# A catalytic mechanism revealed by the crystal structures of the imidazolonepropionase from *Bacillus subtilis**

**Yamei Yu[1†], Yu-He Liang[1†], Erik Brostromer[1], Jun-Min Quan[2], Santosh Panjikar[3], Yu-Hui Dong[3] and Xiao-Dong Su[1,2¶]**

[1]The National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, P.R. china
[2]Peking University Shenzhen Graduate School, Shenzhen, Guangdong, China
[3] EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany
[4]Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

Running Title: Crystal Structure of *Bacillus subtilis* Imidazolonepropionase

¶Address correspondence to: Xiao-Dong Su, College of Life Sciences, Peking University, Beijing 100871, P.R. China; Tel.86-10-62759743; Fax. 86-10-62765669; E-Mail: su-xd@pku.edu.cn

†These authors contributed equally to this work.

**Imidazolonepropionase (EC 3.5.2.7) catalyzes the third step in the universal histidine degradation pathway, hydrolyzing the carbon-nitrogen bonds in 4-imidazolone-5-propionic acid (IPA) to yield N-formimino-L-glutamic acid. Here we report the crystal structures at 2.0 Å resolution of the *B. subtilis* imidazolonepropionase and its complex with a substrate analog imidazole-4-acetic acid sodium (I4AA). The structure of the native enzyme contains two domains, a TIM (triose-phosphate isomerase) barrel domain with two insertions and a small β-sandwich domain. The TIM barrel domain is quite similar to the members of the α/β barrel metallo-dependent hydrolase superfamily, especially to *E. coli* cytosine deaminase. A metal ion was found in the central cavity of the TIM barrel and was tightly coordinated to residues His80, His82, His249, Asp324 and a water molecule. X-ray fluorescence scan analysis confirmed that the bound metal ion was a zinc ion. An acetate ion, 6 Å away from the zinc ion, was also found in the potential active site. In the complex structure with I4AA, a substrate analog, I4AA replaced the acetate ion and contacted with Arg89, Try102, Tyr 152, His185 and Glu252, further defining and confirming the active site. The detailed structural studies allowed us to propose a zinc-activated nucleophilic attack mechanism for the hydrolysis reaction catalyzed by the enzyme.**

The histidine degradation pathway is highly conserved from prokaryotes to eukaryotes (1,2 ). In bacteria, the histidine degradation pathway is operated by the *hut* (histidine utilization) genes, and the *Bacillus subtilis hut* operon encodes proteins in the order of HutP, HutH, HutU, HutI, HutG, HutM (3-6).

There are four enzymes (histidase or HutH, urocanase or HutU, imidazolonepropionase or HutI, and formiminoglutamase or HutG) in the histidine degradation pathway in bacteria as well as in mammals, catalyzing four reactions accordingly:

(1) Histidine → urocanate + $NH_4$ (catalyzed by histidase, EC4.3.1.3);

(2) urocanate + $H_2O$ → 4-imidazolone-5-propionic acid (catalyzed by urocanase, EC 4.2.1.49);

(3) 4-imidazolone-5-propionic acid (IPA) + $H_2O$ → N-formimino-L-glutamic acid (catalyzed by imidazolonepropionase, EC 3.5.2.7) (Fig.1A);

(4) N-formimino-L-glutamic acid + $H_2O$ → glutamate + $HCONH_2$ (catalyzed by formiminoglutamase, EC 3.5.3.8).

Among the four enzymes, the first three dimensional structure determined was the 2.1 Å resolution crystal structure of histidase or HutH from *Pseudomonas putida* by Schulz *et al.* in 1999 (7,8). It was found that there was a polypeptide modification, forming a novel catalytically essential electrophilic group. The second structure was of urocanase or HutU, also from *Pseudomonas putida,* and the structure was determined by the same group who solved the histidase in 2004 (9). The third structure solved was the last enzyme in the histidine degradation pathway, formiminoglutamase or HutG from *Vibrio cholerae*, which was determined by Wu *et al.* and deposited to PDB[1] database with access ID 1XFK. The imidazolonepropionase or HutI is the only enzyme left in this pathway that has no published structural information so far, although some properties of this enzyme have been studied in *S. typhimurium* (10), *P. fluorescens* ATCC 11299 (11), rat

liver (12) and other organism during 1960's and 1970's. The maximal activity of this enzyme occurs at pH7.4 with a narrow pH optimum, and its Michaelis constant was calculated to be 7 μM (12). It was also mentioned in the literature that 1 mM EDTA (ethylenediaminetetraacetic acid) could not affect the enzyme activity, while 0.1 mM *p*-Chloromercuribenzoate could inhibit the activity completely, and the inhibition could be reversed 12% by adding 0.3 mM reduced glutathione (12), these observations indicated that the enzyme was not a metallo-protein but a cysteine-dependent enzyme. In the following twenty years, no publication was made on the imidazolonepropionase until 1997 Holm and Sander suggested that this enzyme could form a TIM barrel, and belong to a novel α/β-barrel amidohydrolase superfamily based on the residue-by-residue optimal alignment and superimposition of the 3D-structures of urease, phosphotriesterase and adenosine deaminase (13).

Here we present the crystal structures of an imidazolonepropionase or HutI, from *B. subtilis*, a native-enzyme and an enzyme complex with its substrate analog imidazole-4-acetic acid (Fig.1B), each to 2.0 Å resolution. This work reports the first crystal structures of the imidazolonepropionase family, and a catalytic mechanism has been proposed for this family of enzymes based on the high resolution structures.

