Computational Determination of Fundamental Pathway and Activation Barriers for Acetohydroxyacid Synthase-Catalyzed Condensation Reactions of α-Keto Acids

YING XIONG,1,2* JUNJUN LIU,1,2* GUANG-FU YANG,1 CHANG-GUO ZHAN2
1Key Laboratory of Pesticide and Chemical Biology of the Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, People’s Republic of China
2Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 725 Rose Street, Lexington, Kentucky 40536

Received 11 January 2009; Revised 5 March 2009; Accepted 18 May 2009
DOI 10.1002/jcc.21356
Published online 24 June 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: Acetohydroxyacid synthase (AHAS) is the first common enzyme in the biosynthetic pathway leading to the production of various branched-chain amino acids. AHAS is recognized as a promising target for new antituberculosis drugs, antibacterial drugs, and herbicides. Extensive first-principles quantum mechanical (QM) and hybrid quantum mechanical/molecular mechanical (QM/MM) calculations have enabled us, in this study, to uncover the fundamental reaction pathway, determine the activation barriers, and obtain valuable insights concerning the specific roles of key amino acid residues for the common steps of AHAS-catalyzed condensation reactions of α-keto acids.

The computational results reveal that the rate-determining step of the AHAS-catalyzed reactions is the second reaction step and that the most important amino acid residues involved in the catalysis include Glu144, Gln207, Gly121, and Gly511 that form favorable hydrogen bonds with the reaction center (consisting of atoms from the substrate and cofactor) during the reaction process. In addition, Glu144 also accepts a proton from cofactor thiamin diphosphate (ThDP) through hydrogen bonding during the catalytic reaction. The favorable interactions between the reaction center and protein environment remarkably stabilize the transition state and, thus, lower the activation barrier for the rate-determining reaction step by ~20 kcal/mol. The activation barrier calculated for the rate-determining step is in good agreement with the experimental activation barrier. The detailed structural and mechanistic insights should be valuable for rational design of novel, potent AHAS inhibitors that may be used as promising new antituberculosis drugs, antibacterial drugs, and/or herbicides to overcome drug resistance problem.


Key words: enzyme; catalytic mechanism; reaction pathway; QM/MM calculation

Introduction

Acetohydroxyacid synthase (AHAS; also known as acetylactate synthase; EC 2.2.1.6; formerly EC 4.1.3.18) is the first common enzyme in the biosynthetic pathway leading to the production of branched-chain amino acids, including valine, leucine, and isoleucine, in plants and a wide range of microorganisms (such as Escherichia coli, Saccharomyces cerevisiae, and Arabidopsis thaliana). This essential enzyme catalyzes the condensation of two α-keto acid molecules, e.g., the condensation of two pyruvate molecules to form (S)-α-acetolactate and the condensation of one pyruvate molecule with one α-ketobutyrate molecule to form (S)-α-aceto-α-hydroxybutyrate. In addition, AHAS has been proven to be an efficient catalyst in chiral synthesis and a potent target in biochemistry engineering. AHAS is recognized as a promising target for new antituberculosis drugs, antibacterial drugs, and herbicides. For example, Mycobacterium tuberculosis is the pathogen of tuberculosis that remains a major threat to the human population. This human disease is responsible for 2–3 millions deaths per year worldwide. The emergence of drug-resistant and multidrug-resistant strains of tuberculosis is a significant public health threat, particularly in low-income countries. Therefore, the development of new, potent AHAS inhibitors that can overcome drug resistance is crucial.

*These two authors contributed equally to this work.

Correspondence to: C.-G. Zhan; e-mail: zhan@uky.edu; G.-F. Yang; e-mail: gfyang@mail.ccnu.edu.cn
Contract/grant sponsor: Kentucky Science and Engineering Foundation; contract/grant number: KSEF-925-RDE-008
Contract/grant sponsor: National Natural Science Foundation of China; contract/grant numbers: 20602014, 20503008, 20432010

© 2009 Wiley Periodicals, Inc.
resistant tuberculosis has greatly increased the need to identify new antituberculosis target proteins. It was shown that branched amino acids auxotrophic strain of Mycobacterium failed to proliferate because of the inability to use amino acids from their hosts, indicating that inhibitors for the branched-chain amino acid biosynthesis could be used as an antituberculosis agent. These observations suggested that AHAS could be a potential target of new antituberculosis drugs. It was found that the recombinant AHAS from M. avium, which is another human tuberculosis pathogen, was efficiently inhibited by sulfonylureas (SU), a herbicide that inhibits plant AHAS.

A number of structurally diverse, potent AHAS inhibitors, have been reported in literature. Some of the known AHAS inhibitors, e.g. SU, imidazolinones (IM), triazolopyrimidine sulfonanilides (TP), and pyrimidinylxybenzoates (PB), have been used commercially as herbicides. Since their first introduction in the early 1980s, AHAS inhibitors have been used worldwide and now constitute the most important herbicide family in use due to their unique advantages such as broad-spectrum weed control at very low application rates, extremely low toxicity to animals, and flexible application timing in a wide variety of crops. However, constantly extensive use of AHAS-inhibiting herbicides has resulted in about 95 tolerant weeds worldwide, and this number has shown to be increasing at an exponential rate. Most seriously, because the structurally diverse AHAS inhibitors are competitive with one another with respect to the binding with the enzyme, tolerance toward a particular herbicide has resulted in varying degrees of cross-tolerance to the other chemicals. Therefore, because of the widespread occurrence of AHAS inhibitor-resistant weed species, the AHAS enzyme has received considerable attention.

However, the currently known AHAS inhibitors, including all the existing AHAS-inhibiting herbicides, have been proved to be "extraneous site inhibitors." That is to say, the binding site of the known AHAS inhibitors is not the active site of AHAS. These compounds just bind within the tunnel connecting to the active site, thus blocking substrate entry into the active site. As a result, the mutations on amino acid residues in the inhibitor-binding site resulted in resistance to the drugs/herbicides while the enzyme activities were kept. An ideal strategy to develop novel AHAS inhibitors for overcoming the drug resistance is to design novel inhibitors that bind to the catalytic active site of the enzyme, as amino acid residues in the active site are rather conservative. With the inhibitor binding to the active site, one would not need to worry about the resistance due to mutations on the active site residues, as the mutations on the active site residues would likely make the enzyme inactive. It is important for rational design of the novel inhibitors of AHAS to understand the detailed structure and fundamental catalytic mechanism of the enzyme.

Recently, six X-ray crystal structures of A. thaliana AHAS in complex with five different SU and one imidazolinone herbicide have been reported. Like almost all thiamin diphosphate (ThDP)-dependent enzymes, the active site of AHAS is formed at the interface between two monomers and, therefore, the minimum quaternary structure for catalysis is a dimer. The enzyme-sequestered ThDP (cofactor) plays a central role in the catalytic mechanism as shown in Figure 1. Initially, ThDP is in its protonated form (R or ES) that then ionizes to the reactive ylide (INT1); A molecule of pyruvate is attacked by the ThDP anion to form the lactyl-ThDP (LThDP) intermediate (INT2) which undergoes decarboxylation. In subsequent steps, the hydroxyethylthiamin diphosphate (HETHDP) enamine/carbonium (INT3) formed from the decarboxylation attacks the carbonyl of a second substrate (pyruvate or 2-ketobutyrate) to form the product-ThDP adduct. Finally, the product-ThDP adduct dissociates to form ThDP and the free product. Direct competition between alternative substrates for the bound HETHDP determines the ratio of products formed. In the wild-type enzyme, the product ratio depends only on the relative amounts of the substrates present and $k_{cat}$ is nearly independent of the amount or identity of the substrates. This implies that $k_{cat}$ is determined...
by the reaction step preceding the product-determining, carbobiligation step so that the steps contributing to discrimination among “second substrates” have little effect on the overall rate of reaction. The ThDP adducts of LThDP, HETHDP, and the product acetohydroxyacid (AlTHDP and/or AlH ThDP) can be detected quantitatively in the mixture after rapid quenching. Determination of the distribution of the key reaction intermediates by $^1$H NMR analysis then makes it possible to calculate the forward unimolecular rate constants for individual microsomal steps. The forward (net) rate constant for formation of LThDP, the first detectable step, is overwhelmingly rate determining for catalysis, whereas the subsequent steps occur with similar, but much larger rates (e.g., the rates of decarboxylation, carbobiligation, and product release are comparable). Such relative rates explain why $k_{cat}$ for a given enzyme is essentially the same for acetohydroxybutyrate (AHB) and acetolactate (AL) formation, despite the 60-fold preference for 2-ketobutyrate as substrate.

The present computational study was aimed to understand the fundamental reaction pathway and energetics for the common reaction steps concerning the formation of HETHDP enamine/carbanion intermediate before the further reaction with a second substrate, particularly for the rate-determining reaction step. The purpose of this computational study is twofold. One is to provide fundamental mechanistic insights for a generally interesting type of chemical and biochemical reactions, because the condensation reactions catalyzed by an enzyme leading to the eventual formation of branched-chain amino acids, such as valine, leucine, and isoleucine, are a type of fundamentally interesting chemical/biochemical reactions in biological systems. In addition, the mechanistic insights obtained from the detailed computational study can provide a solid, valuable mechanistic base for rational design of the highly desirable novel AHAS inhibitors that can tightly bind with the catalytic residues in the active site. The novel AHAS inhibitors may be used as promising new antituberculosis drugs, antibacterial drugs, and/or herbicides to overcome the drug resistance. For these purposes, we have carried out detailed reaction coordinate calculations, for the first time, on the common steps of AHAS-catalyzed condensation reactions by using first-principles quantum mechanical (QM) and hybrid quantum mechanical/molecular mechanical (QM/MM) methods. The reaction coordinate calculations have led to valuable mechanistic insights concerning the detailed reaction pathway and the specific role of some key residues, particularly for the rate-determining step.

**Computational Methods**

**QM Reaction Coordinate Calculations**

Geometries of all molecular structures, including reactants, transition states, and intermediates, involved in this study were optimized by using density functional theory (DFT) using Becke’s three-parameter hybrid exchange functional and the Lee-Yang-Parr correlation functional (B3LYP) with the 6-31+G(d) basis set. Vibrational frequency calculations were carried out to ensure that the optimized geometries are indeed associated with local minima or saddle points on the potential energy surfaces and to determine the zero-point vibrational energies (ZPVEs) and thermal corrections to Gibbs free energies at 310 K. Intrinsic reaction coordinate (IRC) calculations were performed at the B3LYP/6-31+G(d) level to confirm the optimized transition state geometries. The geometries optimized at the B3LYP/6-31+G(d) level were also used to perform single-point energy calculations with the 6-31+G(d), 6-31+G(d,p), and 6-311+G(d,p) basis sets. All of the QM calculations were performed by using Gaussian03 program.

**Preparation of Initial Structure for QM/MM Calculations**

The initial structure used in the QM/MM geometry optimizations was chosen from the X-ray crystal structure of A. thaliana AHAS in complex with a herbicide molecule (PDB code: 1ZSN). After the herbicide molecule was removed, pyruvate was docked into the active site of AHAS. The missing hydrogen atoms were added by using Leap module of Amber8 program. Histidine residues were assigned to (neutral) protonation state HID or HIE based on the local hydrogen bonding network. The partial atomic charges of substrate pyruvate and cofactors ThDP and flavin adenine dinucleotide (FAD) required for the structural preparation using Amber8 program were determined by performing ab initio QM calculations at the HF/6-31G* level following by restrained electrostatic potential (RESP)-fitting calculations using the RESP module of Amber8 program. All other force field parameters were from the parm99 parameter library.

After the energy minimization of the added atoms using Amber8 program in the gas phase, the AHAS-pyruvate complex was solvated in a rectangular box of TIP3P water molecules with a minimum solute-wall distance of 10 Å and neutralized by addition of counterions (Na+) using the Leap module of Amber8 program. Thus, a system of 95,600 atoms was obtained. To equilibrate the solvated system, first of all, the protein system was frozen and the solvent molecules with counterions were allowed to move during a 3000-step energy minimization process. Secondly, all the atoms were allowed to move during a 5000-step full minimization process. Then, the water molecules beyond 2.5 Å of the complex were removed and the remained system was used as the initial structure of QM/MM calculations described below.

**QM/MM Calculations**

QM/MM calculations were performed by using a pseudobond QM/MM method. The pseudobond QM/MM method was initially implemented in revised Gaussian03 and Tinker programs. The revised Gaussian03 and Tinker programs can be used to carry out the QM and MM parts of the QM/MM calculation iteratively until the full self-consistency is achieved. The pseudobond QM/MM method uses a seven-valence-electron atom with an effective core potential constructed to replace the boundary atom of the environment part and to form a pseudobond with the boundary atom of the active part. The main idea of the pseudobond approach is as follows: one considers that a large molecule is partitioned into two parts, an active part and an environment part, by cutting a covalent σ-bond Y-X. Y and X refer to boundary atoms of the environment part and the...
active part, respectively. Instead of using a hydrogen atom to cap the free valence of X atom as in the conventional link-atom approach, a pseudobond \( Y_{ps} - X \) is formed by replacing the Y atom with a one-free-valence boundary Y atom (\( Y_{ps} \)). The \( Y_{ps} \) atom is parameterized to make the \( Y_{ps} - X \) pseudobond mimic the original Y-X bond with similar bond length and strength, and also to have similar effects on the rest of the active part.\(^{34,35}\) In the pseudobond approach, the \( Y_{ps} \) atom and all atoms in the active part form a well-defined QM subsystem which can be treated by a QM method. Excluding Y atom, the remaining atoms in the environment part form the MM subsystem treated by a MM method. The pseudobond \textit{ab initio} QM/MM approach has been demonstrated to be powerful in studies of enzyme reactions.\(^{36–38}\)

In this study, we used a newly revised version\(^9\) of Gaussian03 and Amber8 programs, instead of the revised Gaussian03 and Tinker programs, to perform the QM/MM calculations, because the Amber program is capable of parallel computing. In all of our QM/MM calculations, the QM subsystem included all atoms of the substrate pyruvate, most atoms of cofactor ThDP, amino acid residues Gly511 and Ala512 (from one AHAS molecule of the homodimer), and residues Gly120', Gly121', Glu144', and Gln207' (from the other AHAS molecule of the homodimer). See below for a figure (Fig. 3) in which the QM atoms are highlighted. The remaining atoms of the protein (including Mg\(^{2+}\) and all atoms of nonreactive cofactor FAD that do not directly participate in the reaction) and water molecules were regarded as MM atoms. All water molecules were included in the MM subsystem, as none of the water molecules are expected to be involved in the covalent bond formation/breaking during the reaction based on the X-ray crystal structure and the structures modeled in the present study.

The boundary between the QM and MM subsystems was treated according to the original pseudobond approach.\(^{34}\) The total energy of the QM/MM system is

\[
E_{\text{Total}} = E_{\text{MM}} + E_{\text{QM}} + E_{\text{QM/MM}}
\]

The QM-MM interactions consist of bonding and nonbonding interactions. The nonbonding interactions between the two subsystems include the van der Waals (vdw) and electrostatic interactions calculated through the Lennard-Jones potential and Coulombic term, respectively, in the effective Hamiltonian. The energy corresponding to the effective Hamiltonian, which is obtained by QM calculation, is the sum of the QM energy of the QM subsystem (\( E_{\text{QM}} \)) and the electrostatic interaction between the QM and MM subsystems.\(^{34}\) We used the reaction coordinate driving method\(^{37,40}\) to search for transition states and intermediates. For example, during the QM/MM reaction coordinate calculations on the second reaction step (rate-determining step) and also on the first step, the reaction coordinate is represented by the C\(^2\)-C internuclear distance, where C\(^2\) refers to the carbon atom at position-2 of ThDP as shown in Figure 1 and C refers to the carbonyl carbon of pyruvate molecule. It should be noted that many other geometric parameters should also change while the C\(^2\)-C distance changes during the reaction. The changes of the other geometric parameters are automatic during the QM/MM reaction coordinate calculations. The same reaction coordinate approach has been used in many other computational studies.\(^{39,41}\)

Our initial QM/MM reaction coordinate calculations were performed at the HF/3-21G*:Amber level, i.e., the QM calculations were carried out at the HF/3-21G* level while the MM calculations were carried out by using the Amber force field implemented in the Amber8 program. The geometry optimizations were converged to the default criteria of the Gaussian03 (for QM part) and Amber8 (for MM part) programs. During the QM geometry optimization process, the pseudobonds were treated with the well-established effective core potential parameters associated with 3-21G* basis set.\(^{35}\) In the MM energy minimization process, only atoms within 20 Å of ThDP were allowed to move. No cutoff for nonbonding interactions was used in the QM/MM calculations and the MM energy minimizations. The iterative, restrained geometry optimization/energy minimization processes were applied repeatedly to different points along the reaction coordinate, resulting in a minimum energy path for the reaction in the enzymatic environment and its associated potential energy surface. Given that the determined minimum energy path is smooth and continuous, Hessian matrices for degrees of freedom involving atoms in the QM subsystem were calculated at stationary points, leading to determination of the corresponding vibrational frequencies.\(^{41}\) An energy maximum on the path with one and only one imaginary frequency is a transition state, whereas an energy minimum along the path without any imaginary frequency is characterized as an intermediate. Finally, the geometries optimized at the HF/3-21G*:Amber level along the reaction path were then used to perform single-point energy calculations at higher levels, including B3LYP/6-31G*:Amber, B3LYP/6-31+G*:Amber, B3LYP/6-31+G**:Amber, MP2/6-31G*: Amber, and MP2/6-31+ +G**:Amber levels.

Most of the QM and QM/MM calculations were performed in parallel on an IBM X-series Cluster (with 340 nodes and 1,360 processors) at the Center for Computational Sciences, University of Kentucky. Some computations were carried out on a 34-processors IBM x335 Linux cluster and SGI Fuel workstations in our own laboratory.

**Results and Discussion**

*Fundamental Reaction Pathway and Activation Barriers from QM Reaction Coordinate Calculations on the Model Reaction System.*

As depicted in Figure 1, the first three reaction steps are common for all of concerned condensation reactions of two \( z \)-keto acid molecules no matter whether the second \( z \)-keto acid molecule involved in the later reaction step is still pyruvate or not. In addition, the aforementioned discussion indicates that the rate-determining step of the condensation reactions is always the second reaction step. So, our QM reaction coordinate calculations were focused on the first three reaction steps depicted in Figure 1. To simplify the QM reaction coordinate calculations, we used a simplified model system. Specifically, we replaced the \(-P_2O_7\) group of ThDP with \(-OH\) and the protein atoms were not included in the reaction model. The fundamental reac-
tion pathway for the initial reaction step determined by the QM reaction coordinate calculations is qualitatively consistent with the mechanistic hypothesis depicted in Figure 1. AHAS is believed to be activated in the same way as other ThDP-dependent enzymes in which Glu144, a highly conserved residue among AHAS and other ThDP-dependent enzymes, protonates the N1 atom of the pyrimidine ring and induces the formation of the 1',4'-iminotautomer. This basic 4'-imino group is in close proximity (3.3 Å in the crystal structure of 1Z8N) to the C2 catalytic centre of ThDP as a result of the cofactor’s V-conformation. The abstraction of the proton from C2 by the 1',4'-iminotautomer (R) generates the highly reactive ylide required for catalysis. According to the QM reaction coordinate calculations, during the first reaction step, the proton on C2 gradually transfers to the nearby N atom (N4') of the 4'-imino group on the pyrimidine ring to form the first intermediate (INT1, which is the highly reactive ylide proposed previously) through the first transition state (TS1) depicted in Figure 2.

Further, in the second step of the catalytic cycle of AHAS, C2 atom of the ylide (INT1) gradually attacks the carbonyl carbon of substrate pyruvate, whereas the proton on N4' gradually transfers to the carbonyl oxygen of the pyruvate molecule. This reaction step gives the second intermediate LThDP (INT2) through the second transition state (TS2). In the third step, LThDP is decarboxylated through the third transition state (TS3) to give the resonating HEThDP-enamine intermediate (INT3), when a C—C bond gradually breaks. Depicted in Figure 2 are the optimized geometries of the reactants, transition states, and intermediates.

Figure 2. Geometries of the reactant (ES), transition states, and intermediates optimized at the B3LYP/6-31+G(d) level for the first three steps of proposed catalytic cycle of AHAS.
The energetic results listed in Table 1 show that the highest activation barrier is associated with the second step, suggesting the second step is rate determining. This is also qualitatively consistent with the aforementioned experimental observation. However, the activation barrier, ~36 kcal/mol calculated at the B3LYP/6-31+G* level is significantly higher than the activation barrier (~16 kcal/mol) derived from experimental kinetic data according to the well-known transition state theory in which \( k(T) = \frac{k_B T}{h} \exp(-\Delta G^*/RT) \). We must account for the effects of the protein environment on the activation barrier.

**Table 1. Activation Barriers (\( \Delta G^* \), in kcal/mol) Obtained from the QM Calculations on the Model System for the First Three Steps of AHAS-catalyzed Condensation Reactions.**

<table>
<thead>
<tr>
<th>Method</th>
<th>( \Delta G_1^* )</th>
<th>( \Delta G_2^* )</th>
<th>( \Delta G_3^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3LYP/6-31+G*</td>
<td>9.27</td>
<td>37.50</td>
<td>0.25</td>
</tr>
<tr>
<td>B3LYP/6-31++G**</td>
<td>7.65</td>
<td>37.29</td>
<td>0.29</td>
</tr>
<tr>
<td>B3LYP/6-311++G**</td>
<td>8.09</td>
<td>35.83</td>
<td>0.00</td>
</tr>
<tr>
<td>Expt.</td>
<td>16.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*All the geometries were optimized at B3LYP/6-31+G* level.

\( \Delta G_1^* \), \( \Delta G_2^* \), and \( \Delta G_3^* \) refer to calculated activation barriers for reaction steps 1, 2, and 3, respectively, in Figure 1. Included also in \( \Delta G_1^*, \Delta G_2^*, \) and \( \Delta G_3^* \) values are thermal corrections to free energies calculated at B3LYP/6-31+G* level.

The experimental activation barrier was based on the experimental data in Ref. 25 and the transition state theory in which \( k(T) = \frac{k_B T}{h} \exp(-\Delta G^*/RT) \). We must account for the effects of the protein environment on the activation barrier.

**Reaction Pathway and Activation Barriers from QM/MM Calculations**

To account for the effects of protein environment on the reaction pathway and the corresponding activation barrier of the enzymatic reaction, we performed QM/MM reaction coordinate calculations on the first two reaction steps, including the rate-determining step (i.e. the second step), of the entire enzymatic reaction system. The QM/MM-optimized geometries of the enzyme-substrate complex (ES), TS1, INT1, TS2, and INT2 are depicted in Figure 3.

As seen in the optimized geometries of ES, TS1, and INT1, during the first reaction step, the proton on C2 of ThDP gradually transfers to the nearby N atom (N4) of the 4-thymine group on the pyridine ring, whereas the proton on N1 atom of the pyrimidine ring of ThDP gradually transfers to an oxygen atom of the Glu144 side chain. So, the first reaction step in the protein environment is a double-proton transfer process, instead of the single-proton transfer process found in the aforementioned model reaction system. The protein environment remarkably changes the chemical nature of the first reaction step through coupling with an additional proton transfer from ThDP to the Glu144 side chain of the protein. The additional proton transfer from ThDP to the Glu144 side chain is expected to significantly stabilize the TS1 structure and, thus, significantly lower the activation barrier for this reaction step. The double-proton process is completed when the first intermediate (INT1) is formed. In fact, the activation barrier determined for the first reaction step by the QM/MM calculations at the B3LYP/6-31+G*:Amber//HF/3-21G*:Amber and B3LYP/6-31++G**:Amber//HF/3-21G*:Amber levels are only 1.90 and 1.94 kcal/mol, respectively. The QM/MM-calculated activation barrier for this reaction step is ~6 to 7 kcal/mol lower than that calculated for the corresponding step of the model reaction system.

Because of the different chemical nature of the first reaction step, the chemical nature of the second reaction step in the protein environment is also remarkably different from that in the model reaction system. Without accounting for the protein environment, the QM reaction coordinate calculations on the aforementioned model reaction system demonstrated that the proton on N4 gradually transfers to the carbonyl oxygen of pyruvate, whereas the C2 atom gradually attacks the carbonyl carbon of substrate pyruvate. However, according to the QM/MM reaction coordinate calculation, the proton on N4 only forms a hydrogen bond with the carbonyl oxygen atom of pyruvate (in INT2), whereas the C2 atom gradually attacks the carbonyl carbon of substrate pyruvate. The proton on N4 does not transfer to the carbonyl oxygen of pyruvate according to the QM/MM calculations.

In the INT1, TS1, and INT2 geometries depicted in Figure 3, pyruvate atoms form two hydrogen bonds with amino acid residues of AHAS. One hydrogen bond exists between the carbonyl oxygen of pyruvate and the NH2 group of Gln207 side chain (through 1HE2 atom on the NH2 group). The other hydrogen bond exists between the carbonyl oxygen of pyruvate and the backbone NH of Gly121'. In addition, ThDP atoms also form two hydrogen bonds with amino acid residues of AHAS. One of the hydrogen bonds exists between the carbonyl oxygen of Gly511 backbone and the N4 atom of ThDP (through an H atom on N4). The other is between the N1 atom of ThDP and the protonated oxygen on the side chain of Ghu144'. The later is consistent with the experimental observation reported by Bar-Ilan. Altogether, the four hydrogen bonds between pyruvate-ThDP and the four residues of AHAS are expected to stabilize the TS2 structure and, therefore, lower the activation barrier.

Further, a sequence alignment of 24 available AHAS proteins from different sources, including plant, fungal, algal, and bacterial species, reveals that the aforementioned four key residues forming hydrogen bonds with pyruvate or ThDP are all highly conserved for AHAS enzymes, suggesting that these four residues (i.e., Glu144, Gln207, Gly121, and Gly511) may play an important role in the catalysis for all of the AHAS enzymes. Our structural results are qualitatively consistent with available experimental kinetic data in literature. For example, Glu47Gln and Glu47Ala mutations on E. coli AHAS II (corresponding to Glu144Gln and Glu144Ala mutations on A. thaliana AHAS II) decreased the catalytic rate constant (kcat) of the enzyme by about 10- and 13-fold, respectively. This is because the mutations destroy the hydrogen bond between the N1' atom of ThDP and the protonated oxygen on the side chain of Glu144 (or Glu47 in E. coli AHAS II). For the same reason, substitution of ThDP with an analog without N1' or the 4'-amino group also considerably decreased the enzymatic activity. These experimental data qualitatively confirm the importance of Glu144' residue in stabilizing the TS2 geometry depicted in Figure 3. Chipman and coworkers also reported that mutation of a highly conserved residue, i.e., Gln110 in E. coli AHAS II correspond-
According to Gln207 in *A. thaliana* AHAS II, led to “a significant loss of activity,” although they did not report specific kinetic parameters for the mutations on the residue. Their experimental observation qualitatively confirms the importance of Gln207 in stabilizing the TS2 geometry depicted in Figure 3. The hydrogen bond between the N4 atom of ThDP and the backbone oxygen atom of Gly511 existed in all of the QM-MM-optimized geometries depicted in Figure 3, which is consistent with the observation by Duggleby et al. that such a hydrogen bond existed in all of the X-ray crystal structures of AHAS.

Depicted in Figure 3(F) are the plots of the relative potential energies calculated at the HF/3-21G*:Amber, B3LYP/6-31G*, B3LYP/6-31+G*:Amber, B3LYP/6-31++G**:Amber, and MP2/6-31G*:Amber levels versus the reaction coordinate (the
Table 2. Activation Barrier ($\Delta G^\ddagger$, in kcal/mol) for Calculated by
Performing QM/MM Calculations on the Second Reaction Step of the
AHAS Catalytic Cycle.

<table>
<thead>
<tr>
<th>Methoda</th>
<th>$\Delta G^\ddagger$ (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QM/MM(HF/3-21G*:Amber)</td>
<td>22.33</td>
</tr>
<tr>
<td>QM/MM(B3LYP/6-31G*:Amber)</td>
<td>18.79</td>
</tr>
<tr>
<td>QM/MM(B3LYP/6-31+G*:Amber)</td>
<td>15.82</td>
</tr>
<tr>
<td>QM/MM(B3LYP/6-31+ +G**:Amber)</td>
<td>15.20</td>
</tr>
<tr>
<td>QM/MM(MPWKCIS1K/6-31G**:Amber)</td>
<td>18.19</td>
</tr>
<tr>
<td>QM/MM(BHandHLYP/6-31+ +G**:Amber)</td>
<td>20.25</td>
</tr>
<tr>
<td>QM/MM/B97-2/6-31+ +G**:Amber)</td>
<td>14.48</td>
</tr>
<tr>
<td>QM/MM(nPW1PW91/6-31+ +G**:Amber)</td>
<td>12.93</td>
</tr>
<tr>
<td>QM/MM/B98/6-31+ +G**:Amber)</td>
<td>14.10</td>
</tr>
<tr>
<td>QM/MM/MVP2/6-31G*:Amber)</td>
<td>15.33</td>
</tr>
<tr>
<td>QM/MM(MP2/6-31G*:Amber)</td>
<td>15.31</td>
</tr>
<tr>
<td>Expt.</td>
<td>16.21</td>
</tr>
</tbody>
</table>

*aCalculated activation barrier ($\Delta G^\ddagger$) is calculated as the energy change from INT1 to TS2 on the corresponding minimum-energy path. Thermal correction to Gibbs free energy at the HF/3-21G* level at T = 310 K was included in the calculation of $\Delta G^\ddagger$.

Q/MM reaction coordinate driving was not carried out at the MP2/6-31+ +G**:Amber level because it would be very time-consuming at this level of theory for this particular enzymatic reaction system. Instead, we only performed the single-point energy calculations on the local minimum (associated with INT1) and first-order saddle point (associated with TS2) geometries determined at the MP2/6-31G*:Amber level.

The experimental activation barrier was based on the experimental data in Ref. 25 (determined at T = 310 K) and the classic transition state theory in which $k(T) = (k_B T/h) \exp(-\Delta G^\ddagger/RT)$.

minimum-energy path). It would be very time consuming to perform QM/MM calculations on all of the structures along the reaction path at the MP2/6-31+ +G**:Amber level. So, single-point QM/MM energy calculations at the MP2/6-31+ +G**:Amber level were only performed on the INT1 and TS2 geometries used for the QM/MM calculations at the MP2/6-31G*:Amber level to estimate the activation barrier at the MP2/6-31+ +G**:Amber level. The QM/MM-calculated values of the activation barrier are summarized in Table 2. As one can see in Table 2, the QM/MM calculations at the HF/3-21G*:Amber level considerably overestimated the activation barrier for this reaction step. The QM/MM calculations at the B3LYP/3-1G*:Amber level also significantly overestimated the activation barrier. Nevertheless, the QM/MM calculations at more sophisticated levels, including the B3LYP/6-31+ +G**:Amber, B3LYP/6-31+ +G**:Amber, MP2/6-31G*:Amber, and MP2/6-31+ +G**:Amber levels, all lead to very close results (15.20 to 15.82 kcal/mol) that are all in good agreement with the experimental activation barrier of 16.21 kcal/mol.25 We note that various new density functionals were reported in the past a few years and that some newly developed functionals could outperform B3LYP.34 To examine the effect of the choice of the density functional on the calculated activation barrier, the activation barrier of the rate-determining step (step 2) was also calculated by using several other functionals available in the Gaussian03 program. As shown in Table 2, the activation barrier calculated with B3LYP is closest to the result calculated with the MP2 method and the experimental activation barrier of 16.21 kcal/mol25 for this particular QM/MM system.

The activation barrier calculated with the QM/MM method considering the effects of protein environment is ~20 kcal/mol lower than the corresponding activation barrier calculated with the QM method neglecting the protein environment. To better understand how the protein environment contributes to the ~20 kcal/mol decrease of the activation barrier, we also determined the single-point energy surface of a tailored QM/MM system in which the MM part includes only the residues having covalent bonds with the QM atoms. The activation barrier calculated at the B3LYP/6-31+ +G**:Amber level for the tailored QM/MM system is 15.97 kcal/mol, which is only slightly higher than 15.20 kcal/mol calculated for the complete QM/MM system. These results suggest that the protein environment affects the reaction and activation barrier mainly through maintaining the active site of the enzyme.

To further understand the importance of Glu144 in catalysis, we also carried out the QM/MM reaction coordinate calculations on the second reaction step (i.e., the rate-determining reaction step) associated with the Glu144Ala mutant at the same QM/MM level of theory used for the native enzyme. The QM/MM-optimized INT1 and TS2 geometries for the mutant are depicted in Figure 4. Compared with the corresponding QM/MM-optimized INT1 and TS2 geometries (Fig. 3) associated with the native enzyme, the hydrogen bonding interactions in the INT1 and TS2 geometries (Fig. 4) associated with the Glu144Ala mutant are all similar, except that the hydrogen bond between residue #144 and ThDP no longer exists in the mutant. As a result, the activation barrier calculated (at the same MP2/6-31+ +G**:Amber//HF/3-21G*:Amber level) for the second reaction step of the reaction catalyzed by the Glu144Ala mutant is 17.0 kcal/mol, which is 1.7 kcal/mol higher than the corresponding activation barrier (15.3 kcal/mol in Table 2) calculated for the second reaction step of the reaction catalyzed by the native enzyme at the same level of theory. The calculated activation barrier increase of 1.7 kcal/mol is in good agreement with the experimental observation that the Glu47Ala mutation on E. coli AHAS II (corresponding to Glu144Ala mutation on A. thaliana AHAS II) decreased the catalytic rate constant ($k_{cat}$) of the enzyme by ~13-fold, as a ~13-fold decrease in $k_{cat}$ corresponds to a ~1.5 kcal/mol increase in activation barrier according to the conventional transition state theory (CTST).55,56

