Enzyme Catalysis and Regulation:

Crystal Structure of *Helicobacter pylori* Formamidase AmiF Reveals a Cysteine-Glutamate-Lysine Catalytic Triad

Chiu-Lien Hung, Jia-Hsin Liu, Wei-Chun Chiu, Shao-Wei Huang, Jenn-Kang Hwang and Wen-Ching Wang

J. Biol. Chem. 2007, 282:12220-12229.
doi: 10.1074/jbc.M609134200 originally published online February 16, 2007

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 16, pp. 12220–12229, April 20, 2007

Crystal Structure of *Helicobacter pylori* Formamidase AmiF Reveals a Cysteine-Glutamate-Lysine Catalytic Triad*

Received for publication, September 26, 2006, and in revised form, January 18, 2007 Published, JBC Papers in Press, February 16, 2007, DOI 10.1074/jbc.M609134200

Chiu-Lien Hung[‡], Jia-Hsin Liu[‡], Wei-Chun Chiu[‡], Shao-Wei Huang[§], Jenn-Kang Hwang[§1], and Wen-Ching Wang[‡2]

From the [‡]Institute of Molecular and Cellular Biology and Department of Life Science, National Tsing Hua University, Hsinchu 300, Taiwan and the [§]Institute of Bioinformatics, National Chiao Tung University, Hsinchu 300, Taiwan

Helicobacter pylori **AmiF formamidase that hydrolyzes formamide to produce formic acid and ammonia belongs to a member of the nitrilase superfamily. The crystal structure of AmiF was solved to 1.75 Å resolution using single-wavelength anomalous dispersion methods. The structure consists of a homohexamer related by 3-fold symmetry in which each subunit has an α-β-β-α four-layer architecture characteristic of the nitrilase superfamily. One exterior α layer faces the solvent, whereas the other one associates with that of the neighbor subunit, forming a tight α-β-β-α-α-β-β-α dimer. The apo and liganded crystal structures of an inactive mutant C166S were also determined to 2.50 and 2.30 Å, respectively. These structures reveal a small formamide-binding pocket that includes Cys^{166}, Glu^{60}, and Lys^{133} catalytic residues, in which Cys^{166} acts as a nucleophile. Analysis of the liganded AmiF and *N*-carbamoyl D-amino acid amidohydrolase binding pockets reveals a common Cys-Glu-Lys triad, another conserved glutamate, and different subsets of ligand-binding residues. Molecular dynamic simulations show that the conserved triad has minimal fluctuations, catalyzing the hydrolysis of a specific nitrile or amide in the nitrilase superfamily efficiently.**

Utilization of nitrogen from nonpeptide carbon-nitrogen compounds usually involves nitrilase or nitrile hydratase coupled with amidase to produce ammonia for further assimilation or transfer (1). Based on multiple sequence analysis, nitrilases, cyanide hydratases, and aliphatic amidases, along with β-alanine synthases that reduce organic nitrogen compounds are found to possess several conserved sequence features, including an invariant cysteine active residue (2).

The first determined structure of the nitrilase (Nit) domain is from worm NitFhit that shows a tetramer, and each Nit domain has a four-layer α-β-β-α sandwich fold (3). Additionally, a novel Cys-Glu-Lys (CEK) triad that contains the highly conserved cysteine is seen in a solvent-accessible pocket. Crystal structures of *N*-carbamoyl D-amino acid amidohydrolase reveal a homologous fold, in which the CEK triad is also found (4, 5). Structural and site-directed mutagenesis results suggest the roles of the CEK residues; the active cysteine acts as the nucleophile, glutamate mediates the proton transfer, and lysine stabilizes a tetrahedral transition state (6, 7). Such a novel catalytic triad had not been observed in any other hydrolytic enzymes.

In 2001, Pace and Brenner (8) defined these proteins having the strictly conserved CEK site as the nitrilase superfamily. On the basis of sequence similarity and the presence of additional domains, 13 branches can be classified, nine of which have known or deduced specificity for specific nitrile or amide hydrolysis or amide condensation reactions, including nitrilase, aliphatic amidase, *N*-terminal amidase, biotinidase, β-ureidopropionase, carbamoylase, glutaminase domain of glutamine-dependent NAD synthetase, and apolipoprotein *N*-acyltransferase. Subsequently, structures from two members of the nitrilase superfamily (a yeast hypothetical protein with sequence homologous to CN hydralase from yeast (9) and a hypothetical protein from *Pyrococcus horikoshii* (10)) show the characteristic α-β-β-α sandwich architecture.

Given the functional diversity and the wide evolutionary distribution of the nitrilase superfamily members, we sought to investigate the structure-function relationship of *Helicobacter pylori* formamidase AmiF that shows restricted substrate specificity. *H. pylori* is a human gastric pathogen that persistently colonizes the mucus layer overlaying the epithelium of the stomach. AmiF and the other paralogue AmiE, along with urease, play a major role in producing ammonia to protect against gastric acidity or as a nitrogen source or a cytotoxic molecule, enabling *H. pylori* to adapt into such an exclusive environment (11, 12). In response to growth at mild acidic conditions, genes encoding these proteins are up-regulated (13, 14). Both AmiF and AmiE belong to members of aliphatic amidases. The presumed invariant active cysteine is identified as Cys^{165} for AmiE and Cys^{166} for AmiF from site-directed mutagenesis studies (12). Additionally, AmiF is noted to demonstrate restricted substrate specificity in which formamide is its only known substrate. Here we report the 1.75 Å crystal structure of the native AmiF. Crystal structures of an inactive mutant AmiF C166S were also determined in its apo and liganded forms. Structural analysis reveals a small binding pocket that has the key CEK catalytic residues to reduce the organic nitrogen compounds efficiently.

* This work was supported by National Science Council, Taiwan, Grants NSC95-2313-B-007-001, NSC94-2313-B-007-002, NSC94-3112-B-007-005, and NSC95-3112-B-007-002. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2DYU, 2E2K, and 2E2L) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[1] To whom correspondence may be addressed. E-mail: jkhwang@faculty.nctu.edu.tw.

[2] To whom correspondence may be addressed. E-mail: wcwang@life.nthu.edu.tw.

EXPERIMENTAL PROCEDURES

Cloning, Site-directed Mutagenesis, Expression, and Purification—The gene (*amiF*) encoding formamidase was amplified from chromosomal DNA of *H. pylori* 26695 by PCR using *pfu* DNA polymerase and inserted into the pQE30 expression vector to generate pQE30-AmiF. Primers (AmiF-F, 5′-CGCGGATCCATGGGAAGTATCGGTAGT-3′); AmiF-R, 5′-CCCAAGCTTGGGTTATTTCCCAAAACG-3′) that contain sequences for the BamHI site and HindIII site, respectively, were designed based on the nucleotide sequence of the reported *amiF* gene of *H. pylori* 26695 (accession number NC_000915) (15).

Site-directed mutagenesis was performed essentially using the method of overlap extension PCR with plasmid pQE30-AmiF as the template (16, 17). All mutations were confirmed by sequencing of the whole ligated PCR fragment.

Expression of wild-type AmiF or the C166S mutant in *Escherichia coli* JM109 cells transformed with pQE30-AmiF or pQE30-C166S was induced at 37 °C. Bacterial pellets were fractionated, and soluble proteins in cytosolic fractions were collected. The expressed protein with a His_6 tag was purified by immobilized nickel-ion chromatography, followed by MonoQ ion exchange chromatography (GE Healthcare), and then analyzed by SDS-PAGE to verify the purity. The protein concentration was assayed according to the Bradford method with bovine serum albumin as a standard (18).