## EXPERIMENTAL PROCEDURES

*Protein Preparation and Crystallization* — Selenomethionine (Se-Met) labeled imidazolonepropionase was prepared using the method described by Doublie (14), and both wild-type and Se-Met substituted proteins were purified and crystallized using the same procedure reported previously (15). Sodium salt of a substrate analog, imidazole-4-acetic acid (I4AA) was purchased from *Sigma-Aldrich Co.,* USA (CAS Number 56368-58-2). Crystals of the enzyme in complex with I4AA were obtained by co-crystallization, adding 50mM I4AA into the mother liquor of 20% (w/v) PEG4000, 0.1M Tris-HCl pH7.5, 0.2M Sodium Acetate and 2% (w/v) benzamidine hydrochloride.

*X-ray Data Collection and Processing* — X-ray diffraction data of Se-Met enzyme crystal were collected on a MAR165 CCD (charge-coupled device) detector at beamline BW7A, EMBL Hamburg outstation, around the K-edge of selenium, the exact wavelengths used in the data collection were determined by the fluorescence scan. The native crystal diffracted to better than 2.0 Å and belonged to the space group $P2_1$ with unit cell parameters a=57.73 Å, b=106.34 Å, c=66.47 Å, β=89.93º. The data of the complex crystal were collected to 2.0 Å resolution on a MAR165 CCD detector at beamline 3W1A, BSRF (Beijing synchrotron radiation facilities). The complex crystal was isomorphous with the native enzyme crystal. Both types of crystals were flash cooled and maintained at 100K by a flow of cold nitrogen gas. All data sets were processed with *DENZO* and *SCALEPACK* packages (16). The statistics of the data collection are listed in Table 1.

*Structure Determination and Refinement* — The structure of Se-Met native enzyme was solved *in-situ*, using the single-wavelength anomalous dispersion (SAD) protocol of Auto-Rickshaw: the EMBL-Hamburg automated crystal structure determination platform (17). The input diffraction data were prepared and

converted for use in Auto-Rickshaw using programs of the CCP4 suite (18). FA (normalized anomalous scattering magnitude) values were calculated using the program SHELXC (19). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 2.5 Å. All of the potential 24 selenium atoms were found using the program SHELXD (20). The correct hand for the substructure was determined using the programs ABS (21) and SHELXE (22). The occupancy of all substructure atoms was refined using the program MLPHARE (CCP4, 1994). The initial phases were improved using density modification and phase extension to 2.0 Å resolution using the program DM (18). 95% of the structure model was built automatically using the program ARP/wARP (23,24), and further model building to completion was done with the program O (25) manually. The refinement of the structure was performed using CNS (26) combined with manual rebuilding of the model. Eight percent of the total reflections were set aside as test set. The $R_{cryst}$ and $R_{free}$ values of the final model were 0.188 and 0.221, respectively. The complex structure was solved by molecular replacement and refined with the same strategy of native-enzyme structure. The coordinates of I4AA was obtained from *IUCR - Crystal Structure Communications* (27) and placed manually into the positive electron density map and the model was further refined. The final $R_{cryst}$ and $R_{free}$ values of the complex structure model were 0.185 and 0.216, respectively. The stereochemistry qualities of the two structures were reasonable as checked by PROCHECK (28). The statistics of the refinements are listed in Table 1.

The figures in the paper were prepared by programs Molscript (29), Pymol (http://pymol.sourceforge.net/), ClustalX (30) and ESPript (31). The secondary structure of the model was assigned using the program DSSP (32).

*X-ray Fluorescence scan Analysis* — To determine what kind of metal ion binds to the protein, fluorescence scan analysis of the protein solution used for crystallization was performed at beamline 911-3, MAX lab, Lund, Sweden. The wavelength ranges of the experiments were selected near the K-edge of Zn, Mn, Fe and Co, respectively.

## RESULTS AND DISCUSSION

*Structure Determination of The Native-enzyme* — The structure of *B. subtilis* imidazolonepropionase could be solved by the SAD method using Se-Met substituted protein. The crystals formed in the space group $P2_1$ with β angle nearly equal to 90º and there are two molecules in the asymmetric unit. The Se-Met crystals diffracted to better than 2.0 Å resolution, and the structure model was finally refined to a crystallographic *R*-value of 18.8% ($R_{free}$= 22.1%), when data from 20.0 to 2.0 Å resolution of the peak dataset were used. The stereochemistry quality of the final model is reasonable as checked by *PROCHECK* (33)，only one residue, His272, is in the disallowed region of the Ramachandran plot, however, this residue is well-defined in the density maps and stabilized by hydrogen bonds with neighboring residues through water bridges. The final model shows residues 3 to 415 in both monomers. The two monomers are quite similar to each other with an r.m.s.d. (root mean square deviations) of 0.292 Å with 413 Cα atoms aligned. The crystallographic and refinement information is summarized in Table 1.

*The Overall Structure* — The *B. subtilis* imidazolonepropionase consists of 421 amino acids and exists as homo-dimer (Fig. 2A), in consistent with the result of gel filtration experiments (data not shown). The two molecules are closely packed around a 2-fold symmetry axis. A total of 1904 $Å^2$ of accessible area per monomer is buried in the dimer interface, accounting for about 12% of the entire surface area. Residues involved in the dimer interface are mainly from α-helices or loops protruding from the molecular surface, the interacting residues are 88-97, 133-141, 299-304, 327-351, 386-401 with mostly hydrophobic and van der Waals interactions.

