a different reaction will modify its functional associations in the metabolic network. We propose that when uncharacterized genes are compared to homologous genes with

dx.doi.org/10.1021/bi4010107 | *Biochemistry* 2014, 53, 735−745

experimentally characterized function, one should also compare, in the frame of the family pylogenetic tree, the genomic distribution of the genes known to be involved in the same cellular process or biochemical pathway as the characterized one (see Figure 1). This analysis could hint to the presence of functionally divergent branches of the gene family, and in particular cases, it could also provide specific suggestions about the different biological roles. More in general, by showing whether genes are expected to be functionally associated (or not associated) with a given gene are present or absent in the same genome, this analysis provides context-dependent information on the gene under study. The resulting evidence can integrate other context analyses, such as the examination of gene physical associations,[52−54] extending the comparison also to genomes that do not have an operon structure.

Evidence from genome analysis suggests that among the activities demonstrated *in vitro* by the UGLYAH2 protein, urea release from allantoate is the relevant physiological activity. This UGLYAH2 activity explains early observations on the ability to utilize purines as nitrogen sources of various bacterial and fungal species[21] whose genomic sequence did not revealed the presence of known genes involved in allantoate degradation (see also Figure 1B). In contrast, ammonia released from ureidoglycine should not have any role *in vivo*, given that the compound will not be present in organisms lacking AAH. Although UGLYAH2 proteins are bidomain proteins constituted by two cupin folds, there is sequence and structure evidence that only the carboxy-terminal domain is catalytically active. However, the association of allantoicase and ureidoglycine hydrolase domains would not have a functional meaning. As suggested by the analysis of the active site, the UGLYAH2 proteins can accommodate both allantoate and the smaller ureidoglycine molecule. Thus, the ureidoglycine hydrolase activity of UGLYAH2 can be considered a nonphysiological activity resulting from the configuration of the active site and the similarity of ureidoglycine with the enzyme natural substrate. However, in a plausible evolutionary scenario (see Figure 4A), ureidoglycine hydrolysis gained functional relevance in a group of organisms in which the presence of an activity forming ureidoglycine made the exploitation of this accessory activity advantageous.

■ ASSOCIATED CONTENT

Supporting Information

GenBank/PDB accession numbers of the reference proteins used in the reference set for sequence identification in complete genomes; accession numbers for UGLYAHs and functionally associated proteins identified in complete genomes; comparison of the genomic context of UGLYAH genes identified in bacteria lacking allantoate amidohydrolase; recombinant expression and purification of Atu3205; biochemical activities of UGLYAH and UGLYAH2 proteins; steady-state kinetic plots for UGLYAH reactions with allantoate; time-course of the UGLYAH2 reaction as monitored by 1H NMR spectroscopy; N- and C-terminal domains in UGLYAH2 proteins; sequence similarity plot of aligned UGLYAH proteins; stereoview comparison of UGLYAH and UGLYAH2 structures; similar reaction mechanisms for the UGLYAH2 reactions with ureidoglycine and allantoate substrates; and sequence evidence for the V196M and L242M mutations. This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

The annotated sequence of *A. tumefaciens* UGLYAH2 has been submitted to GenBank with the accession number KF312339.

■ AUTHOR INFORMATION

Corresponding Author

*Laboratory of Biochemistry, Molecular Biology, and Bioinformatics, Department of Life Sciences, Parco Area delle Scienze 23/a, University of Parma, Parma, Italy. Phone: +39 0521 905140. Fax: +39 0521 905151. E-mail: riccardo.percudani@unipr.it

Funding

This work was supported by the University of Parma.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Thelma Pertinhez for help with NMR measurements, Cecilia Nolli and Gianluca Paredi for technical assistance, and Claudio Scazzocchio for comments and suggestions on the manuscript.

■ REFERENCES

(1) Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M., and Yuan, Y. (1998) Predicting function: from genes to genomes and back. *J. Mol. Biol. 283*, 707−725.

(2) Gabaldon, T., and Huynen, M. A. (2004) Prediction of protein function and pathways in the genome era. *Cell. Mol. Life Sci. 61*, 930−944.

(3) Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crecy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Ruckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res. 33*, 5691−5702.