Expression of selenomethionine-labeled AmiF (Se-Met-AmiF)[3] was done essentially described by Yang *et al.* (19) and Van Duyne *et al.* (20). In brief, an overnight culture of transformed *E. coli* JM109 was grown in 1 liter of LB medium and 100 mg/liter ampicillin. The cells from the overnight culture were harvested and resuspended in 1 liter of M9 medium (12.8 g/liter Na_2HPO_4, 3 g/liter K_2HPO_4, 0.5 g/liter NaCl, 1 g/liter NH_4Cl). Twenty ml of M9-add (20% glucose, 1 M $MgSO_4$, 1 M $CaCl_2$, 0.5% thiamine, 100 mg/liter ampicillin) were then added per liter of culture. When the absorbance at 600 nm reached $A_{600\ nm} = 0.6$, 10 ml of Met-shutdown mix (100 mg/liter lysine hydrochloride, 100 mg/liter threonine, 100 mg/liter phenylalanine, 50 mg/liter leucine, 50 mg/liter isoleucine, 50 mg/liter valine, 120 mg/liter seleno-L-methionine) were added per liter of culture. The culture was induced by isopropyl 1-thio-β-D-galactopyranoside and incubated overnight at 37 °C. Purification of Se-Met-AmiF was performed based on the procedures for the native AmiF.

Crystallization—All crystallization was performed by the hanging drop vapor diffusion method at 20 °C. Initial crystallization conditions were screened using Crystal Screen I and II kits (Hampton Research) and Clear Strategy Screen I and II kits (Molecular Dimension). One μl of protein solution (10 mg/ml) in 50 mM Tris-HCl, pH 8.0, was mixed with an equal volume of the well precipitant. The best diffracting AmiF crystals were found in a modified solution containing 0.1 M sodium cacodylate (pH 6.5), 0.15 M potassium thiocynate, and 20% polyethylene glycol (PEG) 550 MME. The crystal grew as rods and reached a maximum size of about 0.6 × 0.1 × 0.1 mm within 2 days at 20 °C. The native crystal was characterized as belonging to space group R3, with the unit cell dimensions $a = b = 147.21$, $c = 72.27$ Å. There are two molecules per asymmetric unit.

Crystals of the Se-Met-AmiF protein were obtained in a condition (0.1 M sodium cacodylate (pH 6.0), 0.8 M sodium formate, 8% PEG 20,000, and PEG 550 MME) using Se-Met-AmiF (20 mg/ml). Rod-shaped crystals with a maximum size of about 0.6 × 0.1 × 0.1 mm grew within 4 days at 20 °C. The Se-Met-AmiF crystal also belongs to space group R3 with unit cell dimensions of $a = b = 147.80$ Å and $c = 72.78$ Å.

Crystals of C166S were obtained in a solution containing 0.1 M sodium acetate (pH 5.0), 0.2 M lithium sulfate, and 14% PEG 2000 MME. Of two crystal forms (thin plates or rods), only the thin plate crystal diffracted beyond 3 Å using the in-house x-ray source. The thin plate crystal was characterized as space group C2, with the unit cell dimensions of $a = 117.72$, $b = 130.53$, $c = 144.59$ Å, and $\beta = 99.44°$. There are six molecules per asymmetric unit.

Attempts to obtain liganded C166S crystals using the soaking method failed. After extensive trials, the C166S complex crystal was finally obtained by the co-crystallization method in a solution containing 0.1 M sodium acetate (pH 5.0), 0.2 M lithium sulfate, 6% PEG 1500, and 3 mM formamide. The complex crystal grew as a rod and was characterized as the space group $P2_1$ with cell dimensions of $a = 83.09$, $b = 151.80$, $c = 89.08$ Å, and $\beta = 114.99°$. There are six molecules per asymmetric unit.

Data Collection—Prior to data collection, crystals were dipped into Fomblin® cryoprotectant oil for several seconds and then flash-frozen in a liquid nitrogen stream. Diffraction data were collected using a MSC X-Stream cryosystem and an R-AXIS IV++ image plate system with double mirror-focused CuKα x-ray radiation generated from a Rigaku RU-300 rotating anode generator at the Macromolecular X-ray Crystallographic Laboratory of National Tsing Hua University, Hsinchu, Taiwan. The 1.75 Å native data set, the 2.27 Å Se-Met-AmiF data set, the 2.50 Å C166S data set, and the 2.30 Å C166S complex data set were collected on the BL12B2 Taiwan beamline at SPring-8, Japan, using an ADSC Quantum 4R CCD detector. All data sets were collected at −165 °C and processed with the HKL/HKL2000 suite (21). Data collection statistics are shown in Table 1.

Structure Determination and Refinement—The AmiF structure was solved by single wavelength anomalous dispersion methods. Twelve selenium sites were located with the program SOLVE (22). Phase determination, density modification, and automated model building with SOLVE/RESOLVE (22, 23) led to an interpretable density map and initial model. The atomic model was further built using O (24). Crystallographic refinement was carried out using the maximum likelihood target function embedded in program REFMAC5 (25, 26). The native AmiF model was obtained using MOLREP (27) and refined using REFMAC5 (25) coupled to ARP/wARP (28), which was used to add water molecules automatically. The protein model was assessed using the program PROCHECK (29). The apo and liganded C166S structures were determined by the molecular

[3] The abbreviations used are: Se-Met, selenomethionine; Nit, nitrilase; PEG, polyethylene glycol; GNM, Gaussian network model; r.m.s., root mean square; MME, monomethyl ether.

TABLE 1
Crystallographic data

Parameters	Values			
	AmiF		C166S	
	Native AmiF	Se-Met-AmiF	C166S	C166S complex
Data collection and phasing				
Space group	R3	R3	C2	$P2_1$
Cell dimensions				
a (Å)	147.21	147.80	117.72	83.09
b (Å)	147.21	147.80	130.53	151.80
c (Å)	72.27	72.78	144.59	89.08
			$\beta = 99.4°$	$\beta = 115.0°$
Wavelength (Å)	0.9800	0.9791	0.9537	1.0000
Resolution (Å)	30.0-1.75	30.0-2.27	30.0-2.50	30.0-2.30
Highest resolution shell (Å)	1.81-1.75	2.35-2.27	2.59-2.50	2.38-2.30
Completeness (%)[a]	99.7 (100.0)	99.1 (95.8)	99.8 (99.8)	91.4 (92.5)
Average I/σ (I)	15.2 (2.7)	10.7 (2.7)	8.4 (4.6)	18.0 (12.4)
Number of unique reflections	55,500		70,942	77,740
Redundancy	3.2	2.5	3.3	3.4
R_{merge} (%)[b]	7.8 (49.8)	6.5 (31.0)	3.6 (10.0)	5.4 (35.4)
Overall figure of merit (acentric/centric)[c]		0.68/0.87		
Solvent content (%)	39.8		51.1	46.1
Refinement				
Resolution range (Å)	30.0-1.75		30.0-2.50	30.0-2.30
Number of atoms				
Protein	5098		14,901	15,007
Solvent	713		850	735
Ligand				9
Average *B*-factor (Å^2)				
Protein	21.8		9.5	36.5
Water	36.1		9.7	35.8
Ligands				41.1
R factor[d]	0.174		0.258	0.253
R_{free}[e]	0.215		0.294	0.285
r.m.s. deviation bond length (Å)	0.006		0.004	0.006
r.m.s. deviation bond angles (degrees)	1.037		0.882	0.947
Estimated coordinate error (Å)	0.083		0.198	0.228
Ramachandran analysis (%)[f]				
Favored/Allowed/Generous/Disallowed	88.2/10.6/0.7/0.4		84.3/13.6/1.7/0.4	89.0/10.3/0.4/0.4

[a] Values in parentheses refer to statistics in the highest resolution shell.
[b] $R_{merge} = \Sigma|I_{obs} - \langle I \rangle|/\Sigma I_{obs}$.
[c] Figure of merit = $|F_{best}|/|F|$.
[d] $R = \Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structure-factor amplitudes, respectively.
[e] R_{free} was computed using 5% of the data assigned randomly.
[f] Estimated standard uncertainties based on maximum likelihood.

replacement methods using the native AmiF structure as a search model by MOLREP (27) and refined by REFMAC5 (25) coupled to ARP/wARP (28).

