

# The origin of the electrostatic perturbation in acetoacetate decarboxylase

Meng-Chiao Ho<sup>1</sup>, Jean-François Ménéret<sup>1</sup>, Hiro Tsuruta<sup>2</sup> & Karen N. Allen<sup>1†</sup>

Acetoacetate decarboxylase (AADase) has long been cited as the prototypical example of the marked shifts in the  $pK_a$  values of ionizable groups that can occur in an enzyme active site. In 1966, it was hypothesized that in AADase the origin of the large  $pK_a$  perturbation ( $-4.5$  log units) observed in the nucleophilic Lys 115 results from the proximity of Lys 116, marking the first proposal of microenvironment effects in enzymology. The electrostatic perturbation hypothesis has been demonstrated in a number of enzymes, but never for the enzyme that inspired its conception, owing to the lack of a three-dimensional structure. Here we present the X-ray crystal structures of AADase and of the enamine adduct with the substrate analogue 2,4-pentanedione. Surprisingly, the shift of the  $pK_a$  of Lys 115 is not due to the proximity of Lys 116, the side chain of which is oriented away from the active site. Instead, Lys 116 participates in the structural anchoring of Lys 115 in a long, hydrophobic funnel provided by the novel fold of the enzyme. Thus, AADase perturbs the  $pK_a$  of the nucleophile by means of a desolvation effect by placement of the side chain into the protein core while enforcing the proximity of polar residues, which facilitate decarboxylation through electrostatic and steric effects.

AADase, a 365 kDa homododecameric enzyme that catalyses the conversion of acetoacetate to acetone, is a key component in the anaerobic metabolism of carbohydrate in solventogenic bacteria. In the early 1960s, Westheimer used AADase to pioneer the application of methods in physical organic chemistry to the study of the chemical and catalytic mechanism of enzymes<sup>1</sup>. These studies revealed that the mechanism of AADase proceeds through a Schiff-base intermediate formed by reaction of Lys 115 with substrate<sup>2–5</sup>. A reporter group used to measure directly the  $pK_a$  of Lys 115 revealed it to be 5.96, a value 4.5 orders of magnitude below that expected<sup>6</sup>. Westheimer hypothesized that the  $pK_a$  of Lys 115 was electrostatically perturbed by charge–charge repulsion due to the proximity of the protonated  $\epsilon$ -amino group of an adjacent Lys 116 (ref. 4). This marked the first appearance of the proposal of microenvironment effects in enzymology, a hypothesis that has since been demonstrated in a number of enzymes<sup>7</sup> but never for the enzyme that inspired its conception, AADase.

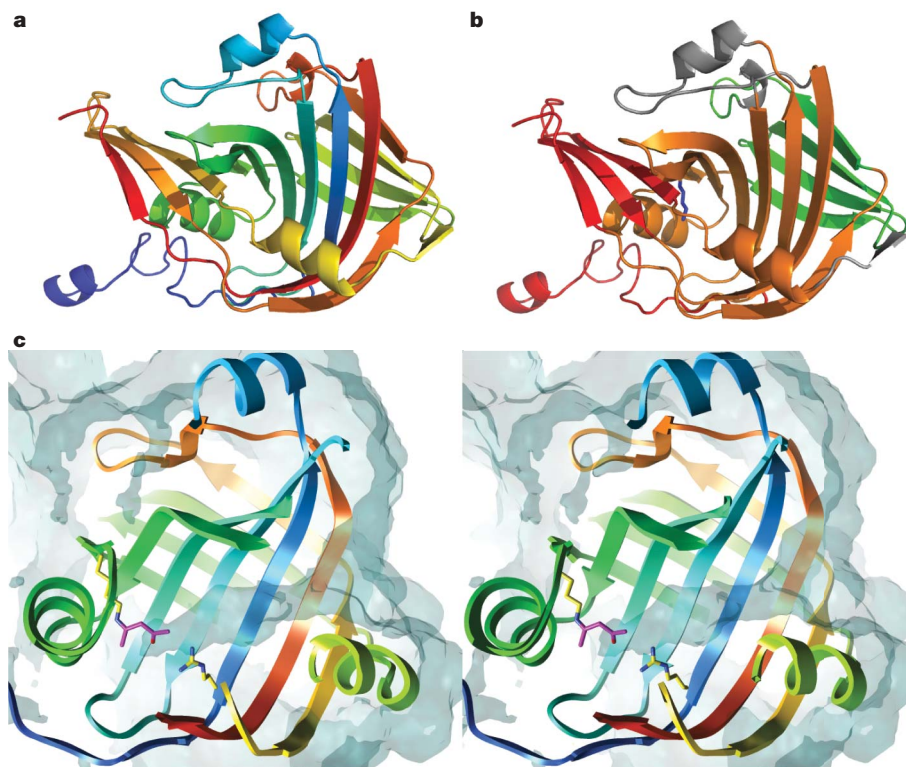
Here we present the X-ray crystal structures of AADase from *Clostridium acetobutylicum* (CaAAD) and *Chromobacterium violaceum* (CvAAD, 50% sequence identity) at 2.4 Å and 2.1 Å resolution, respectively. CaAAD ( $k_{cat} = 165 \text{ s}^{-1}$ ,  $K_m = 4.1 \text{ mM}$ ) and CvAAD ( $k_{cat} = 349 \text{ s}^{-1}$ ,  $K_m = 5.7 \text{ mM}$ ) have the same catalytic efficiency, subunit structures (1.2 Å root mean squared deviation, Supplementary Fig. 1) and overall oligomerization properties. The AADase structure exhibits a previously unknown fold (Fig. 1a and Supplementary Fig. 2) consisting predominately of  $\beta$ -strands (15  $\beta$ -strands and 5 short  $\alpha$ -helices with a  $\beta$ -strand and  $\alpha$ -helical content of 44.3% and 15.6%, respectively). The tertiary structure is formed by three antiparallel  $\beta$ -sheets, dominated by a central seven-stranded cone-shaped  $\beta$ -barrel ( $\beta$ -cone) and flanked by four-stranded and three-stranded  $\beta$ -sheet structures (Fig. 1b). The strands of the  $\beta$ -cone are twisted such that the first and last strands are perpendicular to one another, similar to a single blade of a  $\beta$ -propeller. The order of the three sheets is discontinuous, with the carboxy-terminal sequence of the

protein forming the final strand of all three sheets. Each protomer includes a complete active site, wherein the  $\beta$ -cone encompasses the active site in its hollow core, which extends to a depth of 26.8 Å, with the catalytic Lys 115 positioned at the bottom (CaAAD residue numbering is used throughout). Examination of the solvent-accessible surface shows a single narrow channel (Fig. 1c) leading from bulk solvent at the rim of the cone to the active-site Lys  $\epsilon$ -amino moiety. We suggest that the AADase fold be called the Westheimer fold in honour of the many contributions of Frank H. Westheimer to the field of mechanistic enzymology<sup>8</sup>.

