JBC Papers in Press. Published on March 9, 2018 as Manuscript RA118.001996
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.RA118.001996

An unexpected vestigial protein complex reveals the evolutionary origins of an *s*-triazine catabolic enzyme

Lygie Esquirol*[1,2], Thomas S. Peat*[3], Matthew Wilding[2,3], Jian-Wei Liu[1], Nigel G. French[1], Carol J. Hartley[1], Hideki Onagi[2], Thomas Nebl[3], Christopher J. Easton[2], Janet Newman[3], Colin Scott[1,4]

From: [1] Biocatalysis and Synthetic Biology Team, CSIRO Land & Water, Canberra, ACT, Australia; [2] Research School of Chemistry, Australian National University, Canberra, ACT, Australia; [3] CSIRO Biomedical Manufacturing, Parkville, Melbourne, VIC, Australia; [4]Synthetic Biology Future Science Platform, CSIRO Land & Water, Canberra, ACT, Australia

*Authors contributed equally

To whom correspondence should be addressed: Dr. Colin Scott, CSIRO Land & Water, Foundation Building, Black Mountain Science and Innovation Park, Canberra, ACT 2601, Australia. colin.scott@csiro.au

Running Title: Evolution of AtzE, a 1-carboxybiuret hydrolase

Key words: AtzE; biuret hydrolase; atrazine; cyanuric acid

ABSTRACT

Cyanuric acid is a metabolic intermediate of *s*-triazines, such as atrazine (a common herbicide) and melamine (used in resins and plastics). Cyanuric acid is mineralized to ammonia and carbon dioxide by the soil bacterium *Pseudomonas* sp. strain ADP via three hydrolytic enzymes (AtzD, AtzE, and AtzF). Here, we report the purification and biochemical and structural characterization of AtzE. Contrary to previous reports, we found that AtzE is not a biuret amidohydrolase, but instead catalyzes the hydrolytic deamination of 1-carboxybiuret. X-ray crystal structures of apo AtzE and AtzE bound with the suicide inhibitor phenyl phosphorodiamidate revealed that the AtzE enzyme complex consists of two independent molecules in the asymmetric unit. We also show that AtzE forms an α2β2 heterotetramer with a hitherto unidentified 68-amino-acid-long protein (AtzG) encoded in the cyanuric acid mineralization operon from *Pseudomonas* sp. strain ADP. Moreover, we observed that AtzG is essential for the production of soluble, active AtzE and that this obligate interaction is a vestige of their shared evolutionary origin. We propose that AtzEG was likely recruited into the cyanuric acid–mineralizing pathway from an ancestral glutamine transamidosome that required protein–protein interactions to enforce the exclusion of solvent from the transamidation reaction.

INTRODUCTION

The presence of anthropogenic compounds in the environment provides selection pressures that can lead to the evolution of new metabolic pathways (1). One of the most well studied examples of an evolutionary response by bacterial to the presence of synthetic xenobiotics is that of the *s*-triazines (2). This family of compounds includes fertilizers (e.g., atrazine and ametryn), resins and plastics (e.g., melamine), explosives (e.g., RDX) and disinfectants (e.g. cyanuric acid). Interestingly, many of the *s*-triazine catabolic pathways that have evolved share a common structure, in which specific 'upper pathways' convert different *s*-triazines to cyanuric acid (1,3,5-triazine-2,4,6-triol) by hydrolosis of the three substituents that decorate the heterocyclic ring (3-5). A common 'lower pathway' then mineralizes cyanuric acid to carbon dioxide and ammonia.

The best characterized of the bacterial catabolic pathways that have evolved in response to *s*-triazines in the environment is the atrazine mineralization pathway of *Pseudomonas* sp. strain ADP. This soil bacterium was isolated in the mid-1990s by Wackett and co-workers (6). The atrazine catabolism pathway of *Pseudomonas* sp. strain ADP is comprised of six hydrolases encoded by genes on a self-transmissible plasmid (pADP1) (5). The upper pathway (Fig. 1A) is comprised of three hydrolytic steps: dechlorination, followed by two sequential *N*-alkyl chain hydrolyses. The upper pathway is catalyzed by AtzA (4,7), AtzB (8) and AtzC (9), respectively; all three enzymes are metalloenzymes of the amidohydrolase superfamily (9,10).

The lower pathway from *Pseudomonas* sp. strain ADP (Fig. 1B) is also comprised of three hydrolases: AtzD, AtzE and AtzF. Unlike the upper pathway, these three enzymes are serine hydrolases; AtzD is an unusual Ser-Lys hydrolase and is the archetype of a recently discovered protein fold (the Toblerone fold; (11,12)), while AtzE and AtzF are Ser-*cis*Ser-Lys hydrolases (1,13). The lower pathway progresses *via* the AtzD-mediated ring opening of cyanuric acid to form an unstable product, 1-carboxybiuret, which is known to spontaneously decarboxylate to biuret in an aqueous environment (1). It was therefore thought that the proceeding catabolic steps consisted of the deamination of biuret by AtzE to form allophanate, and allophanate deamination by AtzF to form carbamic acid. Interestingly, these enzymatic products are unstable in an aqueous environment and can undergo spontaneous reactions; allophanate decarboxylates to urea (14), whilst carbamic acid decomposes to ammonia and carbon dioxide in water (13,15). AtzD and AtzF have been overproduced in heterologous expression systems, which has allowed them to be studied biochemically (14-17) and structurally (11-13,18). However, expression and purification of AtzE in a heterologous system has been challenging (1) and structural and biochemical data had not been obtained previously.

The lower pathway has a degree of plasticity in its composition. AtzD can be replaced by homologs, such as TrzD from *Pseudomonas* sp. strain NRRLB-12227 (56 % identity to AtzD; (16,17)). AtzF is sometimes substituted by the homologous TrzF found in *Enterobacter cloacae* strain 99 (70 % identity to AtzF; (14)) and AtzE has been replaced with the non-homologous biuret amidohydrolase (BiuH, a $(\beta/\alpha)_8$ TIM barrel cysteine hydrolase) in *Rhizobium leguminasorum* bv. *viciae* 3841 (Table 1) (1,19).

Herein, we describe the purification and biochemical and structural characterization of AtzE directly from the model organism *Pseudomonas* sp. strain ADP. We show that AtzE forms a stable $\alpha_2\beta_2$ heterotetramer with a hitherto unknown small protein, AtzG, which is required for soluble expression of AtzE. Co-expression of AtzE and AtzG in *E. coli* allows, for the first time, the heterologous over-production of AtzE.

RESULTS

AtzE is a 1-carboxybiuret hydrolase - We attempted to produce active AtzE recombinantly in *E. coli*, but were unsuccessful (consistent with previous attempts; 1, 18). As heterologous expression was unsuccessful, we purified native AtzE from *Pseudomonas* sp. strain ADP. The expression of the *atzDEF* operon is induced under low nitrogen conditions in the presence of cyanuric acid (5), and so we cultured *Pseudomonas* sp. strain ADP on a minimal medium and supplied cyanuric acid as the sole nitrogen source. We were able to detect cyanuric acid amidohydrolase and allophanate amidohydrolase activities in the cell free extracts (CFE) of these cultures. Cyanuric acid added to the CFE was mineralized (i.e., 3 molecules of ammonia were formed for each molecule of cyanuric acid added). However, we were unable to detect any biuret hydrolysis.

We were able to detect and follow AtzE during CFE fractionation and purification by measuring ammonia production when supplementing fractions with purified AtzD and cyanuric acid. Although the cruder samples during fractionation had high background levels of ammonia, additional ammonia produced from cyanuric acid by AtzD and AtzE was distinguishable from this background. Using ammonia production as a proxy for enzyme activity, we were able to isolate 0.125 mg of pure AtzE per liter of *Pseudomonas* sp. strain ADP culture *via* ammonium sulfate precipitation and four chromatography steps (Fig. S1A, Suppl. Table 1).

The mass of the isolated protein was 48,119 Da by mass spectrometry, consistent with the predicted mass of AtzE, and the fragmentation pattern obtained by tryptic digest confirmed that the isolated protein was AtzE. Differential Scanning Fluorimetry revealed that the melting temperature (T_m) for AtzE isolated from *Pseudomonas* sp. strain ADP was greater than 60 °C. Previous reports had suggested that AtzE could not be produced in *E. coli* because of low stability (1); however, the high T_m of purified AtzE suggested that the poor expression of AtzE in *E. coli* was not due to protein instability.

Purified AtzE had no detectable biuret aminohydrolase activity under any condition tested. However, AtzE-mediated ammonia production was observed when incubated with cyanuric acid and pure AtzD. This suggested that the product of cyanuric acid hydrolysis by AtzD, 1-carboxybiuret, was in fact the natural substrate for AtzE. LCMS was subsequently used to follow the reactions directly. There was no detectable biuret amidohydrolase activity with AtzE (Fig. 2A), but it was detected using the biuret amidohydrolase from a *Rhizobium* species, BiuH (19) (Fig. 2B). Moreover, biuret accumulates in reactions in which cyanuric acid is treated with AtzD (Fig. 2C), but does not when AtzE is also included (Fig. 2D), suggesting that AtzE prevents biuret's formation *via* the hydrolytic deamination of 1-carboxybiuret.

Steady-state kinetic data for AtzD, AtzE and the biuret hydrolase were obtained (Table 1). The product inhibition of AtzD reported previously (11) was alleviated by the addition of AtzE, allowing the determination of rates at concentrations above 0.2 mM cyanuric acid for the first time. The kinetic parameters determined for AtzD were found to be: a k_{cat} of 17 s^{-1}, a K_M of 350 µM, and a k_{cat}/K_M of 4.8 x 10^4 $s^{-1}.M^{-1}$. The steady state kinetic parameters for AtzE were found to be similar to those of the *Rhizobium* biuret amidohydrolase for its physiological substrate, with a k_{cat} of 16 s^{-1}, a K_M of 63 µM, and a k_{cat}/K_M of 2.5 x 10^5 $s^{-1}.M^{-1}$ for AtzE, and a k_{cat} of 12 s^{-1}, a K_M of 80 µM, and a k_{cat}/K_M of 1.5 x 10^5 $s^{-1}.M^{-1}$ for BiuH (Table 1).

We also tested AtzE with a number of potential additional substrates, including the product of barbituric acid hydrolase (1-carboxymalonamide), analogs of AtzE's natural substrate (1-nitrobiuret and succinamic acid), structurally related α-amino acids (citrulline, lysine, glutamine, asparagine and 2-amino-3-propionic acid) and the biuret analog malonamide (Table 1, Suppl. Table 2). None of the α-amino acids were substrates for AtzE, nor was the biuret analog malonamide (Suppl. Table 2). However, 1-carboxymalonamide, 1-nitrobiuret, and succinamic acid were all substrates for AtzE with k_{cat}/K_M values of 4.2 x 10^4, 1.2 x 10^4 and 3.6 x 10^4 $s^{-1}.M^{-1}$, respectively (Table 1; Fig. 3).

LCMS was used to identify that the product of AtzE-mediated hydrolysis of 1-carboxymalonamide is 1-carboxymalonamic acid through hydrolysis of the terminal amine. Succinamic acid has a terminal amide rather than

an ureido group, suggesting that AtzE hydrolyzes the amide from this substrate in an equivalent reaction to that of the hydrolysis of 1-carboxymalonamide. We therefore propose an update to the cyanuric acid catabolic pathway, in which AtzE acts as a 1-carboxybiuret amidohydrolase, producing 1,3-dicaboxyurea and ammonia (Fig. 1B, Fig. 2D).