Structural Comparisons—Structure comparisons with *N*-carbamoyl-D-amino acid amidohydrolase (Protein Data Bank entry codes 1ERZ and 1FO6) (4, 5), the Nit domain of NitFhit (Protein Data Bank entry code 1EMS) (3), and the putative CN hydrolase from yeast (Protein Data Bank entry code 1F89) (10), were carried out using the program LSQMAN in O (24) to superimpose Cα atoms.

Combined sequence and secondary structure alignments and figure preparation (Fig. 5*A*) were done with the program ESPript (30). Structural figures (Figs. 1, 2*A*, 3, 4, and 6) were prepared with the program PyMol (42) (available on the World Wide Web at www.pymol.org). Fig. 5*B* was prepared with the programs MOLSCRIPT (31) and RASTER3D (32).

Molecular Dynamic Simulations and Gaussian Network Model—Dynamic simulations were performed with the GROMACS 3.3 program running on a Fedora Linux system essentially as previously described (17, 33). All simulations were carried out for 2 ns at 300 K. The first 1-ns run was used to attain equilibrium, whereas the trajectories of the last 1 ns were used for analysis. The low frequency motions were calculated based on the Gaussian network model (GNM) (34, 35).

RESULTS AND DISCUSSION

Structure Description of the Native AmiF—The 1.75 Å resolution electron density map of AmiF revealed two molecules (AB) per asymmetric unit. Except for the N-terminal segment (residues 1–12) that contained weak or negative density and could not be built into the model, the main-chain and side-chain atoms from 13 to 334 residues are well defined in A and B subunits. Only one residue, Cys^{166}, lies in the disallowed region of the Ramachandran plot, despite its low B factor. The unusual geometry is stabilized by three hydrogen bonds around this peptide unit: Cys^{166} O···N Asp^{168}, Cys^{166} O···N Gly^{169}, and Cys^{166} O···Nη2 Arg^{188}. The final model was refined to an *R* value of 17.4% (R_{free} = 21.5%; Table 1). The average *B*-factor for all polypeptide atoms is 23.6 Å^2.

In the crystal structure, AmiF is a hexamer (Fig. 1): Subunit A and Subunit B associate into an AB dimer related by a noncrystallographic 2-fold axis (Fig. 1*B*). Along the 3-fold axis, three AB dimers assemble into a hexamer (A1B1A2B2A3B3) (Fig. 1*A*). The overall shape of each trimer (A1A2A3 or B1B2B3) is approximately an equilateral triangular prism (~95 Å along the prism axis).

Each monomer folds into a four-layer α-β-β-α structure, characteristic of members in the nitrilase superfamily (Fig. 2, *A*

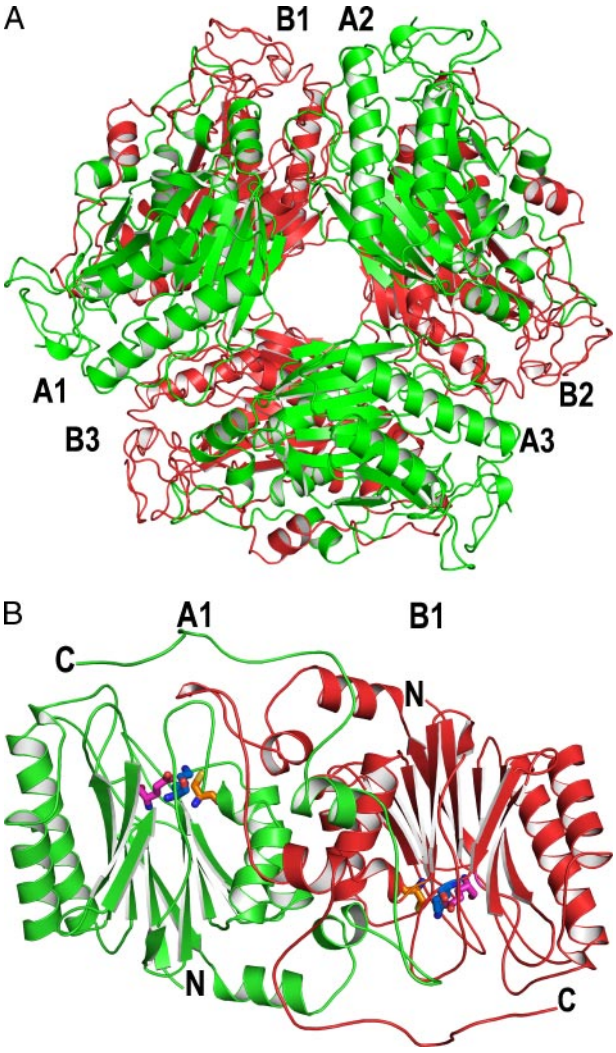

FIGURE 1. **Structure of AmiF.** *A*, *ribbon representation* of the hexameric AmiF structure $(AB)_3$ viewed down the 3-fold axis. Three A subunits (A1, A2, and A3) are depicted in *green*, and three B subunits (B1, B2, and B3) are in *red*. *B*, *ribbon representation* of the dimeric AmiF structure (AB) viewed down a 2-fold axis. E60 (*red*), K133 (*blue*), and C166 (*orange*) are drawn as *stick models*. The oxygen, nitrogen, and sulfur atoms are *colored* in *red*, *dark blue*, and *yellow*, respectively.

and *B*). One exterior α layer (α1 and α2) is solvent-accessible to shield the inner β sheets from the solvent, whereas the other α layer (α3–α7) associates with the neighboring Subunit B. One side of the four-layer sandwich contains short turns or loops to connect the secondary elements, whereas the other side has several long loops, particularly the β2-η1-α2, β5–β6, and β9–β10 segments.

Buried residues at the A-B intersubunit interface are largely from residues of the β5–β6 loop, near or in those of α3–α7 helices and the *C*-terminal region of each monomer. The symmetric assembly (α3 *versus* α3, α4 *versus* α4, and α7 *versus* α7) forms a tightly associated α-β-β-α-α-β-β-α dimer (Fig. 1*B*). In total, there are 456 interactions within 3.8 Å, including 36 hydrogen bonds and 13 salt bridges. The total solvent-accessible surface of the AB dimer is ~29,700 $Å^2$, calculated by protein interfaces, surfaces, and assemblies service PISA at the European Bioinformatics Institute (available on the World Wide Web at www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Of the subunit surface area, 28.8% is used for the formation of the dimer.

Crystal Structures of C166S—To obtain liganded structures, crystallization trials using the co-crystallization or soaking methods in the presence of various concentrations of formamide were attempted. Of various structures, only one showed additional nonpeptide electron density near Cys^{166} (data not shown). Given the relatively large peak, either formamide or a solvent cacodylate molecule could be docked into this density in either subunit. Nonetheless, a piece of residual density that could not be ignored is present in either model, indicating a possibility that formamide and cacodylate co-exist in this density. However, we were not able to model both properly into this density, possibly due to the disordered formamide/cacodylate displacement.

To further clarify the catalytic site, we have obtained the crystal structure of an inactive mutant, C166S, in a crystallization condition without cacodylate. The model consists of six subunits. Like the wild-type structure, the N-terminal segment (residues 1–12) is not visible in each subunit. Disordered regions that could not be defined include the 283–288 region and residues in loops (A194, A249, A299, B194, B249, B298, C194, C226, C300, D194, D232, D248, D296, D298, E194, E298, F194, and F248). Crystallographic data are given in Table 1.

Crystal Structure of C166S·Formamide Complex—After extensive trials, the complex crystal was also obtained by the co-crystallization method in a solution containing 3 mM formamide. The structure was then solved by the molecular replacement methods, showing six subunits. Residues that could not be defined include the following: N-terminal segment (residues 1–12), Ala^{194}, Phe^{195}, and a loop region (residues 223–227).