## A hydrophobic active site

The originally proposed  $pK_a$  perturbation of Lys 115 by means of Coulombic destabilization by a like-charged residue is not supported by this structure, because the  $\epsilon$ -amino groups of Lys 115 and Lys 116 are separated by 14.8 Å (Fig. 2a). Instead, it is the hydrophobic environment of the active site, comprising Phe 26, Leu 71, Tyr 74, Met 96, Leu 98, Tyr 113 and Leu 233 (Fig. 2b), that destabilizes the protonated amine. Furthermore, the side chain of Lys 115 does not form any hydrogen-bonding interactions. Notably, Lys 115 is positioned in parallel with, and in close proximity ( $\sim 4.7$  Å) to, the aromatic ring of Tyr 113, but it is oriented such that there is little potential for stabilization of the lysine ammonium group by cation– $\pi$  interactions<sup>9</sup>. Indeed, the constrained  $\phi/\psi$  dihedral angles of Pro 114 maintain the relative orientation of Tyr 113 and Lys 115. Sequence alignment demonstrates that this (Y/F)PKK motif and neighbouring hydrophobic residues are conserved in all 25 analogues of AADase (Supplementary Table 2). The structure supports a physical role for Lys 116 in the precise positioning of the nucleophilic Lys 115. This is consistent with previous studies including site-directed mutagenesis and chemical rescue<sup>10</sup> that established the essentiality of the  $\epsilon$ -amino group of Lys 116 in catalysis and the maintenance of the depressed  $pK_a$  of the nucleophilic Lys. Lys 116 lies at the subunit interface, pinning the  $\beta$ -strand on which it and Lys 115 reside into position by means of

<sup>1</sup>Department of Physiology and Biophysics, Boston University School of Medicine, Boston, Massachusetts 02118-2394, USA. <sup>2</sup>Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, MS69, Menlo Park, California 94025-7015, USA. †Present address: Department of Chemistry, Boston University, Boston, Massachusetts 02215-2521, USA.



**Figure 1 | The AADase tertiary structure depicted as a ribbon diagram.** **a**, CaAAD ramped from blue to red (N terminus to C terminus) and **b**, coloured to differentiate the central cone shaped seven-stranded  $\beta$ -sheet (orange), the four-stranded sheet that comprises the trimer interface (green) and the three-stranded sheet and helix that is a major component of the dimer interface (red). The lysine nucleophile (blue sticks) lies at the bottom

of the active site. **c**, Stereo-view of the active site funnel highlighted by means of a Connolly surface (blue transparent surface, showing accessibility to a 1.4 Å radius solvent probe) overlaid on a cutaway of the ribbon diagram coloured as in **a**. For reference, Lys 115 and Arg 29 (cpk coloured ball and stick) from the structure of CvAAD liganded to 2,4-pentanedione (magenta) are shown.

hydrogen-bonding interactions with Ser 16 and the carbonyl oxygen of Met 210 (Fig. 2a).

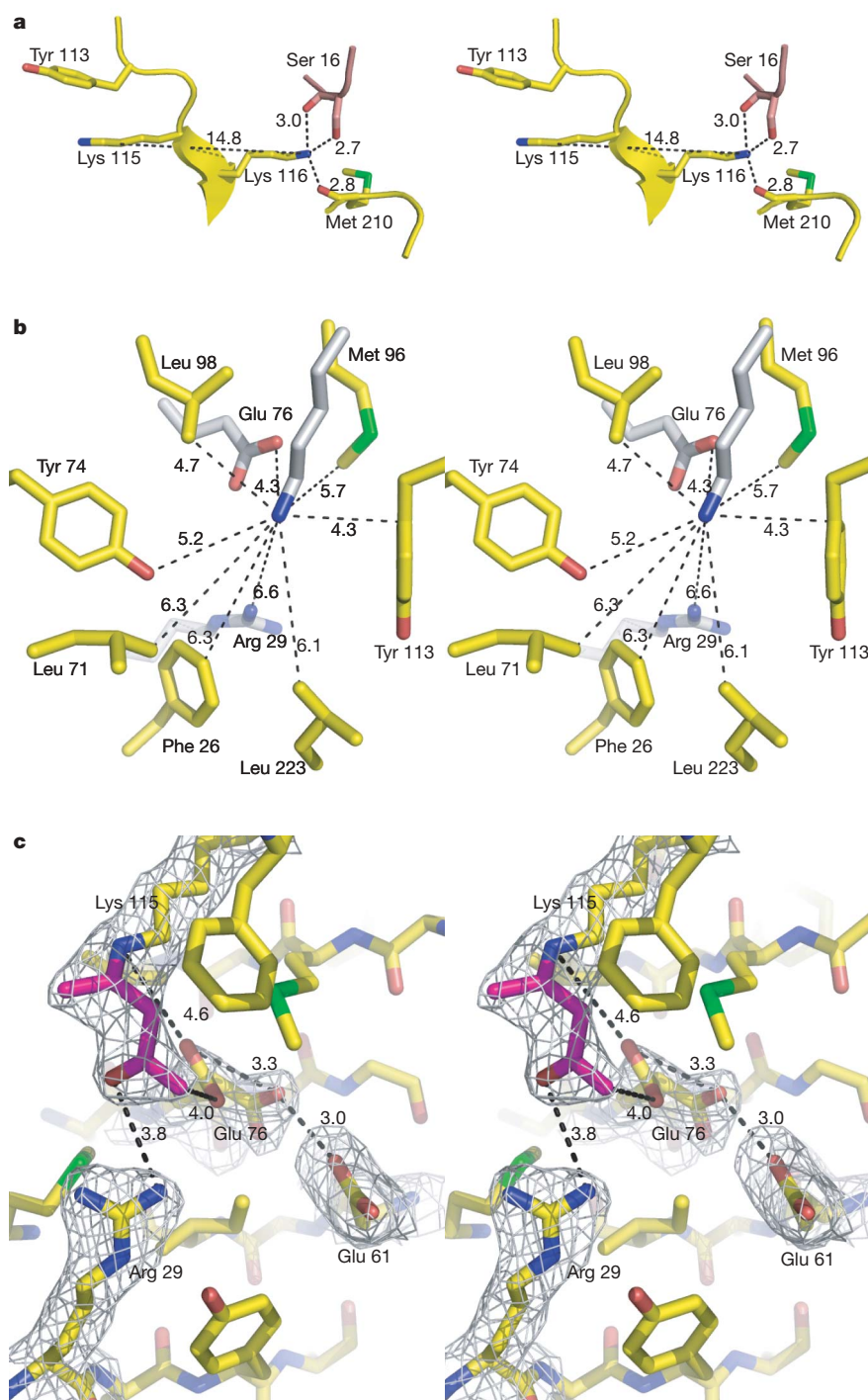
There are only two charged residues, Arg 29 and Glu 76, near the active-site cavity (Fig. 2b); these are strictly conserved in AADase. Arg 29 and Glu 76 are too distant (6.6 Å and 4.3 Å, respectively) from the Lys 115  $\epsilon$ -amino group to form hydrogen bonds or salt bridges, and thus the local environment is non-polar. Therefore, perturbation of the  $pK_a$  of Lys 115 results from the energetically unfavourable process of transferring a charged group from a polar, aqueous solvent (in this case water, with dielectric constant  $\epsilon = 78.54$ ) to the non-polar interior of a protein (with  $\epsilon = 4$ )<sup>11–13</sup>. AADase is not the only example of charge destabilization by means of desolvation in an enzyme; indeed, a similar  $\Delta pK_a$  of  $-4.9$  was determined for a mutant of staphylococcal nuclease in which a Val buried in the hydrophobic core of the protein was replaced by Lys<sup>14</sup>. Likewise, the reactivity of the Schiff base forming Lys in a catalytic antibody with decarboxylase and aldolase activity at neutral pH was demonstrated to be due to the hydrophobicity of the binding site by means of X-ray crystallography and the analysis of the linear free energy relationship between substrate partitioning into *n*-octanol and  $k_{cat}/K_m$ <sup>15</sup>. These cases closely mimic the case of AADase in which the neutral form of the Lys side chain is favoured by placement in a hydrophobic active site. Moreover, binding of a carboxylate group in a site less polar than water has been shown to be an important strategy for catalysis in decarboxylation reactions<sup>16</sup>.