The structure of AtzE reveals the presence of an unexpected, but essential ancillary protein – X-ray crystal structures were obtained for purified AtzE (PDB ID: 6C62; Table 2) and purified AtzE treated with the suicide inhibitor phenyl phosphorodiamidate (PPDI; PDB ID: 6C6G; Table 2). The crystals of AtzE had two independent molecules of AtzE in the asymmetric unit, and the entire chain could be traced for both molecules (Fig. 4). Surprisingly, after positioning the AtzE chains in the structure, additional protein density (Fig. 4). The density was of high enough quality to allow for initial estimation of the unknown sequence, which was identical to the predicted translation product of a short (204 base pairs) unannotated intergenic region between *atzD* and *atzE* in the *atzDEF* operon on the pADP1 plasmid (Fig. 1B).

Examination of purified AtzE by SDS-PAGE indicated a small co-purified protein of approximately the same size as the protein that co-crystallized with AtzE (Fig. S1B). Mass spectrometry of the AtzE purified from *Pseudomonas* confirmed the presence of an unidentified protein having a mass of 7422 Da (Fig. S1C), and analysis of tryptic fragments of the 7.4 kDa protein confirmed that it was encoded by the small unannotated open reading frame located between the *atzD* and *atzE* genes (Fig. S2), which we have termed *atzG* (Fig S1). Along with the two molecules of AtzE, there are two molecules of AtzG in the asymmetric unit (Fig 4B).

Of the protein structures available in the Protein Data Bank (PDB)(20), AtzE is most similar to the glutamine deaminase component (chain A; 35% sequence identity) of 4WJ3, an glutamine transamidosome from *Pseudomonas* (21): 428 residues aligned (out of 457) with a 1.5 Å rmsd over the aligned backbone atoms (Fig. 5). The transamidosome is a three protein complex (GatCAB) that is essential for the production of correctly charged tRNA in bacteria *via* the transfer of an ammonia from gutamine to acidic amino acid-charged tRNA (21). Deamination occurs in the active site of an amidase (GatA). The ammonia is then channelled, *via* an 'ammonia tunnel', to a phosphate-activated, aspartate-charged tRNA in the second active site (in GatB). Channelling prevents loss of ammonia and excludes water from the active site of GatB (22).

AtzG is homologous to the GatC protein in the GatCAB complex and binds AtzE in the same position as GatC binds GatA in the GatCAB complex. The function of GatC is to coordinate the complex by forming appropriate interactions with both GatA and GatB. AtzG appears to fulfil a similar role, gluing together the AtzE dimer resulting in a tightly associated heterotetramer. According to the program PISA (23), the AtzEG heterotetramer sequesters over 12,500 $Å^2$ of buried surface area, giving a ΔG of interaction of -108 kcal/mol. AtzG has extensive interactions with both AtzE and the AtzG of the dimerization partner (Fig. 4). Notably, AtzG is 28 amino acids shorter than GatC (68 *vs.* 96 amino acids), with the additional 28 amino acids of GatC forming the majority of its interactions with GatB (21) (Fig. 5). GatC also participates in the formation of the ammonia tunnel in the GatCAB complex. Interestingly, the ammonia tunnel is retained in AtzEG (Fig. 6). The retention of the ammonia tunnel may facilitate a high reaction rate by provide rapid egress of the ammonia product.

When AtzE and AtzG were co-expressed in *E. coli*, a soluble (Fig. S1B), active enzyme that was catalytically indistinguishable from AtzEG purified from *Pseudomonas* sp. strain ADP was obtained (Table 1). This confirmed that AtzG promotes the production of soluble, active AtzE. Notably, GatA cannot be over-produced in the absence of GatC (24,25).

Substrate specificity and reaction mechanism- Molecular dynamics were used to understand the behaviour of 1-carboxybiuret in the active site. 1-Carboxybiuret binds *via* a series of hydrogen bonding interactions: the terminal carboxylate of the substrate binds to Tyr125, Asn172 and Gln402, the main chain carbonyl of Gly126 binds the terminal adjacent amines and the main chain NH of Gln402 binds the carbonyl between those amines (Fig. 7A). The interactions with the terminal carboxylate substantially strengthen the binding of 1-carboxybiuret relative to biuret, and would explain the lack of activity with biuret. From a 500 ns MD simulation, we found that the ΔG total as a function of time (Fig. 7C) for 1-carboxybiuret reached significantly lower values than biuret (-33 kcal/mole and -12 kcal/mole, respectively). Carboxybiuret was retained by the AtzE active site for the duration of the simulation,

whereas biuret was observed to leave the active site after 400 ns.

The catalytic mechanism of AtzE is likely to resemble that of other enzymes with Ser-*cis*Ser-Lys catalytic triads (e.g., malonamidase, allophanate amidohydrolase, DNA polymerase V accessory protein, and signal peptidase) (13,26-28). The catalytic triad of AtzE, inferred from crystallographic data and homology with other amidases, comprises of Lys74, *cis*Ser150 and Ser174. In the substrate-docked structure, Ser174 is positioned for attack at the carbonyl of the terminal amide of 1-carboxybiuret (Fig. 7B). Moreover, in the crystal structure of AtzE with the suicide inhibitor PPDI, extra density, consistent with a phosphate group was seen associated with Ser174. This suggests that the phosphodiamidate moiety had been transferred to the catalytic serine (Ser174) during the crystallization process, and that Ser174 is the nucleophilic residue in the catalytic triad. Mass spectrometry of proteolyzed, PPDI-treated AtzE showed that 70% of the enzyme was labelled with monoaminophosphate at Ser174, and 30% with diaminophosphate. This suggests that the diaminophosphate is labile, releasing an ammonia in an aqueous solution (consistent with previous reports)(1).

There are three sets of potential reaction products from the deamination of 1-carboxybiuret: allophanate and carbamate, diacarboxyammonia and urea or 1,3-dicarboxyurea and ammonia. These products are difficult to resolve using the GDH-coupled assay that we employed, as carbamate and diacarboxyammonia readily decompose in water to produce ammonia. As GatA produces ammonia, we might expect that AtzE does too; furthermore, the substrate range of AtzE includes succinamic acid from which carbamate cannot be produced. We therefore propose that AtzE deaminates 1-carboxybiuret, producing 1,3-dicarboxyurea (Fig. 1B), which then monodecarboxylates to form allophanate (the substrate for AtzF) (13,15). Currently, it is unclear if the formation of allophanate is spontaneous or enzyme-mediated.

The suggested mechanism for AtzE is described in Fig 8. In the first step, Lys74, Ser150 and Ser174 all form a hydrogen bonding network, where Lys74 acts as a general base to activate the catalytic Ser174 through the cis-Ser150 bridging ligand. The activated Ser174 performs a nucleophilic attack on the terminal amide end of 1-carboxybiuret, leading to the formation of a covalent acyl-enzyme intermediate and ammonia. Lys74 then acts as a general base, activating water and leading to the hydrolysis of the acyl enzyme, releasing 1,3-dicarboxyurea and restoring the active site to its original state.

It is interesting that AtzE has no detectable activity with GatA substrates (e.g., α-amino acids), given their high degree of structural similarity. We compared the structures of AtzE with docked 1-carboxybiuret and GatA with glutamine bound in its active site (Fig. 9). The AtzE active site is highly conserved when compared with that of GatA, albeit with some differences (Fig. 9). The catalytic triad is conserved (Lys79, Ser154 and Ser178 in GatA), as are Gly126 (GatA Gly130), Ser169 (GatA Ser173), Thr171 (GatA Thr175), and Gly173 (GatA Gly177). Among the key differences between the two enzymes are Phe127 and Tyr125 (Gly131 and Met129 in GatA), which appear to stabilize 1-carboxybiuret in the AtzE active site *via* π-stacking interactions to the conjugated carbonyl system in the substrate. Phe127 would also sterically prevent the binding of glutamine and other α-amino acids which cannot adopt a planar orientation like 1-carboxybiuret, and therefore likely has a prominent role in substrate specificity. Additionally, Asp425 in GatA, which is required for hydrogen bonding to the backbone nitrogen of the substrates, is Gly398 in AtzE.

DISCUSSION

Previously, AtzE had been reported to be a biuret hydrolase (1,5,14,19), largely through comparison with the non-homologous biuret hydrolase of *Rhizobium* sp. (Table 1). However, our findings demonstrate that AtzE is a 1-carboxybiuret hydrolase and that the terminal carboxylate is essential for correct substrate binding. Sequence and structural homology suggest that AtzEG may have been 'repurposed' from the bacterial glutamine transamidosome, in which a similar complex, GatAC, is essential in channelling a solvent-labile intermediate between the complex's two active sites (22). AtzG does not appear to fulfil the same complex co-ordinating function as GatC, instead it may be that the obligate AtzEG complex is simply a vestige of their shared evolutionary past. Interestingly, *Pseudomonas* sp. strain ADP possesses genes that encode a predicated GatCAB complex (*gatA*, GenBank KSW28066.1; *gatB*, GenBank KSW26274.1; *gatC*, KSW26275.1). However, as *atzEG* are encoded by a gene cluster on a transposable element, carried on a self-transmissible plasmid (29,30), it seems unlikely

that *atzEG* was recruited from this specific gene cluster.

Despite the structural and mechanistic similarities between AtzEG and its likely ancestor (a GatAC-like protein), the substrate range for AtzEG excludes the GatAC substrates asparagine and glutamine. GatCAB fulfils an essential role in the production of asparagine and glutamine charged tRNA at the expense of glutamine (31), and it is plausible that AtzEG rapidly evolved away from these substrates under selective pressure to eliminate perturbations in core cellular functions (i.e., amino acyl tRNA production and amino acid pools).

Although not itself catalytic, AtzG is essential for correct AtzE function and previous attempts to express AtzE in heterologous hosts were unsuccessful because the gene encoding AtzG had been overlooked by automated annotation. There is an increasing body of evidence that small, overlooked open-reading frames often encode functional low-molecular weight proteins (32,33). It may be that a subset of these have structural roles, like that of AtzG, and are required to promote or enhance soluble protein expression or stability of their partner proteins.

Surprisingly, the cyanuric acid mineralizing operon contains a second unannotated open-reading frame between the genes encoding AtzE and AtzF (Fig. 1B) that is predicted to encode a 129 amino acid protein (GenBank WP_064987550.1), tentatively named AtzH that belongs to the DUF3225 family of uncharacterized proteins. Preliminary proteomics with *Pseudomonas* sp. strain ADP indicate that AtzH is expressed under the same conditions as AtzD, E, F and G. Investigations are underway to determine the role of AtzH.

EXPERIMENTAL PROCEDURES

Purification of AtzE from Pseudomonas sp. strain ADP - *Pseudomonas* sp. strain ADP's growth conditions were optimized to induce expression of the cyanuric acid catabolism operon. 40 L of *Pseudomonas* sp. strain ADP was grown in 500 mL cultures in 2 L flasks. Cultures were grown in minimal medium at 28 °C and shaken at 200 rpm until an OD_{600} of 0.4 was reached. The minimal medium contained 10 mM cyanuric acid as a sole nitrogen source, 26 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 200 µM $MgSO_4$, 2.9 mM sucrose, 3.4 mM trisodium citrate, 44 µM $CaCl_2$, 20 mL/L of vitamin stock and 1 mL/L of trace elements, adapted from Balotra *et al.*, 2015 (13). The cells were harvested by centrifugation (8000 x *g* for 15 minutes), lysed as described in the reference above and the soluble supernatant was used for further purification.

AtzE was purified by a four-step purification process: ammonium sulfate (AS) precipitation, followed by hydrophobic interaction, anion exchange and size exclusion chromatography. A solution of concentrated AS was added to the supernatant to give a final concentration of 0.6 M and stirred for 4 hours at 4 °C. After centrifugation (18000 x *g* for 30 minutes), the soluble fraction was syringe filtered through a 0.22 µm filter.

The supernatant from the AS cut was purified by hydrophobic interaction chromatography using a 65 mL column packed with Phenyl Sepharose preparation grade resin (GE Healthcare Life Sciences), equilibrated with 25 mM potassium phosphate and 0.6 M ammonium sulfate, pH 7.5. After running 4 column volumes (CV) of 0.4 M AS, a reverse gradient from 0.4 M to 0.09 M AS over 10 CV was used to elute the protein. Fractions containing a 48 kDa band (SDS-PAGE) eluted between 0.27-0.14 M AS. These fractions were pooled, and concentrated to 10 mL volume (Amicon Ultra-15 centrifugal filter unit, Ultracel-30 membrane).

A desalting step was performed using a 53 mL HiPrep C26/10 desalting column (GE Healthcare Life Sciences), equilibrated with 25 mM potassium phosphate, pH 8.5 for 1.5 CV.

The crude protein fraction was purified by anion exchange chromatography using a 77 mL column packed with Q Sepharose preparation grade resin (GE Healthcare Life Sciences), equilibrated with 25 mM potassium phosphate, pH 8.5. The protein was eluted with a gradient of 0 to 0.25 M NaCl over 11 CV. Fractions eluted between 0.16-0.23 M NaCl were found to contain a 48 kDa band (SDS-PAGE). After pooling, these fractions were concentrated to 12 mL using an Amicon Ultra-15 centrifugal filter unit.

This combined sample was applied to a 130 mL size exclusion column packed with Superdex 200 preparation grade resin (GE Healthcare Life Sciences), equilibrated with

25mM potassium phosphate and 200 mM NaCl, pH 7.5, and eluted over 1.5 CV with the same buffer.

All chromatography steps were performed using an ÄKTA purifier UPC 10 (GE Healthcare Life Sciences).

Cloning - A synthetic version of the *atzE* gene, coding for the protein NP_862538 (NP_862538.1) was obtained from GenScript Corporation (Piscataway, NJ, USA), inserted into pUC57, flanked by the restriction sites *Nde*I and *Bam*HI. The genes were subcloned into the *Nde*I and *Bam*HI sites of pETcc2, described in Peat *et al.*, 2013 (11). Following the lack of soluble AtzE expression, we reamplified *atzE* flanked with *Nco*I and *Avr*II with the primers (5' to 3': GTACACCCATGGGAATGAAGACAGTAGA AATTATTGAAGG and TTTTTTGAGCTCCCACATTTCAGTCGGGC GATAC). *atzE* was subcloned into the first multiple cloning site of the pACYCDuet-1 vector (Novagen), a low copy number expression vector for co-expression of two genes, with an in-frame N-terminal thrombin cleavable hexahis-tag. *AtzG* coding for the protein WP_082996223 (WP_082996223.1) was cloned directly from *Pseudomonas* sp. ADP genomic DNA with the primers (5' to 3': CGACGACATATGCTCGAGATGACGGAAA CTG and TGCTGCGAGCTCCCTAGGTCAGATATCTT CTGC), before being subcloned into the second cloning site of the pACYCDuet-1 vector using the enzymes *Nde*I and *Xho*I.
The *atzD* coding for the protein NP_862537 (NP_862537.1) had been cloned into pETcc2 previously (11,13,18). Ligations were performed using T_4 ligase. Restriction enzymes and T_4 ligase were obtained from Thermofisher Scientific.
The biuret hydrolase from *Rhizobium leguminasorum* bv. *viciae* 3841 (BiuH) (AM236084.1) and barbituric acid hydrolase from *Nocardioides* sp. JS614 coding for the protein ABL81019 (ABL81019.1) were expressed in *E. coli* as described elsewhere (12,19).

Heterologous protein expression – The expression vectors were used to transform *Escherichia coli* BL21 (λDE3) cells (Invitrogen). Bacteria were grown on Luria-Bertani (LB) medium containing 100 µg/mL ampicillin for the pETcc2 constructs or 34 µg/mL chloramphenicol for the pACYCDuet-1 construct. Cells were grown with shaking at 200 rpm at 28 °C. Protein expression was induced at an OD_{600} of 0.8 by addition of isopropyl-β–D-1-thiogalactopyranoside (IPTG; 1 mM final concentration).

Cells were harvested 24 hours after induction by centrifugation at 5000 x *g* for 15 minutes using an Aventi J-E centrifuge (Beckman Coulter, Indianapolis, USA), resuspended in lysis buffer (25 mM potassium phosphate, 5 mM imidazole, pH 7.5) and lysed by passage through a Microfluidics homogenizer M-110P (Massachusetts, USA) five times at 15000 PSI. The lysis was followed by centrifugation at 18 000 x *g* for 45 minutes to pellet the cellular debris, and the soluble fraction was used for further purification.

The soluble fraction was syringe filtered throughout a 0.22 µm filter. The filtrate was purified using a 5 mL Ni-NTA Superflow cartridge (Qiagen, Maryland, USA) with a gradient from 5 mM imidazole to 500 mM over 10 CV. SDS-PAGE gel analysis was performed to assess the purity of the fractions.

Enzyme assays- Cyanuric acid hydrolysis by AtzD was followed by UV spectrophotometry at 214 nm (13) using 0-0.6 mM of cyanuric acid in 25 mM phosphate buffer pH 9.

A glutamate dehydrogenase (GDH, Sigma Aldrich) coupled reaction was used to measure ammonia release and determine the activity rate in the biuret hydrolase (BH) and AtzE-dependent reactions. GDH catalyzes the NADH-dependent amination of α-ketoglutarate. Ammonia production by BH and AtzE was followed using the decrease of absorbance by UV spectrophotometry at 340 nm that was due to the oxidation of NADH by GDH. 1.25 U of GDH was used in a 250 µL reaction volume, the final concentrations of α-ketoglutarate and NADH were 3.5 mM and 0.2 mM, respectively (13) (Fig. S3).

Monitoring the AtzE activity during purification steps from *Pseudomonas* sp. strain ADP was initially done with its reported substrate, biuret (1,5,13). However, as no ammonia could be detected, the fractions were supplemented with 5.1 nM of AtzD and 0.2 mM of cyanuric acid, in 25 mM phosphate buffer pH 9, in order to generate the substrate *in situ*. There was a significant background of ammonia in the early steps/cruder fractions; however, AtzE-dependent ammonia production was detectable

after the background rate (measured in the absence of cyanuric acid) had been subtracted.

Steady state kinetics for AtzD were obtained by using 41 nM of AtzD, in the presence of various amounts of cyanuric acid substrate ranging from 0-0.6 mM. Cyanuric acid degradation was measured by following the decrease in absorbance at 214 nm, but above 0.2 mM AtzD activity was inhibited. To determine whether the substrate or the product was inhibiting AtzD activity, we pooled 41 nM of AtzD, 24 nM of AtzE and 5 U/µL of GDH in the presence of 0 - 0.6 mM cyanuric acid. This removed the inhibition previously observed at cyanuric acid concentrations above 0.2 mM, and allowed the determination of steady state kinetics for AtzD. Steady state kinetics for AtzD were determined using 41 nM of AtzD, 24 nM of AtzE and 5 U/µL of GDH, after ensuring that in these proportions, the AtzD-mediated reaction was the rate limiting reaction and with the assumption that the rate of ammonia production recorded was proportional to the AtzD activity rate.

Steady state kinetics for AtzE with the product of AtzD or barbituric acid hydrolase (BAH) were obtained under conditions where the AtzE-mediated reaction was the rate limiting reaction (i.e. 3.5 times slower than either the GDH, AtzD, or barbituric acid hydrolase mediated reactions). AtzE steady state kinetics were obtained by using 41 nM of AtzD/ 48 nM barbituric acid hydrolase, 12 nM of AtzE and 5 U/µL of GDH. Substrate (cyanuric acid or barbituric acid) was added over the range of 0-0.6 mM. As AtzE was rate-limiting, it was assumed to be a good estimate of the rate of AtzE.

Other potential AtzE substrates tested were: 1-nitrobiuret, 1-carboxymalonamide, succinamic acid, 2-amino-3-ureidopropionic (Albizziin), citrulline, lysine, glutamine, asparagine, biuret and malonamide. Specific activities were obtained with 12-400 nM of AtzE and 5 U/µL of GDH and 1.5 mM substrates in 25 mM phosphate buffer pH 9. Steady state kinetics were obtained with 1-nitrobiuret using 12 nM of AtzE and 5 U/µL of GDH. Substrate,1-nitrobiuret, was added over the range of 0-2.5 mM. Steady state kinetics were obtained with 1-carboxy-malonamide using 12 nM of AtzE, 48 nM BAH and 5 U/µL of GDH. Substrate (barbituric acid) was added over the range of 0-4 mM.

Biuret hydrolase kinetic data were obtained by using 11 nM of biuret hydrolase and 5 U/µL of GDH in presence of various amount of biuret ranging from 0-0.7 mM.

All the kinetics constants were calculated with Hyper32 software (http://homepage.ntlworld.com/john.easterby/hyper32.html), fitting the rate data to Michaelis-Menten equation:

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{Km + [S]}$$

The steady-state kinetic values for AtzE with 1-nitro-biuret were estimated using Lineweaver-Burk equation:

$$\frac{1}{Vo} = \frac{Km}{V_{max}}\frac{1}{[S]} + \frac{1}{V_{max}}$$

LC-MS/MS multiple reaction monitoring (MRM)-based enzyme assay was used to follow the fate of cyanuric acid and biuret when in presence of either BiuH, AtzD and/or AtzE. Reaction samples were run in triplicates. Four reactions were composed of: 10 mM biuret with 580 nM BiuH, 10 mM biuret with 374 nM AtzE, 10 mM cyanuric acid with 420 nM AtzD and 10 mM cyanuric acid with 420 nM AtzD and 374 nM AtzE. 5 uL of reaction mix was injected every 2 minutes for 30-60 minutes. Analysis was performed using a Waters Alliance 2695 separation module coupled to a Waters TQD detector but injections were made using a manual injector with a 5 µL injection loop. The HPLC Column used was a Phenomenex Develosil RPAQEOUS-AR 5u C30 (dimension: 250 × 4.6 mm), which was heated to 35 ˚C. A flow rate of 0.7 mL/min was used (15:1 flow splitter was used after separation). The eluent used was 15% (v/v) acetonitrile and 0.1% formic acid in water. Two simultaneous but independent MRM analyses were performed as follows: MRM parameters for biuret (Parent m/z = 103.9424): Channel 1 (ES+): Daughter (m/z): 17.9654, Dwell (s): 0.025, Cone (V): 20, Collision (V): 10; Channel 2 (ES+): Daughter (m/z): 43.8956, Dwell (s): 0.025, Cone (V): 20, Collision (V): 20. MRM Parameters for cyanuric acid (Parent m/z = 127.8577): Channel 1 (ES-): Daughter (m/z): 41.9357, Dwell (s): 0.025, Cone (V): 28, Collision (V): 14. MRM peaks were integrated using Waters MassLynx v4.1 software.

Proteomics- In-gel digestion was performed by resuspending pieces of bis-acrylamide gel

containing 50-100 µg protein in 100 µL of 25 mM ammonium bicarbonate. After addition of 5 µL of 15% DTT, the mix was incubated at room temperature for 30 min, followed by addition of 3.5 µL of 40% Acrylamide and another 30 min incubation at room temperature. The sample was digested with 0.5 µg of trypsin at 37 °C overnight. To stop the digestion, formic acid was added to a final concentration of 0.1-1%, then the trypsin digested peptides were eluted from gel pieces by sonication for 15 min and further incubated at room temperature for 30 min. Typically, 1-5 µL of sample was used for LC-MS analysis.

The trypsin-digested peptides were separated by reversed-phase HPLC using an Agilent NanoFlow LC system (1260 Infinity). The peptides were loaded onto a NanoLC trap column (3 µm, ChromXP C18CL, 120Å 0.5 mm x 350 µm from Eksignet Technologies) by autosampler at a flowrate of 5 uL/min and continuously desalted for 5 min. The desalted peptides were eluted from the trap column and separated on a NanoLC column (3 µm, ChromXP C18CL, 120Å, 15 cm x 75 µm from Eksigent) with a 0-40% gradient in 100 min consisting of buffer A (0.1% FA in water) and Buffer B (0.1% FA in acetonitrile). The flowrate of Nano pump was set to 300 nL/min.

Peptides were analyzed using positive and high sensitive mode on an AB Sciex Tripletof 5500 mass spectrometer. The voltage of Nano spray II was set to 2300 kV. In Data Dependent Acquisition (DDA) mode, the mass window for precursor ions of the quadrupole mass analyzer was set to +/- 1 m/z. The precursor irons were fragmented by nitrogen collision gas. The MS1 survey scan (250 msec; mass 350-1500) was carried out. 50 of the most abundant precursors were selected for MS/MS Scan (50 msec; mass 100-1800). MS/MS spectra were obtained for product ions which had charge state of 2 to 5 and were above 10 counts per second. Rolling collision energy was used, and automatic calibration was carried out after every 5 sample runs.

Mass spectrum data was analyzed using the Paragon algorithm of ProteinPilot (AB Sciex). The detected protein threshold was set to >0.05(10%). The matching peptide sequences were identified against the *Pseudomonas* sp. ADP proteome found in Uniprot. The false discovery rate (FDR) analysis tool algorithm of ProteinPilot provided a global FDR of 1% and a local FDR at 1% in all cases.

PPDI-treated AtzE samples were digested with trypsin using the SP3 digestion protocol (34). Peptides were separated using an UltiMate nanoUPLC system, utilizing a 60min gradient on an Acclaim Pepmap 100 column (25 cm × 75µm id with 3 µm particles). High-resolution MS/MS data was obtained on an Orbitrap Fusion Lumos Mass Spectrometer and top-twenty multiply charged species selected for fragmentation in high-high mode with the Orbitrap resolution set at 75,000.

Orbitrap MS/MS data was searched against a focused decoy database containing AtzE, *E. coli* and common contaminant protein sequences using the Byonic search engine (Protein Metrics) with tolerance of 5 ppm for precursor ions and 10 ppm for product ions. Enzyme specificity was tryptic and allowed for up to 2 missed cleavages per peptide. Variable modifications were set for NH2-terminal acetylation or protein N-termini, oxidation of methionine or tryptophan, deamidation of asparagine or glutamine and dehydration of cysteine. A Wildcard search with a range of +75 to +80Da facilitated confident peptide identification (< 1% FDR) and spectrum counting of monoamidophphate or diaminophosphate modified serine residues.

Crystallization and structural determination -The stability of purified AtzE in the phosphate buffer used in the size exclusion chromatography was assessed using Differential Scanning Fluorimetry as implemented in the Collaborative Crystallisation Centre (35). Protein was used at 2.3 mg/mL, and showed a T_m of 64.6 +/- 1.4° C in the phosphate buffer, but also showed that the protein was equally stable in all buffers tested with pH between 6.5 and 8.5, and the stability was essentially independent of salt concentration. An initial bank of 384 crystallization trials (PACT, shotgun and PS gradient at 20 °C, shotgun at 8° C; see c6.csiro.au for details of the screens) showed the formation of crystals overnight in many PEG-containing conditions. All of the crystals were small and rod-shaped, and all diffracted poorly and had a pathological packing problem which made them unsuitable for diffraction analyses. Over the course of a year, over 10,000 droplets were set up, trying various combinations of seeding, *in-situ* proteolysis, concentrations, differing formulations, vapor diffusion, microbatch and temperatures. All

crystals showed the same pathological form as the original crystals that grew overnight. Finally, it was found that the addition of 0.05 % agarose to the protein solution before setup, coupled with seeding from the poor form crystals gave a new crystal form that showed diffraction to 2 Å, and which did not have the same packing problems. The well-behaved crystal form appeared overnight and grew to full size in 5 days, and grew at both 8° C and 20° C. The crystals used in the structure determination were grown from sitting drop trials with 250 nL of 1.1 mg/mL protein in 25 mM Hepes pH 7.5, 100 mM NaCl, with 0.05% agarose, 250 nL crystallant (0.1 M bis-tris pH 6.04, 0.276 M $MgCl_2$, 17.6 % w/v PEG 8000), incubated at 8° C. Crystals of the protein treated with an inhibitor, phenyl phosphorodiamidate (PPDI), grew under similar conditions to the native protein (reservoir 0.1 M bis-tris pH 5.46, 0.128 M $MgCl_2$, 21 %w/v PEG 8000) at 20° C. A crystal of the protein/PPDI complex was cryoprotected with reservoir solution supplemented with glycerol to 20% final concentration and flash cooled in liquid nitrogen. This crystal was used to collect 360 degrees of data to 2.1 Å at beamline MX1 of the Australian Synchrotron. The data were initially processed with XDS (36), and the structure was solved by molecular replacement using MoRDa (37), which output two copies each of the A domains from 3ip4 and 3dha in the asymmetric unit. After some rebuilding, the new model was used in Phaser with reprocessed data (using Xia2 and DIALS) (38-42) and this was followed by manual building (COOT) and refinement with Refmac (43). Two complete chains of AtzE (residues 1-457) were located in the asymmetric unit. Additional protein density was clearly visible in the maps and another protein chain was modelled into the density. The sequence was confirmed by searching the Pseudomonas database and finding that the AtzG sequence matched the sequence of the density (with 66 of 68 residues visible). Subunit interactions were analysed using PISA (23).

A crystal of the native AtzE without inhibitor was harvested and data collected that extended to just beyond 2 Å at the MX2 beamline of the Australian Synchrotron. These data were processed using Xia2/DIALS and the AtzE/PPDI model was used to phase the data with Phaser.

Computational methods- Models of AtzEG containing biuret and 1-carboxybiuret were prepared in Accelrys Discovery Studio v4.1 using the experimentally derived structure as a starting point. Models for the substrates were relaxed using the Full Minimization tool in Discovery Studio v4.1 using the default settings (CHARMm force field) and positioned in the receptor cavity (active site) using CDOCKER with the default parameters.

Ligands were prepared for MD using the Antechamber module in AMBER16 (44) using the GAFF2 force field. The protein models were prepared for MD simulations using xLeap applying the ff14SB force field and charge-neutralized by the addition of Na+ ions. The proteins were solvated in a TIP3P truncated octahedral solvent box with a minimum 12 Å periodic boundary distance from the solute.

Initial minimization of both systems was performed using AMBER 16 over 10,000 steps under a constant pressure of 1 bar (Berendsen barostat). Bonds lengths on bonds involving hydrogen were constrained using SHAKE, and force evaluation on these bonds was not performed. MD simulations of 500 ns with a step-size of 0.002 ps were performed at 310 K and 1 bar pressure with a 1 ps relaxation time. Long range electrostatic interactions were treated with the particle mesh Ewald method beyond 12 Å. Simulations were analysed using VMD (v. 1.9.2) (45).

Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) continuum solvation models were calculated using the MM-PBSA module in AMBER 16 over the course of the entire simulation. Entropy approximations were calculated with ptraj.

The ammonia tunnel in the GatCAB complex (PDB: 2G5H) and the equivalent tunnel in AtzEG were visualized using CAVER 3 (46) using the default parameters (Clustering threshold of 3.5, probe radius of 0.9. Approximation of 12).

ACKNOWLEDGEMENTS

We would like to thank the beamline scientists at the Australian Synchrotron for their help in data collection. Crystals were grown in the CSIRO Collaborative Crystallisation Centre (http://www.

csiro.au/C3) at CSIRO Manufacturing. We would also like to thank Drs Andrew Warden and Thomas Walsh (CSIRO Land & Water) for their constructive comments while preparing this manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

AUTHOR CONTRIBUTIONS

CS conceived and coordinated the study. LE, CS, TSP and JN wrote the paper. LE, CJH, CJE and MW designed and interpreted the biochemical characterization experiments. TSP and JN did the crystallization and structural biology. LE, HO and JWL designed and interpreted the mass spectrometry experiments. TN designed and performed the mass spec experiments on the inhibited protein complex. NGF provided technical assistance. LE and MW did the substrate docking. MW did the molecular dynamics. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Cameron, S. M., Durchschein, K., Richman, J. E., Sadowsky, M. J., and Wackett, L. P. (2011) A new family of biuret hydrolases involved in *s*-triazine ring metabolism. *ACS Catal* **2011**, 1075-1082
2. Udikovic-Kolic, N., Scott, C., and Martin-Laurent, F. (2012) Evolution of atrazine-degrading capabilities in the environment. *Appl Microbiol Biotechnol* **96**, 1175-1189
3. Jutzi, K., Cook, A. M., and Hutter, R. (1981) The degradative pathway of the *s*-triazine melamine. *Experientia* **37**, 1231-1232
4. Mandelbaum, R. T., Allan, D. L., and Wackett, L. P. (1995) Isolation and characterization of a *Pseudomonas* sp that mineralizes the *s*-triazine herbicide atrazine. *Applied and Environmental Microbiology* **61**, 1451-1457
5. Martinez, B., Tomkins, J., Wackett, L. P., Wing, R., and Sadowsky, M. J. (2001) Complete nucleotide sequence and organization of the atrazine catabolic plasmid pADP-1 from *Pseudomonas* sp. strain ADP. *J Bacteriol* **183**, 5684-5697
6. de Souza, M. L., Wackett, L. P., Boundy-Mills, K. L., Mandelbaum, R. T., and Sadowsky, M. J. (1995) Cloning, characterization, and expression of a gene region from *Pseudomonas* sp. strain ADP involved in the dechlorination of atrazine. *Appl Environ Microbiol* **61**, 3373-3378
7. Peat, T. S., Newman, J., Balotra, S., Lucent, D., Warden, A. C., and Scott, C. (2015) The structure of the hexameric atrazine chlorohydrolase AtzA. *Acta Crystallogr D Biol Crystallogr* **71**, 710-720
8. Seffernick, J. L., Aleem, A., Osborne, J. P., Johnson, G., Sadowsky, M. J., and Wackett, L. P. (2007) Hydroxyatrazine *N*-ethylaminohydrolase (AtzB): an amidohydrolase superfamily enzyme catalyzing deamination and dechlorination. *J Bacteriol* **189**, 6989-6997
9. Balotra, S., Warden, A. C., Newman, J., Briggs, L. J., Scott, C., and Peat, T. S. (2015) X-Ray structure and mutagenesis studies of the *N*-isopropylammelide isopropylaminohydrolase, AtzC. *PLoS One* **10**, e0137700
10. Sadowsky, M. J., Tong, Z. K., de Souza, M., and Wackett, L. P. (1998) AtzC is a new member of the amidohydrolase protein superfamily and is homologous to other atrazine-metabolizing enzymes. *J. Bacteriol.* **180**, 152-158

11. Peat, T. S., Balotra, S., Wilding, M., French, N. G., Briggs, L. J., Panjikar, S., Cowieson, N., Newman, J., and Scott, C. (2013) Cyanuric acid hydrolase: evolutionary innovation by structural concatenation. *Mol Microbiol* **88**, 1149-1163
12. Peat, T. S., Balotra, S., Wilding, M., Hartley, C. J., Newman, J., and Scott, C. (2017) High-resolution X-ray structures of two functionally distinct members of the cyclic amide hydrolase family of Toblerone fold enzymes. *Appl Environ Microbiol* **83**, 16
13. Balotra, S., Newman, J., Cowieson, N. P., French, N. G., Campbell, P. M., Briggs, L. J., Warden, A. C., Easton, C. J., Peat, T. S., and Scott, C. (2015) X-ray structure of the amidase domain of AtzF, the allophanate hydrolase from the cyanuric acid-mineralizing multienzyme complex. *Appl Environ Microbiol* **81**, 470-480
14. Cheng, G., Shapir, N., Sadowsky, M. J., and Wackett, L. P. (2005) Allophanate hydrolase, not urease, functions in bacterial cyanuric acid metabolism. *Appl Environ Microbiol* **71**, 4437-4445
15. Shapir, N., Sadowsky, M. J., and Wackett, L. P. (2005) Purification and characterization of allophanate hydrolase (AtzF) from *Pseudomonas* sp. strain ADP. *J Bacteriol* **187**, 3731-3738
16. Eaton, R. W., and Karns, J. S. (1991) Cloning and comparison of the DNA encoding ammelide aminohydrolase and cyanuric acid amidohydrolase from three *s*-triazine-degrading bacterial strains. *J Bacteriol* **173**, 1363-1366
17. Seffernick, J. L., Erickson, J. S., Cameron, S. M., Cho, S., Dodge, A. G., Richman, J. E., Sadowsky, M. J., and Wackett, L. P. (2012) Defining sequence space and reaction products within the cyanuric acid hydrolase (AtzD)/barbiturase protein family. *J Bacteriol* **194**, 4579-4588
18. Balotra, S., Newman, J., French, N. G., Briggs, L. J., Peat, T. S., and Scott, C. (2014) Crystallization and preliminary X-ray diffraction analysis of the amidase domain of allophanate hydrolase from *Pseudomonas* sp. strain ADP. *Acta Crystallogr F Struct Biol Commun* **70**, 310-315
19. Esquirol, L., Peat, T. S., Wilding, M., Lucent, D., French, N. G., Hartley, C. J., Newman, J., and Scott, C. (2018) Structural and biochemical characterization of the biuret hydrolase (BiuH) from the cyanuric acid catabolism pathway of *Rhizobium leguminasorum* bv. *viciae* 3841. *PLoS ONE* **13**, e0192736
20. Berman, H. M. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235-242
21. Suzuki, T., Nakamura, A., Kato, K., Soll, D., Tanaka, I., Sheppard, K., and Yao, M. (2015) Structure of the *Pseudomonas aeruginosa* transamidosome reveals unique aspects of bacterial tRNA-dependent asparagine biosynthesis. *Proc Natl Acad Sci U S A* **112**, 382-387
22. Nakamura, A., Yao, M., Chimnaronk, S., Sakai, N., and Tanaka, I. (2006) Ammonia channel couples glutaminase with transamidase reactions in GatCAB. *Science* **312**, 1954-1958
23. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797
24. Curnow, A. W., Hong, K. W., Yuan, R., Kim, S. I., Martins, O., Winkler, W., Henkin, T. M., and Soll, D. (1997) Glu-tRNA(Gln) amidotransferase: A novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 11819-11826
25. Skouloubris, S., de Pouplana, L. R., de Reuse, H., and Hendrickson, T. L. (2003) A noncognate aminoacyl-tRNA synthetase that may resolve a missing link in protein evolution. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11297-11302

26. Shin, S., Lee, T. H., Ha, N. C., Koo, H. M., Kim, S. Y., Lee, H. S., Kim, Y. S., and Oh, B. H. (2002) Structure of malonamidase E2 reveals a novel Ser-*cis*Ser-Lys catalytic triad in a new serine hydrolase fold that is prevalent in nature. *EMBO J* **21**, 2509-2516
27. Paetzel, M., Dalbey, R. E., and Strynadka, N. C. (1998) Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature* **396**, 186-190
28. Peat, T. S., Frank, E. G., McDonald, J. P., Levine, A. S., Woodgate, R., and Hendrickson, W. A. (1996) Structure of the UmuD' protein and its regulation in response to DNA damage. *Nature* **380**, 727-730
29. de Souza, M. L., Wackett, L. P., and Sadowsky, M. J. (1998) The atzABC genes encoding atrazine catabolism are located on a self-transmissible plasmid in Pseudomonas sp. strain ADP. *Applied and Environmental Microbiology* **64**, 2323-2326
30. Martinez, B., Tomkins, J., Wackett, L. P., Wing, R., and Sadowsky, M. J. (2001) Complete nucleotide sequence and organization of the atrazine catabolic plasmid pADP-1 from Pseudomonas sp. strain ADP. *J. Bacteriol.* **183**, 5684-5697
31. Min, B., Pelaschier, J. T., Graham, D. E., Tumbula-Hansen, D., and Soll, D. (2002) Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. *Proc Natl Acad Sci U S A* **99**, 2678-2683
32. Baek, J., Lee, J., Yoon, K., and Lee, H. (2017) Identification of unannotated small genes in *Salmonella*. *G3 (Bethesda)* **7**, 983-989
33. Hucker, S. M., Ardern, Z., Goldberg, T., Schafferhans, A., Bernhofer, M., Vestergaard, G., Nelson, C. W., Schloter, M., Rost, B., Scherer, S., and Neuhaus, K. (2017) Discovery of numerous novel small genes in the intergenic regions of the *Escherichia coli* O157:H7 Sakai genome. *PLoS One* **12**, e0184119
34. Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., and Krijgsveld, J. (2014) Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular Systems Biology* **10**
35. Rosa, N., Ristic, M., Seabrook, S. A., Lovell, D., Lucent, D., and Newman, J. (2015) Meltdown: A tool to help in the interpretation of thermal melt curves acquired by differential scanning fluorimetry. *J Biomol Screen* **20**, 898-905
36. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
37. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242
38. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
39. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* **63**, 32-41
40. Parkhurst, J. M., Winter, G., Waterman, D. G., Fuentes-Montero, L., Gildea, R. J., Murshudov, G. N., and Evans, G. (2016) Robust background modelling in DIALS. *J Appl Crystallogr* **49**, 1912-1921
41. Waterman, D. G., Winter, G., Gildea, R. J., Parkhurst, J. M., Brewster, A. S., Sauter, N. K., and Evans, G. (2016) Diffraction-geometry refinement in the DIALS framework. *Acta Crystallogr D Struct Biol* **72**, 558-575
42. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190
43. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255

44. Case, D. A., Cerutti, D. S., Cheatham, I., Darden, T., A.R.E., D., Giese, T. J., Gohlke, H., Goetz, A. W., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D., M. Merz, K. M., Monard, G., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Simmerling, C. L., Botello-Smith, W. M., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L., York, D. M., and Kollman, P. A. (2017) AMBER16. University of California, San Francisco
45. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38
46. Pavelka, A., Sebestova, E., Kozlikova, B., Brezovsky, J., Sochor, J., and Damborsky, J. (2016) CAVER: Algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE-ACM Trans. Comput. Biol. Bioinform.* **13**, 505-517

Table 1: Steady-state kinetic parameters for AtzE. The structures of the substrates and non-substrates tested (biuret, malonamide, asparagine, glutamine, lysine, citrulline and 2-amino-3-ureidopropionic acid) are shown in Supplementary Table 2. The standard errors for the K_M and k_{cat} values are given.

Enzyme	Substrate	k_{cat} (s^{-1})	K_M (µM)	k_{cat}/K_M ($s^{-1}.M^{-1}$)
AtzD	cyanuric acid	17.0 ± 1.7	350 ± 42	4.8 x 10^4
BiuH	biuret	11.5 ± 0.2	80 ± 7	1.5 x 10^5
AtzE (*Pseudomonas* sp. ADP)	1-carboxybiuret	15.5 ± 0.1	63 ± 3	2.5 10^5
AtzE (*E. coli*)	1-carboxybiuret	14.1 ± 0.2	61 ± 3	2.3 x 10^5
	1-nitrobiuret	9.4 ± 3.3	785 ± 217	1.2 x 10^4
	1-carboxymalonamide	7.0 ± 0.1	167 ± 17	4.2 x 10^4
	succinamic acid	19.0 ± 0.4	523 ± 30	3.6 x 10^4

Table 2: Crystal structure parameters.

	6C62 (Native)	6C6G (PPDI)
Data collection		
Space group	I2	I2
Cell dimensions		
a, *b*, *c* (Å)	79.5, 89.0, 141.7	78.6, 88.9, 141.9
α, β, γ (°)	90, 101.9, 90	90, 101.3 , 90
Resolution (Å)	1.95 (1.99 - 1.95)	2.10 (2.16 - 2.10)
R_{merge}	0.233 (1.545)	0.326 (1.282)
R_{pim}	0.095 (0.622)	0.126 (0.496)
$I / \sigma I$	6.9 (2.1)	6.7 (2.0)
CC1/2	0.991 (0.804)	0.982 (0.656)
Completeness (%)	100 (100)	100 (100)
Redundancy	6.9 (7.1)	7.6 (7.6)
Refinement	2 dimers	2 dimers
Resolution (Å)	69.3 - 1.95	69.6 - 2.10
Unique reflections	66,810	53,225
R_{work} / R_{free}	15.3 / 19.1	16.7 / 21.0
No. atoms	8659	8523
Protein	7930	7882
Inhibitor	n/a	8
Water	726	639
B-factors (Å^2)	17.5	16.8
Protein	17.4	16.9
Inhibitor	n/a	12.4
Water	24.6	21.1
R.m.s deviations		
Bond lengths (Å)	0.017	0.014
Bond angles (°)	1.726	1.643

Note: Values in parentheses are for the highest resolution shell.

Figure Legends

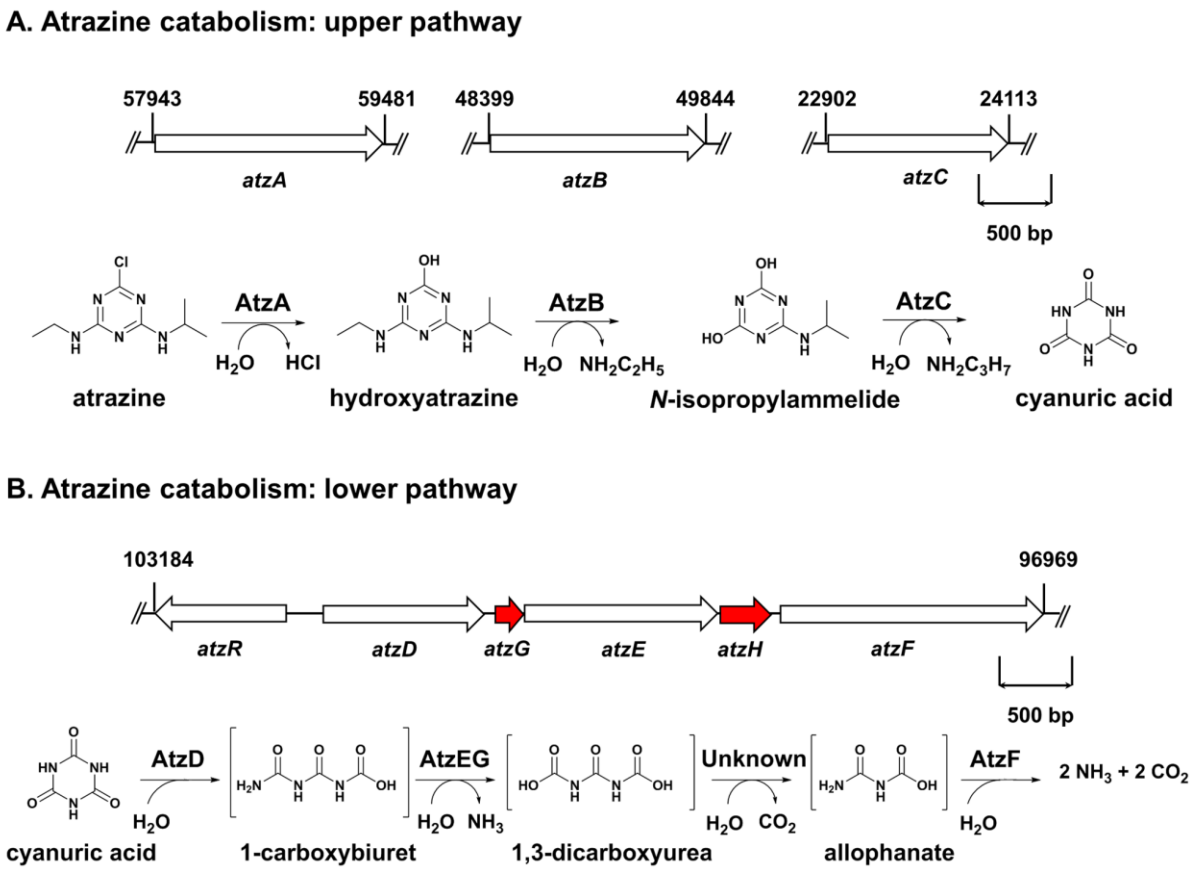

Figure 1. Catabolism of the *s*-triazine herbicide atrazine by *Pseudomonas* sp. strain ADP. The atrazine catabolic pathway is comprised of an 'upper pathway' (A), which transforms atrazine to cyanuric acid, and a lower pathway (B), which catabolizes cyanuric acid to ammonia and carbon dioxide. The entire *Pseudomonas* sp. strain ADP atrazine catabolic pathway is encoded on a single conjugative plasmid (pADP1; GeneBank NZ_CM003636.1), and the co-ordinates of each gene are shown. The genes of the upper pathway (*atzA*, *atzB*, and *atzC*) encode three hydrolases (AtzA, AtzB, and AtzC). The genes of the lower pathway (*atzD*, *atzG*, *atzE*, *atzH* and *atzF*) are found as a single operon: *atzG* and *atzH* (shown in red) were identified in the work reported here. AtzE had previously been reported as a biuret amidohydrolase (1); however in the work presented here we have demonstrated that it is a 1,3-dicaboxyurea forming 1-carboxybiuret amidohydrolase (as shown in the updated lower pathway).

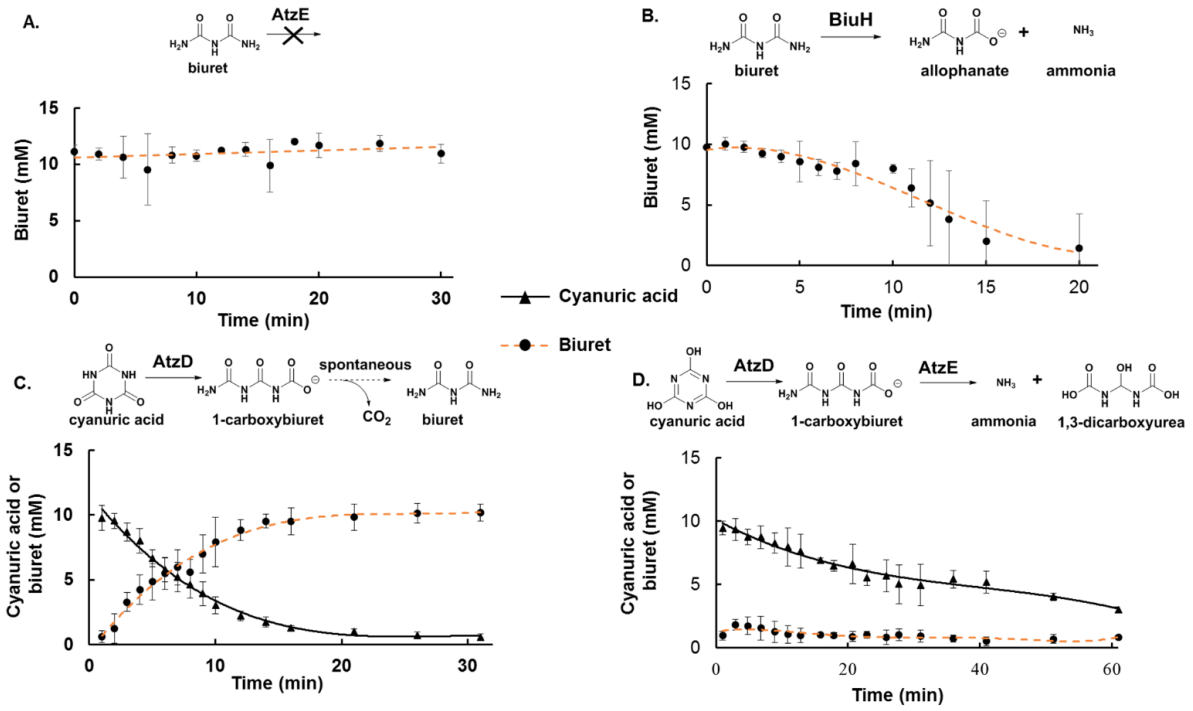

Figure 2. LC-MS analysis of AtzE activity with biuret and 1-carboxybiuret. The concentrations of cyanuric acid (triangles) and biuret (circles) present in each reaction is shown. Four reaction conditions were tested: A) biuret in presence of AtzE; B) biuret in presence of the biuret hydrolase BiuH; C) cyanuric acid in presence of AtzD; and, D) cyanuric acid in presence of AtzD and AtzE. Reactions were done in triplicate and the standard deviations are shown.

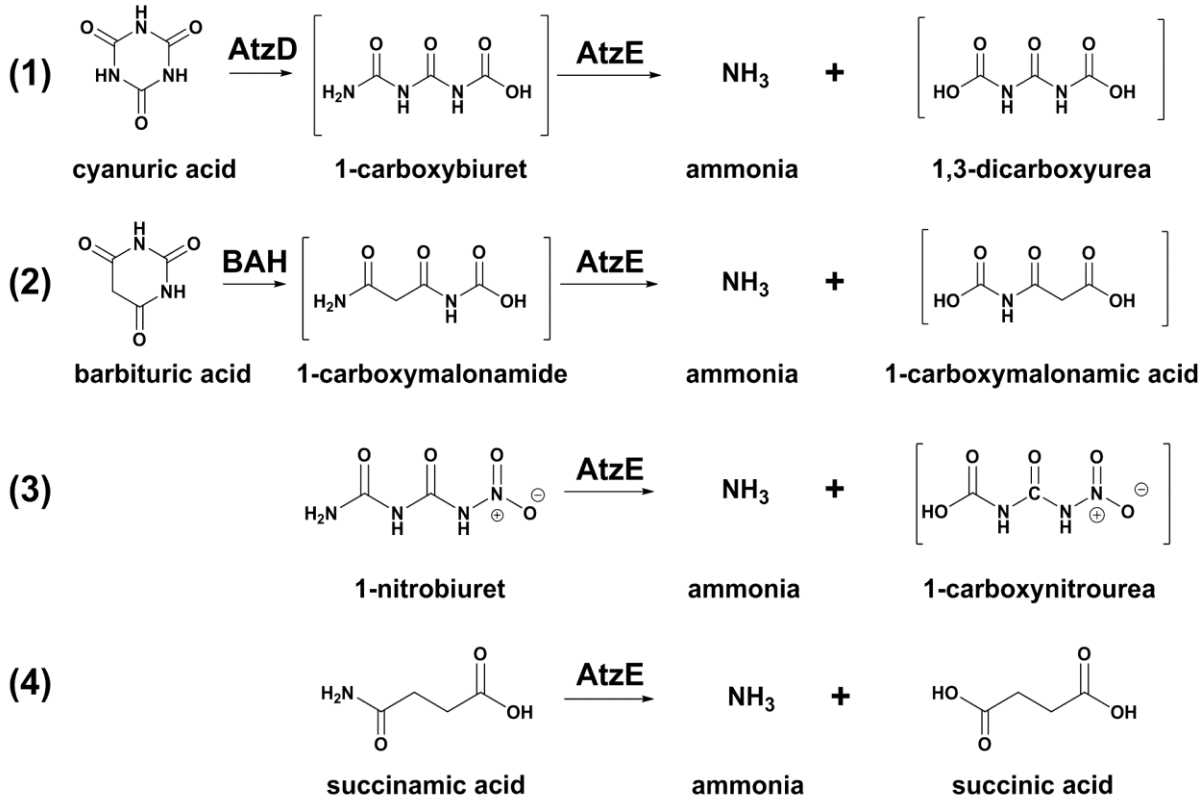

Figure 3. AtzE-dependent deamination reactions. The schemes for AtzE-dependent deamination of: 1) 1-carboxybiuret, 2) 1-carboxymalonamide, 3) 1-nitrobiuret and 4) succinamic acid are shown. 1-Carboxybiuret and 1-carboxymalonamide were produced *in situ* by cyanuric acid amidohydrolase (AtzD) and barbituric acid amidohydrolase (BAH), respectively. Square brackets indicate compounds that spontaneously decarboxylate under the reaction conditions used.