As shown in Fig. 3, the electron density maps show a nonpeptide electron density nearby the catalytic triad of the C166S complex structure as compared with a water molecule in the apo structure (Fig. 3). This density is defined as a formamide molecule, in which two possible directions of formamide are considered (model 1 and model 2; Fig. 3, *B* and *C*). In either model, formamide is modeled in three subunits (ACF) and is omitted in the other subunits due to the weak density.

Given the opposite orientation in two models, formamide interacts with nearby polar groups distinctively (Table 2). It is noted that the oxygen atom of formamide is in close proximity to the main-chain nitrogen atom of His^{167} (3.02 Å in Subunit A) and the Lys^{133} Nζ atom (3.15 Å in Subunit A) in model 2. These interactions may together stabilize the transition state intermediate that bears a negative charge (discussed under "Proposed Mechanism"). Model 1, in contrast, lacks such contacts. Additionally, superposition between liganded C166S and *N*-carbamoyl D-amino acid amidohydrolase (6) structures reveals that formamide of model 2 superimposes relatively well with the reactive atoms of the other substrate, *N*-carbamoyl-D-*p*-hydroxyphenylglycine (discussed under "Conserved Catalytic CEK and Glutamate Residues and Distinct Ligand-binding Residues between AmiF and N-Carbamoyl D-Amino Acid Amidohydrolase"). These results together suggest that model 2 has a correct assignment for formamide. Subsequent analysis is thus conducted based on this model.

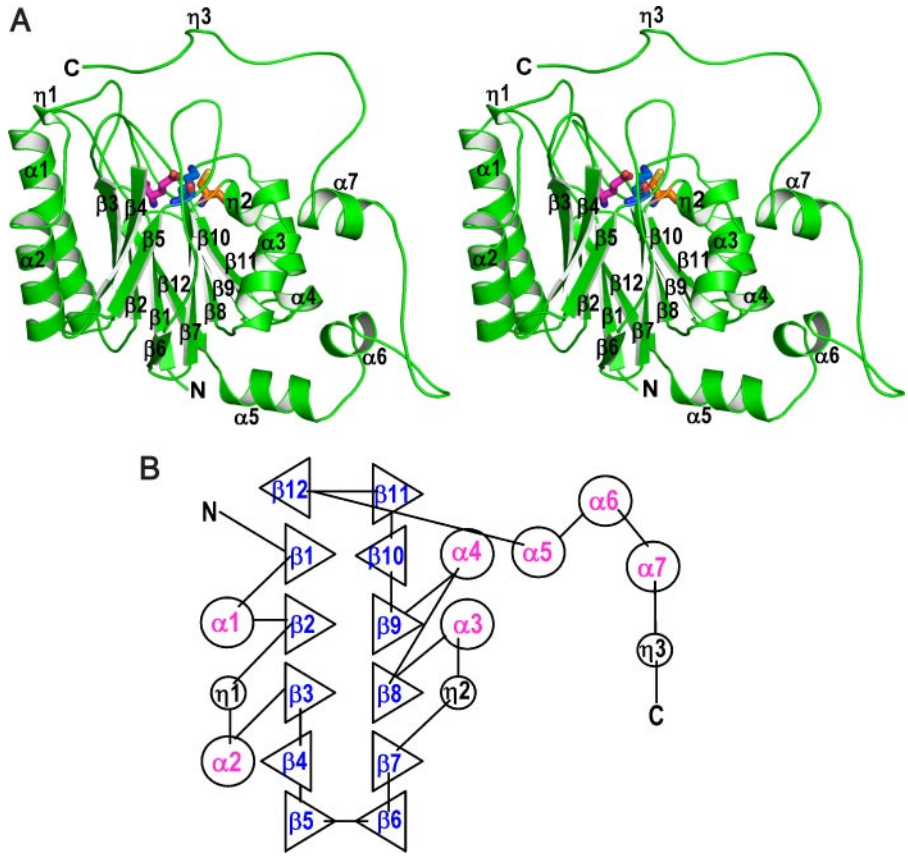

FIGURE 2. *A*, stereoview of the monomeric AmiF. Glu^{60} (*red*), Lys^{133} (*blue*), and Cys^{166} (*orange*) are drawn as *stick models*. The oxygen, nitrogen, and sulfur atoms are *colored* in *red*, *dark blue*, and *yellow*, respectively. *B*, topology of the monomeric AmiF.

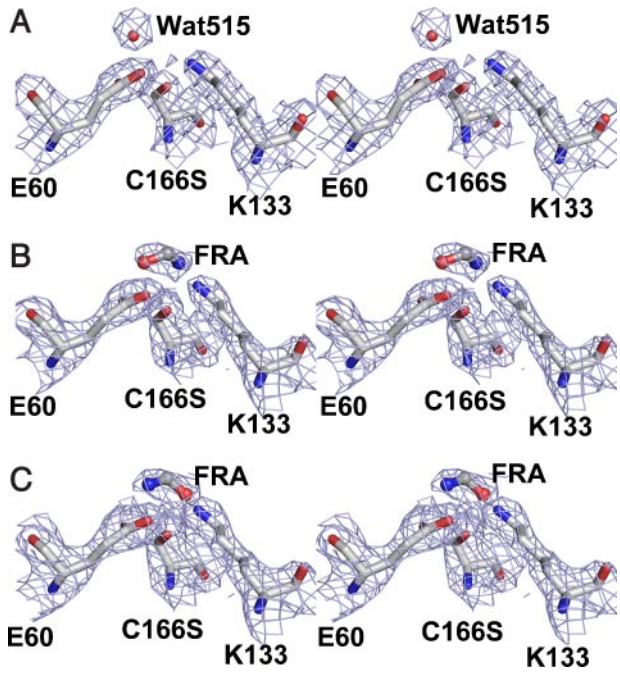

FIGURE 3. **The $2F_o - F_c$ electron density map of the inactive mutant C166S in its apo form (*A*) and model 1 (*B*) and model 2 (*C*) of the liganded C166S complex around C166S, Glu^{60}, and Lys^{133}, contoured at the 1.0-σ level.** The bound formamide (*FRA*) is drawn as *ball-and-stick models*. The carbon, oxygen, and nitrogen atoms are colored in *white*, *red*, and *blue*, respectively.

Like the native AmiF, the C166S crystal structure shows a tight dimer (AB) related by a noncrystallographic 2-fold axis. Along a noncrystallographic 3-fold axis, three AB dimers pack as a hexamer. Similarly, the C166S-formamide structure reveals a tight AB dimer and a loosely associated hexamer $(AB)_3$.

Each subunit with or without substrate demonstrates the characteristic α-β-β-α architecture, whereas the AB dimer shows the α-β-β-α-α-β-β-α fold. These results suggest that C166S shares an overall identical wild-type structure. Superposition between the apo and liganded C166S monomers (Subunit A) reveals a higher deviation in three loops (191–195, 221–232, and 295–300) that involve intersubunit interactions and/or crystal packing. The r.m.s. deviation of the superimposed Cα atoms with or without these regions is 0.45 and 0.27 Å, respectively.

The Binding Pocket—Formamide is positioned near the bottom of a pocket surrounded by three long segments (β2-η1-α2-(59–82), β5–β6-(133–152), and β8-α4-(191–197)) and a short 3_{10} turn (η2-(166–170)) between β7 and α3 (Fig. 4*A*). Glu^{60}, Lys^{133}, Trp^{137}, Glu^{141}, C166S, His^{167}, and Tyr^{192} from these regions make contacts with formamide ($\leq$3.8 Å). Three strictly conserved residues, Glu^{60}, Lys^{133}, and C166S, are situated at or close to the C-terminal anchoring point of β2, β5, and β7, respectively.