### The catalytic mechanism

The structure of AADase bound to a ligand was also determined to investigate potential conformational changes upon substrate binding by enzyme and to provide insight into the interactions of the substrate with the active-site residues. The intermediate analogue 2,4-pentanedione acts as a potent inhibitor of AADase ( $K_i = 7 \times 10^{-7}$  M

for CaAAD)<sup>17</sup> by forming an enamine adduct with the enzyme-based nucleophile, mimicking the substrate Schiff base. The crystal structure of CvAAD complexed with 2,4-pentanedione (2.1 Å resolution) shows no major conformational differences when compared to the unliganded enzyme (0.5 Å root mean squared deviation). The position of the ligand with respect to the solvent-accessible surface (Fig. 1c) shows that the acyl group of the inhibitor sits at the bottom of the active site, with the  $\beta$ -carbonyl group corresponding to the carboxylate of acetoacetate pointing towards the solvent channel. This finding indicates that the funnel shape of the  $\beta$ -cone guides the substrate to the active site. There are only two polar residues in proximity to the bound inhibitor, Arg 29 and Glu 76 (Fig. 2c). The position of Arg 29 in the analogous substrate complex would allow a weak ionic interaction (3.8 Å) with the substrate  $\beta$ -carboxylate, favouring a productive binding mode wherein the carbonyl group of the substrate points towards the catalytic lysine and carboxylate group towards solvent, but without stabilizing the carboxylate ground state. A role in promoting productive substrate binding for Arg 29 is consistent with the results of kinetic analysis of CaAAD bearing an Arg29Gln mutation (Supplementary Table 3). The activity of the mutant was too low to accurately determine the kinetic constants; however, the upper limit for activity shows a diminution of  $>2,000$ -fold in maximal rate, which could only be measured at high substrate concentrations ( $>10$  times  $K_m$  of wild type). The fact that the catalytic activity of Arg29Gln CaAAD did not increase at pH values above the wild-type optimum for CvAAD of  $\sim 5.4$  (pH range tested 4.2–6.6, Supplementary Fig. 3) indicates that Arg 29 does not cause the  $pK_a$  perturbation of Lys 115.

In the decarboxylation of  $\beta$ -keto acids, the bond undergoing cleavage must be out of the plane of the imine  $\pi$  bond<sup>18</sup>. In the AADase enamine complex, the observed electron density for the 2,4-pentanedione shows that the proximity of Glu 76 forces the group corresponding to the carboxylate group of substrate out of the plane of the enolic oxygen



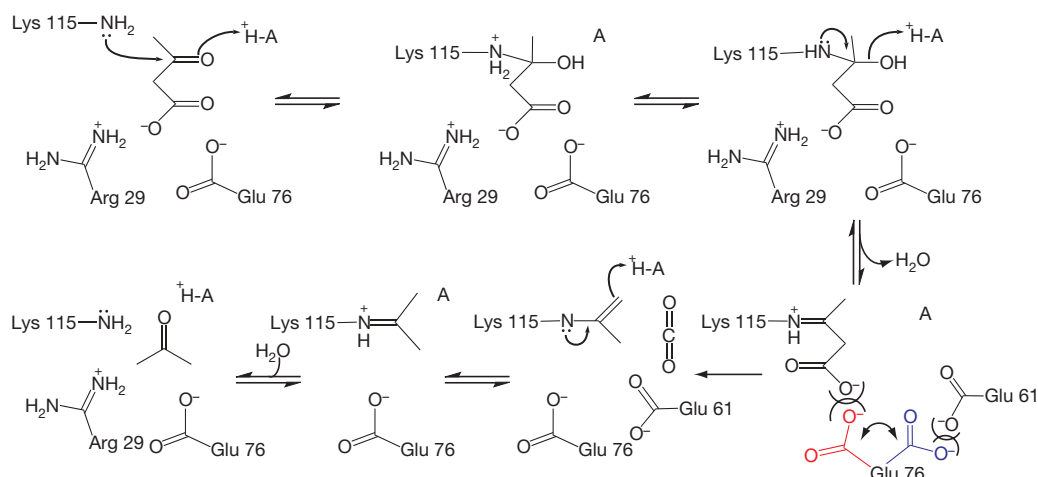
**Figure 2 | The positions of Lys 115 and Lys 116 and surrounding environment.** **a**, Distances (in Å) between  $\epsilon$ -amino groups of Lys 115 and Lys 116 and between Lys 116 and neighbouring residues from the same subunit (yellow) and adjacent subunit (pink) are shown in stereo and labelled. **b**, Stereo view of the active site of AADase with Lys 115 (light blue) and surrounding hydrophobic (yellow) and conserved polar (grey) residues

depicted as sticks (distances in Å). **c**, Stereo diagram of the active site of the CvAAD intermediate analogue complex with active site residues (yellow) and 2,4-pentanedione (magenta) shown as sticks overlaid with a  $2F_o - F_c$  (contoured at  $1\sigma$ ) electron density map (grey cages). Note that there are two conformations for Glu 76. The distances (in Å) are shown.

and Schiff-base nitrogen (Fig. 2c and Supplementary Fig. 4), favouring the expected geometry for decarboxylation. This role for the carboxylate group in substrate alignment is consistent with the observed 250-fold decrease in  $k_{cat}$  (with no significant change in  $K_m$ ) measured for the Glu76Gln mutant. Notably, the lack of a change in pH optimum for this mutant enzyme argues against an effect of Glu 76 on the ionization of the nucleophilic Lys (Supplementary Fig. 3).

