Molecular Recognition of Aldehydes by Aldehyde Dehydrogenase and Mechanism of Nucleophile Activation

Troy Wymore,[1]* John Hempel,[2] Samuel S. Cho,[3] Alexander D. MacKerell, Jr.,[3] Hugh B. Nicholas, Jr.,[1] and David W. Deerfield II[1]

[1]*Pittsburgh Supercomputing Center, Biomedical Initiative Group, Pittsburgh, Pennsylvania*
[2]*Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania*
[3]*University of Maryland, School of Pharmacy, Baltimore, Maryland*

ABSTRACT **Experimental structural data on the state of substrates bound to class 3 Aldehyde Dehydrogenases (ALDH3A1) is currently unknown. We have utilized molecular mechanics (MM) simulations, in conjunction with new force field parameters for aldehydes, to study the atomic details of benzaldehyde binding to ALDH3A1. Our results indicate that while the nucleophilic Cys243 must be in the neutral state to form what are commonly called near-attack conformers (NACs), these structures do not correlate with increased complexation energy calculated with the MM-Generalized Born Molecular Volume (GBMV) method. The negatively charged Cys243 (thiolate form) of ALDH3A1 also binds benzaldehyde in a stable conformation but in this complex the sulfur of Cys243 is oriented away from benzaldehyde yet yields the most favorable MM-GBMV complexation energy. The identity of the general base, Glu209 or Glu333, in ALDHs remains uncertain. The MM simulations reveal structural and possible functional roles for both Glu209 and Glu333. Structures from the MM simulations that would support either glutamate residue as the general base were further examined with Hybrid Quantum Mechanical (QM)/MM simulations. These simulations show that, with the PM3/OPLS potential, Glu209 must go through a step-wise mechanism to activate Cys243 through an intervening water molecule while Glu333 can go through a more favorable concerted mechanism for the same activation process. Proteins 2004;57:758–771.** © 2004 Wiley-Liss, Inc.

Key words: enzyme mechanism; thiolate; aldehydes; general base; molecular dynamics simulation; molecular mechanics; hybrid QM/MM potentials; force field parameterization

INTRODUCTION

Detoxification of aldehydes by Aldehyde Dehydrogenases (ALDHs) is a critical function in healthy cells since aldehydes are implicated in cytotoxicity, mutagenicity, genotoxicity, and carcinogenesis.[1] ALDHs occur in a wide variety of phyla and most organisms have multiple ALDHs that act on different sets of substrates oxidizing them to their corresponding carboxylic acid. Human class 3 ALDH (ALDH3A1; referred to as ALDH3 in this text) has been shown to detoxify a metabolite of the cancer chemotherapeutic agent cyclophosphoamide (CP), leading to drug resistance.[2] Therefore, inhibitors of ALDH3 taken with CP drugs may lead to more effective treatments for those experiencing drug resistance to CP. Designing inhibitors of ALDH3 should be based on a thorough knowledge of the reaction mechanism. Several insights into the ALDH mechanism of converting benzaldehyde to benzoic acid have been provided by the crystal structure,[3] site-directed mutagenesis,[4–6] and NMR.[7] These studies provide the basis by which to carry out advanced molecular modeling studies on the enzyme.

ALDH3 is a homodimeric enzyme; each subunit contains a nicotinamide adenine dinucleotide (NAD)-binding domain, a catalytic domain, and an oligomerization domain.[3] A 15-Å-long funnel-shaped passage formed between these domains leads to the active site. NAD binds to the face of the dimer opposite of the funnel entrance. The catalytic cysteine (Cys243) is located in a loop structure connecting helix-9 with beta strand-7. This loop structure starts with Asn238 and ends with Asp247. Two water molecules lie within 5.0 Å of the cysteine sulfur atom and allow the possibility of hydrogen bonds with the backbone amide protons of Cys243 and Val244. The side-chain amide protons of Asn114 also may favorably interact with Cys243. There do not appear to be any obvious direct interactions that would stabilize the thiolate form of Cys243. Two highly conserved acidic residues are located near the catalytic cysteine, Glu209 and Glu333. Either one of these residues could serve the function as a general base to activate Cys243. The side-chain of Glu209 is directed somewhat away from Cys243 yet still has two water molecules within 4.0 Å of both carboxylate oxygen atoms. These water molecules are separated from the sulfur of

---

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html

Grant sponsor: NIH-NIAAA; Grant number: AA-12753; Grant sponsor: NIH-NCRR; Grant number: RR06009; Grant sponsor: NIH; Grant number: GM51501.

*Correspondence to: Troy Wymore, Pittsburgh Supercomputing Center, Biomedical Initiative Group, 4400 Fifth Avenue, Pittsburgh, PA 15213. E-mail: wymore@psc.edu

Received 3 November 2003; Accepted 8 June 2004

Published online 10 August 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20256

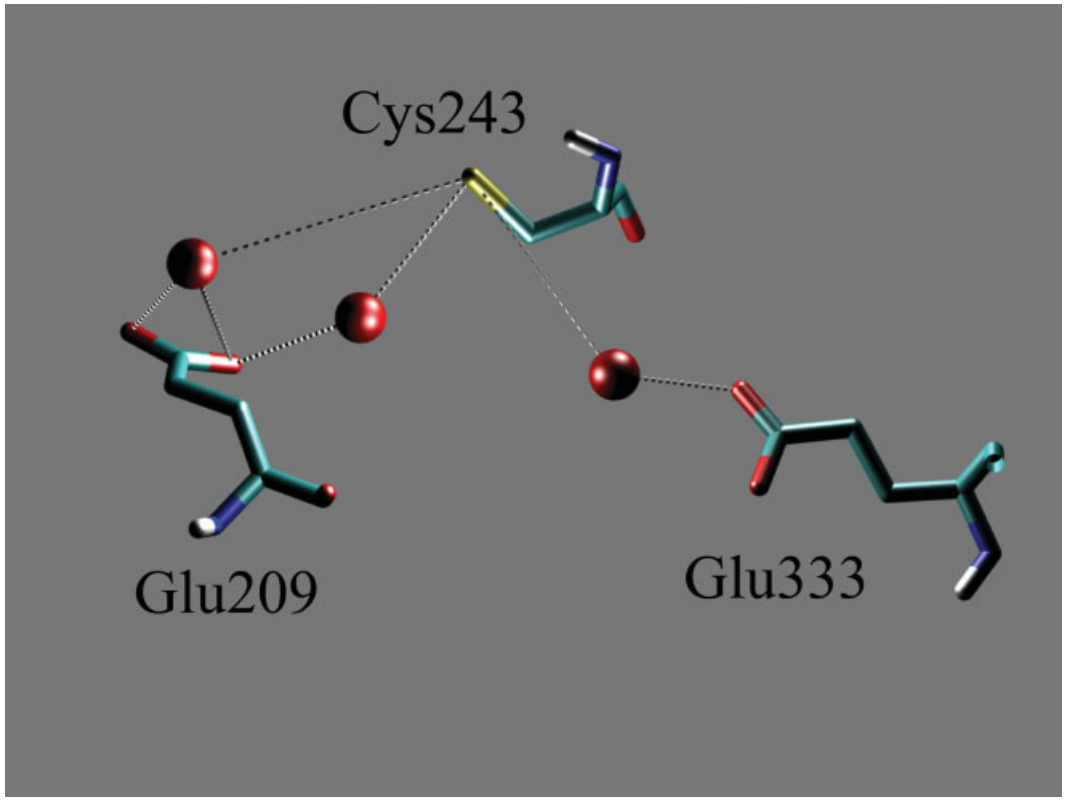

Fig. 1. Illustration of the active site Cys243, surrounding water molecules (red spheres), and the two possible general bases, Glu209 and Glu333, from X-ray structure.

Cys243 by 4.0 and 6.0 Å. The side-chain of Glu333 is directed more toward Cys243 than Glu209 is and has one water molecule within both 2.7 Å of a carboxylate oxygen atom and 5.0 Å of the sulfur atom of Cys243. Glu333 is also involved in interactions with the NAD cofactor. Figure 1 illustrates these residues in the crystal structure of rat ALDH3. The temperature factors for Glu209 are much higher than Glu333,[3] which may serve as evidence that the side-chain experiences larger scale motions, some of which may activate Cys243. Neither residue can be excluded from being the general base by inspection of the crystal structure. Furthermore, the substrate binding mode is not easily deduced from the holo form nor is the stereospecificity of hydride transfer explained.

Whether Glu209 or Glu333 functions as the general base has been the basis for some controversy. There are actually two reaction steps in which participation of a general base is expected: for activation of the nucleophilic cysteine to attack the substrate (Reaction I, see Fig. 2) and for activation of a water molecule for hydrolysis of the thioester intermediate (Reaction III, Fig. 2). The NAD cofactor accepts a hydride anion from the thiohemiacetal intermediate to produce NADH and a thioester intermediate. It is uncertain to what extent NAD is involved in other steps of the ALDH cycle. A proposal for the catalytic mechanism of ALDH3 suggested that Glu333 was the general base used for Rxn I and Rxn III based on its strict sequence conservation and less-than-strict conservation of Glu209.[8,9] Chemical modification of the glutamate sequentially adjacent to

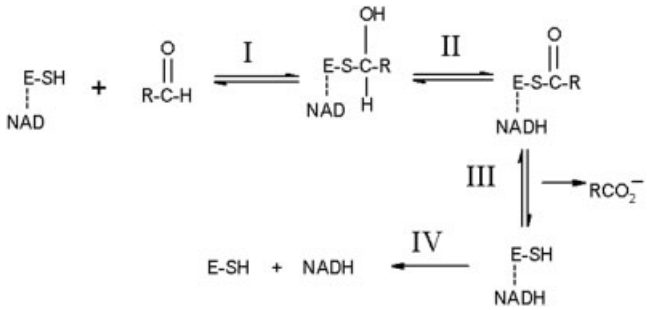

Fig. 2. Reaction mechanism for ALDH. E-SH indicates the enzyme with Cys243 as the nucleophile and the dashed line to NAD indicates that it is bound prior to reacting with the aldehyde substrate. Reaction I: Nucleophilic activation and attack on substrate by Cys243 to give a thiohemiacetal intermediate. Reaction II: Hydride transfer from the intermediate to NAD to form NADH and a thioester. Reaction III: Hydrolysis of the thioester with release of the carboxylic acid. Reaction IV: Release of NADH.