Superposition of the Cα atoms of the CEK triad between apo and liganded C166S structures shows virtually identical Cα conformation in the binding site region (Glu^{60}, Lys^{133}, Trp^{137}, Glu^{141}, C166S, His^{167}, and Tyr^{192}) with or without the ligand (Fig. 4*B*). Moreover, the side chains of Glu^{60}, Lys^{133}, and C166S superimpose relatively well between apo and liganded forms of C166S. It is noted that a water molecule is positioned between Glu^{60} and C166S (Glu^{60} (ϵ1)-Wat^{515} (O), 2.84 Å; Glu^{60} (ϵ2)-Wat^{515} (O), 3.53 Å; C166S (Oγ)-Wat^{515} (O), 3.24 Å in Subunit A) in the apo form, which is replaced by formamide in the bound form (Fig. 3). The wild-type apo structure also includes a water molecule at the same site. These results suggest that Glu^{60}, Lys^{133}, Trp^{137}, Glu^{141}, C166S, His^{167}, and Tyr^{192} are the ligand-binding residues. Upon binding to a formamide molecule, the reactive atom (the Sγ atom of Cys^{166} in AmiF) may then react toward the carbon atom of formamide for the presumed nucleophilic attack.

Comparison with Homologous Structures—Based on analysis with DALI (34), all members of the nitrilase superfamily exhibit structural homology, including a hypothetical protein from *P. horikoshii* that shares 24% sequence identity (Protein Data

Bank code 1F89, *Z* score = 29.1, r.m.s. deviation = 2.8 Å for 251 residues) (10), NitFhit from *Caenorhabditis elegans* that shares 20% identity (Protein Data Bank code 1EMS, *Z* score = 27.9, r.m.s. deviation = 2.9 Å for 245 residues) (3), *N*-carbamoyl D-amino acid amidohydrolase that shares 29% identity (Protein Data Bank code 1ERZ, *Z* score = 25.4, r.m.s. deviation = 2.8 Å for 239 residues) (4) (Fig. 5*A*). All of these proteins have the same α-β-β-α fold (Fig. 5*B*). Significant structural conservation is seen in the CEK triad among nitrilase members in which each catalytic residue is located in or adjacent to the C-terminal anchoring point of a β strand (Fig. 5*B*) (Glu^{60}, β2; Lys^{133}, β5; Cys^{166}, β7). Moreover, they exist as a dimer or a higher symmetric form in the crystal. These dimers make intersubunit contacts largely from an exterior α layer and the C-terminal region, forming an α-β-β-α-α-β-β-α architecture (Fig. 5*B*).

TABLE 2
Binding of formamide in Subunit A of the C166S complex

Formamide	Model 1		Model 2	
	Atom	Distance (Å)	Atom	Distance (Å)
N	Lys^{133} Nζ	3.80	Glu^{60} Oϵ1	2.98
	Ser^{166} Oγ	2.97	Glu^{60} Oϵ2	3.66
	His^{167} N	3.43	Lys^{133} Nζ	3.45
			Glu^{141} Oϵ2	2.98
			Ser^{166} Oγ	2.87
C	Lys^{133} Nζ	3.40	Lys^{133} Nζ	3.66
	Glu^{141} Oϵ2	3.17	Glu^{141} Oϵ2	3.47
	Ser^{166} Oγ	3.04	Ser^{166} Oγ	3.12
O	Glu^{60} Oϵ1	2.74	Lys^{133} Nζ	3.15
	Glu^{60} Oϵ2	3.69	Ser^{166} Oγ	2.65
	Lys^{133} Nζ	3.75	His^{167} N	3.02
	Glu^{141} Oϵ2	3.39		
	Ser^{166} Oγ	2.83		
	Tyr^{192} N	3.76		

Conserved Catalytic CEK and Glutamate Residues and Distinct Ligand-binding Residues between AmiF and N-Carbamoyl D-Amino Acid Amidohydrolase—Other liganded structures in the nitrilase superfamily are only available for *N*-carbamoyl D-amino acid amidohydrolase (6). It was thus of interest to compare their binding pockets. As seen in Figs. 5*B* and 6*A*, the presumed catalytic CEK triad of AmiF superimposes relatively well with that of *N*-carbamoyl D-amino acid amidohydrolase. Comparable interactions between the polar groups that may facilitate polarization of the Sγ atom of the cysteine residue are found in the bound form (in Subunit A of the liganded C166S form: Glu^{60} (Oϵ2)-Lys^{133} (Nζ), 2.95 Å; Glu^{60} (Oϵ1)-C166S (Oγ), 2.70 Å; Glu^{60} (Oϵ2)-C166S (Oγ), 3.42 Å; Lys^{133} (Nζ)-C166S (Oγ), 2.95 Å) (6). These interactions together enable the active nucleophile to polarize so as to attack the susceptible C–N bond of formamide or the carbamoyl moiety efficiently for the hydrolytic reaction.

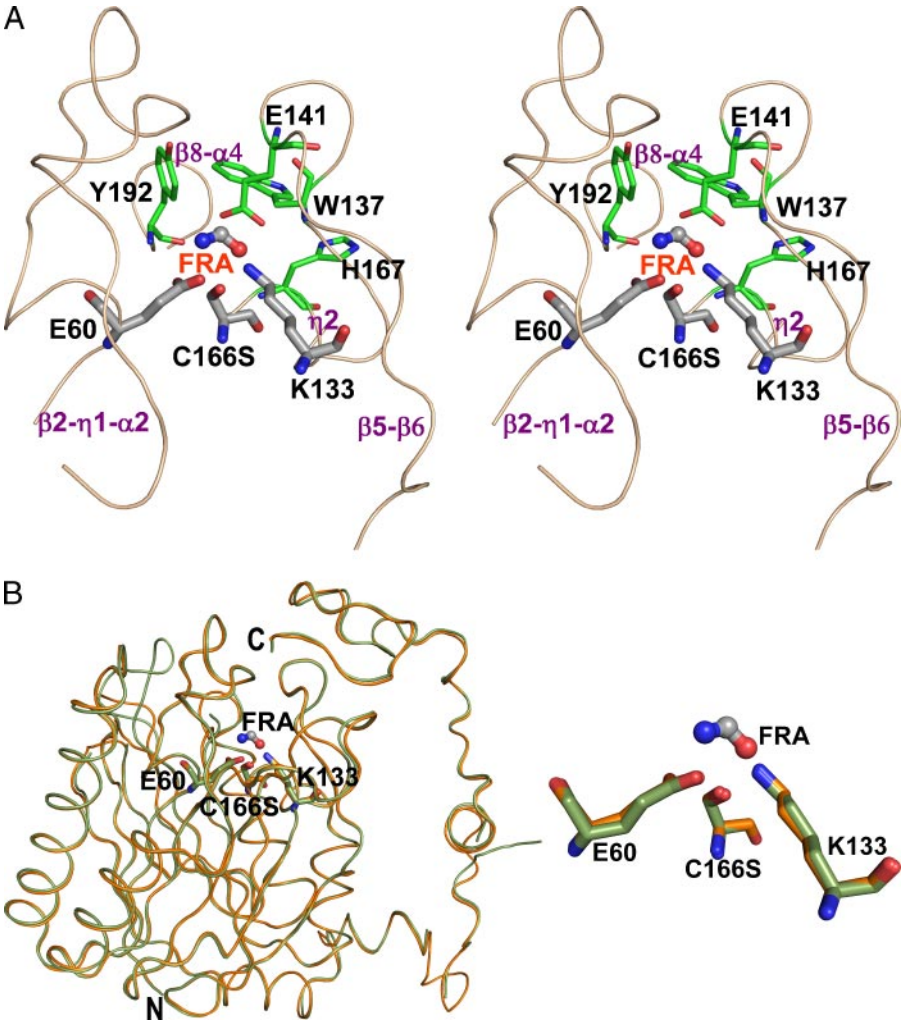

FIGURE 4. **The binding pocket of the C166S·formamide complex.** *A*, stereoview of the C166S·formamide binding pocket. Loops (β2-η1-α2 (residues 58–82), β5–β6 (residues 133–152), and β8-α4 (residues 191–197)) and a short 3_{10} turn between β7 and α3 (residues 166–170) enclosing formamide are colored *beige*. Glu^{60}, Lys^{133}, and C166S are shown as *heavy gray sticks*, whereas other ligand-binding residues are shown as *green sticks*. *B*, superposition between apo-C166S (*green*) and C166S·formamide (*orange*) subunits. C166S·Glu^{60}-Lys^{133} residues are shown as *heavy sticks*. Formamide (*FRA*) in *A* and *B* is drawn as *ball-and-stick models*. The carbon, oxygen, and nitrogen atoms are *colored white*, *red*, and *blue*, respectively.