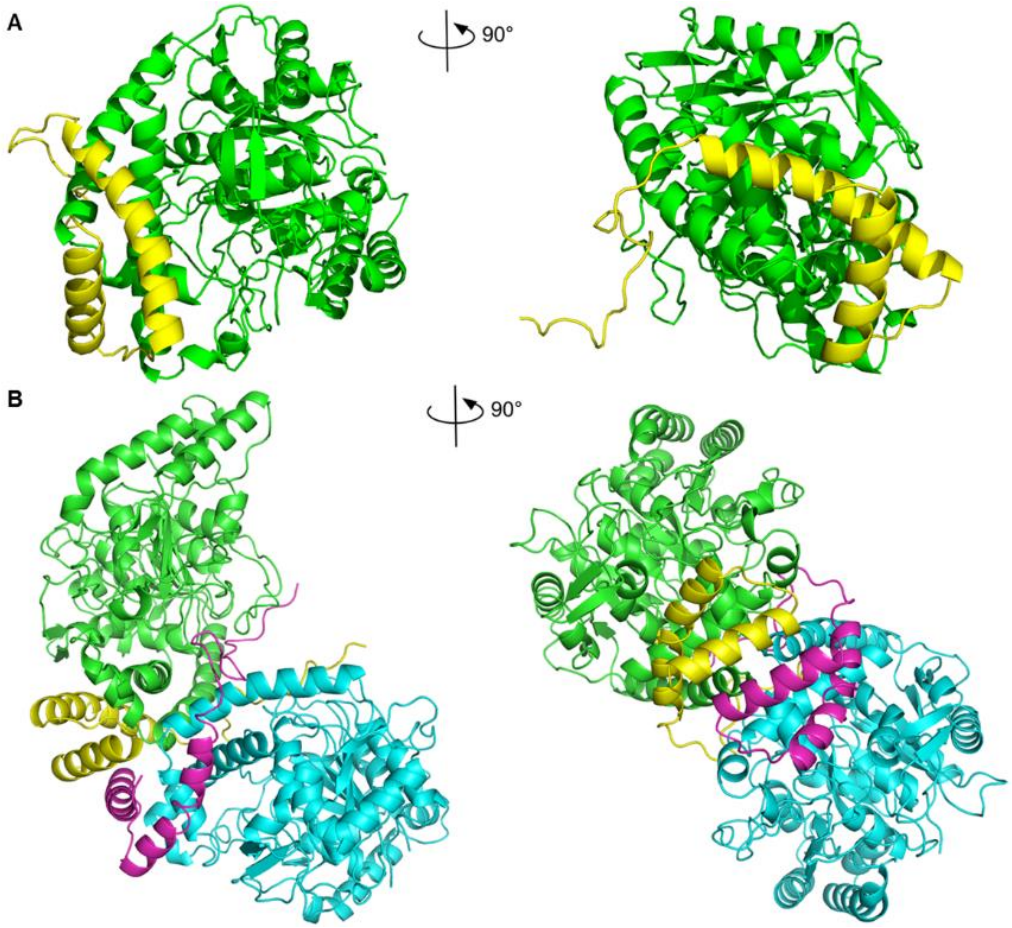

Figure 4. X-ray crystal structure of the unexpected AtzE complex. A) The AtzE monomer (green) has a typical amidase fold. Surprisingly, it is associated with a small (68 amino acid) protein (yellow); B) The $\alpha_2\beta_2$ heterotetramer, containing two molecules of AtzE (green and cyan) and two molecules of the 68 amino acid protein (yellow and magenta).

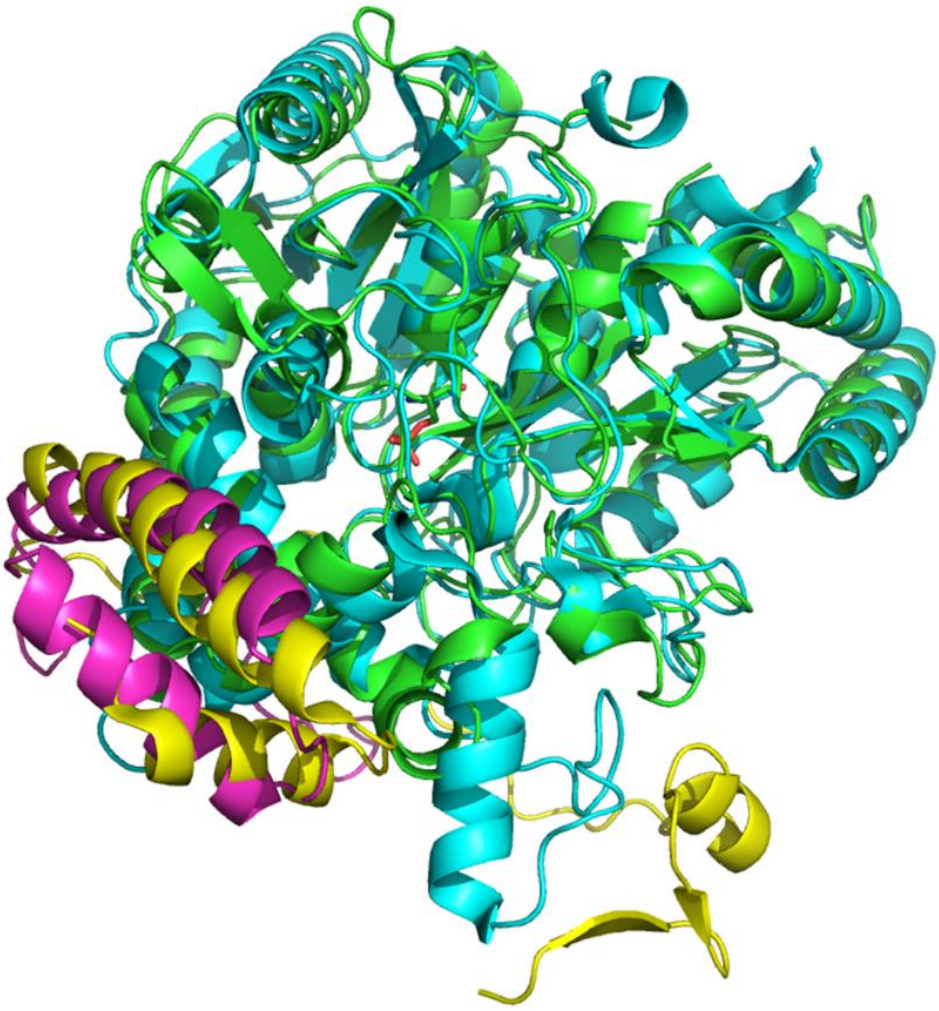

Figure 5. Superposition of the AtzEG heterodimer and the GatA:GatC complex. AtzE and AtzG are shown in green and magenta, respectively. GatA and GatC are shown in cyan and yellow, respectively. The phosphoserine from the AtzE x-ray crystal structure (Ser174) is shown in stick representation.

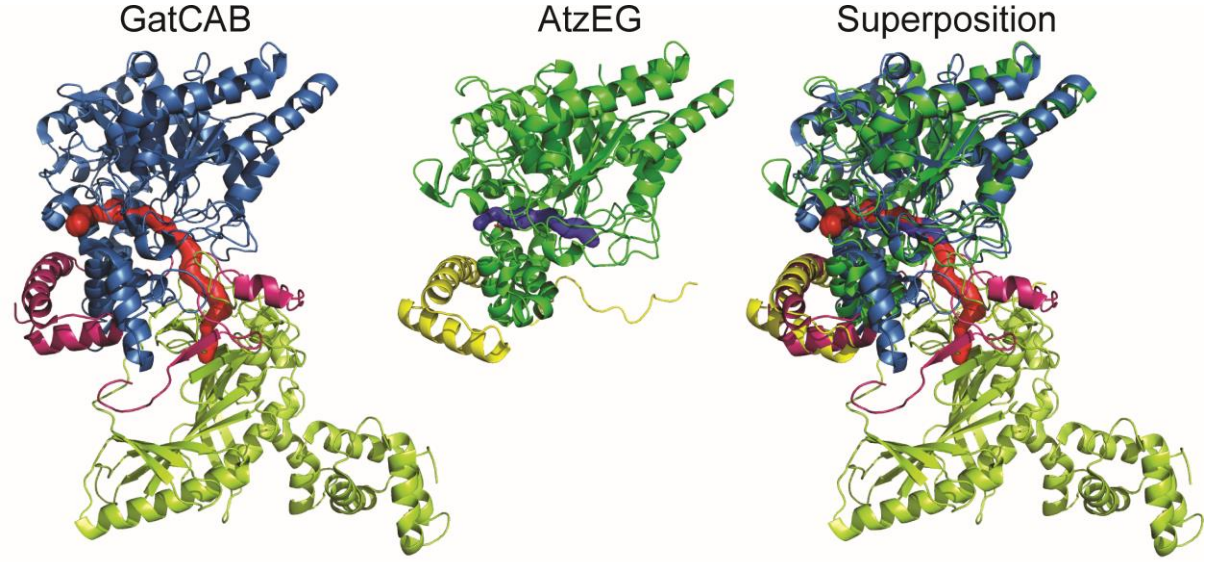

Figure 6. Conservation of the transamidosome ammonia tunnel in AtzEG. The ammonia tunnel is shown in the GatCAB complex and a comparable tunnel is seen in AtzEG. Superposition of the two protein complexes show that the tunnels overlap for the whole length of the AtzEG tunnel. CAVER (46) was used to generate this figure.

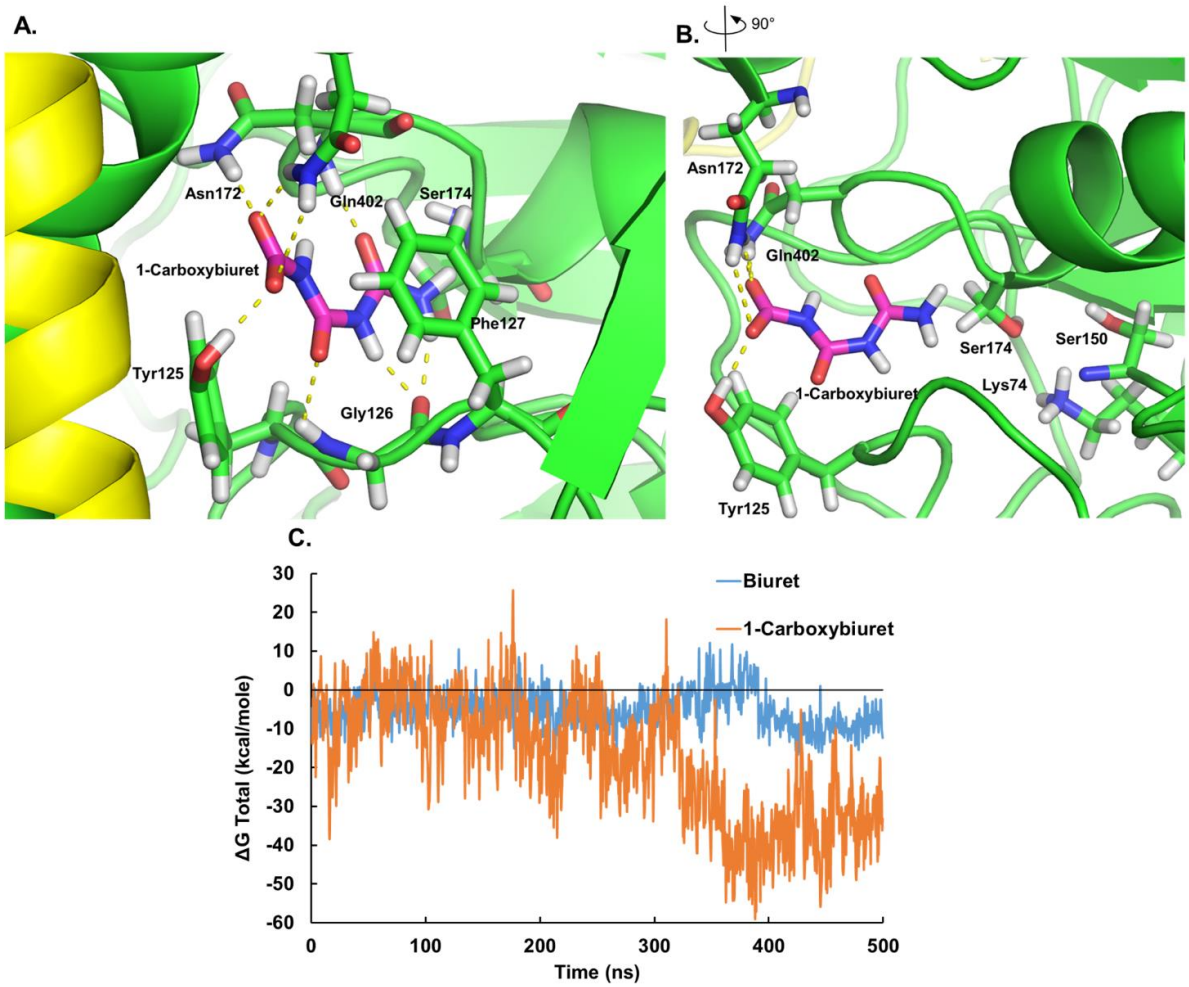

Figure 7. Docking and molecular dynamics of substrate in the AtzE active site. A) 1-carboxybiuret docked into the AtzE active site in its most stable configuration (ΔG minimal). 1-carboxybiuret is shown in magenta, AtzE is shown in green and AtzG is shown in yellow. Hydrogen bonds between the protein and substrate are shown as dashed lines; B) A 90° rotation of panel A showing the position of the catalytic amino acids Ser174, Ser150 and Lys74, for clarity the stick representation of amino acids Gly126, Phe127 have been removed; C) Variation in ΔG (kcal/mole) as a function of time (ns) during a molecular dynamics simulation of biuret and 1-carboxybiuret docked into the active site of AtzEG. The lowest ΔG total reached during the 1-carboxybiuret MMPBSA is -59.0 at 388 ns of simulation, whereas the lowest ΔG total for biuret is only of -16.7 at 425 ns of simulation. While 1-carboxybiuret adopts a stable conformation over time, biuret does not and escapes the active site at t=400 ns.

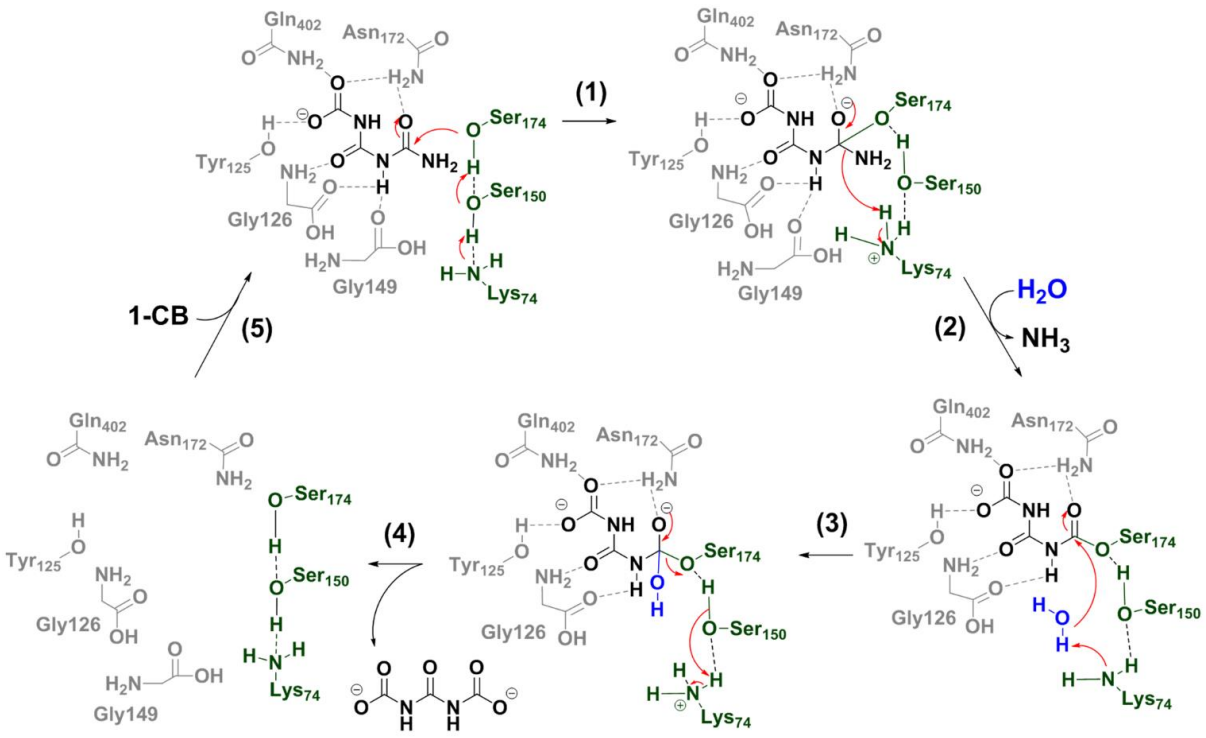

Figure 8. The AtzE reaction mechanism. The catalytic triad (Lys74, Ser150 and Ser174; shown in green) form a hydrogen bonding network with the substrate, Lys74 activates the catalytic (Ser174) *via* the cis-Ser150 bridging ligand. Ser174 performs a nucleophilic attack on the terminal amide of the substrate, forming the covalent acyl-enzyme intermediate and releasing ammonia. Lys74 then acts as a general base, activating water (blue) and leading to the hydrolysis of the acyl enzyme, releasing 1,3-dicarboxyurea and regenerating the active site.

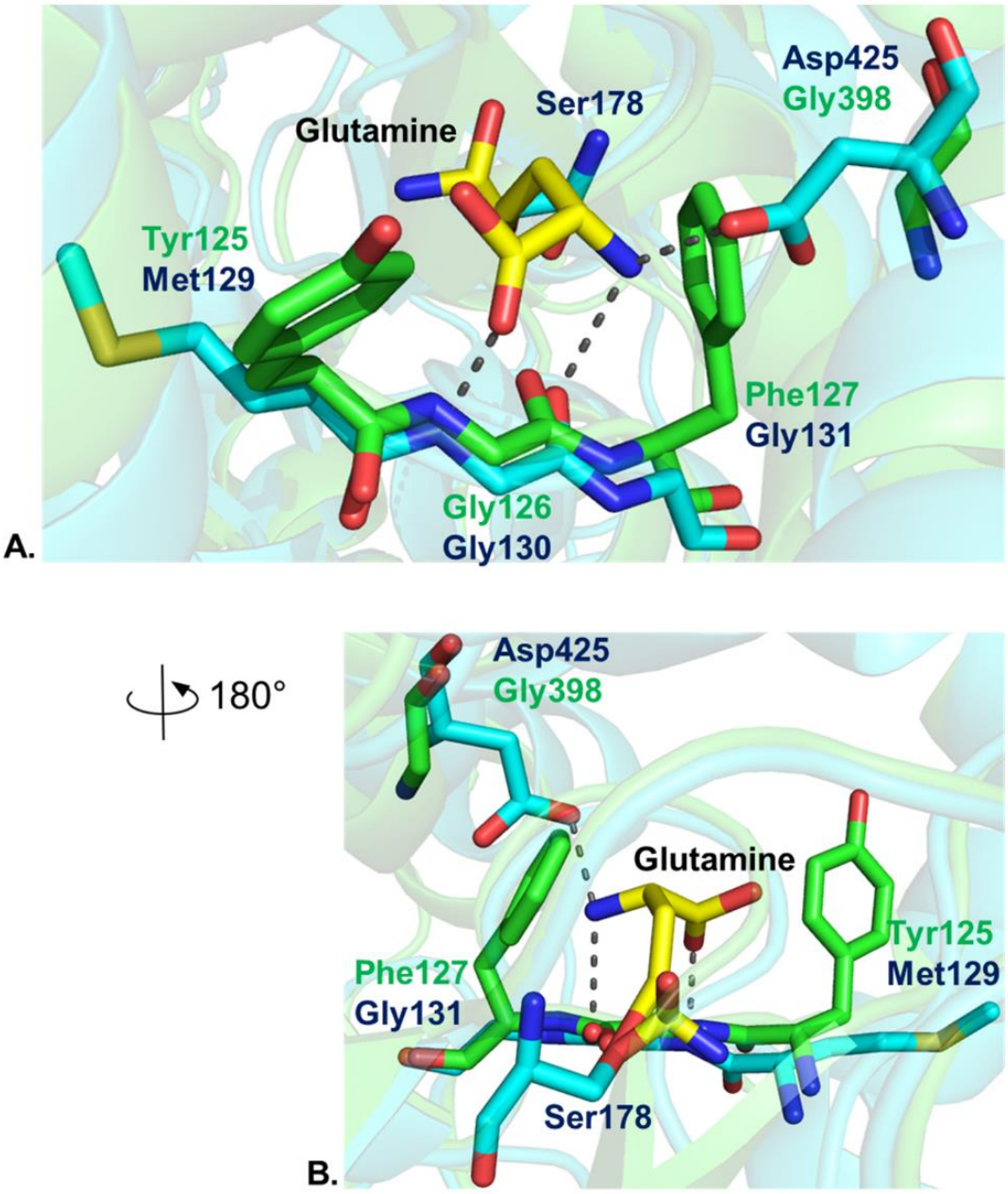

Figure 9. Comparison of the active sites of AtzE and GatA bound to 1-carboxybiuret and glutamine, respectively. The AtzE active site (green) is shown superposed on the GatA active site (cyan). The GatA substrate (glutamine; yellow) is also shown. Hydrogen bonds between GatA and glutamine are shown as dotted lines. Panels A and B are rotated 180° relative to each other to highlight the positions of the alpha amine of glutamine and Phe127 in AtzE.

An unexpected vestigial protein complex reveals the evolutionary origins of an s-triazine catabolic enzyme

Lygie Esquirol, Thomas S. Peat, Matthew Wilding, Jian-Wei Liu, Nigel G. French, Carol J. Hartley, Hideki Onagi, Thomas Nebl, Christopher J. Easton, Janet Newman and Colin Scott

J. Biol. Chem. published online March 9, 2018