In addition, in the unliganded and liganded complexes the side chain of Glu 76 has two alternative conformations (Fig. 2c). In one

conformation Glu 76 (20% occupancy in the liganded enzyme subunit B and 100% occupancy in all subunits of the unliganded enzyme) is 4.0 Å from the  $\beta$ -carbonyl group of the inhibitor, whereas in the second it is 3.3 Å from the carboxylate of Glu 61 (positioned closer to the opening of the solvent channel). By analogy, in the substrate complex the side chain of Glu 76 may alternate between two positions, one proximal to Glu 61 (where it would be destabilized by like-charge repulsion) and one near to the  $\beta$ -carboxylate of the substrate Schiff base (favouring decarboxylation; Fig. 3). This model is supported by



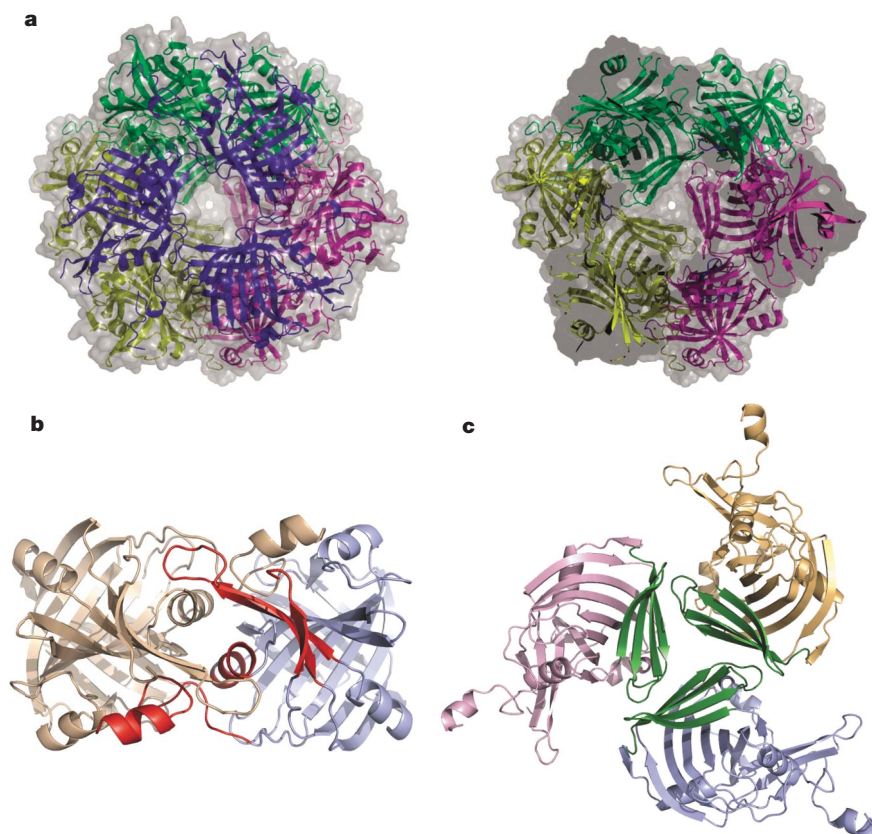
**Figure 3 | The proposed mechanism of AADase.** The two alternative conformations of the side chain of Glu 76 are shown in blue and red. A, acid.

the catalytic activity of the Glu61Gln mutant, which shows a decrease in  $k_{\text{cat}}$  ( $\sim 20$ -fold with no change in  $K_m$ ) similar in magnitude to that caused by the Glu76Gln mutation. The slight downward shift of the pH optimum of the Glu61Gln mutant (Supplementary Fig. 3) is not in the direction expected if this residue had an effect on the  $pK_a$  of Lys 115.

#### Oligomeric structure

The AADase homododecamer known to exist in solution<sup>19</sup> was generated by crystallographic symmetry from the protomers in the

asymmetric unit; the resultant assembly did not differ between the structures from the two clostridial species. The 12 monomers form a tetrahedron, with axial lengths of 78 Å and longest dimension of 118 Å, enclosing a central tetrahedral cavity with axial lengths of 31 Å (Fig. 4a). The crystallographic dodecamer is consistent with dimensions determined for CaAAD by means of small angle X-ray scattering of  $117 \pm 3$  Å in maximal diameter (Supplementary Figs 5 and 6). Additionally, two-dimensional cryo-electron microscopy projections of single CaAAD particles yielded a three-dimensional



**Figure 4 | The biological dodecamer of AADase built from the crystallographic asymmetric unit.** **a**, Shown is the view parallel to the three-fold of the tetrahedral macromolecular assembly, with each trimer coloured differently (left) and cutaway to show the hollow interior (right). **b**, View of the dimer, with each subunit coloured separately and the helix and

three-stranded sheet comprising the interface from a single subunit depicted in red. **c**, The structure is shown as a trimer, with each subunit coloured separately and the four-stranded sheet comprising the interface depicted in green (also shown in green in Fig. 1b).

volume with inner cavity dimensions ( $\sim 30 \text{ \AA}$ ) that did not differ from those determined from the crystallographic dodecamer (Supplementary Fig. 7). Thus, the crystallographic model appears to be an accurate representation of the state of the enzyme in solution. Macromolecular assemblies with tetrahedral geometry are extremely rare. The basic symmetry of a tetrahedron is that of four trimers, with each three-fold symmetry axis at a vertex, or that of six dimers with each two-fold at the centre of the axes connecting the vertices. The dimer interface comprising one three-stranded  $\beta$ -sheet and the amino-terminal  $\alpha$ -helix of each monomer lies at the bottom of the  $\beta$ -cone and interdigitates with the identical interface of the neighbouring subunit (Fig. 4b). The trimer interface is formed from the four-stranded  $\beta$ -sheets, which form a platform with intersubunit interactions at the three-fold symmetry axis (Fig. 4c) by means of hydrophobic (CaAAD) or ionic (CvAAD) interactions.