Apart from the triad, Glu^{141} of AmiF superimposes relatively well with Glu^{146} of *N*-carbamoyl D-amino acid amidohydrolase (Fig. 6), although their sequence is at a nonconserved segment between β5 and β6 in the nitrilase superfamily (Fig. 5*A*). For Glu^{141} of AmiF, it sits just next to a β turn, and its carboxyl group of the side chain contacts with the polar groups of Glu^{60} and Lys^{133}. It is noted that Glu^{141} also contacts with formamide (Table 2). Similarly, for Glu^{146} of *N*-carbamoyl D-amino acid amidohydrolase, it is just next to a 3_{10} helix and makes hydrogen bonds with side chains of Glu^{47} and Lys^{127} and the ligand. These results suggest that this conserved glutamate is crucial to maintain the side-chain geometry of the catalytic CEK triad as well as to facilitate the docking of a substrate

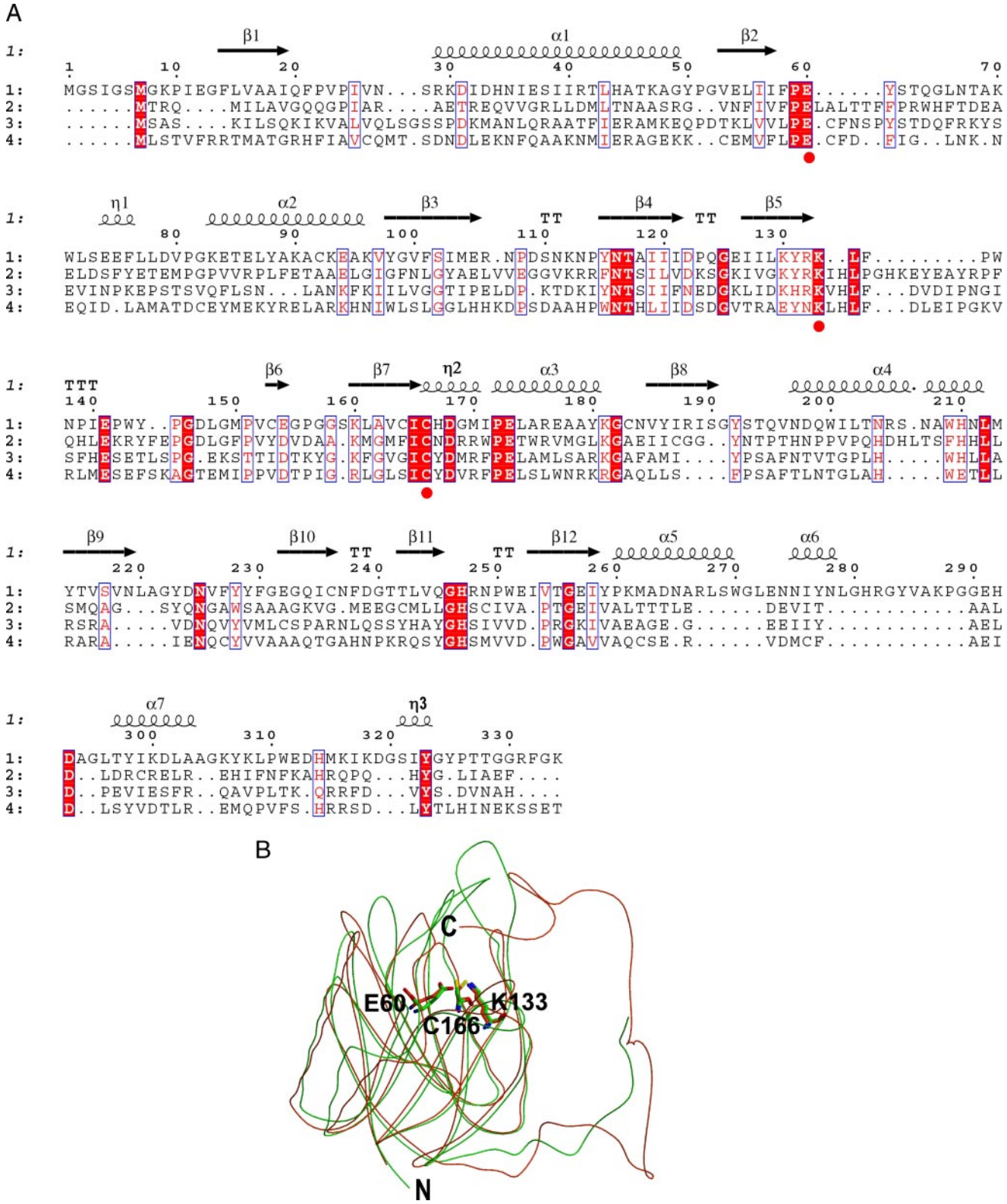

FIGURE 5. **Structural comparison of AmiF with nitrilases.** *A*, structure-based alignment of members in the nitrilase superfamily. *1*, AmiF from *H. pylori*; *2*, *N*-carbamoyl-D-amino acid amidohydrolase from *Agrobacterium radiobacter*; *3*, a putative CN hydrolase from yeast; *4*, the Nit domain of NitFhit from *C. elegans*. The locations of the secondary structural elements are shown *above* the sequence. The $\beta1$–$\beta12$ strands and α-helices ($\alpha1$–$\alpha6$) are numbered from the N terminus. *TT*, β-turns; η ($\eta1$–$\eta3$), 3_{10} helix. Catalytic residues are indicated as *red circles*. *B*, superimposed structures of AmiF (*red*) and *N*-carbamoyl-D-amino acid amidohydrolase (*green*). Presumed cysteine-glutamate-lysine residues corresponding to Glu^{60}, Lys^{133}, and Cys^{166} in AmiF are shown by *stick structures* (Glu^{60}, Lys^{133}, and Cys^{166} in AmiF; Glu^{47}, Lys^{127}, and Cys^{172} in *N*-carbamoyl-D-amino acid amidohydrolase).