Glu399, Glu398 in ALDH2 (corresponding to Glu332 in ALDH3), revealed that this residue is in a hydrophobic environment and ionized.[10] Mutation of Glu209 to glutamine showed a 10-fold larger K*m* for NAD compared to wild-type while the K*m* for benzaldehyde was similar to wild-type.[4] This mutant also displayed a slightly decreased k*cat* for both dehydrogenase and esterase activity. Glu268 in ALDH2 (corresponding to Glu209 in ALDH3) was initially shown to be essential for the catalytic mechanism of ALDH2 via chemical modification studies.[11,12]

Later, Wang and Weiner[13] found further support that Glu268 in ALDH2 served as the general base since both dehydrogenase and esterase activity were abolished when this residue was changed to an aspartate, glutamine, or lysine residue and no pre-steady state burst of NADH was found in either reaction of these mutants. Therefore, Glu268 was ascribed the role of general base in both Rxn1 and Rxn2. Yet, it was also shown that E399Q in ALDH2 showed no pre-steady state burst of NADH either, while its kcat was reduced 10-fold.[14] Mann and Weiner[5] later presented experimental evidence that Glu333 served as the general base in ALDH3. Since this evidence was based on the fact that the E333Q mutant destabilized the protein as well as the results mentioned above, the general base may be different depending on the form of ALDH. One argument against this would be the high degree of sequence and structural similarity between different isozymes (ALDH2 vs. ALDH3, for example).[15]

In this report, we detail our efforts using molecular modeling methods to characterize the initial events in the catalytic cycle of ALDH3. For these studies, we considered several possibilities for modeling the Michaelis complex of benzaldehyde (BA) with ALDH3. A major challenge in this process is assigning protonation states to ionizable residues, since the ionization states remain constant during an MD simulation using a molecular mechanics (MM) force field. While continuum methods exist for calculating pKa's in proteins,[16] these methods are best at detecting large shifts in pKa. Their accuracy suffers at modeling small shifts in pKa. Thus, in the present study an approach is applied in which MD simulations were performed with different ionization states for Cys243 and Glu333 (see Methods). These simulations and the subsequent post-processing of the trajectories were used to determine the most energetically favorable conditions for substrate binding. The ligand-binding energies were calculated with the MM-Generalized Born Molecular Volume (GBMV) method.[17] Burgi and Dunitz have previously shown that there is a preferred orientation between nitrogen- and oxygen-nucleophiles and the plane of the electrophilic carbonyl group.[18,19] This Burgi-Dunitz approach trajectory (BDat) can be defined as an angle of 109° between nucleophile-carbonyl carbon-carbonyl oxygen. Therefore, we examined the orientation of the Cys243 *sulfur* nucleophile with respect to the plane of the substrate carbonyl group to determine if this preferred orientation was present in ALDH3 as well. A study by Chakrabarti and Pal[20] of this nucleophile-electrophile interaction within cysteine residues of metalloprotein structures has been reported. Structures that lie somewhat closer to covalent bonding along the BDat have been called near-attack conformers (NACs)[21] and we examined their occurrence in the simulations as well. A full assessment of any *NAC effect* would require simulations to estimate the free energy of moving from the bound state to a NAC as well as the free energy profile for formation of the thiohemiacetal intermediate from the NAC and also the reference solution reactions.[22,23] A complete account of the cysteine nucleophilic attack on benzaldehyde is forthcoming.

We also employed a hybrid quantum mechanical (QM)/molecular mechanical (MM) potential as the energy function[24,25] to simulate possible Cys243 activation mechanisms. In the QM/MM approach, atoms making up the active site are treated explicitly by a QM method and, therefore, can be used to model bond breaking, bond formation, and charge transfer while the surroundings are represented by a MM force field. A term in the hybrid potential is used to couple the two regions so that the QM region is affected by the enzyme environment. This methodology has been used in several investigations of enzyme reactions.[26,27]

METHODS

Optimization of the Aldehyde Parameters

Parameters were developed for a set of common, simple aldehydes: acetaldehyde, propionaldehyde, chloroacetaldehyde, and benzaldehyde. The CHARMM potential energy function, consisting of external (nonbonded) and internal (bonded) interaction terms, has the general form[28]:

$$E = \sum_{\text{bonds}} \mathrm{K_b}(\mathrm{b}-\mathrm{b_0})^2 + \sum_{\text{angles}} \mathrm{K_\theta}(\theta-\theta_0)^2 + \sum_{\text{torsions}} K_\phi[1+\cos(n\phi-\delta)] + \sum_{U-B} \mathrm{K_{UB}}(S-S_0)^2 + \sum_{\text{impropers}} K_\omega(\omega-\omega_0)^2 + \sum_{\text{nonbonded pairs}} \left\{\varepsilon_{ij}\left[\left(\frac{\mathrm{R}_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\mathrm{R}_{\min,ij}}{r_{ij}}\right)^{6}\right] + \frac{qiqj}{4\pi Dr_{ij}}\right\}$$

where $\mathrm{K_b}$, $\mathrm{K_{UB}}$, $\mathrm{K_\theta}$, $\mathrm{K_\chi}$, and $\mathrm{K_{imp}}$ are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; b, S, θ, ϕ, and ω are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium values for the individual terms. Coulomb and Lennard-Jones 6–12 terms contribute to the nonbonded interactions; ϵ is the Lennard-Jones well depth and $\mathrm{R_{min}}$ is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, D is the effective dielectric constant, and r_{ij} is the distance between atoms i and j. The Lennard-Jones parameters between pairs of different atoms are obtained from the Lorentz-Berthelodt combination rules, in which ϵ_{ij} values are based on the geometric mean of ϵ_i and ϵ_j and $\mathrm{R}_{\min,ij}$ values are based on the arithmetic mean between $\mathrm{R}_{\min,i}$ and $\mathrm{R}_{\min,j}$.

The optimization of the force field for the aldehydes was performed using the same iterative approach previously employed for proteins, nucleic acids, lipids, and other biomolecules in the CHARMM force field[29–31] insuring their compatibility. All of the data obtained for parameter optimization were determined either via QM calculations or, when appropriate, from already existing parameters. Initially, the nonbonded interaction parameters were opti-

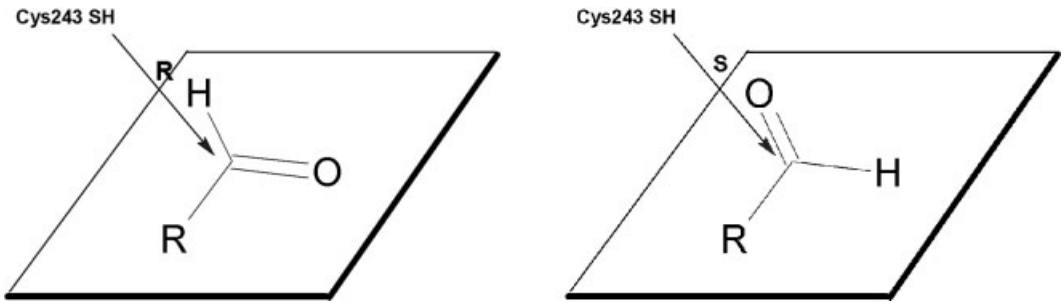

Fig. 3. Two possibilities are shown for benzaldehyde binding in the ALDH3 active site, one in which the Cys243 attack on benzaldehyde would lead to a R-thiohemiacetal (labeled R, **left**) and one in which the attack would lead to a S-thiohemiacetal (labeled S, **right**)

mized, followed by the internal parameters for the bonds, angles, and torsions. Details of the parameter optimization are included in the Supplemental Material.

Molecular Dynamics Simulation Using a MM Force Field

The starting coordinates for the rat ALDH3 (**1ad3**[3]) were taken from the RCSB Protein Data Bank.[32] This structure was solved to a resolution of 2.6 Å and contains the position of the NAD cofactors. CHARMM[33] was used for all MD simulations using a MM force field. We employed the all-atom force field in CHARMM that contains parameters for the protein[29] and NAD.[34] The MM parameterization for benzaldehyde is described in the Supplemental Material. Hydrogens were added to the protein and NAD using the HBUILD algorithm.[35] All histidine residues were assigned to be neutral with their tautomeric state determined by the proximity of a hydrogen bond acceptor to either Nε or Nδ of histidine. Cys243 was assigned to its neutral state while Glu209 and Glu333 were negatively charged. All other residues were represented in their standard protonation states. The protein-cofactor system was then solvated by TIP3P water molecules and the system was minimized for 2,000 steps with harmonic restraints on all solute heavy atoms. Sodium ions were placed uniformly around the protein with at least 3 water molecules between each ion and the protein to neutralize the system. The system contained 13,914 protein atoms, 140 NAD atoms (2 NAD molecules, one per monomeric unit), 41,538 water atoms, and 14 sodium counterions for a total of 55,606 atoms. SHAKE[36] restraints were used for bonds to hydrogen, which removes this high-frequency stretching motion. Periodic boundary conditions were used with the system simulated in the NVT (constant volume, constant temperature) ensemble. The minimization was followed by 30 picoseconds (ps) of MD in which the velocities were reassigned every 3 ps to correspond to a temperature that started at 10 K and ended at 310 K. A 1-femtosecond (fs) timestep was used at this stage. The remainder of all simulations were performed in the NPT ensemble, a 2-fs timestep and the use of particle mesh Ewald (PME)[37] for calculating long-range electrostatic interactions. During the final part of equilibration (last 120 ps of 200 ps) and during the entire production run, the harmonic restraints on the protein/NAD solute for the holo form were released. The average dimensions of the simulation cell were $104.8 \times 71.6 \times 72.4$ Å.

Our benzaldehyde (BA) model was placed in the active site of the solvated ALDH system with the following restraints that are consistent with observations from X-ray models of substrates complexed with other ALDH family members.[38,39] In all simulations, the carbonyl carbon of BA was restrained to be within 2–4 Å of the sulfur of Cys243 and the carbonyl oxygen of BA was restrained to be within 2–4 Å of the sidechain nitrogen of Asn-114 during the first 83 ps of each simulation and removed during the production phase. Two simulations were performed with BA in the active site with a neutral Cys243 and ionized Glu333. Of these, one simulation had BA oriented for nucleophilic attack in which the product would be a thiohemiacetal in the R-configuration whereas the other simulation was oriented for production of a thiohemiacetal in the S-configuration (see Fig. 3). Two additional simulations were performed with BA in the active site; each with Cys243 in the negatively charged thiolate form. Of these, one had Glu333 negatively charged (labeled S-1-1) and the other had a neutral protonated form (labeled S-1-0) as shown in Figure 4. The simulations were performed for 2.2 nanoseconds (ns) for the holo form and 1.063 ns each for the four possible Michaelis complex forms. Analysis of substrate-bound conformations was performed in two ways. First, we wanted to determine if the enzyme would orient the substrate along the BDat for nucleophilic addition to the carbonyl group and, if so, then what percentage of these structures could be classified as a NAC. A NAC is defined here as a cysteine sulfur–BA carbonyl carbon distance of 3.5 Å or less and a sulfur-carbonyl carbon (BA)-carbonyl oxygen (BA) of $109.5 \pm 15°$.[40] In principle, these geometric criteria could be met with the cysteine sulfur interacting with the aldehydic proton. This would represent an in-plane approach of the nucleophile, which is unfavorable. All NACs identified were out of the aldehyde plane and thus were truly oriented for reaction.

Next, to assess the relevance of forming structures along the BDat and within the NAC region, we divided the trajectory structures into complex (ALDH, NAD, and BA), receptor (ALDH and NAD), and substrate BA systems. We then calculated substrate-binding energies with the MM-GBMV analytical method[17] with an added term that

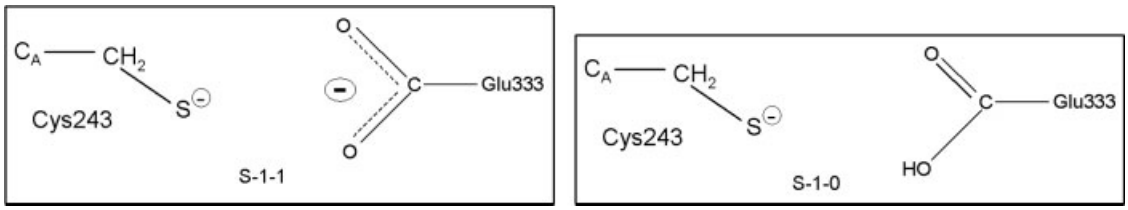

Fig. 4. Protonation state of Cys243 and Glu333 shown for the S-1-1 and S-1-0 simulation.

calculates the non-polar contribution to solvation through the solvent-accessible surface area (available in CHARMM version c30b1). This was done over the last 600 ps (6,200 structures) of each trajectory. The $\Delta G_{MM\text{-}GBMV}$ energies were then calculated in the standard way by subtracting the average ligand and receptor energy from the average complex energy. GBMV solvation energies have been shown to highly correlate with results from the Poisson-Boltzmann (PB) method.[17] The resulting energies are not absolute binding energies because they neglect the effect of protein relaxation on loss of substrate, loss of rotational and translation motion to the ligand, and other effects.[41] Yet, these effects should cancel since we are not looking at different ligands, just slightly different protonation states of active site residue(s). Hydrogen bond analysis was performed with the CORREL module and the hydrogen-bonding parameters in CHARMM.

Modeling Reactions With QM/MM potential

Snapshots from the MM simulations in the presence of substrate that showed the thiol proton of Cys243 pointing in the direction of a carboxylate oxygen of Glu333 or Glu209 that would most support either residue as the general base were used as starting points for simulations with the QM/MM potential. Each structure had the thiol proton hydrogen bonded to an intervening water molecule, which was hydrogen bonded to either Glu, respectively. In addition, the conformations had the substrate carbonyl carbon within 3.5 Å of the sulfur atom of Cys243 along the BDat.[18,19] The selected snapshots appeared most relevant for the reactions we studied but we cannot rule out that other structures with similar features but with different "second-sphere" geometries would give different results. Water molecules outside 30 Å of the sulfur atom of Cys243 and all sodium ions were deleted. We divided up the system into the following regions for both simulations: (1) the QM region consisted of the side-chains of Cys243 and Glu333 or Glu209, the BA substrate, and 6 or 7 surrounding water molecules (46 or 49 QM atoms); (2) the inner MM region in which atoms that were allowed to move freely were those residues with one or more atoms within 20 Å of the sulfur atom of Cys243; and (3) the outer MM region that remained fixed included atoms in residues with no atoms within 20 Å of the sulfur atom of Cys-243. These simulations were performed with the program DYNAMO.[42]

In the QM/MM simulations, we used the OPLS-AA force field[43] for the MM partition with charges for NAD from the CHARMM distribution[34] and the PM3 Hamiltonian for the QM atoms.[44] The PM3 method was chosen because it gives the correct orientation for the water dimer and solvent structures are critical for the reactions studied. The heat of formation, ΔH_f, is underestimated for the glutamate side-chain in the AM1 method. The ΔH_f for the hydroxide anion and hydronium ion are underestimated in both AM1 and PM3.[44] So the magnitude of the energy differences for these reactions will not be accurate. We could make them more accurate by the use of reaction-specific parameters.[45] However, since our goal was to investigate the difference between two likely mechanisms, it is unlikely that the qualitative conclusions would change by making the potential more accurate. The use of reaction-specific parameters also can change the magnitude of interactions that were parameterized correctly.[46] To deal with the case where a covalent bond traverses the QM/MM boundary, a hydrogen link-atom approximation was used.[42] Since periodic boundary conditions are not used in the QM/MM simulations, nonbonding interactions were calculated using an atom-based switching function with inner and outer cutoffs of 9.5 and 13.5 Å, respectively, which has shown to give results comparable to no-cutoff methods.[42]

An appropriate reaction coordinate, q1, for the proton transfer between the intervening water molecule and glutamate (atom type of donor and acceptor are the same) is the symmetric stretch:

$$q1 = r_{OwH} - r_{OGH}$$

where r_{OWH} and r_{OGH} are the distances of the proton from the donor water molecule and the acceptor carboxylate oxygen of each respective glutamate. The minimum structure for the hydroxide anion and protonated glutamate can be used as the starting point for the second proton transfer from Cys243 to the hydroxide anion. An appropriate reaction coordinate, q2, for the proton transfer between these two residues (atom type of donor and acceptor are different) is the asymmetric stretch:

$$q2 = [1/(m_s + m_o)](m_s r_{SH} - m_o r_{OH})$$

where r_{SH} and r_{OH} are the distances of the proton from the donor sulfur and the acceptor oxygen, respectively, with m_s and m_o as their masses.[47] Umbrella sampling molecular dynamics was performed with 100 windows harmonically constrained to reaction coordinate values from −1.00 to 1.00 with spacing of 0.02 Å for the first reaction (q1). A force constant of 2,000 kJ mol^{-1} $Å^{-2}$ was used for the umbrella potentials. Each window was minimized for 2,000 steps. Then 2 ps of equilibration was performed with the velocity Verlet algorithm followed by 10 ps of sam-

pling. A 1-fs time-step is used in all QM/MM simulations. Therefore, each free energy profile consisted of 1.2 ns of simulation. The sampled values of q were used with the weighted histogram analysis method (WHAM) to derive the potential of mean force (free energies) along the reaction coordinate.[48]

Simulations along the q1 reaction coordinate for producing a hydroxide anion and a protonated Glu333 resulted in a spontaneous proton transfer from Cys243 to the hydroxide anion. This result suggested a possible lower energy concerted mechanism to activate Cys243. Therefore, we simulated the 2D free energy profile (q1 × q2). Umbrella sampling MD was performed with 672 windows (32 windows in q1 and 21 windows in q2) harmonically constrained to reaction coordinate values from −1.00 to 1.00 with spacing of 0.0625 Å in q1 and reaction coordinate values of 0.10 to 1.40 with spacing of 0.065 Å in q2. Each window was simulated as in the 1D simulation totaling 8 ns of simulation. The activation of Cys243 through Glu209 indicated a step-wise mechanism; therefore, we performed umbrella sampling MD along q2 (no q1 coupling) ranging from −0.32 to 1.40 with spacing of 0.02 Å totaling 960 ps of simulation. The free energies for the 2D profile were determined from Alan Grossfield's 2D-WHAM program (freely available on his web site, dasher.wustl.edu/alan). PSI-PLOT[49] version 7.0 was used to display the free energy surface. VMD[50] was used extensively to view the trajectories and create Figures 1 and 11.

RESULTS AND DISCUSSION

MM Simulations of Holo Form ALDH3

The results showing that our simulation protocol produces structures along the trajectory that only deviate slightly from the crystal structure and that the dynamics appear reasonable have been reported in a preliminary publication.[51] Briefly, the NAD-binding domains (209 residues each) average Cα root-mean-square-coordinate deviations (RMSCD) from the X-ray structure of 1.23 and 0.97 Å while the catalytic domains (169 residues each) average 0.97 Å for both subunits over the last 2 ns of simulation. Since active sites are often made up of residues within loop structures and the side-chains themselves can undergo conformational changes, a single experimental or simulation structure may not provide a full account of the interactions between residues. Several conformational states should be examined for relevance to the enzyme mechanism. This 2-ns MD simulation is too short in length to definitively determine whether or not we have sampled all the relevant conformations. Still, the consistency between the results of the simulations suggests that some experimentally relevant conformations have been identified.

A charged glutamate side-chain buried in the active site should participate in stabilizing intramolecular interactions or intermolecular interactions with water molecules or the NAD cofactor. The MD simulation of the holo form ALDH3 protein reveals that Glu209 has one carboxylate oxygen atom in hydrogen bonding contact with the C-terminal end of a segment forming a characteristic loop

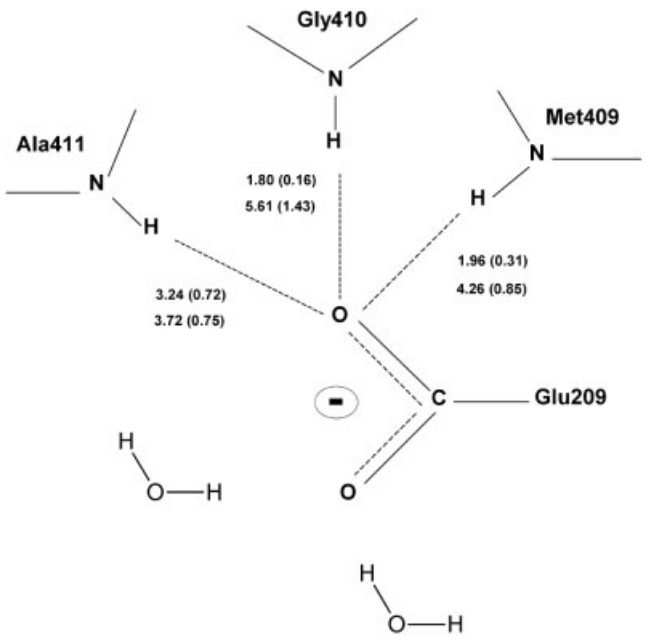

Fig. 5. Hydrogen-bonding interactions between one carboxylate oxygen atom of Glu209 and protein backbone amide protons of Met409-Ala411. The average distances and rms fluctuations in parentheses are shown for the A subunit (**top**) and for the B subunit (**bottom**).

structure in all ALDHs (residues Met409-Gly410-Ala411, the U-turn[15]). This hydrogen bonding with the amide backbone hydrogens is shown in Figure 5 with average distances and fluctuations of these values. In the A subunit, the hydrogen bond between the Ala411 amide proton and the carboxylate oxygen of Glu209 is broken at approximataely 500 ps and remains broken for the rest of the simulation. In the B subunit, these hydrogen-bonding interactions are also present but at approximately 400 ps a water molecule wedges in between these hydrogen-bonded partners and thus they are separated a bit more. The other carboxylate oxygen atom of Glu209 is solvated by approximately 5 water molecules and is directed toward the nucleophilic Cys243. The average distance of this atom from the sulfur of Cys243 is 11.4 and 9.3 Å in subunit A and B, respectively. Distances ranging from 6.9 to 7.0 Å are observed in the B subunit throughout the trajectory. These distances in the crystal structure are 7.6 and 8.1 Å, respectively. Furthermore, most of these also have the thiol proton pointing in the direction of Glu209 forming a hydrogen bond bridge with Glu209 through a water molecule. Other nearby water molecules help fill the active-site funnel.

The only significant direct interaction of Glu333 side-chain with the rest of the system comes from one carboxylate oxygen atom with a hydroxyl group on the NAD ribose as shown in Figure 6. This interaction is formed near the 500-ps mark and remains for the rest of the simulation with an average distance of 1.8 Å. The crystal structure shows this carboxylate oxygen bifurcating the two ribose hydroxyl groups with distances of approximately 5.5 Å for both subunits. This binding mode is observed in the unoccupied subunit of the ALDH3 dimer for some sub-

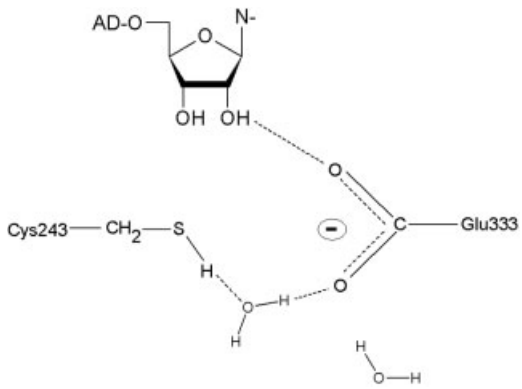

Fig. 6. Hydrogen bonding interactions between Glu333 to ribose hydroxyl group of NAD and to Cys243 through an intervening water molecule.

strate-bound simulations. The distance from the other carboxylate oxygen atom of Glu333 to the sulfur of Cys243 averages 7.3 and 8.6 Å for subunit A and B, respectively, compared with the crystal structure distance of 7.1 Å. Both of these average distances are shorter than the averages for Glu209. Furthermore, the hydrogen-bonded bridge from Cys243 to Glu333 running through a water molecule is also seen several times in the trajectory. As with the holo form crystal structure solved at medium resolution, it is difficult to determine the identity of the general base from the holo form simulation.

MM Simulations of Substrate Bound Forms of ALDH3

Thiol forms of Cys243

Benzaldehyde (BA) can be docked into the active site in two ways such that the carbonyl oxygen of BA would interact with the highly conserved Asn114 while the sulfur of Cys243 is oriented to attack the carbonyl carbon of BA. One of these orientations would lead to a thiohemiacetal intermediate in the R-configuration with the other in the S-configuration (see Fig. 3). The R-configuration binds in the active site forming structures located along the BDat and more NACs (4.0 vs. 0.0%) than the S-configuration as shown in Figure 7. Yet this ensemble of structures shows a less favorable MM-GBMV binding energy by 3.1 kcal/mol (see Table I). The distributions are also very different with the R-configuration average much closer to the NAC region than the S-configuration. Interestingly, these distributions do not change throughout the simulation and conversions between the R- and S-configuration do not occur. Both simulations show fluctuations where the carbonyl carbon of BA is 6–7 Å from the sulfur of Cys243 indicating that the substrate is not locked into the active site. These results suggest that either BA configuration could be present in the active site and that some non-negligible barrier exists preventing conversion of one configuration to the other. The S-configuration places the carbonyl *oxygen* atom of the substrate in close contact with Cys243. This configuration is held in the active site by enhanced hydrogen bonding from the thiol of Cys243 compared to the R-configuration where this interaction is minimal and sporadic. Therefore, despite favorable binding characteristics this hydrogen bonding may prevent Cys243 from being activated by a general base. The backbone amide proton of Cys243 also interacts strongly with the substrate in the S-configuration. A recent crystal structure of ALDH2 shows crotonaldehyde in the active site with the nucleophilic Cys302 aligned to form a thiohemiacetal in the R-configuration.[38] Finally, BA in the R-configuration places the hydride of BA closer to the C-4 of NAD than in the S-configuration. Whether this closer placement is favorable for hydride transfer, considering the dynamic movements of the NAD cofactor, will be investigated in our future studies.

The Michaelis complex from the MM simulation is drawn in Figure 8. This complex shows that the ALDH3 active site forms a hydrophobic region into which the benzene moiety of BA fits and a polar region made up of hydrogen bonds to the aldehyde moiety. The hydrophobic pocket consists of Val244, Leu361, and Ile394 below the BA and Tyr115 above. There is likely a π-stacking interaction between Tyr115 and BA as well, although this cannot be considered to be a determinant of the aromatic substrate specificity of ALDH3 since this residue type (Tyr and Phe) is highly conserved throughout the family. These calculations suggest that there would be a significant entropic effect to substrate binding since this binding would displace some water molecules occupying the hydrophobic pocket. Hydrogen bonds to the carbonyl oxygen of BA are primarily donated by the side-chain of Asn114 and the backbone amide of Cys243. Energy analysis of these interactions shows the backbone amide on average hydrogen bonding more to the substrate than Asn114. This is due to the fact that during the 400–800-ps section of the trajectory, the Asn114 hydrogen bond is lost before returning during the latter part of the simulation. In fact, there are 190 NACs during the last 85 ps of this simulation out of a total 369 produced during the 1-ns production simulation. Therefore, 22% of the conformations (860 analyzed) were NACs in the latter part of the simulation compared to 2% (8,340 analyzed) in the first 1 ns of the simulation. Yet during this time, the running average of the MM-GBMV-binding energy decreases slightly from −9.9 kcal/mol to −9.6 kcal/mol. Clearly, there is some enhancement of NACs that coincides with the greater contribution of Asn114 side-chain hydrogen bonding to BA but it does not correlate with more favorable MM-GBMV-binding energies. The thiol proton of Cys243 hydrogen bonds to the substrate only in a few instances, and even then is not as energetically favorable as the other two donors. These hydrogen-bond donors make up what is commonly called the oxyanion hole since they are positioned to stabilize a developing charge on the carbonyl oxygen atom of the thiohemiacetal intermediate.[52]

Thiolate forms of Cys243

It is commonly assumed in several enzyme systems containing cysteine nucleophiles that the protein environment stabilizes the thiolate form of this nucleophile. This

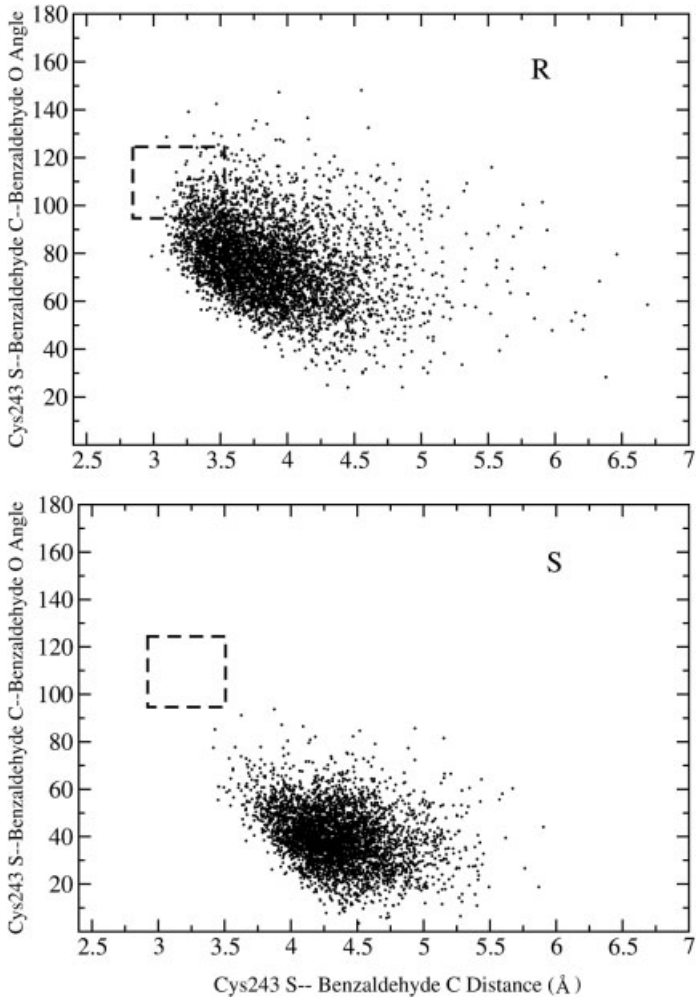

Fig. 7. Conformations of benzaldehyde relative to the Cys243 nucleophile. Each point represents snapshots taken from the MD simulation with the substrate in the respective "R" (**top**) or "S" (**bottom**) orientation. The box represents near-attack conformers. See Methods for a detailed description.

TABLE I. Substrate Binding Energies Calculated With the MM-GBMV Analytical Method

System	$\Delta G_{MM\text{-}GBMV}$ (kcal/mol)
Cys243SH-S configuration	−12.7
Cys243SH-R configuration	−9.6
Cys243S-	−14.5

thiolate then binds the substrate and reacts with it. In order to determine if a thiolate form of Cys243 could bind the BA substrate and since the substrate-bound simulations and experiments gave some support for Glu333 as the general base (see below), we undertook two additional simulations.

The first one, designated S-1-1 (see Fig. 4), was performed to determine if a Cys243 thiolate, stabilized by the protein environment, would bind the substrate. This simulation has several structures located along the BDat but only produces 4 NACs out of the 4,200 analyzed and these occur during the first 426 ps of simulation (see Fig. 9). Yet, the substrate position and orientation remain throughout the simulation, suggesting an energetically favorable interaction. In fact, the MM-GBMV-binding energies are more favorable for this complex than the other Cys243 protonated forms. Again, formation of NACs does not correlate with increased binding energies. The average Cys243 sulfur to carbonyl carbon of BA has increased from 3.86 Å in the protonated "R" simulation to 4.88 Å in this simulation. This orientation is different from the Michaelis complex shown in Figure 8 in the following ways: (1) the sulfur of Cys243 is always oriented away from the BA substrate and (2) the hydrogen bonds making up the oxyanion hole are from the backbone amide protons of Cys243 and Val244; thus, the hydrogen bond from the side-chain of Asn114 has been replaced. The thiolate anion of Cys243 interacts with surrounding water molecules and the backbone amide proton of Gly212, one of two "hinge region" glycines connecting the coenzyme and substrate-binding domains. In this orientation, it appears that the active site can close in around the hydrophobic benzene moiety of BA much better than in Cys243 protonated forms since water radial distribution functions (rdf) are

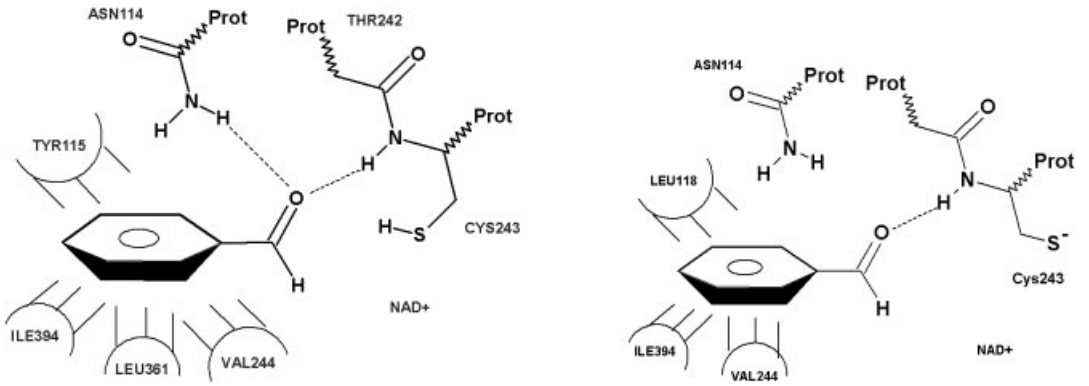

Fig. 8. Schematic representation of Michaelis complex showing both hydrophobic and polar interactions that contribute to substrate binding for thiol form of Cys243 (**left**) and thiolate form of Cys243 (**right**).

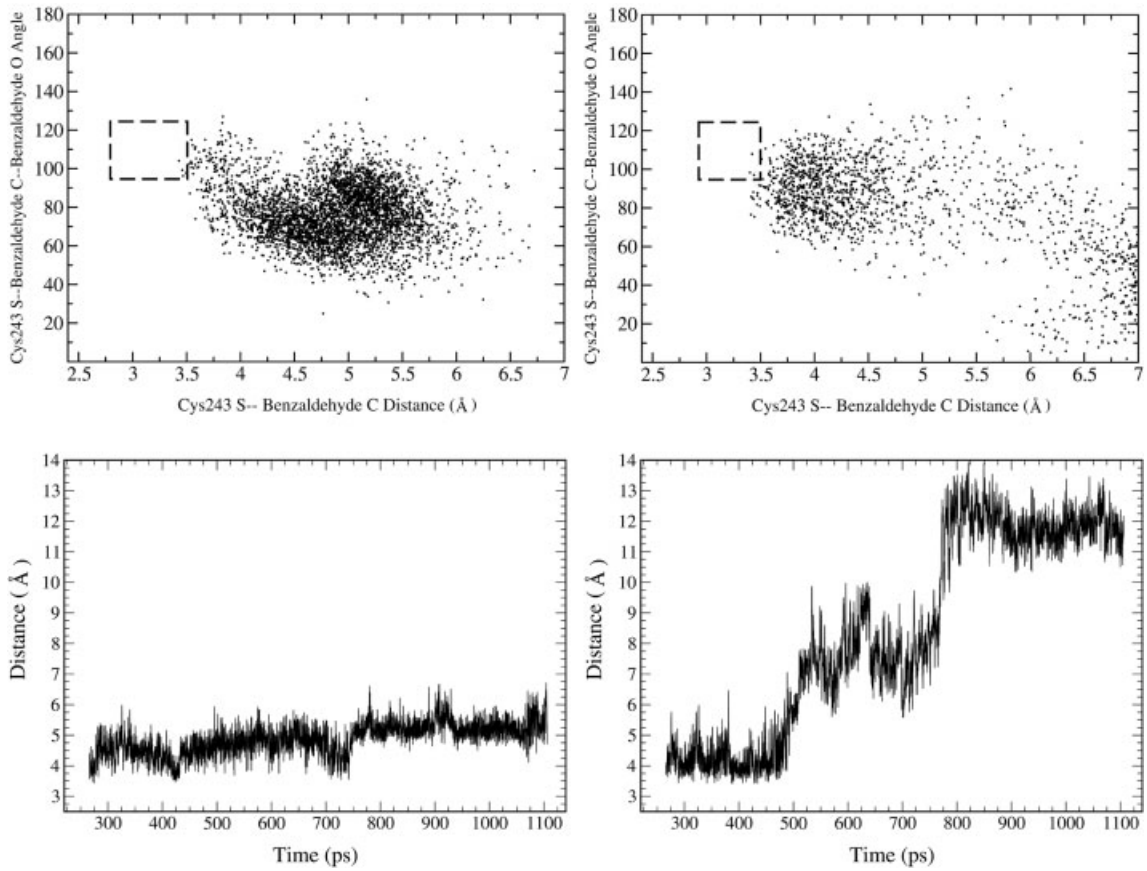

Fig. 9. Conformations of benzaldehyde relative to the Cys243 nucleophile. Each point represents snapshots taken from the S-1-1 (**top left**) and S-1-0 (**top right**) MD simulation with the substrate. The box represents near-attack conformers. Distance plots between the sulfur of Cys243 and the carbonyl carbon of BA are shown for the S-1-1 (**bottom left**) and S-1-0 (**bottom right**) simulation.

decreased in this form (see Supplementary Material for rdf graphs). The cysteine side-chain just needs to rotate about its Cα-Cβ bond to bring the thiolate anion closer to the carbonyl carbon to react and form the thiohemiacetal intermediate. Bringing this thiolate within 3.5 Å of the carbonyl carbon of BA costs approximately 5.4 kcal/mol (unpublished results). We are currently examining the energetics of forming the thiohemiacetal intermediate from this conformation.

The second simulation designated S-1-0 (see Fig. 4) represents a state in which Glu333 activates Cys243 prior to substrate binding by transfer of a proton. Surprisingly, this resulted in the substrate becoming expelled from the active site during the simulation. Large shifts occur near the 500-ps mark, which suggests that this state is only marginally stable. It is likely that the enzyme required some small collective reorientations before the substrate could exit the active site. The aldehyde moiety of BA

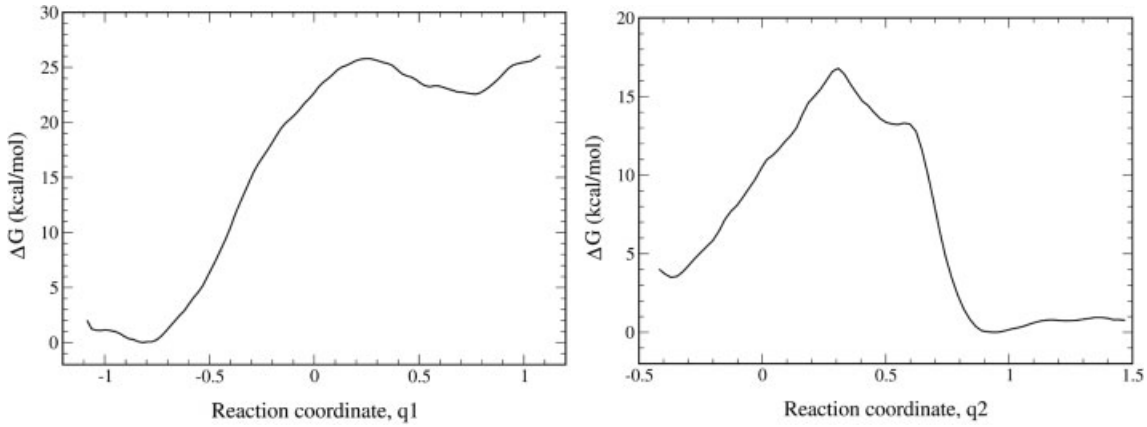

Fig. 10. Free energy profile for proton transfer from water to carboxylate oxygen atom of Glu209 (**left**). Free energy profile for proton transfer from thiol of Cys243 to hydroxide anion (**right**).

actually rotates around to face away from the active site and is solvent separated from Cys243. The few NACs that are formed in the initial part of the simulation show the carbonyl oxygen atom of BA interacting primarily with the amide group of Asn114 and the amide group of NAD. Thus, the aldehyde has left the binding region observed in the other stable substrate-bound forms. The thiolate hydrogen bonds almost exclusively with the protonated Glu333 throughout the entire trajectory with some interaction with the face of the NAD ring. The position of the thiolate and its interaction partners appear crucial to binding the substrate if indeed Cys243 does bind BA as a thiolate. It is interesting to note that the E333Q mutant causes serious stability problems with ALDH3.[5] Perhaps it is important for this position to have two hydrogen bond acceptor sites, and if Glu333 is the general base, it is important that it have a mechanism for fast removal of the transferred proton.

Glu209 Versus Glu333 in Substrate-Bound MM Simulation

As shown above in the holo form simulation, Glu209 makes significant interactions with the three backbone amide protons of Met409-Gly410-Ala411 in the substrate-bound simulation (the simulation that produced the most NACs). The hydrogen bond between Glu209 and Ala411 is also broken, this time near the 600-ps mark (plots shown in Supplemental Material). In contrast to the holo form simulation, this hydrogen-bond breaking leads to a slightly greater separation from this loop structure (5–6 Å) and allows the other carboxylate oxygen to move closer to the thiol proton of Cys243. Indeed, during the latter part of the simulation, this distance gets as close as 6.1 Å. Interestingly, the substrate-bound form does not block a possible route for Glu209 to act as a general base as was once proposed.[8] In the unoccupied unit of the dimer, the hydrogen bonds to the backbone loop remain for the entire simulation. The distance from the carboxylate oxygen to Cys243 sulfur averages 10.3 Å and only gets as close as 7.7 Å.

As in the holo form simulation, one carboxylate oxygen atom of Glu333 hydrogen bonds with the hydroxyl group of the NAD ribose in the substrate bound simulation. This interaction is formed at the 650-ps mark (plots shown in Supplemental Material). The other carboxylate oxygen atom makes significant interactions with the thiol proton of Cys243. During the 420–580-ps section, Cys243 and Glu333 are directly hydrogen bonded while at other times they are solvent separated. Some of the larger distances above 7 Å are the result of both solvent separation and the thiol proton not being directly oriented toward Glu333. Several conformations show a strong hydrogen-bond bridge between Glu333 and Cys243 through an intervening water molecule. Many of these conformations also have the substrate positioned for attack by the sulfur of Cys243.

QM/MM Simulations of Nucleophilic Cysteine Activation

While ground-state structures seen in the MM simulations slightly favor Glu333 versus Glu209 as the general base for Cys243 activation, simulations that reveal the energetic differences between the protonated and activated state of Cys243 as well as the barrier between the two are required to distinguish between the two possibilities. To obtain such information, umbrella sampling simulations using a hybrid QM/MM potential were performed for both Glu209 and Glu333 to abstract a proton from a water molecule, thereby forming a hydroxide anion positioned to abstract the thiol proton from Cys243. The results for this process with Glu209 creating a hydroxide anion are shown in Figure 10. The transition state barrier is 25.8 kcal/mol occurring late in the process with the proton from the water molecule mostly on Glu209. The free energy difference between the two states is 22.6 kcal/mol. In order to assess the magnitude of the sampling error, we ran this same simulation starting from a structure obtained from the results of the first simulation, which contained the protonated Glu209 and a hydroxide anion. The transition state barrier was decreased slightly by 0.5 kcal/mol and the free energy difference was decreased by 2.9 kcal/mol. A representative structure for the product is shown in Figure 11 and reveals that the hydroxide anion is stabilized by surrounding water molecules but is still

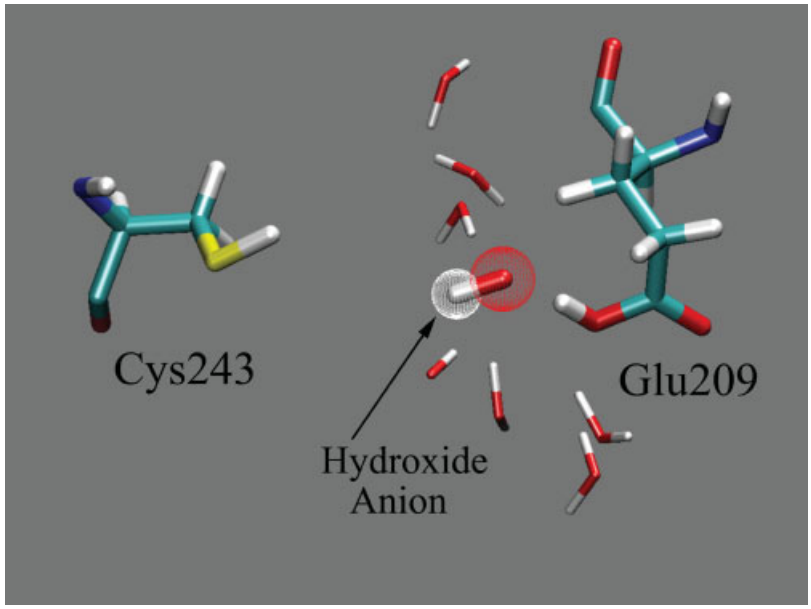

Fig. 11. The product structure of the proton transfer from water to carboxylate oxygen atom of Glu209 showing the stabilizing interactions with the hydroxide anion.

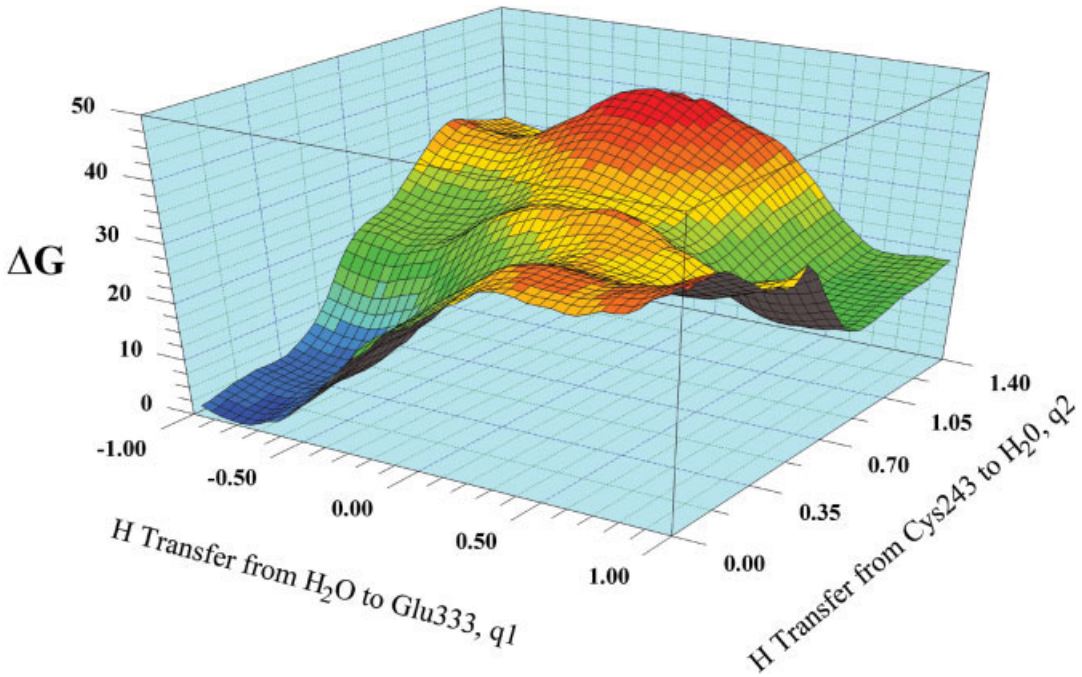

Fig. 12. 2D Free energy surface for transferring a proton from Cys243 to Glu333 through an intervening water molecule. The lowest free energy pathway from reactants (**bottom left**, blue graph) to the product (**top right**, green graph) is along the diagonal of the surface.

separated from Cys243. Thus, to activate Cys243 this hydroxide anion must either diffuse a few Ångströms to abstract the thiol proton or abstract a proton from an intervening water molecule with the resulting hydroxide anion abstracting the thiol proton. Either way, the process requires traversing another energetic barrier due to rearrangement of the water molecules surrounding the hydroxide anion. Indeed, our calculations along the q2 reaction coordinate reveal that the barrier leading to the activated Cys243 and the protonated Glu209 is 13.3 kcal/mol with a free energy difference of −3.6 kcal/mol. Thus, the overall free energy difference ($\Delta G = \Delta G_{q1} + \Delta G_{q2}$) is in the range of 17.1–19.0 kcal/mol. The reverse barrier leading back to charged Glu209 and water from the protonated Glu209

and hydroxide anion is only 3.2–6.1 kcal/mol. Therefore, the intermediate state is more likely to return to reactants than to proceed to the product. It should be noted that in this case, q2 describes both hydroxide anion diffusion along the S-H bond vector of Cys243 and the asymmetric stretch that transfers the proton. This reaction coordinate may not be optimal for describing the small diffusion part of the reaction. While a more descriptive reaction coordinate would lower the energy barrier, the difference from the results obtained here is anticipated to be small.

In contrast, if the same reaction coordinate (q1) is followed with Glu333 abstracting a proton from an intervening water molecule to create a hydroxide anion, it leads directly to an activated Cys243 without any additional steps. This result suggested that another reaction coordinate may be more applicable to this process, one that would go through a concerted mechanism. Therefore, we simulated the 2D reaction profile. The results (see Fig. 12) do indeed show a concerted mechanism as being the lowest energy pathway for activation of Cys243 by Glu333 through an intervening water molecule. The transition state barrier is 29.2 kcal/mol and occurs very close to where both protons are halfway between the donor and acceptor atoms in mass-weighted coordinates. The free energy difference between the two states is 14.7 kcal/mol. The sampling error for this reaction is likely similar to the one calculated for activation of Cys243 by Glu209. The surface also shows a barrier-less downhill path from the hydroxide anion/protonated Glu333 state to the activated Cys243. Along this path, the carbonyl carbon of benzaldehyde to sulfur of Cys243 fluctuates between distances of 2.9 to 4.5 Å and remains along the BDat. Obviously, this 2D free energy profile is dependent on the placement of benzaldehyde from the zone of the reaction, yet the placement of BA is very similar in both simulations.

CONCLUSIONS

We have used several different MM and QM/MM simulations to model the initial stages of the ALDH3 catalytic cycle. Since it was not known a priori how the substrate is bound in the active site, we performed several long MM simulations. These MM simulations allowed the BA substrate to exit the active site if placed in an energetically unfavorable location or protein state. Simulations shorter than about 500 ps would have resulted in different conclusions. Furthermore, due to a large favorable solvation energy of methyl thiolate (−73.7 kcal/mol),[53] we ran the simulations with different protonation states for Cys243. In these, we hypothesized that the thiolate forms in contact with the aldehyde would result in large unfavorable forces and expulsion of the substrate from the active site. Indeed, when modeling Cys243 as a thiolate, the average distance between the Cys243 sulfur atom and the carbonyl carbon of BA was increased by 1.0 Å over the neutral form. Based on this analysis, it would appear that a neutral Cys243 is required to bind BA. Yet, we discovered a stable conformation of BA in the S-1-1 simulation with the thiolate pointing away from the substrate. Furthermore, our MM-GBMV analysis revealed this to be the most favorable binding mode by 4.9 kcal/mol due primarily to more expulsion of water around the benzene moiety of BA and shielding of the thiolate anion by a dihedral rotation. A rotation about the Cα-Cβ bond would bring the thiolate in contact with the carbonyl carbon of BA forming the thiohemiacetal intermediate. These calculations do not reveal a favored protonation state of Cys243 in the presence of BA but they do reveal the relevant configurations one should use in calculations of this property. Such calculations represent an enormous challenge in this field.[54]

Both BA-bound conformations exhibit hydrogen bonding from the backbone amide protons of Cys243 and Val244 and side-chain amide protons of Asn114, though to different extents. The thiol form prefers hydrogen bonding to Asn114 versus Val244 for the thiolate form. Both exhibit a conformation such that if Cys243 were to attack the carbonyl carbon of BA, a thiohemiacetal of the R-configuration would be formed. This assumes that there is not a configuration change upon activation of Cys243 by the general base. Indeed, the QM/MM simulations of this activation process from either Glu209 or Glu333 did not result in any changes in the BA orientation with respect to Cys243. The R-configuration also brings the hydride of the thiohemiacetal intermediate closer to the NAD cofactor than if it were to form the S-configuration. Both simulations contained structures of the electrophilic BA substrate located along the BDat, which suggests that the structure of ALDH3 is pre-organized for the nucleophilic attack. How the structure changes to catalyze other reactions in the catalytic cycle will be of interest to determine. Hydride transfer is the next step in the catalytic cycle after thiohemiacetal formation and has been shown to be stereospecific.[7] Our simulations do not offer any obvious clues as to why hydride transfer is stereospecific. Since this step is rate-limiting in ALDH3,[4] it is possible that some protein reorganization takes place before this step is favorable. Also, the protein reorganization may only take place once the thiohemiacetal intermediate is formed, which was not simulated here.

While the productive substrate-bound MM simulation with Cys243 in thiol form favored Glu333 over Glu209 as the general base due to closer contacts and several more structures that were hydrogen bonded through a water molecule to Cys243, the differences still appeared minor. This led us to speculate that if Cys243 required activation, either the free energy barrier or possibly the free energy difference for the respective activation mechanisms were deciding factors. Therefore, we simulated the Cys243 activation mechanism using a hybrid QM/MM potential to distinguish between the two possibilities. While we cannot strictly compare the energetics of the two reactions since they use slightly different QM mechanical zones, the results show two different pathways for activating Cys243 through an intervening water molecule. Very different quantum regions can be used if one first calibrates reference solution reactions as demonstrated by the Warshel group.[55] The activation of Cys243 by Glu209 occurs stepwise through a hydroxide anion that is stabilized by

several surrounding water molecules. This hydroxide anion must then traverse another significant barrier to abstract the thiol proton. In contrast, a concerted mechanism was found for Glu333 in which, as the proton from the intervening water molecule donates a proton to Glu333, the thiol proton of Cys243 transfers to the water molecule. The barrier for the concerted mechanism did not differ greatly from the stepwise mechanism yet the hydroxide anion was not a minimum on q2 with Glu333.

Since PM3 underestimates the ΔH_f of the hydroxide anion and the hydronium cation, we might expect the stability of structures along the diagonal of the 2D surface to be underestimated to a greater extent than along the step-wise pathways. Furthermore, this transition state structure is very delocalized and may require the use of high level ab initio methods or an extensive reparameterization of PM3 for accurate results. Given these caveats, the fact that Glu333 can activate Cys243 through an intervening water molecule in a concerted manner may be considerably more favorable than a step-wise mechanism and, hence, be crucial to the function of Glu333 as the general base in ALDH3. The results also support the view that small changes in the active site could easily shift the identity of the general base to Glu209. This appears to be the case with ALDH2 where Glu268 (Glu209 in ALDH3) has been identified as the general base.

This study supports a research plan in which MM simulations and subsequent postprocessing of the trajectory through MM-GBMV calculations are used to determine favorable substrate-bound conformations. Some of these structures may be oriented in a manner that would lead to a chemical reaction. These structures can then be further examined by a QM/MM potential to determine their relevance to the enzyme mechanism. Finally, these simulations support the importance of examining different ionization states for active-site cysteine nucleophiles when modeling the substrate bound form.

ACKNOWLEDGMENTS

The calculations were done in part on the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputing Center with an allocation from the National Resource Advisory Committee. A.D.M. gratefully acknowledges support from NIH (GM51501). We also thank Martin J. Field for several useful discussions.

REFERENCES

1. Lindahl R. Aldehyde dehydrogenases and their role in carcinogenesis. Crit Rev Biochem Mol Biol 1992;27:283–335.
2. Sladek NE. Aldehyde dehydrogenase-mediated cellular relative insensitivity to the oxazaphosphorines. Curr Pharm Des 1999;5: 607–625.
3. Liu Z, Sun Y, Rose J, Chung Y, Hsiao C, Chang W, Kuo I, Perozich J, Lindahl R, Hempel J, Wang B-C. The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossman fold. Nat Struct Biol 1997;4:317–326.
4. Hempel J, Kuo I, Perozich J, Wang B-C, Lindahl R, Nicholas H. ALDH: maintaining critical active site geometry at motif 8 in the class 3 enzyme. Eur J Biochem 2001;268:722–726.
5. Mann CJ, Weiner H. Differences in the roles of conserved glutamic acid residues in the active site of human class 3 and class 2 aldehyde dehydrogenases. Prot Sci 1999;8:1922–1929.
6. DeLaurenzi V, Rogers GR, Hamrock DL, Marekov LN, Steinert PM, Compton JG, Marekova N, Rizzo WB. Sjögren-Larsson syndrome is caused by mutations in the fatty acid aldehyde dehydrogenase gene. Nature Genet 1996;12:52–57.
7. Jones KH, Lindahl R, Baker DC, Timkovich R. Hydride transfer stereospecificity of rat liver aldehyde dehydrogenases. J Biol Chem 1987;262:10911–10913.
8. Hempel J, Perozich J, Chapman T, Rose J, Boesch JS, Liu Z-J, Lindahl R, Wang BC. Aldehyde Dehydrogenase catalytic mechanism: a proposal. In: Weiner H, editor. Enzymology and molecular biology of carbonyl metabolism 7. New York: Plenum Press; 1999. p. 53–59.
9. Hempel J, Nicholas H, Lindahl R. Aldehyde dehydrogenases: widespread structural and functional diversity within a shared framework. Prot Sci 1993;2:1890–1900.
10. Dryjanski M, Kosley LL, Pietruszko R. N-Tosyl-L-phenylalanine chloromethyl ketone, a serine protease inhibitor, identifies glutamate 398 at the coenzyme-binding site of human aldehyde dehydrogenase. Evidence for a second "naked anion" at the active site. Biochemistry 1998;37:14151–14156.
11. MacKerell AD Jr, MacWright RS, Pietruszko R. Bromoacetophenone as an affinity reagent for human liver aldehyde dehydrogenase. Biochemistry 1986;25:5182–5189.
12. Abriola DP, Fields R, Stein S, MacKerell AD Jr, Pietruszko R. Active site of human liver aldehyde dehydrogenase. Biochemistry 1987;26:5679–5684.
13. Wang, X, Weiner H. Involvement of glutamate 268 in the active site of human liver mitochondrial (class 2) aldehyde dehydrogenase as probed by site-directed mutagenesis. Biochemistry 1995; 34:237–243.
14. Ni L, Sheikh S, Weiner H. Involvement of Glutamate 399 and Lysine 192 in the mechanism of human liver mitochondrial aldehyde dehydrogenase. J Biol Chem 1997;272:18823–18826.
15. Perozich J, Nicholas H, Wang B, Lindahl R, Hempel J. Relationships within the aldehyde dehydrogenase extended family. Prot Sci 1999;8:137–146.
16. Yang A, Gunner M, Sampogna R, Sharp K, Honig B. On the calculation of pKas in proteins. Proteins 1993;15:252–265.
17. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks III CL. Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comp Chem 2004;25:265–284.
18. Burgi HB, Dunitz JD, Shefter EJ. Geometrical reaction coordinates. II. Nucleophilic addition to a carbonyl group. J Am Chem Soc 1973;95:5065–5067.
19. Burgi HB, Dunitz JD, Shefter E. Chemical reaction paths. IV. Aspects of O—C=O interactions in crystals. Acta Cryst B 1974;30: 1517–1527.
20. Chakrabarti P, Pal D. An electrophile-nucleophile interaction in metalloprotein structures. Prot Sci 1997;6:851–859.
21. Bruice TC, Benkovic SJ. Chemical basis for enzyme catalyis. Biochemistry 2000;39:6267–6274.
22. Shurki A, Štrajbl M , Villá J, Warshel A. How much do enzymes really gain by restraining their reacting fragments? J Am Chem Soc 2002;124:4097–4107.
23. Štrajbl M, Shurki A, Kato M, Warshel A. Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. J Am Chem Soc 2003;125;10228–10237.
24. Warshel A, Levitt M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biol 1976;103:227–249.
25. Field MJ. Simulating enzyme reactions: challenges and perspectives. J Comput Chem 2002;23:48–58.
26. Gao J, Truhlar DG. Quantum mechanical methods for enzyme kinetics. Annu Rev Phys Chem 2002;53:467–505.
27. Warshel A. Computer simulations of enzyme catalysis:methods, progress, and insights. Ann Rev Biophys Biomol Struct 2003;32: 425–443.
28. MacKerell, AD Jr, Banavali NB, Foloppe N. Development and current status of the CHARMM force field for nucleic acids. Biopolymers 2001;56:257–265.
29. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack Jr. RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher III WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical

potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
30. Foloppe N, MacKerell AD Jr. All-atom empirical force field for nucleic acids: 1. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comp Chem 2000;21:86–104.
31. Feller SE, MacKerell AD Jr. An improved empirical potential energy function for molecular simulations of phospholipids. J Phys Chem B 2000;104:7510–7515.
32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
33. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
34. Pavelites JJ, Bash PA, Gao J, MacKerell AD Jr. Molecular mechanics force field for NAD+, NADH, and the pyrophosphate groups of nucleotides. J Comp Chem 1997;18:221–239.
35. Brunger AT, Karplus M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. Proteins 1988;4:148–156.
36. Ryckaert J-P, Cicotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.
37. 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.
38. Perez-Miller SJ, Hurley TD. Coenzyme isomerization is integral to catalysis in aldehyde dehydeogenase. Biochemistry 2003;42: 7100–7109.
39. Cobessi D, Tete-Favier F, Marchal S, Branlant G, Aubry A. Structural and biochemical investigations of the catalytic mechanism of an NADP-dependent aldehyde dehydrogenase from streptococcus mutans. J Mol Biol 2000;300:141–152.
40. Reddy SY, Kahn K, Zheng Y, Bruice TC. Protein engineering of nitrile hydratase activity of papain: molecular dynamics study of a mutant and wild-type enzyme. J Am Chem Soc 2002;124:12979–12990.
41. Swanson JMJ, Henchman RH, McCammon JA. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys J 2004;86: 67–74.
42. Field MJ, Albe M, Bret C, Proust-De Martin F, Thomas A. The DYNAMO library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. J Comput Chem 2000;21:1088–1100
43. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 1996;118: 11225–11236.
44. Stewart JJP. Optimization of parameters for semiempirical methods. I. Methods. J Comp Chem 1989;10:209–220.
45. Gonzalez-Lafront A, Truong TN, Truhlar DG. Direct dynamics calculations with neglect of diatomic differential overlap molecular orbital theory with specific reaction parameters. J Phys Chem 1991;95:4618–4627.
46. Repasky MP, Chandrasekhar J, Jorgensen WL. PDDG/PM3 and PDDG/MNDO: improved semiemperical methods. J Comput Chem 2002;23:1601–1622.
47. Karplus M. Aspects of protein reaction dynamics: deviations from simple behavior. J Phys Chem B 2000;104:11–27.
48. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. J Comp Chem 1992;13:1011–1021.
49. PSI-Plot, Version 7.0. New York: Poly Software International; 2002.
50. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38.
51. Wymore T, Deerfield II DW, Field MJ, Nicholas HB, Hempel J. Initial events in class 3 aldehyde dehydrogenase: MM and QM/MM simulations. Chem Biol Int 2003;143–144:75–84.
52. Kraut DA, Carroll KS, Herschlag D. Challenges in enzyme mechanism and energetics. Ann Rev Biochem 2003;72:517–571.
53. Pliego Jr. J, Riveros JM. Gibbs energy of solvation of organic ions in aqueous and dimethyl sulfoxide solutions. Phys Chem Chem Phys 2002;4:1622–1627.
54. Olsson MHM, Hong G, Warshel A. Frozen density functional free energy simulations of redox proteins: computational studies of the reduction potential of plastocyanin and rusticyanin. J Am Chem Soc 2003;125:5025–5039.
55. Štrajbl M, Florian J, Warshel A. Ab initio evaluation of the free energy surfaces for the general base/acid catalyzed thiolysis of formamide and the hydrolysis of methyl thiolformate: a reference solution reaction for studies of cysteine proteases. J Phys Chem B 2001;105:4471–4484.