(4) Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Guldener, U., Mannhaupt, G., Munsterkotter, M., and Mewes, H. W. (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res. 32*, 5539−5545.

(5) Galperin, M. Y., and Koonin, E. V. (1998) Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operon disruption. *In Silico Biol. 1*, 55−67.

(6) Gerlt, J. A., and Babbitt, P. C. (2000) Can sequence determine function? *Genome Biol. 1*, REVIEWS0005.

(7) Punta, M., and Ofran, Y. (2008) The rough guide to in silico function prediction, or how to use sequence and structure information to predict protein function. *PLoS Comput. Biol. 4*, e1000160.

(8) Rost, B. (2002) Enzyme function less conserved than anticipated. *J. Mol. Biol. 318*, 595−608.

(9) Tian, W., and Skolnick, J. (2003) How well is enzyme function conserved as a function of pairwise sequence identity? *J. Mol. Biol. 333*, 863−882.

(10) Peterson, M. E., Chen, F., Saven, J. G., Roos, D. S., Babbitt, P. C., and Sali, A. (2009) Evolutionary constraints on structural similarity in orthologs and paralogs. *Protein Sci. 18*, 1306−1315.

(11) Page, R. D. (1998) GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics 14*, 819−820.

(12) Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D. (2012) Inferring duplications, losses, transfers and incomplete

lineage sorting with nonbinary species trees. *Bioinformatics 28*, i409–i415.
(13) Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011) Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinf. 12*, 124.
(14) Li, L., Stoeckert, C. J., Jr., and Roos, D. S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res. 13*, 2178–2189.
(15) Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families. *Science 278*, 631–637.
(16) Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T., and White, O. (2001) TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res. 29*, 41–43.
(17) Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res. 37*, D233–238.
(18) Percudani, R., and Peracchi, A. (2009) The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. *BMC Bioinf. 10*, 273.
(19) Selengut, J. D., Haft, D. H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W. C., Richter, A. R., and White, O. (2007) TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res. 35*, D260–264.
(20) Wong, S., and Wolfe, K. H. (2005) Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat. Genet. 37*, 777–782.
(21) Vogels, G. D., and Van der Drift, C. (1976) Degradation of purines and pyrimidines by microorganisms. *Bacteriol Rev 40*, 403–468.
(22) Werner, A. K., and Witte, C. P. (2011) The biochemistry of nitrogen mobilization: purine ring catabolism. *Trends Plant Sci. 16*, 381–387.
(23) 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.
(24) Werner, A. K., Romeis, T., and Witte, C. P. (2010) Ureide catabolism in Arabidopsis thaliana and Escherichia coli. *Nat. Chem. Biol. 6*, 19–21.
(25) Bowers, P. M., Cokus, S. J., Eisenberg, D., and Yeates, T. O. (2004) Use of logic relationships to decipher protein network organization. *Science 306*, 2246–2249.
(26) Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A. 96*, 4285–4288.
(27) 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.
(28) Ramazzina, I., Costa, R., Cendron, L., Berni, R., Peracchi, A., Zanotti, G., and Percudani, R. (2010) An aminotransferase branch point connects purine catabolism to amino acid recycling. *Nat. Chem. Biol. 6*, 801–806.
(29) Serventi, F., Ramazzina, I., Lamberto, I., Puggioni, V., Gatti, R., and Percudani, R. (2010) Chemical basis of nitrogen recovery through the ureide pathway: formation and hydrolysis of S-ureidoglycine in plants and bacteria. *ACS Chem. Biol. 5*, 203–214.
(30) 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.
(31) Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4*, 406–425.
(32) Young, E. G., and Conway, C. F. (1942) On the estimation of allantoin by the Rimini-Schryver reaction. *J. Biol. Chem. 142*, 839–853.
(33) Kalia, A., Rattan, A., and Chopra, P. (1999) A method for extraction of high-quality and high-quantity genomic DNA generally applicable to pathogenic bacteria. *Anal. Biochem. 275*, 1–5.
(34) 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.
(35) Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2013) TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res. 41*, D387–395.
(36) Deluca, T. F., Wu, I. H., Pu, J., Monaghan, T., Peshkin, L., Singh, S., and Wall, D. P. (2006) Roundup: a multi-genome repository of orthologs and evolutionary distances. *Bioinformatics 22*, 2044–2046.
(37) Huerta-Cepas, J., Capella-Gutierrez, S., Pryszcz, L. P., Denisov, I., Kormes, D., Marcet-Houben, M., and Gabaldon, T. (2011) PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic Acids Res. 39*, D556–560.
(38) Chen, F., Mackey, A. J., Stoeckert, C. J., Jr., and Roos, D. S. (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res. 34*, D363–D368.
(39) Whiteside, M. D., Winsor, G. L., Laird, M. R., and Brinkman, F. S. (2013) OrthologueDB: a bacterial and archaeal orthology resource for improved comparative genomic analysis. *Nucleic Acids Res. 41*, D366–D376.
(40) Trijbels, F., and Vogels, G. D. (1967) Allantoate and ureidoglycolate degradation by Pseudomonas aeruginosa. *Biochim. Biophys. Acta 132*, 115–126.
(41) Lee, H., Fu, Y. H., and Marzluf, G. A. (1990) Molecular cloning and characterization of alc the gene encoding allantoicase of Neurospora crassa. *Mol. Gen. Genet. 222*, 140–144.
(42) Vigetti, D., Monetti, C., Pollegioni, L., Taramelli, R., and Bernardini, G. (2000) Xenopus allantoicase: molecular cloning, enzymatic activity and developmental expression. *Arch. Biochem. Biophys. 379*, 90–96.
(43) Yoo, H. S., and Cooper, T. G. (1991) Sequences of two adjacent genes, one (DAL2) encoding allantoicase and another (DCG1) sensitive to nitrogen-catabolite repression in Saccharomyces cerevisiae. *Gene 104*, 55–62.
(44) 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.
(45) Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., Hau, D. D., Psychogios, N., Dong, E., Bouatra, S., Mandal, R., Sinelnikov, I., Xia, J., Jia, L., Cruz, J. A., Lim, E., Sobsey, C. A., Shrivastava, S., Huang, P., Liu, P., Fang, L., Peng, J., Fradette, R., Cheng, D., Tzur, D., Clements, M., Lewis, A., De Souza, A., Zuniga, A., Dawe, M., Xiong, Y., Clive, D., Greiner, R., Nazyrova, A., Shaykhutdinov, R., Li, L., Vogel, H. J., and Forsythe, I. (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res. 37*, D603–610.
(46) Shin, I., Percudani, R., and Rhee, S. (2012) Structural and functional insights into (S)-ureidoglycine aminohydrolase, key enzyme of purine catabolism in Arabidopsis thaliana. *J. Biol. Chem. 287*, 18796–18805.
(47) Wunderlich, Z., Acton, T. B., Liu, J., Kornhaber, G., Everett, J., Carter, P., Lan, N., Echols, N., Gerstein, M., Rost, B., and Montelione, G. T. (2004) The protein target list of the Northeast Structural Genomics Consortium. *Proteins 56*, 181–187.
(48) Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774–797.
(49) Tchertanov, L., and Pascard, C. (1997) Statistical Analysis of noncovalent interactions of anion groups in crystal structures. III. Metal complexes of thiocyanate and their hydrogen-donor accepting function. *Acta Crystallogr., Sect. B*, 904–915.
(50) Martinez-Nunez, M. A., Perez-Rueda, E., Gutierrez-Rios, R. M., and Merino, E. (2010) New insights into the regulatory networks of paralogous genes in bacteria. *Microbiology 156*, 14–22.

(51) Ribard, C., Rochet, M., Labedan, B., Daignan-Fornier, B., Alzari, P., Scazzocchio, C., and Oestreicher, N. (2003) Sub-families of alpha/beta barrel enzymes: a new adenine deaminase family. *J. Mol. Biol. 334*, 1117−1131.
(52) Galperin, M. Y., and Koonin, E. V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat. Biotechnol. 18*, 609−613.
(53) Kolesov, G., Mewes, H. W., and Frishman, D. (2001) SNAPping up functionally related genes based on context information: a colinearity-free approach. *J. Mol. Biol. 311*, 639−656.
(54) Rogozin, I. B., Makarova, K. S., Murvai, J., Czabarka, E., Wolf, Y. I., Tatusov, R. L., Szekely, L. A., and Koonin, E. V. (2002) Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res. 30*, 2212−2223.