The dimer interface forms extensive hydrophobic, hydrogen-bonding and ionic interactions involving more than 10% of the amino-acid residues of AADase (Supplementary Table 4 and Supplementary Fig. 8). It is these interactions that establish the dimeric form of CaAAD even in the presence of 4 M urea<sup>19</sup>. The conversion of acetoacetate to acetone is a key process in the anaerobic metabolism of carbohydrate in solventogenic bacteria such as *C. acetobutylicum*. The oligomeric state of AADase may be used to stabilize the enzyme at the very low pH and high solvent environment found in the cell during solventogenesis, or to colocalize the enzymes in the solventogenic pathway to facilitate the diffusion of product to the next enzyme in the pathway. The high  $K_m$  values for acetyl CoA acetyl transferase and AADase in the solventogenesis pathway (0.27 and 8 mM, respectively)<sup>10,20</sup> support this concept.

## METHODS SUMMARY

The structure of AADase was determined by X-ray crystallography using phases determined experimentally by means of single wavelength anomalous diffraction on the selenomethionine-substituted protein. The dodecameric state of AADase greatly increased the complexity of *de novo* structure determination of the enzyme from *C. acetobutylicum* owing to the presence of near perfect merohedral twinning in many crystals and the presence of two dodecamers in the asymmetric unit (216 Se sites) in other crystal forms. The final successful structure analysis depended on determination of the structure of AADase from another bacterial source, CvAAD (Supplementary Table 2), which crystallizes with a smaller number of protomers in the asymmetric unit, and the use of the resulting model to determine the phases by means of molecular replacement of the native CaAAD and liganded CvAAD structures (see Supplementary Table 1 and Methods). Steady-state-kinetic analyses were performed on wild-type and mutant enzymes using a simple spectrophotometric assay. The dimensions of the oligomer of CaAAD in solution were determined by means of small angle X-ray scattering and further verified using three-dimensional particle reconstruction of cryo-electron microscopy images.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 19 September 2008; accepted 25 February 2009.**

- Westheimer, F. H. Coincidences, decarboxylation and electrostatic effects. *Tetrahedron* **51**, 3–20 (1995).
- Hamilton, G. A. & Westheimer, F. H. On the mechanism of the enzymatic decarboxylation of acetoacetate. *J. Am. Chem. Soc.* **81**, 6332–6333 (1959).
- Fridovich, I. & Westheimer, F. H. On the mechanism of the enzymatic decarboxylation of acetoacetate. II. *J. Am. Chem. Soc.* **84**, 3208–3209 (1962).
- Laursen, R. A. & Westheimer, F. H. The active site of acetoacetate decarboxylase. *J. Am. Chem. Soc.* **88**, 3426–3430 (1966).
- Warren, S., Zerner, B. & Westheimer, F. H. Acetoacetate decarboxylase. Identification of lysine at the active site. *Biochemistry* **5**, 817–823 (1966).
- Kokesh, F. C. & Westheimer, F. H. A reporter group at the active site of acetoacetate decarboxylase. II. Ionization constant of the amino group. *J. Am. Chem. Soc.* **93**, 7270–7274 (1971).
- Harris, T. K. & Turner, G. J. Structural basis of perturbed  $pK_a$  values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98 (2002).
- Gerlt, J. A. Obituary, Frank H. Westheimer. *Nature* **447**, 543 (2007).
- Gallivan, J. P. & Dougherty, D. A. Cation– $\pi$  interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464 (1999).
- Highbarger, L. A., Gerlt, J. A. & Kenyon, G. L. Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the  $pK_a$  of active-site lysine 115. *Biochemistry* **35**, 41–46 (1996).
- García-Moreno, E. B. et al. Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **64**, 211–224 (1997).
- Lee, J. K. & Houk, K. N. A proficient enzyme revisited: the predicted decarboxylase mechanism for orotidine monophosphate. *Science* **276**, 942–945 (1997).
- Rashin, V. & Honig, B. Reevaluation of the Born model of ion hydration. *J. Phys. Chem.* **89**, 5588–5593 (1985).
- Stites, W. E., Gittis, A. G., Lattman, E. E. & Shortle, D. In a staphylococcal nuclease mutant the side-chain of a lysine replacing valine66 is fully buried in the hydrophobic core. *J. Mol. Biol.* **221**, 7–14 (1991).
- Barbas, C. F. III et al. Immune versus natural selection: antibody aldolases with enzymic rates but broader scope. *Science* **278**, 2085–2092 (1997).
- Crosby, J., Stone, R. & Liehard, G. E. Mechanisms of thiamine-catalyzed reactions. Decarboxylation of 2-(1-carboxy-1-hydroxyethyl)-3,4-dimethylthiazolium chloride. *J. Am. Chem. Soc.* **92**, 2891–2900 (1970).
- Fridovich, I. A study of the interaction of acetoacetate decarboxylase with several inhibitors. *J. Biol. Chem.* **243**, 1043–1051 (1968).
- O'Leary, M. H. in *Mechanisms of Catalysis* (ed. Sigman, D. S.) 239 (Academic, 1992).
- Tagaki, W. & Westheimer, F. H. Acetoacetate decarboxylase. Reassociation of subunits. *Biochemistry* **7**, 891–894 (1968).
- Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. *Appl. Environ. Microbiol.* **54**, 2717–2722 (1988).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Murzin for provisional classification of the AADase fold and C. Akey for valuable advice on the execution and interpretation of the electron microscopy work. We also thank H. Robinson and A. Soares for help with data collection. This work was supported by a grant to K.N.A. from the National Science Foundation. Data for this study were measured at Beamlines X12B, X25 and X29A of the National Synchrotron Light Source. Financial support comes principally from the Offices of Biological and Environmental Research (BER) and of Basic Energy Sciences (BES) of the US Department of Energy (DOE), and from the National Center for Research Resources (NCRR) of the National Institutes of Health (NIH). Small-angle X-ray scattering analyses were carried out at the Stanford Synchrotron Radiation Lightsource, funded by DOE, BES. The SSRL Structural Molecular Biology Program is supported by DOE, BER, and by NIH, NCRR. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH.

**Author Contributions** M.-C.H. cloned, expressed, purified, crystallized, collected data and performed crystal structure determination, refinement and model analysis. H.T. designed, executed, analysed and wrote the description of the small-angle X-ray scattering analysis. J.F.M. designed, executed, analysed and wrote the description of the electron microscopy experiments. K.N.A. conceived of and designed the project. M.-C.H. and K.N.A. wrote the manuscript; all authors discussed the results and commented on the manuscript.