The molecule folds into two structural domains (Fig. 2B), a TIM barrel domain and a small β-sandwich domain. The TIM barrel domain, which is composed of eight quite parallel β-strands and eight α-helices, is somewhat distorted and interrupted by two insertions: residues 79-122 (Insertion I) between β1-strand and α1-helix, forming a coil and three helices, and residues 294-308 (Insertion II) between β7-strand and α7-helix, forming a helix and a coil. The β-strands and α-helices in the barrel are connected by two kinds of loops, the β-α loops (from β-strand to α-helix) and the α-β loops (from α-helix to β-strand). Comparatively, the β-α loops were much longer and more flexible than the α-β loops, and located at the entrance of the central cavity of the TIM barrel domain, it is thus likely that the β-α loops are more important for the enzyme activity. The structure based sequence alignment and the topology diagram of the protein are presented as Fig.2C, 2D.

The small β-sandwich domain, lying at the bottom of the barrel far from the central cavity as shown in Fig. 2B, is composed of two separate peptide segments from both N and C termini (residues 3-70 and 365-415). This domain is connected to the TIM barrel domain by two linkers: the first linker is a β strand (residues 71-74, βE) and the second linker is an α-helix (residues350-364, αE), an extra insertion (residues 389-398) in this domain including αF-helix interacts tightly with the neighbor subunit.

*The Substrate Binding Sites and Putative Active Center*— Although the previous studies reported that EDTA did not affect the activity of the imidazolonepropionase (12), implicating the enzyme is not a metalloprotein, we did observe a penta-coordinated metal ion in the density map (Fig. 3A), and further confirmed by the anomalous difference Fourier map (data not shown). The metal ion is located in the central cavity of the TIM barrel, and directly coordinated by His80-$N^{\varepsilon 2}$, His82-$N^{\varepsilon 2}$, His249-$N^{\varepsilon 2}$, Asp324-$O^{\delta 1}$ and a water molecule (W1 in Fig. 3A)with distances of 2.19 Å, 2.18 Å, 2.28 Å, 2.32 Å and 2.02 Å, respectively (marked in Fig. 3C). These residues are fully conserved across different organisms (Fig. 2C). The metal bound water also forms a hydrogen bond with Glu252 through a water bridge (W2 in Fig. 3A). Later, the X-ray fluorescence scan of the enzyme gave clear absorption jump only at the spectrum over K-edge of Zn (data not shown), which unambiguously demonstrated this metal ion is a zinc ion.

Close to the cavity center of the TIM barrel, an acetate ion was also observed (Fig. 3A), this acetate ion is most likely from the crystallization buffer containing 0.2M Sodium Acetate. This acetate is quite close to the zinc ion site (about 6 Å away), and forms very nice salt-bridges with Arg89-$N^{\eta 1}$

(2.95Å) and Arg89-$N^{\eta 2}$ (2.93Å). This scenario of zinc ion plus the acetate ion around the TIM barrel cavity center immediately prompted us that this place is the active site since the acetate binding site could well be the binding site for the carboxyl group of IPA (see Fig. 1A), the substrate of the enzyme. Furthermore, the zinc ion with the coordinated water is positioned well for catalysis.

In order to confirm the active site and elucidate the catalytic mechanism of the imidazolonepropionase, the structure of the enzyme in complex with its substrate analog I4AA, was made and solved. The *Fo - Fc* density omit map clearly showed that I4AA molecule is bound to the enzyme at the same place where acetate ion is bound in the native structure (Fig. 3B, Fig. 3C). The overall structure of the complex is very similar to that of the native-enzyme with an r.m.s.d. of 0.13 Å among 413 Cα atoms aligned in subunit A between the two structures. The only notable differences between the two structures are mainly from residues Arg89，Glu252，Tyr102 and His272 (Fig. 3C), caused by the insertion of I4AA molecule into the active site. In the complex structure, the zinc ion and the coordinated water molecule located exactly at the same place as in the native-enzyme structure. The carboxyl group of I4AA is located very close to the acetate binding site in the native-enzyme structure, with about 0.6 Å shift towards the zinc site, so is the side chain of Arg89, which still forms nice salt bridges with the carboxyl group of I4AA. Besides the salt bridges with Arg89, I4AA forms hydrogen bonds with Tyr102, Tyr152, His185 and Glu252, these bonds keep the I4AA molecule in a correct orientation in the central cavity. The imidazole ring of the I4AA is positioned very close to the water molecule coordinated to zinc, and replaced the water molecule (see Fig. 3C) bridging zinc-bound water and Glu252 in the native structure, indicating a possible nucleophilic attack by zinc-activated water in the hydrolysis reaction.

Among the residues involved in zinc and I4AA binding, consequently important to the enzyme function, most of them are located close to the central cavity of the TIM barrel as expected from many other TIM barrel enzymes with known structures and catalytic mechanisms. His80, His82, Arg89 and Tyr102 are from the Insertion I region, His 185, His249 and His272 are located at the C terminus of β-strands 3, 5 and 6, respectively, and Tyr152, Glu252 and Asp324 are located in the β-α loops 2, 5 and 8 respectively (Fig. 2C). Insertion I (αA, αB and αC-helices) lies as a lid on the entrance of central cavity (Fig 2A), making the zinc and substrate binding site deeply buried in the molecule. The tight coordination and deep buried location could be the reason for the zinc ion hard to be dissociated from the enzyme by EDTA. Although the Insertion I covers the central cavity, there is a narrow hydrophilic tunnel leading to the barrel core in the structure, which could be the entrance path for the substrate. Since the inserted helices connect to the TIM barrel through long loops, the entire region could be flexible and change to different conformations during hydrolysis reaction.

