Published in final edited form as:

J Struct Biol. 2010 March ; 169(3): 304–311. doi:10.1016/j.jsb.2009.11.008.

Crystal structure and molecular modeling study of *N*-carbamoylsarcosine amidase Ta0454 from *Thermoplasma acidophilum*

Hai-Bin Luo^{1,2,4}, Heping Zheng^{2,4}, Matthew D. Zimmerman^{2,4}, Maksymilian Chruszcz^{2,4}, Tatiana Skarina^{2,4}, Olga Egorova^{2,4}, Alexei Savchenko^{3,4}, Aled M. Edwards^{3,4}, and Wladek Minor^{2,4,*}

¹ School of Pharmaceutical Sciences, East campus, Sun Yat-Sen University, Guangzhou 510006, China

² Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA

³ Banting and Best Department of Medical Research and Structural Genomics Consortium, 112 College Street, University of Toronto, Toronto, Ontario M5G 1L6, Canada

⁴ Midwest Center for Structural Genomics

Abstract

A crystal structure of the putative *N*-carbamoylsarcosine amidase (CSHase) Ta0454 from *Thermoplasma acidophilum* was solved by single-wavelength anomalous diffraction and refined at a resolution of 2.35 Å. CSHases are involved in the degradation of creatinine. Ta0454 shares a similar fold and a highly conserved C-D-K catalytic triad (Cys123, Asp9, and Lys90) with the structures of three cysteine hydrolases (PDB codes 1NBA, 1IM5, and 2H0R). Molecular dynamics (MD) simulations of Ta0454/*N*-carbamoylsarcosine and Ta0454/pyrazinamide complexes were performed to determine the structural basis of the substrate binding pattern for each ligand. Based on the MD simulated-trajectories, the MM/PBSA method predicts binding free energies of −24.5 and −17.1 kcal/mol for the two systems, respectively. The predicted binding free energies suggest that Ta0454 is selective for *N*-carbamoylsarcosine over pyrazinamide, and zinc ions play an important role in the favorable substrate bound states.

Keywords

N-carbamoylsarcosine amidase; C-D-K catalytic triad; creatinine degradation; crystal structure; MM/PBSA; molecular dynamics simulations

---

*Corresponding author. Fax: 434-982-1616, Phone: 434-243-6865, and wladek@iwonka.med.virginia.edu.

Supplemental information

Supplemental information, comprising three figures with 3D binding models of the Ta0454/*N*-carbamoylsarcosine (*N*-CS), Ta0454/pyrazinamide (PZA), and PH999/PZA complexes, is available.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Introduction

Creatinine is a breakdown product of creatine phosphate in muscle. Retention of creatinine in the bowel of patients with chronic renal failure has been suggested to induce bacterial creatininase activity, which increases the rate of creatinine degradation by different microbes and may affect the reliability of the compound as a renal function indicator (Dunn et al., 1997). *N*-carbamoylsarcosine amidase (CSHase, EC3.5.1.59), also called *N*-carbamoylsarcosine amidohydrolase, is involved in one of the alternative metabolic pathways for the degradation of creatinine detected in several bacterial genera, such as *Arthrobacter*, *Brevibacterium*, *Clostridium*, *Moraxella*, *Micrococcus*, *Pseudomonas*, and *Tissierella* (Farrow et al., 1995; Hong et al., 1998; Nishiya et al., 1998). In this pathway, CSHase catalyzes its natural substrate *N*-carbamoylsarcosine (*N*-CS) to sarcosine, carbon dioxide, and ammonia (Figure 1). CSHase is highly specific for *N*-CS as a substrate (Kim et al., 1986). The mechanism of CSHase has been studied in *Arthrobacter* sp. (Romao et al., 1992; Zajc et al., 1996), *Pseudomonas putida* 77 (Yamada et al., 1985), and other microorganisms.

Although CSHase has been studied for over 20 years, only one crystal structure, which does not contain an ordered substrate, has been deposited in the Protein Data Bank (PDB). The structure of CSHase (CSHase1, PDB code: 1NBA) from *Arthrobacter* sp. was determined at 2.0 Å resolution (Romao et al., 1992). The asymmetric unit contains a tetramer, and a sulfate ion is bound to the catalytic site of each monomer. Some structures of cysteine hydrolases with a similar fold and *N*-carbamoylamide hydrolysis activity have been determined, such as pyrazinamidase PH999 (PDB code: 1IM5) from *Pyrococcus horikoshii* (Du et al., 2001) and nicotinamidase Pnc1p (PDB code: 2H0R) from *Saccharomyces cerevisiae* (Hu, et al., 2007). The reactions catalyzed by these related enzymes are shown in Figure 1.

In this work, we report the crystal structure of the putative CSHase Ta0454 from *Thermoplasma acidophilum*. Since the active site of Ta0454 is more structurally similar to that of PH999 than that of CSHase1 (as well as higher overall sequence similarity to PH999 versus CSHase1), the substrate preference of Ta0454 for *N*-CS versus pyrazinamide (PZA) is unclear. Molecular modeling studies were performed to analyze the binding of either *N*-CS or PZA to Ta0454 structurally, using flexible-docking and molecular dynamics (MD) simulations, followed by free energy of binding calculations using the molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) approach (Stoica et al., 2008; Kuhn et al., 2000). MM/PBSA combines molecular mechanics energies with continuum solvent models to post-process a series of representative snapshots from MD trajectories. The analyses of the MD structural fluctuation and the predicted binding free energies of the Ta0454/*N*-CS, Ta0454/PZA, and PH999/PZA complexes should permit us to (1) obtain a complete picture of the structural characteristics of the three complexes; (2) understand the differences and similarities between the substrate binding patterns of these complexes; and (3) evaluate whether Ta0454 prefers binding *N*-CS versus PZA (or vice versa) by comparing their predicted binding free energies.

2 Materials and Methods

2.1 Protein cloning, expression, purification, and crystallization

The small regulatory subunit of CSHase Ta0454 from *Thermoplasma acidophilum* was cloned, expressed, and purified by protocols developed by the Midwest Center for Structural Genomics (Zhang et al., 2001). Selenomethionine (Se-Met) substituted protein was produced from cells grown in M9 SeMET High-Yield growth media (Shanghai Medicilon). Cleavage of the N-terminal hexahistidine affinity tag from the protein by treatment with recombinant His-tagged TEV protease (Kapust et al., 2002) was unsuccessful (data not shown). Therefore *in situ* proteolysis with trypsin protease was performed in the subsequent crystallization experiments both to attempt to cleave the affinity tag and to improve the overall packing of the crystal lattice.

The *in situ* proteolysis technique uses a trace concentration of nuclease in the crystallization solution and relies upon the crystallization process itself to exclude heterogenous (e.g. uncleaved) protein from the lattice (Dong et al., 2007). Trypsin (T8003) was ordered from Sigma-Aldrich, and a trypsin stock solution was prepared (1.5 mg/mL in 1 mM HCl and 2 mM CaCl₂), which was mixed 1:100 (v/v) with a solution containing 5 mg/mL purified Ta0454 protein. Crystals suitable for X-ray diffraction experiments were obtained by hanging-drop vapor diffusion at 20°C. Crystals were grown from 2 µL droplets of protein dissolved in 0.3 M NaCl and 10 mM HEPES-Na (pH=7.5) mixed 1:1 with a well solution composed of 2 M (NH₄)₂SO₄ and 5% isopropanol. After 3 months of growth, protein crystals were selected and harvested for data collection. The crystals were flash cooled in a cryogenic stream of nitrogen, after being transferred into a cryoprotectant composed of 3 M (NH₄)₂SO₄, 10% isopropanol, 5% glycerol and 5% sucrose.

2.2 Data collection, structure determination, and refinement

SAD data from Se-Met substituted protein were collected at a temperature of 100 K on beamline 19-ID (Rosenbaum et al., 2006) of the Structural Biology Center at Argonne National Laboratory and processed with HKL-2000 (Otwinowski et al., 1997). Data collection, structure determination, and refinement statistics are summarized in Table 1. An initial electron density map and model was obtained with HKL-3000 (Minor et al., 2006), which integrates with the family of SHELX programs (Sheldrick 2008), MLPHARE (Otwinowski et al., 1991), DM (Cowtan 1994), SOLVE/RESOLVE (Terwilliger 2002), CCP4 (Cowtan 1994), ARP/wARP (Perrakis et al., 1999), O (Jones et al., 1991), and COOT (Emsley and Cowtan 2004). The initial 2.35 Å model was extended by cycles of iterative rebuilding with RESOLVE followed by ARP/wARP, as implemented in HKL-3000. The resulting model was iteratively improved by cycles of manual model building with COOT followed by maximum-likelihood refinement with REFMAC (Murshudov et al., 1997). MOLPROBITY (Lovell et al., 2003) and PROCHECK (Laskowski, et al., 1993) were used for structure validation. In the crystal structure, 9 residues on the C-termini of both chains could not be observed in the final electron density map. The atomic coordinates for the structure, along with the structure factors, were deposited to the PDB with accession code 3EEF.

2.3 Molecular docking

Since it is unclear how either *N*-CS or PZA binds to Ta0454, the initial configuration of the conserved substrate binding site (Asp9, Lys90, Leu118, Asp119, and Cys123) was prepared using *CDOCKER* (CHARMM-based *DOCKER*) (Wu et al., 2003; Erickson et al., 2004) calculations. The structure of Ta0454 was used as an initial model, where all histidines were considered to be protonated on the the delta nitrogen (i.e., HIS was mutated to HID). The water molecules and zinc ions in the crystal structure were removed, and disordered side chains and hydrogen atoms were added to the system using *Discovery Studio 2.1* molecular modeling software (Accelrys, San Diego, CA). Docking studies were then performed using the *CDOCKER* docking protocol implemented in the same software.

CDOCKER generated several poses of either *N*-CS or PZA to populate the conserved putative binding site in Ta0454. Each pose was subjected to high-temperature MD using a modified version of *CHARMM* (Brooks et al., 1983). Each of the structures from the MD run were then located and fully minimized. The solutions were clustered according to position and conformation and ranked by energy. *CHARMM* partial atomic charges were used for the protein structure. In addition, MD simulated annealing was used to generate substrate flexibility in the algorithms, and residues in the active site (i.e., within 10 Å of the center of the putative site) were allowed to move. The poses with the highest score for each complex were selected as starting models for the subsequent MD simulations.

2.4 MD simulations

The *AMBER* 10 program suite (Case et al., 2005) was used to perform all the MD simulations in this study. Based on the electrostatic potential (ESP) calculations at the *ab initio* HF/6-31G* level (Frisch et al., 2004), the partial atomic charges for the *N*-CS and PZA were calculated by using the restricted electrostatic potential (RESP) fitting protocol implemented in the *Antechamber* module of the *AMBER* 10 package. The parameters for the two substrates were taken from the generalized *AMBER* force field (GAFF). The complexes were neutralized by adding sodium counterions, and were solvated in a rectangular box of water molecules with solvent layers 10 Å between the box edges and the solute surface.

Initially, 1000 steps of conjugated-gradient energy minimizations with the monomer Ta0454 and the inner substrate constrained at a force constant of 100 kcal mol⁻¹ Å⁻² were performed, followed by another 1000 steps of minimization for the entire system without the restraints. The system after minimization was heated from 0 K to 300 K for 50 ps with constant volume dynamics using the Langevin temperature equilibration scheme. For non-bonded interactions, a residue-based cutoff was employed with a cutoff distance of 10 Å. Another 50 ps simulation was performed to equilibrate the density of the system.

Periodic boundary dynamics simulations of 1 ns were performed for the production step in an NPT (constant composition, pressure, and temperature) ensemble at 1 atm and 300 K. The distances between the Zn²⁺ ion and the closest four protein atoms (Asp50-O_δ1, Hid52-N_ε2, Glu58-O_δ1, and Hid65-N_ε2) were constrained to maintain a distance of 1.9 Å to 2.3 Å (Zheng et al., 2008). The temperature was maintained at 300 K by means of the weak-coupling algorithm (Berendsen et al., 1984). The SHAKE algorithm (Miyamoto, et al., 1992) was used to constrain all bonds involving hydrogens with a tolerance of 1×10⁻⁵ Å. The Particle-Mesh-Ewald (Darden et al., 2002) method was applied to constrain the long range electrostatic interactions. The output trajectory files were saved every 1 ps for subsequent analysis. Equilibration was monitored by convergence in terms of the temperature, energy, and density of the system and the root-mean-squared deviations (RMSD) of the backbone atoms compared to the X-ray structure.

2.5 MM/PBSA analysis

The absolute binding free energy ($\Delta G_{\text{binding}}$) was calculated using the MM-PBSA procedure in *AMBER* 10. The electrostatic contribution to the solvation free energy was determined using the PBSA program in the *AMBER* 10 suite which numerically solves the Poisson-Boltzmann equations. A grid spacing of 0.5 Å was chosen, and the dielectric constant for the solute and solvent were set to 1 and 80, respectively. The optimized atomic radii set in *AMBER* 10 were used. 100 snapshots from 0.8 ns to 1.0 ns at the production step were collected at time intervals of 2 ps for the energetic analysis.

To check the role of the possible interactions between the Zn²⁺ ions with the *N*-CS and PZA substrates, identical MD simulation and MM/PBSA procedures were carried out for the following four systems: Ta0454/*N*-CS with and without zinc ions, Ta0454/PZA with zinc ions, and PH999/PZA with zinc ions.

3 Results and Discussion

3.1 Overall structure

A ribbon diagram of Ta0454 is shown in Figure 2. Two chains of Ta0454 form a dimer in the asymmetric unit (ASU). Each monomer buries about 2400 Å² of solvent-accessible surface in contact with other copies in the lattice, which corresponds to 29% of the total surface area of a monomer. Based on analysis of the packing of Ta0454 in the crystal structure by the PISA

server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html; Krissinel and Henrick, 2007), the protein is predicted to form either a tetramer (composed of two ASU dimers) or a dimer in solution. PISA's predicted increase in the free energy of solvation upon oligomerization is -88 kcal/mol for the dimer and -189 kcal/mol for the tetramer. A similar analysis by the PITA v0.4.3 service (<http://www.ebi.ac.uk/thornton-srv/databases/pita/>; Ponstingl et al., 2003) predicts that Ta0454 forms a dimer in solution with an assembly score of 106.9 (scores higher than 70 are considered to be significant). Similar analyses suggest that CHSase1 is tetrameric and PH0999 is monomeric in solution.

The monomer contains six β -strands and nine α -helices (Figure 2). The monomer of 173 amino acids forms a Rossmann-type fold similar to those of three other proteins (CSHase1, PH999, and Pnc1p). In this fold, the core of the monomer comprises central β sheets sandwiched between two layers of α -helices. Superposition of the two monomers using the SSM (secondary structure matching) algorithm (Krissinel and Henrick 2004) as implemented in *COOT* shows an RMSD value of $0.5 \text{ \AA}$ for 173 aligned residues.

PH999 is the most similar protein to Ta0454 in the PDB by sequence (34% identity), followed by CSHase1 (29% identity) and Pnc1p (24 % identity). The multiple sequence alignment of Ta0454 to the other three proteins is shown in Figure 3. The RMSD of 156 aligned pairs of C_{α} atoms between Ta0454 and CSHase1 is $1.5 \text{ \AA}$, while that of 165 aligned pairs of C_{α} atoms between Ta0454 and PH999 is $1.3 \text{ \AA}$.

3.2 The active site of Ta0454 by structural study

Ta0454 contains a highly conserved C-D-K catalytic triad (Cys123, Asp9, and Lys90), which is also observed in the other three structures (Figure 4). The triad comprises Cys177, Asp51, and Lys144 in CSHase1; Cys167, Asp8, and Lys122 in PH999; and Cys133, Asp10, and Lys094 in Pnc1p. A *cis*-peptide bond in the active site is also conserved in all three structures, between Leu118 and Asp119 in Ta0454; between Ala172 and Thr173 in CSHase1; between Val128 and Ala129 in PH999; and between Val162 and Ala163 in Pnc1p.

The metal in Ta0454 was identified as a zinc ion by its coordination and anomalous scattering behavior. The zinc ion is coordinated by four residues (Asp50, His52, Glu58, and His65) where the coordinating atoms form a distorted tetrahedron. The Zn^{2+} -binding site is proposed to play both a structural and catalytic role since it abuts the highly conserved binding site and constitutes a physical component of the presumed active site, which may stabilize a particular Ta0454 conformation and may be involved in catalysis. For PH999, the quite rare octahedral coordination in the Zn^{2+} -binding site is formed by residues Asp52, His54, His71, and three water molecules. Only three residues (Asp51, His53, and His94) constitute the Ni^{2+} -binding site in Pnc1p, whereas three residues (Trp56, Thr173, and Arg202) and two water molecules form the sulfate-binding site in CSHase1.

3.3 Molecular dynamics (MD) simulations

3.3.1 The Ta0454/*N*-CS complexes with and without zinc ions—Figure 5 (and Supplemental Figure S1) shows a schematic of the binding site of a simulated model of the Ta0454/*N*-CS complex with a Zn^{2+} ion after 1.1 ns of MD simulation, which can be used to elucidate the intermolecular interactions in the active site. To explore the dynamic stability of the Ta0454/*N*-CS complex, the RMSD fluctuations of backbone atoms compared to the X-ray structure were obtained over a 1 ns NPT trajectory and plotted in Figure 6a. The RMSD of the Ta0454/*N*-CS system indicates that the solvated system achieved equilibrium after 600 ps. Compared to the substrate *N*-CS (Figure 6b), the equilibrium of the conserved binding site was more readily established (Figure 6a) at the beginning of the production step.

Relative to the initial docking pose, *N*-CS rotates in a clockwise direction (about 35 degrees) to form a coordination bond (**CB1** = 2.13 Å) with the zinc ion. (All distances described are shown schematically on Figure 5.) A water molecule enters the substrate binding site after 175 ps (Figure 7) and contributes to the binding of the substrate to Ta0454. The S_γ atom of Cys123, the O_{δ1} atom of Asp9, and the entering water molecule form H-bonds (**H1** = 3.24 Å, **H2** = 3.50 Å, and **H3** = 3.37 Å, respectively) with the N1 atom of *N*-CS. The water molecule and the NH₂ atom of Arg148 make an H-bond (**H4** = 3.29 Å and **H5** = 2.85 Å respectively) with the O3 atom of *N*-CS. Another strong H-bond (**H6** = 2.67 Å) is also observed between the O_{δ2} atom of Asp119 and the O2 atom of *N*-CS. Besides the conserved five residues (Asp9, Lys90, Leu118, Asp119, and Cys123), Arg148 abuts the substrate binding site, which is not observed in CSHase1 (Romao et al., 1992; Zajc et al., 1996). The distance **CB1** for the coordination bond and two H-bonds (**H1** and **H3**) are monitored in Figure 7. It is clear that the coordination bond is maintained at a length of around 1.98 Å. In addition to the six H-bonds, electrostatic interactions between *N*-CS and the four residues (Asp50, Glu58, Hid65, and Asp119; shown as bonds **E1**, **E2**, **E3**, **E4**, and **E5** on Figure 5) are observed with all interaction distances less than 3.0 Å. The two different kinds of interactions between Ta0454 and *N*-CS (coordination bonds and H-bonds) are the major forces in stabilizing the favorable *N*-CS binding pattern.

Removal of the zinc ion from the initial docking model for the Ta0454/*N*-CS system did not lead to a significantly different configuration for the system under identical MD conditions.

3.3.2 The Ta0454/PZA and PH999/PZA complexes with zinc ions—Figure 8 (and Supplemental Figure S2) shows a schematic of the binding site of the simulated Ta0454/PZA complex with zinc ions after 1.1 ns of MD simulation. Comparing it with the schematic of the Ta0454/*N*-CS complex in Figure 5, the H-bonds **H1**, **H3**, **H4**, **H5**, and **H6** are not conserved. The distance between the S_γ atom of Cys123 and the N1 atom of PZA is 3.72 Å, which suggest there is only a weak interaction between these two atoms. Lys90 and another water molecule make two strong H-bonds (**H7** = 3.01 Å and **H8** = 2.76 Å) with PZA. However, there is not an H-bond interaction between Lys90 and *N*-CS in Figure 5. The electrostatic interactions (**E1**, **E2**, **E3**, and **E4**) and the coordination bond (**CB1**) in the catalytic site are identified in both the Ta0454/PZA and Ta0454/*N*-CS complexes. Unlike the Ta0454/PZA complex, the H-bond interaction and the electrostatic interaction formed in the Ta0454 complex between *N*-CS and Asp119, which forms a rare *cis*-peptide, could possibly stabilize *N*-CS in the active site. π - π stacking interactions are observed between PZA and the neighboring Phe14 and Trp62 residues in Ta0454, with planar distances between the cores of the aromatic rings of less than 3.9 Å (Figure 8). As a result, these energetically favorable interactions stabilize PZA in the substrate binding pocket.

Like CSHase, a crystal structure of PH999 with an ordered substrate is not available in the PDB database. Except for the π - π stacking interactions, the simulated PH999/PZA model (seen in Supplemental Figure S3) indicates a similar substrate binding pattern (data not shown) to the Ta0454/PZA complex. While residues Phe14 and Trp67 are conserved in PH999 (Figure 3), the initial configuration of the PH999/PZA complex after molecular docking placed the sidechain of Trp67 in an orientation rotated away from the aromatic ring of PZA, and a dihedral angle of 28° between the phenyl rings of PZA and Phe14 was observed. The relative orientations of Trp67 and Phe14 to the phenyl ring of PZA (by monitoring the dihedral angles) did not change over the course of the 1 ns NPT trajectory.

3.4 MM/PBSA analysis

Based on the MD simulated-trajectories, the MM/PBSA method predicts free energies of binding of -24.5 and -15.2 kcal/mol for the Ta0454/*N*-CS complexes with and without zinc

ions, respectively (Table 2). According to the energy components of the binding free energies, the major favorable contributions to the substrate binding are the electrostatic energies, whereas the electrostatic contributions to the solvation oppose the binding. Removal of the zinc ion leads to a significant decrease in the absolute free energy of binding, which suggests that the coordination bond **CB1** between the zinc ion and *N*-CS plays an important role in the favorable binding pattern. For the zinc-bound case, only the zinc ion and *N*-CS in the complex are used to carry out the MM-PBSA procedure, which results in a free energy of binding of -7.6 kcal/mol.

The resulting free energies of binding for the Ta0454/PZA and PH999/PZA systems with zinc ions are -17.1 and -14.3 kcal/mol, respectively. The absolute free energy of binding of the former is slightly higher than that of the latter. It is clear that the two π - π stacking interactions can account for the difference in the binding free energies for two systems.

For the PH999/PZA system, the experimentally measured k_{cat} for the catalytic activity of PH999 increased 19-fold to about $333 \mu\text{mol/mg/minute}$ in the presence of 2 mM ZnSO_4 (Du et al., 2001). The zinc ions were removed from our PH999/PZA simulated complex, and the resulting binding free energy for the new counterpart was -8.8 kcal/mol after the MD simulation and MM/PBSA procedures. The increase in the predicted absolute binding free energy by the zinc ion in our PH999/PZA system is consistent with the previous experimental results (Du et al., 2001). Removal of the zinc ion from the Ta0454/PZA system also results in a lower binding free energy (-6.8 kcal/mol), which demonstrates that zinc ions play a crucial role in the substrate binding pattern for PZA as well.

4 Conclusions

The present study provides a structural insight into the substrate binding patterns for the Ta0454/*N*-CS and Ta0454/PZA complexes. The predicted binding free energies using the MM/PBSA method suggest that Ta0454 binds *N*-CS more strongly than PZA. The Ta0454/PZA complex shows a slightly higher binding free energy than the PH999/PZA counterpart, which can be attributed to the additional π - π stacking interactions between PZA and Phe14/Trp62 in Ta0454.

Our results highlight that the comparisons between the zinc-bound and unbound proteins is crucial in evaluating the contributions of key residues and enzymatic factors such as zinc ions. Further crystal and structural studies are required to validate the Ta0454/*N*-CS binding pattern suggested by our docking and MD simulations, as well as to elucidate the catalytic mechanism of *N*-CS hydrolysis by Ta0454.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Alexander Wlodawer, Andrzej Joachimiak and the members of the Structural Biology Center at the Advanced Photon Source and the Midwest Center for Structural Genomics for help and discussions. The work described in the paper was supported by NIH PSI grants GM62414 and GM074942. The results shown in this report are derived from work performed at Argonne National Laboratory, at the Structural Biology Center of the Advanced Photon Source. Argonne is operated by University of Chicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

References

- Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–3690.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy minimization and dynamics calculations. *J Comp Chem* 1983;4:187–217.
- Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem* 2005;26:1668–1688. [PubMed: 16200636]
- Cowtan K. Joint CCP4 and ESF-EACBM newsletter. *Protein Crystallogr* 1994;31:34–38.
- Darden T, York D, Pedersen L. Particle mesh Ewald: An $N\log(N)$ method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–10092.
- Dong A, Xu X, Edwards AM. Midwest Center for Structural Genomics, and Structural Genomics Consortium. In situ proteolysis for protein crystallization and structure determination. *Nat Methods* 2007;4(12):1019–21. [PubMed: 17982461]
- Du XL, Wang WR, Yakota H, Nguyen H, Kim SH. Crystal structure and mechanism of catalysis of a pyrazinamidase from *Pyrococcus horikoshii*. *Biochemistry* 2001;40:14166–14172. [PubMed: 11714269]
- Dunn SR, Gabuzda GM, Superdock KR, Krolecki RS, Schaedler RW, Simenhoff ML. Induction of creatininase activity in chronic renal failure: Timing of creatinine degradation and effect of antibiotics. *Am J Kidney Dis* 1997;29:72–76. [PubMed: 9002532]
- Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr* 2004;D60:2126–2132.
- Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in Molecular Recognition: The effects of ligand and protein: Flexibility on molecular docking accuracy. *J Med Chem* 2004;47:45–55. [PubMed: 14695819]
- Farrow JA, Lawson PA, Hippe H, Gauglitz U, Collins MD. Phylogenetic evidence that the gram-negative nonsporulating bacterium *Tissierella* (*Bacteroides*) *praeacuta* is a member of the *Clostridium* subphylum of the gram-positive bacteria and description of *Tissierella creatinini* sp. nov. *Int J Syst Bacteriol* 1995;45:436–40. [PubMed: 8590669]
- Frisch, MJ.; Trucks, GW.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA., Jr; Vreven, T., et al. Gaussian 03M, Revision E.01. Gaussian, Inc; Wallingford, CT: 2004.
- Giessing M, Liefeldt L. Getting the measure of renal function: Determination of best predicted creatinine in living-donor kidney renal transplant recipients based on donor kidney function. *Transplantation* 2008;86:513–514. [PubMed: 18724218]
- Hong MC, Chang JC, Wu ML, Chang MC. Expression and export of *Pseudomonas putida* NTU-8 creatinase by *Escherichia coli* using the chitinase signal sequence of *Aeromonas hydrophila*. *Biochem Genet* 1998;36:407–415. [PubMed: 10230521]
- Hu G, Taylor AB, McAlister-Henn L, Hart PJ. Crystal structure of the yeast nicotinamidase Pnc1p. *Arch Biochem Biophys* 2007;461:66–75. [PubMed: 17382284]
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr* 1991;A47:110–119.
- Kapust RB, Tozser J, Copeland TD, Waugh DS. The P1' specificity of tobacco etch virus protease. *Biochem Biophys Res Commun* 2002;294:949–955. [PubMed: 12074568]
- Kim JM, Shimizu S, Yamada H. Purification and characterization of a novel enzyme, N-carbamoylsarcosine amidohydrolase from *Pseudomonas putida* 77. *J Biol Chem* 1986;261:1832–1839.
- Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr* 2004;D60:2256–2268.
- Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007;372:774–797. [PubMed: 17681537]
- Kuhn B, Kollman PA. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J Med Chem* 2000;43:3786–3791. [PubMed: 11020294]

- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283–291.
- Lovell SC, Davis IW, Arendall WB III, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by $\text{C}\alpha$ geometry: ϕ , ψ and $\text{C}\beta$ deviation. *Proteins* 2003;50:437–450. [PubMed: 12557186]
- Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: The integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallogr* 2006;D62:859–866.
- Miyamoto S, Kollman PA. Settle - An analytical version of the shake and rattle algorithm for rigid water models. *J Comput Chem* 1992;13:952–962.
- Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr* 1997;D53:240–255.
- Nishiya Y, Toda A, Imanaka T. Gene cluster for creatinine degradation in *Arthrobacter* sp TE1826. *Mol Gen Genet* 1998;257:581–586. [PubMed: 9563845]
- Otwinowski, Z. Proceedings of the CCP4 study weekend. In: Wolf, W.; Evans, PR.; Leslie, AGW., editors. *Isomorphous replacement and anomalous scattering*. Daresbury Laboratory; Warrington, UK: 1991. p. 80-86.
- Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, CW., Jr; Sweet, RM., editors. *Methods in enzymology: Macromolecular crystallography*. Vol. 276. Academic Press; New York: 1997. p. 307-326.part A
- Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 1999;6:458–463. [PubMed: 10331874]
- Ponstingl H, Kabir T, Thornton JM. Automatic inference of protein quaternary structure from crystals. *J Appl Crystallogr* 2003;36:1116–1122.
- Romao MJ, Turk D, Gomisruth FX, Huber R, Schumacher G, Mollering H, Russmann L. Crystal-structure analysis, refinement and enzymatic-reaction mechanism of N-carbamoylsarcosine amidohydrolase from *Arthrobacter* sp at 2.0-angstrom resolution. *J Mol Biol* 1992;226:1111–1130. [PubMed: 1381445]
- Schneider TR, Sheldrick GM. Substructure solution with SHELXD. *Acta Crystallogr* 2002;D58:1772–1779.
- Sheldrick GM. A short history of SHELX. *Acta Crystallogr* 2008;A64:112–122.
- Stoica I, Sadiq SK, Coveney PV. Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. *J Am Chem Soc* 2008;130:2639–2648. [PubMed: 18225901]
- Terwilliger TC. Automated structure solution, density modification and model building. *Acta Crystallogr* 2002;D58:1937–1940.
- Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot Eng* 1995;8:127–134.
- Wu G, Robertson DH, Brooks CL III, Vieth M. A detailed analysis of grid-based molecular docking. A case study of CDOCKER- A CHARMM based MD docking algorithm. *J Comput Chem* 2003;24:1549–1562. [PubMed: 12925999]
- Yamada H, Shimizu S, Kim JM, Shinmen Y, Sakai T. A novel metabolic pathway for creatinine degradation in *Pseudomonas putida* 77. *FEMS Microbiol Let* 1985;30:337–340.
- Zajc A, Romao MJ, Turk B, Huber R. Crystallographic and fluorescence studies of ligand binding to N-carbamoylsarcosine amidohydrolase from *Arthrobacter* sp. *J Mol Biol* 1996;263:269–283. [PubMed: 8913306]
- Zhang RG, Skarina T, Katz JE, Beasley S, Khachatryan A, Vyas S, Arrowsmith CH, Clarke S, Edwards A, Joachimiak A. Structure of *Thermotoga maritima* stationary phase survival protein SurE: A novel acid phosphatase. *Structure* 2001;9:1095–1106. [PubMed: 11709173]
- Zheng H, Chruszcz M, Lasota P, Lebioda L, Minor W. Data mining of metal ion environments present in protein structures. *J Inorg Biochem* 2008;102:1765–1776. [PubMed: 18614239]

Figure 1. Reactions catalyzed by *N*-carbamoylsarcosine amidase (CSHase), pyrazinamidase (PZAase), and nicotinamidase.

Figure 2. Ribbon diagram of the dimer structure of the *N*-carbamoylsarcosine amidase (CSHase) Ta0454 from *Thermoplasma acidophilum* (PDB access code: 3EEF). The ribbon is colored by B factor, except the residues coordinating zinc cation.

Figure 3.

Multiple sequence alignment of the presumed *N*-carbamoylsarcosine amidase (CSHase, 3EEF) Ta0454 from *Thermoplasma acidophilum*, CSHase1 (1NBA) from *Arthrobacter* sp., pyrazinamidase PH999 (1IM5) from *Pyrococcus horikoshii*, and nicotinamidase Pnc1p (2HOR) from *Saccharomyces cerevisiae*. The structure-based alignment of the four sequences was made with MultiProt (<http://bioinfo3d.cs.tau.ac.il/MultiProt/>).

Figure 4.

Alignment of the selected five residues in the conserved binding sites of the presumed *N*-carbamoylsarcosine amidase (CSHase, 3EEF) Ta0454 from *Thermoplasma acidophilum*, CSHase1 (1NBA) from *Arthrobacter* sp., pyrazinamidase PH999 (1IM5) from *Pyrococcus horikoshii*, and nicotinamidase Pnc1p (2H0R) from *Saccharomyces cerevisiae*. The residues of 3EEF, 1NBA, 2H0R, and 1IM5 are colored red, green, orange, and cyan, respectively.

Figure 5.

A binding model made with LIGPLOT (Wallace et al., 1995) of the Ta0454/*N*-carbamoylsarcosine (*N*-CS) complex after 1.1 ns of MD simulation, used to elucidate the interactions in the active site. The green numbers indicate in Å the distances (CB1 for the coordination bond; H1, H2, H3, H4, H5, and H6 for the H-bond interactions; E1, E2, E3, E4, and E5 for the electrostatic interactions) between pairs of atoms. Asp9, Lys90, Leu118, Asp119, Cys123, and Arg148 are involved in the substrate binding site, while Asp50, Glu58, Hid65 are involved in the catalytic site. The coordination bonds between the zinc ion and residues Asp50, Glu58, and Hid65 are omitted for clarity. Oxygen, nitrogen, carbon, and sulfur atoms are colored red, blue, black, and yellow, respectively. Please refer to the web version of this paper for an alternative color interpretation.

Figure 6.

Plots of RMSD (Å) versus simulation time in the MD-simulated Ta0454/ *N*-carbamoylsarcosine (*N*-CS) complex with zinc ions. RMSD represents the root mean-square deviation (Å) of the simulated positions of the backbone atoms (C, N, and C_α) of Ta0454/the conserved binding site from those in the initial x-ray crystal structure (a), and the non-hydrogen atoms of the substrate *N*-CS (b) from those in the initial docking structure, respectively.

Figure 7.

Plots of the key intermolecular distances (Å) versus the simulation time in the MD-simulated Ta0454/*N*-carbamoylsarcosine (*N*-CS) complex with zinc ions. *N*-CS(O1)-ZNA(Zn) refers to the distance between the O1 atom of *N*-CS and the zinc ion in Ta0454. *N*-CS(N1)-Cys123 (S_γ) refers to the distance between the N1 atom of *N*-CS and the S_γ atom in residue Cys123. *N*-CS(N1)-Wat(O) refers to the distance between the N1 atom of *N*-CS and the oxygen atom in the nearest water molecule.

Figure 8.

A binding model made with LIGPLOT (Wallace et al., 1995) of the Ta0454/pyrazinamide (PZA) complex after 1.1 ns of MD simulation, used to elucidate the interactions in the active site. The green numbers in Å indicates the distances (CB1 for the coordination bond; H2, H7, and H8 for the H-bond interactions; E1, E2, E3, E4, and E5 for the electrostatic interactions) between two atoms. Asp9, Phe14, Trp62, Lys90, Leu118, Asp119, and Cys123 are involved in the substrate binding site, while Asp50, Glu58, and Hid65 are involved in the catalytic site. Two π - π stacking interactions are identified between Phe14/Trp62 and PZA. The coordination bonds between the zinc ion and Asp50, Glu58, and Hid65 are omitted for clarity. The oxygen, nitrogen, carbon, and sulfur atoms are colored red, blue, black, and yellow, respectively. Please refer to the web version of this paper for an alternative color interpretation.

Table 1

Summary of data collection, phasing, and refinement for *N*-carbamoylsarcosine amidase (CSHase) Ta0454 from *Thermoplasma acidophilum*

Organism	<i>Thermoplasma acidophilum</i>
	PDB code: 3EEF
Crystal	Ta0454
Space group	P6 ₂
Unit cell	a = b = 84.0 Å, c = 98.7 Å $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$
Solvent content (%)	49.0
Matthews coefficient (Å ³ /Da)	2.45
No. of molecules in asymmetric unit	2
Data collection	
Diffraction protocol	SAD
Wavelength (Å)	0.9793
Resolution (Å)	40.85-2.35
Highest resolution shell (Å)	2.39-2.35
Observed reflections	235900
Unique reflections	16369
Redundancy	14.3 (14.2)
Completeness (%)	99.6 (100.0)
<i>I</i> / σ (<i>I</i>)	65.8 (7.6)
R _{merge} anom (%)	7.0 (43.0)
R _{merge} (%)	8.8 (47.1)
Refinement	
R _{work} (%)	19.4 (20.2)
R _{free} (%)	24.1 (27.8)
RMSD for bond length (Å)	0.014
RMSD for bond angles (°)	1.32
Number of total atoms	2687
Number of protein atoms	2633
Number of water molecules	52
Average B factor (Å ²)	39.9
Ramachandran statistics	
Most favored regions (%)	96.8
Additional allowed regions (%)	3.2

Data for the highest resolution shell are given in parentheses.

Ramachandran statistics were calculated with *MOLPROBITY*.

Table 2

The calculated binding free energies ($\Delta G_{\text{binding}}$) for the four systems using the MM/PBSA approach (all energies $\pm$ standard error of the mean in kcal/mol).

No.	Protein	Substrate	$\Delta G_{\text{binding}}$	ΔELE	ΔVDW	ΔPBCAL
1	Ta0454	N-CS	-24.5 (± 3.7)	-66.2 (± 7.3)	-10.0 (± 3.8)	54.4 (± 4.2)
		N-CS	-7.6 (± 6.6)	-90.58 (± 7.3)	3.0 (± 1.5)	80.7 (± 3.5)
2	Ta0454 ^b	N-CS	-15.2 (± 1.8)	-32.0 (± 5.6)	-13.3 (± 3.4)	32.6 (± 4.4)
3	Ta0454	PZA	-17.1 (± 3.7)	-35.4 (± 5.3)	-17.4 (± 2.6)	38.3 (± 2.5)
4	PH999	PZA	-14.3 (± 2.5)	-41.3 (± 5.4)	-13.1 (± 3.0)	42.9 (± 2.8)

Ta0454 = *N*-carbamoylsarcosine amidase (3EEF) from *Thermoplasma acidophilum*,

PH999 = pyrazinamidase (IIM5) from *Pyrrococcus*,

N-CS = *N*-carbamoylsarcosine and PZA = pyrazinamide,

MM/PBSA = molecular mechanics/Poisson-Boltzmann surface area,

ELE = electrostatic energy as calculated by the MM force field,

VDW = van der Waals contribution from MM,

PBCAL = the electrostatic contribution to the solvation free energy calculated by PB.

^a only the zinc ion and N-CS in the complex are used to calculate the binding free energy.

^b the TA0454/N-CS complex without the zinc ion.