**Author Information** The coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 3BH2, 3BGT and 3BH3 depending, respectively, to *C. acetobutylicum* acetoacetate decarboxylase and *C. violaceum* acetoacetate decarboxylase in the unliganded form and complexed with 2,4-pentanedione. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.N.A. ([drkallen@bu.edu](mailto:drkallen@bu.edu)).

## METHODS

**Materials.** Except where indicated, all chemicals were purchased from Sigma-Aldrich or American Bioanalytical. DNase I was from Worthington Biochemical Corp. All enzymes and primers for PCR, T4 DNA ligase and restriction enzymes were from Invitrogen. The DNA cleanup kit and miniprep kit were from Qiagen. Host cells and pET vectors were purchased from Novagen.

**Cloning, expression and purification.** The plasmid encoding the AADase gene of *C. acetobutylicum* (CaAAD) was purified from host cells<sup>10</sup> provided by J. Gerlt. Primers (5'-GCGCCATGGTAAAGGATGAAGTAAT-3' and 5'-CGCGGATCCTTACTTAAGATAATCATATAT-3') containing NCOI and BamHI endonuclease cleavage sites (underlined) were used to amplify the gene by PCR. The genome of *C. violaceum* was purchased from American Type Culture Collection (ATCC 12472D) and treated with the endonuclease BamHI before PCR. The AADase from *C. violaceum* (CvAAD) gene was amplified using the primers (5'-TGAAGGGAGTTACATATGAAGCAACAGGAAGTCCAG-3' and 5'-GGGCGCGTTTCTCGGATCGC-GCCGTTATTGCCGAGG-3') containing NdeI and BamHI restriction sites (underlined). The CaAAD and CvAAD were cloned into pET3d and pET15b vectors, respectively. The ligation products were transformed into *Escherichia coli* DH5 $\alpha$  and sequenced by the Tufts University Core Facility.

CaAAD was expressed and purified from *E. coli* BL21 (DE3) cells using a modification of the procedure described previously<sup>10</sup>. The transformed cells were grown in Luria Bertani medium at 37 °C and induced by 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside for 20 h at room temperature (25 °C). The cells were collected by centrifugation. The cell pellet was suspended in 40 ml ice-cold lysis buffer (50 mM phosphate buffer, pH 7.4, 2 mM DTT, 1 mM PMSF and 1 mg DNase I) and lysed by sonication (5  $\times$  6 min, 80% power and 60% duty cycle, Branson Sonifier 250). The cell lysate were centrifuged at 4 °C for 1 h at 147,000g. The protein was precipitated by dropping the pH to  $\sim$ 3.8–3.9, and the precipitated material was dissolved in 150 ml 200 mM phosphate buffer, pH 5.95, at 4 °C overnight (12 h). The remaining impurities were removed by precipitating with 55% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> followed by centrifugation at 3,700g for 30 min. CaAAD was precipitated with 75% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, dissolved in 3 ml of 50 mM phosphate buffer, pH 5.95, and loaded directly onto a S-200 gel filtration column (GE Healthcare) pre-equilibrated with buffer A (50 mM phosphate buffer, pH 5.95). The fractions containing CaAAD, detected using a kinetic assay (see below) and 12% SDS-PAGE, were collected. If the CaAAD fractions from S-200 gel filtration were >90% pure, the next purification step, using DEAE-Sepharose, was omitted. If not, the collected fractions were loaded onto a DEAE-Sepharose column (GE Healthcare) equilibrated with buffer A. The column was washed with buffer A and eluted with a 0.5 l linear gradient of 0–0.5 M NaCl in buffer A. All fractions containing CaAAD, detected by activity assays and 12% SDS-PAGE, were collected and dialysed against 5 mM phosphate buffer, pH 5.95, and 2 mM DTT at 4 °C overnight, followed by concentration to  $\sim$ 10 mg ml<sup>-1</sup> using an Amicon ultracentrifugal filter device (Millipore).

The expression and purification of CvAAD was similar to CaAAD with some modification. Forty minutes before induction, 1  $\times$  Augmedium (AthenaES) was added to reduce the percentage of protein expressed as insoluble inclusion bodies. The cells were collected by centrifugation and lysed by sonication in 50 ml ice-cold lysis buffer (50 mM HEPES, pH 7.4, 30 mM NaCl, 1 mM PMSF and 1 mg DNase I). The cell lysate was centrifuged at 4 °C for 1 h at 147,000g and the supernatant loaded onto a column packed with 15 ml TALON cobalt resin (Clontech). The column was washed with 50 ml wash buffer (50 mM HEPES, pH 7.4, 30 mM NaCl and 2 mM  $\beta$ -mercaptoethanol) and 50 ml wash buffer plus 30 mM imidazole, pH 7.6, and eluted with 50 ml elution buffer (50 mM HEPES pH 7.4, 30 mM NaCl, 150 mM imidazole pH 7.6 and 2 mM  $\beta$ -mercaptoethanol). The protein was dialysed against 5 mM imidazole buffer, pH 7.6, and 2 mM DTT at 4 °C overnight and concentrated to  $\sim$ 20 mg ml<sup>-1</sup> using an Amicon Ultra centrifugal filter device.

Selenomethionine-labelled CvAAD (SeMet CvAAD) was expressed in *E. coli* (B834) cells. The cells were grown in SelenoMet medium base (AthenaES). A custom Augmedium formulation (0.5 g l<sup>-1</sup> methionine was replaced by 0.15 g l<sup>-1</sup> selenomethionine) was added 2 h before induction. The cells were induced with 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside at 18 °C for 48 h and collected by centrifugation. The purification of selenomethionine-labelled CvAAD was the same as that of native CvAAD.

Mutagenesis of CvAAD was accomplished by means of quick-change site-directed mutagenesis (Invitrogen) using CvAAD/pET3A as the template. The R29Q mutant was created using the primers (5'-GCCTTATCGTTTCGTCACCAAGGAATACATGATCATCACCTA-3' and 5'-TAGGTGATGATCATGTTCTCCTGGTTGACGAAACGATAAGGC-3'). The E61Q mutant was created using the primers (5'-GGATTTGGCGATTACTCGCAAAGCGGGCAGG-3' and 5'-CCTAAACCGCTAATGAGCGTTTCGCCCGTCC-3'). The E76Q mutant

was created using the primers (5'-GCCTTATCGTTTCGTCACCAAGGAATACATGATCATCACCTA-3', and 5'-TAGGTGATGATCATGTTCTCCTGGTTGACGAAACGATAAGGC-3'). The mutated sites are underlined. The resulting PCR products were digested with NdeI and BamHI endonuclease, and cloned into pET15b vectors that were cut by the same enzymes. The ligation products were transformed into *E. coli* DH5 $\alpha$ , followed by DNA sequencing by Agencourt Bioscience. The plasmid, encoding His<sub>6</sub>-tagged CvAAD point mutants, were purified from *E. coli* DH5 $\alpha$  and transformed into *E. coli* BL21(DE3) for protein expression. The expression and purification of CvAAD mutants was identical to that of wild-type CvAAD.

**Crystallization and data collection.** The initial crystallization conditions of CaAAD and CvAAD were identified by high-throughput screening (Hauptman Woodward Medical Research Institute) and refined by the hanging-drop vapour-diffusion method. The final crystallization condition of CaAAD was 18–20% glycerol, 40 mM phosphate buffer, pH 5.95, 100 mM sarcosine and 14–15% PEG 3350 (Hampton Research). Drops were formed by mixing 1–2  $\mu$ l of protein solution (10 mg ml<sup>-1</sup>) with 1  $\mu$ l of well solution. Crystals of CaAAD were flash-frozen in a gaseous N<sub>2</sub> stream at 100 K. X-ray diffraction data were collected at the National Synchrotron Light Source (Brookhaven National Laboratory, New York), Beamline X25. CaAAD crystallized in space group R32 with unit-cell dimensions  $a = b = 104.07$  Å,  $c = 578.09$  Å and diffracted to 2.4 Å resolution.

Crystals of CvAAD and SeMet CvAAD were grown by hanging-drop vapour diffusion over a well solution containing 0.4–0.5 M K<sub>2</sub>HPO<sub>4</sub>/NaH<sub>2</sub>PO<sub>4</sub> at pH 7.5–7.9. Drops consisted of 1–2  $\mu$ l of protein solution (8–12 mg ml<sup>-1</sup>) and 1  $\mu$ l of well solution. Crystals were transferred to well solution with 30% glucose as a cryoprotectant and crystals were frozen in liquid N<sub>2</sub>. The automounter at BNL, Beamline X12B, was used to screen crystals for diffraction and data were collected at Beamline X29A. SeMet CvAAD crystallized in space group R3 ( $a = b = 105.45$  Å,  $c = 252.38$  Å). A 2.1 Å multiple wavelength anomalous diffraction (MAD) data set was collected (0.9791 Å, 0.9793 Å and 0.9754 Å), but only the data collected at 0.9791 Å were used to solve the phases. For the 2,4-pentanedione complex structure with CvAAD, the enzyme (10 mg ml<sup>-1</sup>) was co-crystallized with 2 mM potassium acetyl acetonate under the same condition as unliganded CvAAD. The crystal was additionally soaked in mother liquid plus 120 mM potassium acetyl acetonate for 1 h at room temperature and transferred to well solution with 30% glucose plus 4 mM potassium acetyl acetonate as cryoprotectant and frozen in liquid N<sub>2</sub>. The inhibitor-soaked crystals diffracted to 2.1 Å and were isomorphous to the SeMet CvAAD. A 2.1 Å data set was collected at 0.900 Å at Brookhaven National Laboratories, beamline X12B. All data sets were indexed and scaled using HKL2000 (ref. 21). Data collection statistics are summarized in Supplementary Table 1.

**Structure determination and model refinement.** The structure of CvAAD was phased by the single wavelength anomalous diffraction method, using the 2.1 Å SeMet CvAAD data set collected at the peak wavelength 0.9791 Å. The Se substructure was solved using HKL2MAP/SHELXD<sup>22,23</sup>, which identified 24 of 32 selenomethionine sites. The 24 sites were used to phase the protein structure in SOLVE<sup>24</sup> and generated a high-quality electron density map. The map was improved using RESOLVE<sup>25</sup>, by means of NCS averaging (4 monomers per asymmetric unit). An initial model including 44% of the structure was built automatically by RESOLVE. The experimentally phased NCS averaged electron density map permitted further manual rebuilding of one of the four monomers using COOT<sup>26</sup>. This monomer was then used to perform molecular replacement using MOLREP<sup>27</sup> to find the remaining three monomers in the asymmetric unit. Iterative rounds of model building and refinement were performed using COOT, CNS and REFMAC5<sup>26–28</sup>.

The structure of CaAAD was phased by molecular replacement with the program PHASER using SeMet CvAAD as the search model<sup>29</sup>. The process of refinement and rebuilding was the same as that used for SeMet CvAAD. The structure of CaAAD and 2,4-pentanedione liganded CvAAD was phased by molecular replacement with the program PHASER<sup>29</sup> using the structure of SeMet CvAAD as the search model. The process of refinement and rebuilding was the same as that used for SeMet CvAAD. The model of the inhibitor was not added to the refinement until the value of  $R_{work}$  dropped below 28% to avoid phase bias. The topology restraints for 2,4-pentanedione adduct of Lys 115 for use in REFMAC5 were generated using SKETCHER in the CCP4 program suite<sup>30</sup>. The Ramachandran plot, as assessed by Procheck<sup>31</sup>, had 99.5% of the residues of unliganded CaAAD and CvAAD and 99.2% of the residues of liganded CvAAD in the allowed regions with 1 residue (Ser 126 in CaAAD/Gln 127 in CvAAD) in the disallowed regions for all three structures.

**Small-angle X-ray scattering measurements.** Solution X-ray scattering measurements were performed at the Stanford Synchrotron Radiation Lightsource Beamline 4-2 (ref. 32). Each protein sample solution containing 50 mM phosphate (pH 5.95) and 2 mM DTT was held in a sample cuvette, maintained at 20 °C and located at 2.5 m from a MarCCD165 detector

(MarUSA). The detector pixel numbers were converted to the momentum transfer  $Q = 4\pi\sin(\theta)/\lambda$ , where  $\theta$  is one half of the scattering angle and  $\lambda$  the X-ray wavelength 1.381 Å, using the (100) reflection and related reflections recorded from a cholesterol myristate powder sample. Twenty-four 10-s exposures were acquired in series at 2, 5, 7.5, 10 and 15 mg ml<sup>-1</sup> (Supplementary Fig. 5). No time-dependent radiation-induced protein aggregation was observed and all 24 images were averaged after intensity scaling, with the exception of 10 and 15 mg ml<sup>-1</sup> samples, for which the last few exposures were excluded in the averaging process due to the onset of radiation-induced aggregation. The radial integration, intensity-scaling, statistical analysis, frame-averaging and background subtraction were done by MarParse<sup>32</sup>. Radii of gyration ( $R_g$ ) and forward scattered intensities ( $I_{(Q=0)}$ ) were obtained by Primus<sup>33</sup> in the  $Q$  range 0.0117–0.0297 Å<sup>-1</sup> (Supplementary Fig. 5, inset). The concentration-scaled forward scattered intensity remained within 3% of the mean value between 2 and 10 mg ml<sup>-1</sup>, indicating no evidence of concentration-dependent aggregation. Slightly lower  $R_g$  values at higher protein concentrations indicated mildly repulsive inter-particle interaction. The composite scattering curve with satisfactorily high statistics covering  $Q$  values 0.0117 to 0.25 Å<sup>-1</sup> was thus obtained by scaling and merging the scattering curve recorded at 2 mg ml<sup>-1</sup> (0.0117 <  $Q$  < 0.0611 Å<sup>-1</sup>) with the intermediary angle data recorded at 15 mg ml<sup>-1</sup> (0.0448 <  $Q$  < 0.25 Å<sup>-1</sup>) by Primus<sup>33</sup> to minimize weak effects of the particle structure factor due to the repulsive interaction (Supplementary Fig. 6). The overlapping  $Q$  range contained 30 common data points to assure accuracy of the scaling. The X-ray scattering profile and the electron pair distance distribution function  $P(r)$  were computed for the crystallographic structure model using ORNL\_SAS<sup>34</sup>. The electron density and thickness of the hydration layer were altered in the range 3–15% higher electron density of water and 2.5–3.5 Å. The best fit was obtained with 8% excess density and 2.5 Å layer thickness, although similarly satisfactory fits could be obtained with lower hydration layer density and a slightly thicker hydration layer.  $Q$  values in the experimental data were also adjusted by 1–3% to account for the small unit cell parameter variations observed in the crystallographic study. The lowest chi square value we have obtained with 1%  $Q$  value adjustment was 2.445, which confirms the presence of the crystallographic dodecamer in solution but reflects the clearly recognizable deviation in the  $Q$  range 0.06–0.10 Å<sup>-1</sup>. This deviation may suggest minor structural differences between solution and crystal structures at the tertiary and/or quaternary structural level because experimental conditions can not be identical. Similar differences between experimental and computed X-ray scattering curves have been observed for other oligomeric enzyme systems<sup>35</sup>. The indirect Fourier transform analysis of the composite scattering curve by GNOM<sup>36</sup> yielded the electron pair distance distribution function, giving the maximum particle dimension of 117 (± 3 (± s.d.)) Å and  $R_g$  of 42.9 Å, both of which are highly consistent with the corresponding values of 118 and 43.3 Å obtained for the crystal structure with the hydration layer (Supplementary Fig. 6). The nearly perfect match of the experimental and computed  $P(r)$  indicates that the solution and crystal structures are essentially identical at the quaternary structure level (Supplementary Fig. 6, inset).

**Cryo-electron microscopy and image processing.** The cryo-electron microscopy and image processing of CaAAD were performed at Boston University School of Medicine. CaAAD (10 mg ml<sup>-1</sup>, 1 mM phosphate buffer, pH 5.95) was loaded onto perforated carbon film on 400 mesh copper grids. The specimens were plunge-frozen in liquid ethane cooled by liquid nitrogen<sup>37</sup> and data collected on a Tecnai TF20 with Kodak SO163 film at Boston University School of Medicine. Electron micrographs were recorded at 200 kV at ×62,000 with a defocus range of -1.0 to -2.5 μm. Negatives were digitized on a CreosCitex EVERSMART scanner using a 4.54 μm raster and binned 4 × 4 to 2.93 Å per pixel. Particles were picked with boxer in EMAN<sup>38</sup> and classified using refine2d in EMAN to avoid bias from a three-dimensional model<sup>39</sup>.

Additionally, the three-dimensional structure was refined in EMAN using ~8,700 particles and by imposing the tetrahedral symmetry found in the crystal

structure<sup>39</sup>. The resolution was estimated by splitting the data set into two halves (according to even and odd numbered particles), generating an even and an odd volume and then calculating their cross-correlation coefficient. Plotting the correlation coefficient versus resolution gave the Fourier shell correlation (FSC) curve. Using the FSC 0.5 criteria, it was concluded that the resolution of the three-dimensional volume was 12 Å.

**Steady-state kinetics.** The decarboxylation of acetoacetate by CvAAD and CaAAD was assayed by monitoring the disappearance of the enolate form of acetoacetate spectrophotometrically at 270 nm ( $\Delta\epsilon = 26.7 \text{ M}^{-1} \text{ cm}^{-1}$ ) at 25 °C in 50 mM phosphate buffer, pH 5.95, on a Beckman Coulter DU-800 spectrophotometer<sup>10</sup>. The same assay was used for CvAAD mutants in 50 mM phosphate/citrate buffer at various pH values on a Varian Carey-300 spectrophotometer. Data were fitted to the Michaelis–Menten equation  $V = V_{\text{max}}[S]/(K_m + [S])$ , where  $V$  is the initial rate at substrate concentration  $S$ , and  $K_m$  is the Michaelis constant.  $k_{\text{cat}}$  is obtained from  $V_{\text{max}}$  and total enzyme concentration in the reaction. The  $k_{\text{app}}$  for wild-type and mutant CvAAD proteins versus pH was obtained from the initial rate and enzyme concentration at the saturating substrate concentration of 20 mM, with the exception of the CvAAD R29Q mutant for which activity could only be measured using 40 mM substrate.

- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
- Usón, I. & Sheldrick, G. M. Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643–648 (1999).
- Terwilliger T.C. & Berendzen J.. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Terwilliger, T. C. Maximum likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphic. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Collaborative computation project number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brünger, A. T. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation function. *Acta Crystallogr. D* **60**, 432–438 (2004).
- Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59**, 1131–1137 (2003).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Smolsky, I. L. et al. Biological small-angle x-ray scattering facility at the Stanford Synchrotron Radiation Laboratory. *J. Appl. Crystallogr.* **40**, S453 (2007).
- Konarev, P. V. et al. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
- Tjioe, E. & Heller, W. T. ORNL\_SAS: software for calculation of small-angle scattering intensities of proteins and protein complexes. *J. Appl. Crystallogr.* **40**, 782–785 (2007).
- Svergun, D. I. et al. Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state. *Proteins* **27**, 110–117 (1997).
- Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).
- Ménétre, J. F. et al. The structure of ribosome-channel complexes engaged in protein translocation. *Mol. Cell* **6**, 1219–1232 (2000).
- Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
- Ludtke, S. J., Chen, D. H., Song, J. L., Chuang, D. T. & Chiu, W. Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. *Structure* **12**, 1129–1136 (2004).