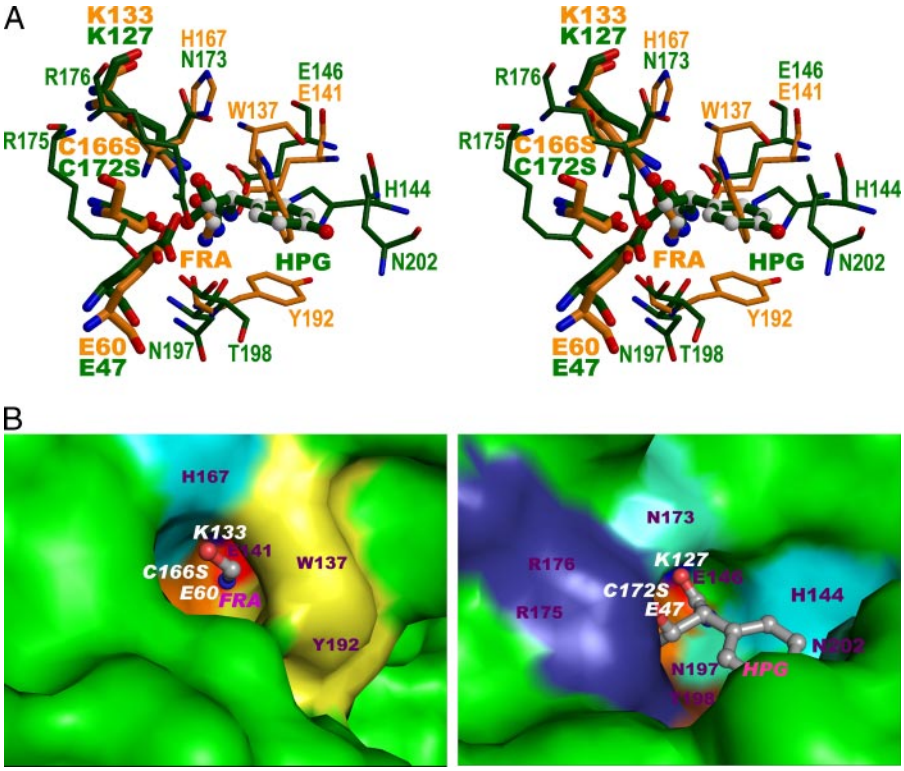

FIGURE 6. *A*, superposition of the binding sites between AmiF (C166S)·formamide (*orange*) and *N*-carbamoyl D-amino acid amidohydrolase (C172S)·*N*-carbamoyl-D-*p*-hydroxyphenylglycine (*HPG*) (*green*) complex structures. Catalytic CEK residues are shown as *heavy sticks*, whereas other ligand-binding residues are shown as *thin sticks*. Formamide (*FRA*) and hydroxyphenylglycine are drawn as *ball-and-stick models*. The carbon, oxygen, nitrogen, and sulfur atoms are *colored white*, *red*, *blue*, and *yellow*, respectively. *B*, molecular surfaces of AmiF (C166S)·FRA (*left*) and *N*-carbamoyl D-amino acid amidohydrolase (C172S)·HPG (*right*) complexes. Surfaces are *colored* by the type of residues: glutamate (*red*), cysteine (*orange*), lysine (*blue*), histidine (*cyan*), arginine (*dark blue*); serine (*yellow orange*), asparagine (*green*), threonine (*cyan*), tryptophan (*brown*), and tyrosine (*yellow*). FRA and HPG are drawn as ball-and-stick models. The carbon, oxygen, and nitrogen atoms are *colored white*, *red*, and *blue*, respectively.

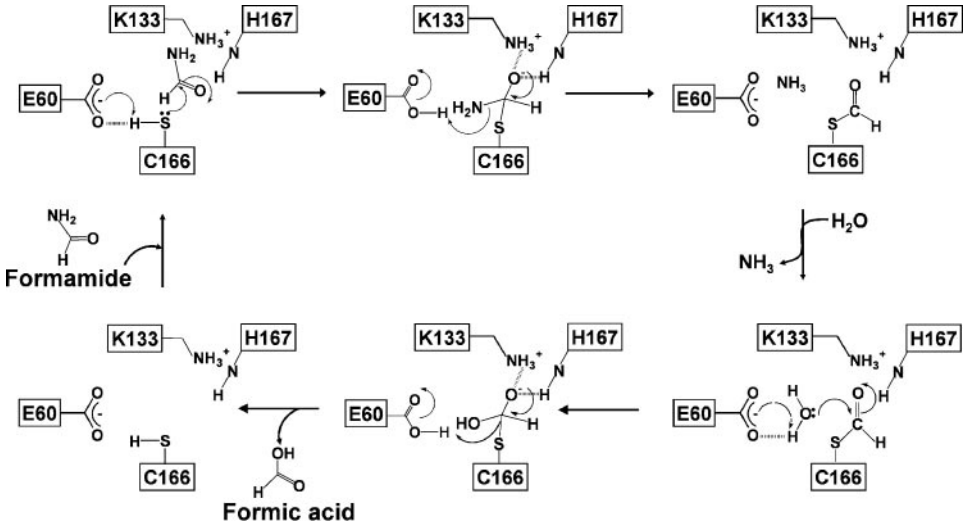

FIGURE 7. **Proposed acylation-deacylation catalytic mechanism of AmiF.** The carbon atom of formamide is attacked by a nucleophilic catalyst (Cys^{166}). An acyl intermediate is formed, and an ammonia molecule is released via the collaboration by Glu^{60} and Lys^{133}. The subsequent nucleophilic attack is presumed by a water molecule to begin the deacylation reaction.

into the right site. Structural comparison reveals that Glu^{143} of the Nit domain of NitFhit from *C. elegans* also occupies at the similar position and contacts with Lys^{127} via a hydrogen bond (Glu^{143} (Oϵ2)-Lys^{127} (Nζ), 2.75 Å), which indicates a similar role of Glu^{143} in NitFhit. On the other hand, the side chain of Glu^{144} in a putative yeast CN hydrolase points away from the catalytic triad and may possess a different function.

Apart from the CEK and Glu^{146}, *N*-carbamoyl D-amino acid amidohydrolase has a subset of residues contacting with the carboxyl group of the ligand (Asn^{173}, Arg^{175}, Arg^{176}, Asn^{197}, and Thr^{198}) (Fig. 6, *A* and *B*, *right*). This enzyme also possesses a relatively large space for a bulky side-chain substrate (D-phenylglycine, D-methionine, and D-leucine) (37, 38).

In AmiF, residues corresponding to bind to the carboxyl moiety in *N*-carbamoyl D-amino acid amidohydrolase (His^{167}, Gly^{169}, Met^{170}, Gly^{191}, and Tyr^{192}) are not conserved (Fig. 6*A*). Such a structural disparity may explain the distinct substrate specificity between *N*-carbamoyl-D-amino acid amidohydrolase and AmiF. Moreover, Trp^{137} and Tyr^{192} form an exterior wall in the other side (Fig. 6*B*, *left*), forming a much smaller room of the active site. As a result, AmiF can only accommodate a small ligand like formamide.

Proposed Mechanism—Given the common CEK residues that point toward the presumed susceptible C–N bond of their ligands between AmiF and *N*-carbamoyl D-amino acid amidohydrolase, these catalytic residues are proposed to share the same catalytic roles: Cys^{166} is the nucleophile; the nearby Lys^{133} plays an essential role in stabilizing the transition state intermediate; and Glu^{60} mediates the proton transfer. A water molecule is noted to sit between Glu^{60} and Cys^{166}, which may be coupled with Glu^{60} for the proton transfer given the low pK_a of the glutamate. *N*-carbamoyl D-amino acid amidohydrolase also possesses a water molecule at a similar site.

The first stage to hydrolyze formamide by AmiF is an acylation reaction (Fig. 7). (i) When a formamide molecule diffuses into the pocket, the substrate would bind onto the CEK triad of AmiF, positioning the formamide in a right orientation via interacting with ligand-binding residues. Glu^{60} may then mediate the pro-

ton transfer from the -SH group of C166, yielding a strongly nucleophilic -S^- ion to initiate the first attack to the carbon atom of formamide. (ii) A tetrahedral transition state negatively charged intermediate is formed, which is surrounded with polar atoms (Table 2). Notably, the carbonyl oxygen makes a hydrogen bond with the peptide H atom of His^{167}, which may stabilize the transition state intermediate in a manner similar to the oxyanion hole in the serine proteases (39). The Lys^{133} Nζ atom is also nearby. Hence, these interactions together reduce the activation energy barrier. (iii) Instability of the negative charge on the substrate carbonyl oxygen is likely to result in collapse of this intermediate within a short period of time, which leads to the production of an acyl-enzyme intermediate, breaking the C–N bond, and the release of an NH_3 molecule.

The second phase is to deacylate the acyl-enzyme intermediate. An incoming water molecule is deprotonated by Glu^{60}, generating a nucleophilic hydroxide ion to start the second nucleophilic attack. Like the acylation stage, a tetrahedral intermediate is formed and may be stabilized by Lys^{133} and His^{167}. The subsequent collapse from the unstable intermediate yields a formic acid molecule. After release from the active site, a free enzyme is regenerated.