It should be pointed out that the starting structures used for the above QM/MM reaction coordinate calculations were all based on the X-ray crystal structure. Warshel and coworkers noted that QM/MM reaction coordinate calculations on an enzymatic reaction system using different starting structures could lead to significantly different activation barrier values. More recently, Zhang and coworkers performed pseudobond QM/MM reaction coordinate calculations on an enzymatic reaction by using multiple initial structures, i.e., multiple snapshots of a molecular dynamics (MD) simulation of the enzyme-substrate complex. According to the QM/MM calculations by Zhang and coworkers, the fluctuation of the calculated energy barriers is only ±1.1 kcal/mol and the barrier fluctuation has a strong correlation with the values of the geometric parameters associated
with the reaction coordinate. When the values of the geometric parameters associated with the reaction coordinate in a snapshot are closer to the corresponding average values of the geometric parameters in the simulated enzyme-substrate complex, the energy barrier calculated with this snapshot is closer to the average value of the energy barriers calculated with all of the chosen snapshots. The pseudobond QM/MM reaction coordinate approach used in the present study is very similar to that used by Zhang and coworkers. Hence, the similar performance and similar structure-barrier fluctuation correlation, and hopefully the similarly small energy barrier fluctuation, may be expected for the present pseudobond QM/MM reaction coordinate calculations. To examine the possible barrier fluctuation for the current reaction system, we further carried out an MD simulation on the INT1 structure for 2 ns and obtained a stable MD trajectory. The MD simulation was performed in the same way as we did in our previous computational studies on other protein systems. The final snapshot of the MD-simulated INT1 structure was used as the initial structure to repeat the QM/MM reaction coordinate calculations and calculate the activation barrier for the second reaction step. The activation barrier calculated at the MP2/6-31+G**:Amber//HF/3-21G*:Amber level starting from the MD-simulated INT1 structure is 15.1 kcal/mol, which is close to the activation barrier of 15.3 kcal/mol calculated at the same level starting from the X-ray crystal structure. The close activation barrier values calculated by using different starting structures suggest that the energetic results calculated in this study are reliable.

The good agreement between the QM/MM-calculated activation barrier for the rate-determining step and the corresponding experimental kinetic data suggests that the reaction pathway identified in this study is reasonable and that the QM/MM protocol used is satisfactory. As the activation barrier calculated by the QM/MM method accounting for the effects of protein environment is ~20 kcal/mol lower than the corresponding activation barrier calculated by the QM method neglecting the protein environment, the protein environment stabilizes the rate-determining transition state TS2 more favorably than stabilizing the intermediate INT1. In light of these new insights, it should be very interesting to design stable analogs of the substructure of transition state TS2. The substructure of transition state TS2 may include all atoms of pyruvate and ThDP and there is no covalent bond between the substructure and the remaining part of the protein. The stable analogs of the substructure are expected to be novel, potent AHAS inhibitors that may be used as promising new antituberculosis drugs, antibacterial drugs, and/or herbicides to overcome the drug resistance problem.

Conclusions

This is the first computational study to uncover the detailed catalytic mechanism for the common steps of AHAS-catalyzed condensation reactions of α-keto acids. The first-principles QM and hybrid QM/MM reaction coordinate calculations have led us to uncover the fundamental reaction pathway, determine the activation barriers, and obtain valuable insights concerning the specific roles of several key amino acid residues for the common steps of AHAS-catalyzed condensation reactions of α-keto acids. The QM/MM reaction coordinate calculations have provided valuable details concerning the reaction process and role of the protein environment. The most important amino acid residues involved in the catalysis include Glu144, Gln207, Gly121, and Gly511; these residues form favorable hydrogen bonds with the reaction center (consisting of atoms from substrate pyruvate and cofactor ThDP) during the catalytic reaction process to stabilize the transition states. In addition, Glu144 also accepts a proton from ThDP through hydrogen bonding during the catalytic reaction process. The optimized geometries of the transition states are qualitatively consistent with available experimental observations, including data from site-directed mutagenesis, and provide more detailed structural and mechanistic information for transition states and their evolution during the key reaction steps. Compared to the high activation barrier calculated for the model...
reaction system neglecting the protein environment, the favorable interactions between the reaction center and protein environment remarkably stabilize the transition state and, thus, lower the activation barrier for the rate-determining reaction step by \(-20\) kcal/mol. All of the computational results indicate that the rate-determining step of the AHAS-catalyzed reactions is the second reaction step and the activation barrier calculated for the rate-determining step is in good agreement with the experimentally derived activation barrier. The detailed structural and mechanistic insights should be valuable for rational design of novel, potent AHAS inhibitors that may be used as promising new anti-tuberculosis drugs, antibacterial drugs, and/or herbicides to overcome the drug resistance problem.

Acknowledgment

The authors acknowledge the Center for Computational Sciences (CCS) at University of Kentucky for supercomputing time on IBM X-series Cluster with 340 nodes and 1.360 processors.

References