*Structural and Functional Homologues* — In a search for related homologs of *B. subtilis* imidazolonepropionase using *BLAST* (34), it is observed that the enzyme existed widely in various organism. We picked HutI proteins from representative gram-negative, gram-positive bacteria and eukaryote

respectively and aligned these sequences with ClustalX and ESPript (Fig. 2C). The highest percentage of identical amino acid residues among the six proteins is 73%, which is between *X. tropicalis* and *H. sapiens*, while the lowest is 32%, between *S. typhimurium* and *X. tropicalis*. It was noticed that the conserved regions are distributed to the whole length of the proteins except at the N and C termini. The residues suggested to be important to the substrate binding and catalytic activity are highly conserved. The most significant difference between the prokaryotic and eukaryotic sequences is that there are two gaps in the prokaryotic sequences. One gap was at the βD-strand from the small β-sandwich domain and the other small gap is at the third β-α loop in the TIM barrel domain. The two gaps are far from the active center and probably do not interfere with the activity of the enzyme.

Structural similarity search by DALI (35) revealed that the catalytic domain (TIM barrel domain) of imidazolonepropionase belonged to the α/β barrel metal-dependent hydrolase superfamily. The structure with the highest similarity to imidazolonepropionase is that of cytosine deaminase (CDA) from *E. coli* (36), with an r.m.s.d. of 2.8 Å (358 Cα atoms with 18% sequence identity). CDA is confirmed to be a ferrous-dependent enzyme, while the coordination geometry of $Fe^{2+}$ is quite similar to that of imidazolonepropionase (36,37). Another enzyme in this superfamily, adenosine deaminase (ADA), is discovered to be a zinc-binding metallo-protein; ADA has also shown very similar coordination geometry at the metal binding site in the active center (38). Recently, crystal structure of imidazolonepropionase from *Agrobacterium tumefaciens* was deposited to PDB database (Tyagi et al., PDB id: 2GOK). *A. tumefaciens* imidazolonepropionase shows 41% of sequence identity to that of *B. subtilis*, and the similarity between the two structures is high (r.m.s.d. of 1.7 Å among 393 Cα atoms). Significant differences between the two structures locate at the loop connecting β-strands A and B which is in the small β-sandwich domain, and the helices of Insertion I region, the average shift of αB helix is about 10 Å. In the model of the structure of *A. tumefaciens* imidazolonepropionase, a similar metal ion binding site was also observed, however, with a ferric iron put in the binding site.

*Catalytic Mechanism* — Based on the above structural studies and analyses, we can propose a hydrolysis mechanism for the imidazolonepropionase family as outlined in Fig. 4. In the original state in the "empty enzyme" (labeled as "Enzyme" in Fig. 4), the zinc-bound water forms strong hydrogen bonds with Asp324, and with Glu252 through another "bridging" water molecule. The enzyme adopts an open conformation in the Insertion I region, ready for accepting a substrate molecule to enter through the hydrophilic tunnel. Next step as indicated by the arrowhead in Fig. 4, a substrate molecule 4-imidazolone-5-propionic acid (IPA) is bound into the active site, one of the protons of the zinc-bound water was shifted to Glu252, and the water bridge between the zinc-bound water and Glu252 was replaced by the substrate. The IPA molecule is stabilized and correctly oriented by a strong salt bridge with residue Arg89, and hydrogen bond network with residues Glu252, Ser329 and His185. Additionally, the effect of salt bridge on Arg89 will induce a "closed confirmation" of Insertion I region, and isolate the reaction from the "water solvent" outside. The carbonyl

oxygen of the substrate will be activated by the protonated carboxyl group of Glu252, then the zinc-bound hydroxide attacks the carbonyl carbon of the substrate to form the tetrahedral intermediate (EI-1, as shown in Fig. 4), the amide nitrogen changes its hybridization from sp2 to sp3 during the nucleophilic attack and protonated by the proton transferred from Glu252 (EI-2, as shown in Fig. 4), facilitating the cleavage of the N-C bond of the substrate, to form the enzyme-product complex (EP, as shown in Fig. 4), finally the reaction cycle goes back to the original state as in "Enzyme" with the release of product and the uptake of a new water molecule.

*Conclusions* — In summary, *B. subtilis* imidazolonepropionase represents a widely existed protein family, either viewed from the sequential homology analysis or from structural similarity analysis. The structural study and the X-ray fluorescence scan experiments reported here revealed that the imidazolonepropionase is a zinc-enzyme, and adopts a zinc-activated nucleophilic attacking mechanism. An insertion region on the entrance of the central cavity might be important to the enzyme activity.

## REFERENCES


1. Revel, H. R., and Magasanik, B. (1958) *J Biol Chem* **233**(4), 930-935
2. Brown, D. D., Silva, O. L., Gardiner, R. C., and Silverman, M. (1960) *J Biol Chem* **235**, 2058-2062
3. Chasin, L. A., and Magasanik, B. (1968) *J Biol Chem* **243**(19), 5165-5178
4. Kimhi, Y., and Magasanik, B. (1970) *J Biol Chem* **245**(14), 3545-3548
5. Oda, M., Sugishita, A., and Furukawa, K. (1988) *J Bacteriol* **170**(7), 3199-3205
6. Yoshida, K., Sano, H., Seki, S., Oda, M., Fujimura, M., and Fujita, Y. (1995) *Microbiology* **141 ( Pt 2)**, 337-343
7. Schwede, T. F., Badeker, M., Langer, M., Retey, J., and Schulz, G. E. (1999) *Protein Eng* **12**(2), 151-153
8. Schwede, T. F., Retey, J., and Schulz, G. E. (1999) *Biochemistry* **38**(17), 5355-5361
9. Kessler, D., Retey, J., and Schulz, G. E. (2004) *J Mol Biol* **342**(1), 183-194
10. Smith, G. R., Halpern, Y. S., and Magasanik, B. (1971) *J Biol Chem* **246**(10), 3320-3329
11. Rao, D. R., and Greenberg, D. M. (1960) *Biochim Biophys Acta* **43**, 404-418
12. Snyder, S. H., Silva, O. L., and Kies, M. W. (1961) *J Biol Chem* **236**, 2996-2998
13. Holm, L., and Sander, C. (1997) *Proteins* **28**(1), 72-82
14. Doublie, S. (1997) *Methods Enzymol* **276**, 523-530
15. Yu, Y., Li, L., Zheng, X., Liang, Y. H., and Su, X. D. (2006) *Biochim Biophys Acta* **1764**(1), 153-156
16. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
17. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., and Tucker, P. A. (2005) *Acta Crystallogr D Biol Crystallogr* **61**(Pt 4), 449-457
18. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38
19. Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R., and Usón, I. (2001) Ab

initio phasing. In: Rossmann, M. G., and Arnold, E. (eds). *International Tables for Crystallography*, IUCr and Kluwer Academic Publishers, Dordrecht
20. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr D Biol Crystallogr* **58**(Pt 10 Pt 2), 1772-1779
21. Hao, Q. (2004) *J. Appl. Cryst.* **37**, 498-499
22. Sheldrick, G. M. (2002) *Z. Kristallogr.* **217**(644-650)
23. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) *Acta Crystallogr D Biol Crystallogr* **53**(Pt 4), 448-455
24. Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A., and Lamzin, V. S. (2004) *J Synchrotron Radiat* **11**(Pt 1), 56-59
25. Jones, T. A., Bergdoll, M., and Kjeldgaard, M. (1990) O: A macromolecular modeling environment. In: Bugg, C., and Ealick, S. (eds). *Crystallographic and Modeling Methods in Molecular Design.*, Springer-Verlag Press Berlin
26. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54 ( Pt 5)**, 905-921
27. Okabe, N., and Hayashi, T. (1999) *Acta Crystallographica Section C* **55**(7), 1142-1144
28. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallog.* **26**, 283-291
29. Kraulis, P. J. (1991) *Journal of Applied Crystallography* **24**, 946-950
30. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res* **25**(24), 4876-4882
31. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* **15**(4), 305-308
32. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**(12), 2577-2637
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* **26**(2), 283-291
34. Lopez, R., Silventoinen, V., Robinson, S., Kibria, A., and Gish, W. (2003) *Nucleic Acids Res* **31**(13), 3795-3798
35. Dietmann, S., Park, J., Notredame, C., Heger, A., Lappe, M., and Holm, L. (2001) *Nucleic Acids Res* **29**(1), 55-57
36. Ireton, G. C., McDermott, G., Black, M. E., and Stoddard, B. L. (2002) *J Mol Biol* **315**(4), 687-697
37. Porter, D. J. (2000) *Biochim Biophys Acta* **1476**(2), 239-252
38. Wilson, D. K., Rudolph, F. B., and Quiocho, F. A. (1991) *Science* **252**(5010), 1278-1284
39. Bond, C. S. (2003) *Bioinformatics* **19**(2), 311-312

## FOOTNOTES

*This work was supported by grant from National High Technology and Development Program of China (863 program 2002BA711A13). Peking

University's 985 and 211 grants are also greatly acknowledged. XDS is a recipient of National Science Fund for Distinguished Young Scholars of NSFC (30325012). We would also like to thank the generous help with our data collection from staffs of SR sources at BSRF, MAX-lab and DESY at EMBL Hamburg outstation.

*The atomic coordinates and structure factors (codes 2BB0 for native-enzyme and 2G3F for enzyme-I4AA complex, have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).*

[1]The abbreviations used are: IPA, 4-imidazolone-5-propionic acid; I4AA, imidazole-4-acetic acid; TIM, triose-phosphate isomerase; EDTA, ethylenediaminetetraacetic acid; Se-Met, Selenomethionine; CCD, charge-coupled device; BSRF, Beijing Synchrotron Radiation Facility; SAD, single anomalous dispersion; r.m.s.d., root mean square deviations; CDA, cytosine deaminase; ADA, adenosine deaminase; PDB, Protein Data Bank, SR, Synchrotron Radiation.

## FIGURE LEGENDS

Fig. 1 A, The chemical reaction catalyzed by the imidazolonepropionase. B, Imidazole-4-acetic acid sodium salt (I4AA), an analog of 4-imidazolone-5-propionic acid, the substrate of imidazolonepropionase.

Fig. 2 A, The structure of *B. subtilis* imidazolonepropionase dimer. B, The overall structure of a monomer with the secondary structure elements indicated, the zinc ion is shown in CPK mode in yellow, and the acetate is shown in ball-stick mode. The illustrations were made by the program Molscript (29). C, Structural based sequence alignment of representative HutI (imidazolonepropionase) proteins. The primary sequences of HutI from gram-negative bacteria *P. putida*, *A. tumefaciens* and *S. typhimurium*, gram-positive bacteria *B. subtilis*, *S. aureus* and eukaryote *X. tropicalis*, *H. sapiens* were aligned. The secondary structure of *B. subtilis* and *A. tumefaciens* imidazolonepropionase was placed on the top and the bottom, respectively. The alignment was generated with ClustalX (30) and drawn with ESPript (31). Highly conserved residues were marked with red background (completely identical) and blue boxes (partially identical). The red upward triangles indicate residues interact with the zinc. Black stars highlight the residues forming hydrogen bonds with the substrate analogue. D, The topology diagram of the imidazolonepropionase structure generated by TopDraw (39), the secondary structure distribution of the model was given by program DSSP (32).

Fig. 3 The zinc and substrate binding sites of the enzyme. A, The native-enzyme structure. B, The complex structure. The key residues and ligands were marked and represented in stick mode (C, cyan; N, blue; O, red; Zn, yellow). The *Fo - Fc* maps were calculated with zinc, water bound to zinc, acetate and I4AA molecules omitted, and drawn at a contour level

of 3σ. C, The superposition of the native-enzyme structure (carbon atoms colored in yellow) and complex structure (carbon atoms colored in cyan), the distances marked are from the native-enzyme structure, indicating the zinc coordination and the salt-bridge between acetate ion and the residue Arg89.

Fig. 4 The schematic diagram of a proposed mechanism for the hydrolysis reaction catalyzed by the family of imidazolonepropionase, see text for the detailed description.

# Table 1 Data collection and refinement statistics

| Parameters | Native-enzyme crystal | Enzyme-I4AA Complex |
|---|---|---|
| ***Data collection*** | | |
| Wavelength | 0.9792 | 1.0 |
| Space group | $P2_1$ | $P2_1$ |
| Unit cell parameters (Å,º) | a=57.73 b=106.34 c=66.47 β=89.92 | a=57.53 b=106.080 c=66.31 β=90.20 |
| Resolution Range (Å) | 20-2.0 (2.03-2.0)* | 20-2.0 (2.03-2.0) |
| Completeness (%) | 99.8 (99.8) | 98.9 (97.3) |
| $R_{sym}$(%)[a] | 9.8 (25.6) | 6.7 (31.0) |
| I/σ(I) | 8.3 (3.5) | 10.7 (2.7) |
| Unique Reflections | 54728 | 52978 |
| ***Refinement*** | | |
| No. of non-H protein atoms | 6287 | 6344 |
| No. of waters | 532 | 405 |
| $R_{cryst}$[b] | 0.188 | 0.185 |
| $R_{free}$[c] | 0.221 | 0.216 |
| Average B-factors (Å$^2$) | | |
| All atoms | 17.14 | 16.01 |
| Main chain | 15.50 | 14.67 |
| Side chain | 17.88 | 16.90 |
| Solvent | 23.23 | 20.54 |
| Ramachandran plot statistics (%) | | |
| Most favored region | 88.4 | 88.8 |
| Additional allowed region | 10.9 | 10.7 |
| Generously allowed region | 0.4 | 0.3 |
| Disallowed region | 0.3 | 0.3 |

*Values in parentheses refer to the highest resolution shell.

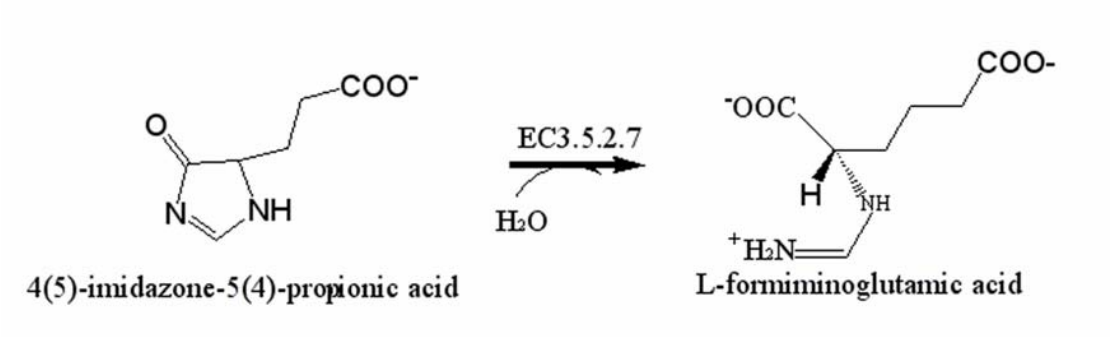
[a] $R_{sym} = \sum|I_{obs} - I_{avg}|/\sum I_{obs}$ where the summation is over all reflections

[b] $R_{sym} = \sum(||F_o| - |F_c||)/\sum|F_o|$

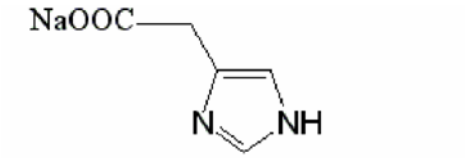
[c] $R_{free}$ is the R-factor for a selected subset of the reflections which are not included in refinement calculations.

# Fig. 1

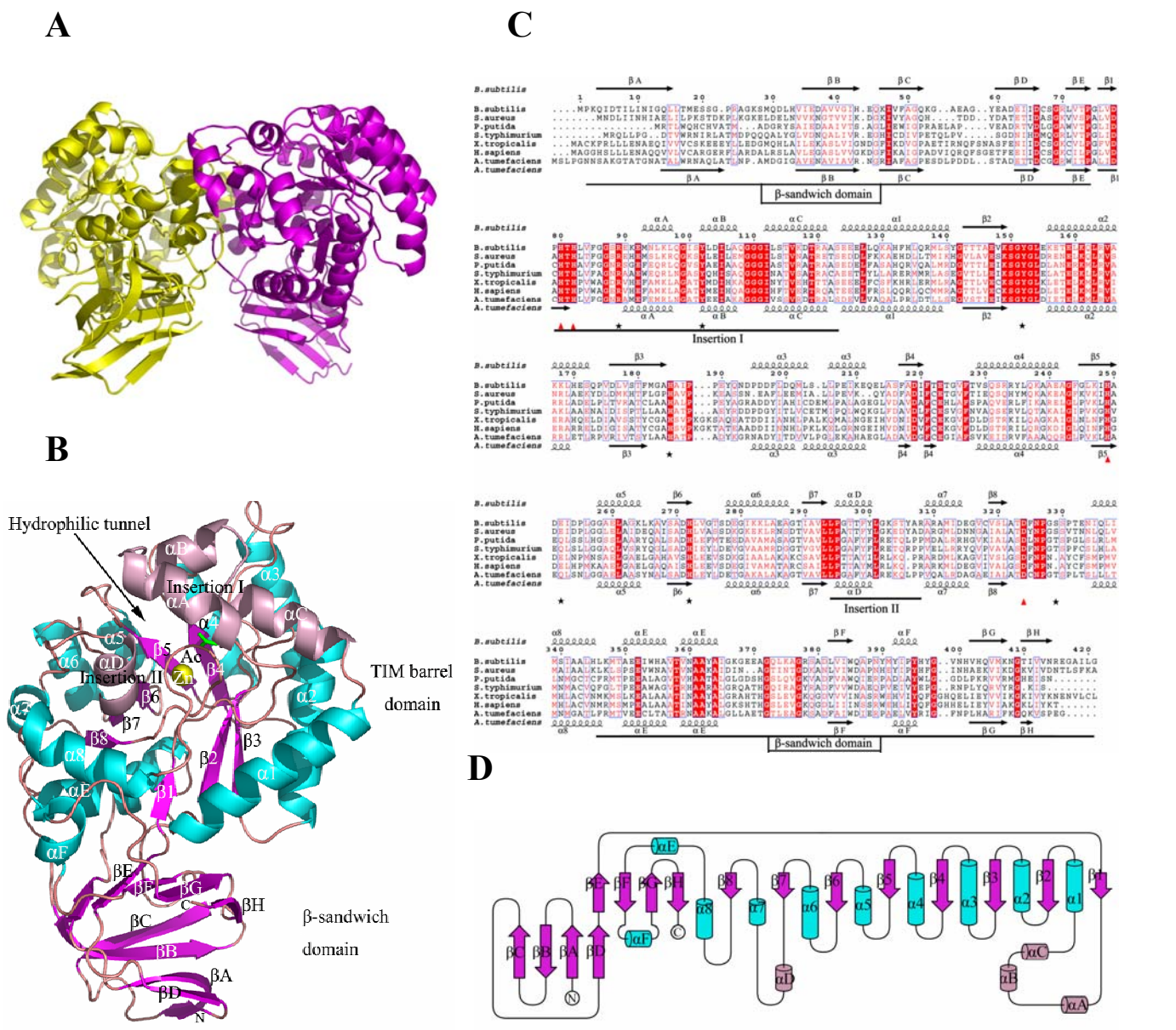
**A**


COO-
O
N
NH
EC3.5.2.7
H2O
COO-
-OOC
H
NH
+H2N
4(5)-imidazone-5(4)-propionic acid
L-forminoglutamic acid

**B**

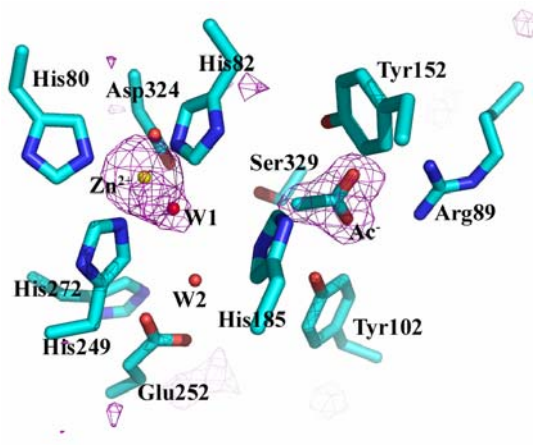

NaOOC
N
NH
Imidazole-4-acetic acid sodium salt

# Fig. 2

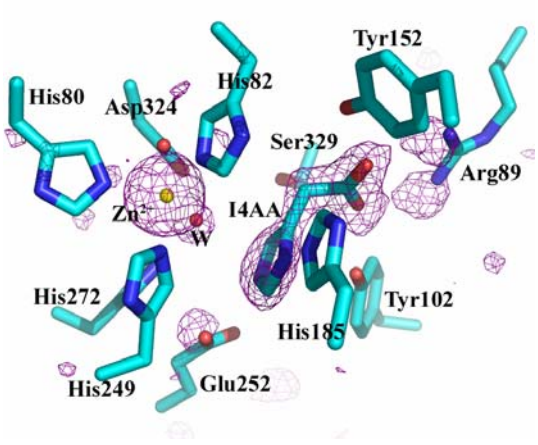

A
B
C
D
Hydrophilic tunnel
Insertion I
Insertion II
TIM barrel domain
β-sandwich domain
β-sandwich domain

# Fig 3

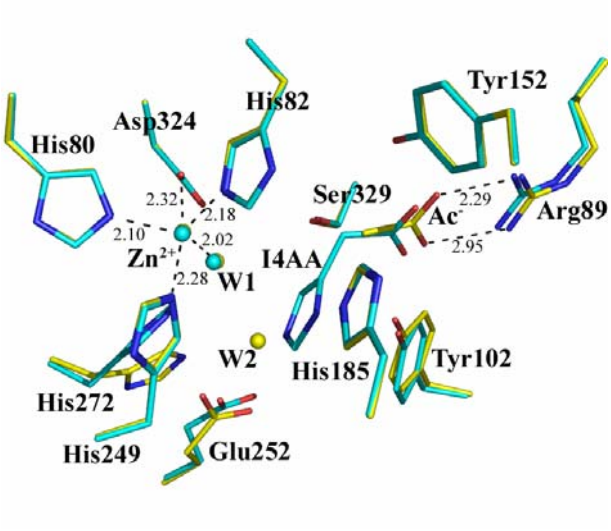
**A**


His80
Asp324
His82
Tyr152
Ser329
Zn2+
W1
Ac-
Arg89
His272
W2
His185
Tyr102
His249
Glu252

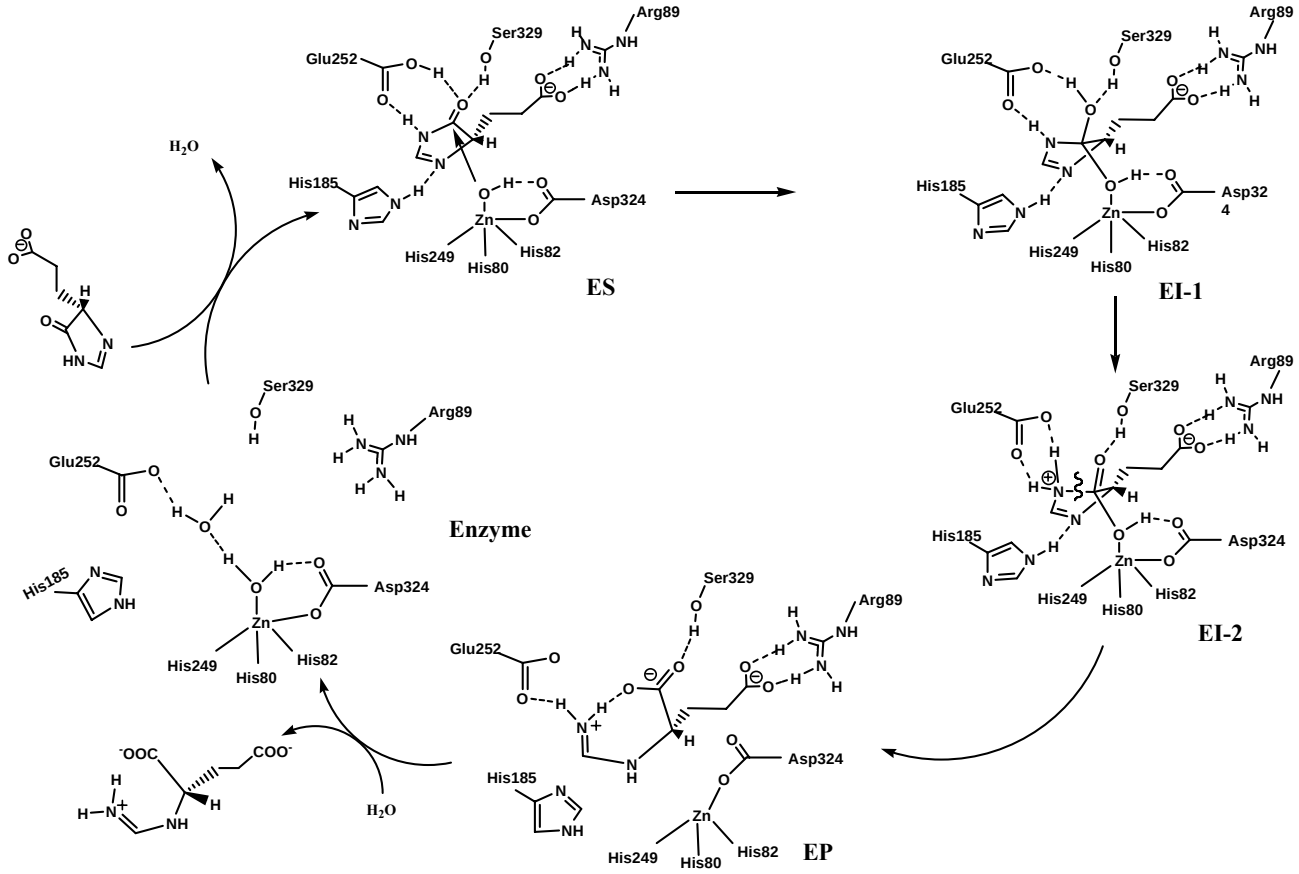
**B**


Tyr152
His80
Asp324
His82
Ser329
Arg89
Zn2+
W
I4AA
His272
Tyr102
His185
His249
Glu252

**C**


Tyr152
His82
Asp324
His80
2.32
2.18
Ser329
2.29
Ac-
Arg89
2.10
2.02
Zn2+
2.28
W1
I4AA
2.95
W2
His185
Tyr102
His272
His249
Glu252

# Fig4


Arg89
Ser329
Glu252
His185
Asp324
Zn
His249
His80
His82
ES
EI-1
EI-2
Enzyme
EP
$H_2O$
$^-OOC$
$COO^-$