CEK Residues Have Significantly Lower Translational Fluctuations and Are Positioned in or near the Key Mechanical Sites—We previously found that the CEK residues coincide with or are near to the minima in the fluctuation profile computed by molecular dynamic simulation (17). In this investigation, the AmiF dimer (apo) shows fluctuation profiles comparable with those of *N*-carbamoyl-D-amino acid amidohydrolase (data not shown). Moreover, structurally conserved CEK residues (Glu^{60}, Lys^{133}, and Cys^{166}) display nearly minimal translational fluctuations (Glu^{60}, 0.048; Lys^{133}, 0.050; Cys^{166}, 0.048). For ligand-binding residues of AmiF, there is larger degree of distribution in the fluctuation profiles (Trp^{137}, 0.049; Glu^{141}, 0.061; His^{167}, 0.046; Tyr^{192}, 0.079). Our results thus support the view of Yang and Bahar that most catalytic residues are located at or near the so-called global hinge regions and have lower translational fluctuations, whereas some subsets of residues may include more flexible sites for more efficient recognition of a diffusing substrate (40).

Analysis of the low frequency motions based on GNM model (34) reveals that the first or the slowest mode of the dimeric AmiF moves symmetrically with respect to extended C-terminal regions between the two monomers (data not shown), indicating a global hinge region lying at the interface. The second mode of AmiF describes the global motion within each monomer. This mode restrains the CEK catalytic residues to the key mechanical sites running through one side of the inner β sheets, as evidenced by their low mobility scores (data not shown). These results thus suggest a close coupling of the CEK residues to a mechanically key site for an efficient catalytic hydrolysis, since small reorganization energy will reduce activation free energy and, hence, accelerate chemical reactions (41).

In conclusion, we have determined the 1.75 Å structure of a novel formamidase from *H. pylori*. The four-layer α-β-β-α topology falls into the category of the nitrilase superfamily. We also determined the C166S apo and complex structures, enabling us identify CEK catalytic residues as well as other ligand-binding residues. In parallel with its restricted substrate specificity, AmiF possesses a relatively small binding pocket. An acyl-enzyme intermediate-mediated catalytic mechanism that involves the CEK triad is also proposed here. Structural comparison between *N*-carbamoyl D-amino acid amidohydrolase and AmiF reveals a common CEK triad but distinctive subsets of ligand-binding residues. Furthermore, molecular simulation and GNM analysis results demonstrate minimal fluctuations of the triad that sits in a mechanically key region for both *N*-carbamoyl D-amino acid amidohydrolase and AmiF. These results together support the catalytic roles of the conserved CEK triad in the nitrilase superfamily.

Acknowledgments—We thank the Macromolecular X-ray Crystallographic Center of NTHU Instrument Center at Hsinchu, National Tsing Hua University, Taiwan; BL12B2 Taiwan beamline at SPring-8, Japan; and BL13B1 beamline at the National Synchrotron Radiation Research Center (Hsinchu, Taiwan) for access to facilities for data collection. We thank Dr. J. Y. Lin for helpful discussions.

REFERENCES

1. Willison, J. C. (1993) *FEMS Microbiol. Rev.* **10,** 1–38
2. Bork, P., and Koonin, E. V. (1994) *Protein Sci.* **3,** 1344–1346
3. Pace, H. C., Hodawadekar, S. C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C. M., and Brenner, C. (2000) *Curr. Biol.* **10,** 907–917
4. Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikinaka, Y., Takahashi, S., Sato, M., and Tsukihara, T. (2000) *Structure* **8,** 729–739
5. Wang, W. C., Hsu, W. H., Chien, F. T., and Chen, C. Y. (2001) *J. Mol. Biol.* **306,** 251–261
6. Chen, C. Y., Chiu, W. C., Liu, J. S., Hsu, W. H., and Wang, W. C. (2003) *J. Biol. Chem.* **278,** 26194–26201
7. Grifantini, R., Pratesi, C., Galli, G., and Grandi, G. (1996) *J. Biol. Chem.* **271,** 9326–9331
8. Pace, H. C., and Brenner, C. (2001) *Genome Biol.* **2,** 1–9
9. Kumaran, D., Eswaramoorthy, S., Gerchman, S. E., Kycia, H., Studier, F. W., and Swaminathan, S. (2003) *Proteins* **52,** 283–291
10. Sakai, N., Tajika, Y., Yao, M., Watanabe, N., and Tanaka, I. (2004) *Proteins* **57,** 869–873
11. Skouloubris, S., Labigne, A., and De Reuse, H. (1997) *Mol. Microbiol.* **25,** 989–998
12. Skouloubris, S., Labigne, A., and De Reuse, H. (2001) *Mol. Microbiol.* **40,** 596–609
13. van Vliet, A. H., Kuipers, E. J., Stoof, J., Poppelaars, S. W., and Kusters, J. G. (2004) *Infect. Immun.* **72,** 766–773
14. Bury-Mone, S., Thiberge, J. M., Contreras, M., Maitournam, A., Labigne, A., and De Reuse, H. (2004) *Mol. Microbiol.* **53,** 623–638
15. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) *Nature* **388,** 539–547
16. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene* (*Amst.*) **77,** 51–59
17. Chiu, W. C., You, J. Y., Liu, J. S., Hsu, S. K., Hsu, W. H., Shih, C. H., Hwang, J. K., and Wang, W. C. (2006) *J. Mol. Biol.* **359,** 741–753
18. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254
19. Yang, W., Hendrickson, W. A., Kalman, E. T., and Crouch, R. J. (1990) *J. Biol. Chem.* **265,** 13553–13559

20. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J. Mol. Biol.* **229,** 105–124
21. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276,** 307–326
22. Terwilliger, T. C., and Berendzen, J. (1999) *Acta. Crystallogr. Sect. D* **55,** 849–861
23. Terwilliger, T. C. (2000) *Acta. Crystallogr. Sect. D* **56,** 965–972
24. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta. Crystallogr. Sect. A* **47,** 110–119
25. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta. Crystallogr. Sect. D* **53,** 240–255
26. Collaborative Computational Project 4 (1994) *Acta. Crystallogr. Sect. D* **50,** 760–763
27. Vagin, A., Teplyakov, A., and (1997) *J. Appl. Crystallogr.* **30,** 1022–1025
28. Lamzin, V., and Wilson, K. S. (1993) *Acta. Crystallogr. Sect. D* **49,** 127–147
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
30. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* **15,** 305–308
31. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
32. Merrit, E. A., and Murphy, M. E. P. (1994) *Acta. Crystallogr. Sect. D* **50,** 869–873
33. Berendsen, H. J. C., v. d. S. D., and van Drunen, R. (1995) *Comp. Phys. Commun.* **91,** 43–56
34. Bahar, I., Atilgan, A. R., and Erman, B. (1997) *Fold Des.* **2,** 173–181
35. Eichinger, B. E. (1972) *Macromolecules* **5,** 496–505
36. Holm, L., and Sander, C. (1999) *Nucleic Acids Res.* **27,** 244–247
37. Ogawa, J., Shimizu, S., and Yamada, H. (1993) *Eur. J. Biochem.* **212,** 685–691
38. Ikenaka, Y., Nanba, H., Yajima, K., Yamada, Y., Takano, M., and Takahashi, S. (1998) *Biosci. Biotechnol. Biochem.* **62,** 1672–1675
39. Kraut, J. (1977) *Annu. Rev. Biochem.* **46,** 331–358
40. Yang, L. W., and Bahar, I. (2005) *Structure* **13,** 893–904
41. Warshel, A., Naray-Szabo, G., Sussman, F., and Hwang, J. K. (1989) *Biochemistry* **28,** 3629–3637
42. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA