CO_2 Binding to Human Carbonic Anhydrase II

Kenneth M. Merz, Jr.

Contribution from the Department of Chemistry and the Department of Molecular and Cell Biology, The Pennsylvania State University, University Park, Pennsylvania 16802.
Received February 15, 1990

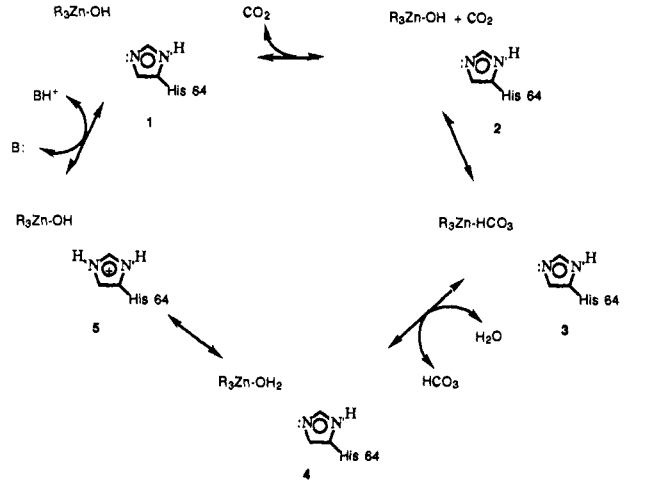
Abstract: We report molecular dynamics and free energy perturbation simulations aimed at studying the binding of CO_2 with human carbonic anhydrase II (HCAII). We first describe a novel approach by which metal ions can be realistically incorporated into molecular mechanical force fields, by simply using electrostatic potential derived point charges. These charges better represent the active site environment and eliminate the large electrostatic interactions that result from the use of formal charges (e.g. +2 for zinc) for metal ions. From our molecular dynamics simulations we have identified two CO_2 binding sites. The first binding pocket is about 3-4 Å away from the zinc ion (Zn···C distance) and is surrounded by Val 121,143,207, Trp 209, Leu 198, His 119, Thr 199, and the residue consisting of Zn-OH. The second pocket is about 5-6 Å away from zinc and is formed by His 64,94,96, Ala 65, Asn 244, Tyr 7, Phe 93, and Thr 200. Further analysis of the MD trajectories yields information regarding the fluctuations of a CO_2 molecule bound in these two sites and the conformational flexibility of His 64. Using free energy perturbation techniques we have also evaluated the absolute free energies of binding for these two sites. The pocket closer to the zinc ion has a higher affinity for CO_2 than does the more remote site. The calculated binding energies are -3.37 ± 1.14 and -2.17 ± 0.83 kcal/mol, respectively. Our computed free energies of binding are in reasonable agreement with the experimental value (-2.2 kcal/mol), but importantly our computed values indicate that the closer of the two pockets has a higher affinity for CO_2. These observations as well as the proximity of CO_2 to the functionally important zinc ion suggest that the closer site is catalytically important, while the second site serves to increase the surface area by which a CO_2 molecule is recognized and bound by HCAII. By taking advantage of both molecular dynamics and free energy perturbation techniques new insights into the dynamics and energies of protein-substrate interactions can be readily obtained.

Introduction

The precise way in which a substrate molecule interacts and then binds to its protein host is very difficult to characterize, but it is crucial to have this knowledge in order for us to extend our understanding of protein structure and function.[1] Kinetics has given us free energies, in the form of binding constants, for the association of an inhibitor or substrate molecule with a protein.[1] This allows us to build up an "energetic database" which provides us with information that correlates the structure (or a structural feature) of an inhibitor or substrate molecule with binding efficacy. This is turn gives us information regarding the disposition of residues in the active site of an enzyme. X-ray crystallography has provided us with numerous insights into protein-ligand interactions by providing us with information regarding the spatial interaction of an inhibitor or a substrate molecule with its protein host.[2] Modern NMR techniques are providing us with this same sort of information.[3] Furthermore, modern theoretical techniques, such as molecular dynamics (MD)[4] and MD free energy perturbation (MDFEP) simulations[4,5] have started to provide us with both structural and energetic information regarding protein-substrate interactions. Herein, we present results from MD and MDFEP simulations that probe the noncovalent interaction of the substrate molecule CO_2 with its protein host, human carbonic anhydrase II (HCAII). The combined use of these two types of simulations allows us to provide both qualitative structural and energetic insights into CO_2/HCAII interactions.

HCAII is a small monomeric protein consisting of 260 residues with a total molecular weight of about 30 kd. The zinc ion in all known carbonic anhydrases is of catalytic importance.[6] The coordination of the zinc ion is to three histidines, His 94,96,119.[6-9] The fourth coordination site is occupied by a water molecule in the low-pH form of the enzyme (<7), while in the high-pH form this site is occupied by hydroxide ion.[6-9] There is a potential fifth

Scheme I



(1) Fersht, A. R. *Enzyme Structure and Mechanism*; W. H. Freeman and Co.: New York, 1985.
(2) Blundell, T. L.; Johnson, L. N. *Protein Crystallography*; Academic Press: New York, 1976.
(3) Braun, W. *Q. Rev. Biophys.* **1987**, *19*, 115. Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
(4) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics. *Advances in Chemical Physics*; John Wiley and Sons: New York, 1988; Vol. LXXI. McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: New York, 1987.
(5) van Gunsteren, W. F. *Protein Eng.* **1988**, *2*, 5. Mezei, M.; Beveridge, D. L. *Ann. N.Y. Acad. Sci.* **1986**, *482*, 1. Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184. Kollman, P. A.; Merz, K. M., Jr. *Acc. Chem. Res.* **1990**, *23*, 246.
(6) Merz, K. M., Jr.; Hoffmann, R.; Dewar, M. J. S. *J. Am. Chem. Soc.* **1987**, *111*, 5636. For a selection of reviews see: Silverman, D. N.; Lindskog, S. *Acc. Chem. Res.* **1988**, *21*, 30. Silverman, D. N.; Vincent, S. H. *CRC Crit. Rev. Biochem.* **1983**, *14*, 207. Lipscomb, W. N. *Annu. Rev. Biochem.* **1983**, *52*, 17. Lindskog, S. In *Zinc Enzymes*; Spiro, T. G., Ed.; John Wiley & Sons: New York, 1983. *Metal Ions in Biological Systems*; Sigel, H., Ed.; Marcel Dekker: New York, 1983; Vol. 15; p 77. *Biophysics and Physiology of Carbon Dioxide*; Bauer, C., Gros, G., Bartels, H., Eds.; Springer-Verlag: New York, 1980. Prince, R. H. *Adv. Inorg. Chem. Radiochem.* **1979**, *22*, 349. Chlebowski, J. F.; Coleman, J. B. *Metal Ions in Biological Systems*; Sigel, H., Ed., Marcel Dekker: New York, 1976; Vol. 6. Pocker, Y.; Sarkanen, S. *Adv. Enzymol.* **1978**, *47*, 149. Bertini, I.; Luchinat, C.; Scozzafava, A. *Struct. Bonding (Berlin)* **1981**, *48*, 45.
(7) Kannan, K. K.; Ramanadham, M.; Jones, T. A. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 49.
(8) Kannan, K. K.; Liljas, A.; Waara, I.; Bergsten, P.-C.; Lövgren, S.; Strandberg, B.; Bengtsson, U.; Carlborn, U.; Fridborg, K.; Jarup, L.; Petef, M. *Cold Spring Harbor Symp. Quantum Biol.* **1971**, *36*, 221.
(9) Eriksson, E. A.; Jones, T. A.; Liljas, A. In *Zinc Enzymes*; Bertini, I., Luchinat, C., Maret, W., Zeppezauer, M., Eds.; Birkhäuser: Boston, 1986; p 317. Eriksson, A. E.; Jones, A. T.; Liljas, A. *Proteins* **1989**, *4*, 274. Eriksson, A. E.; Kylsten, P. M.; Jones, T. A.; Liljas, A. *Proteins* **1989**, *4*, 283.

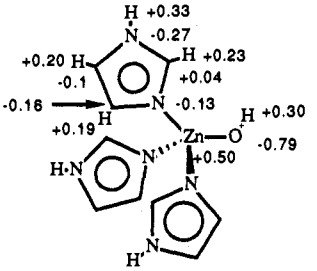
0002-7863/91/1513-406$02.50/0

coordination site that may or may not be occupied depending on the identity of the moiety occupying the fourth coordination site.[6,9] Hydrogen bound to the zinc bound water (or hydroxide) is the hydroxyl oxygen of Thr 199, which is itself hydrogen bound, via its hydroxyl hydrogen, to Glu 106.[6-9] The only known physiological role that this protein has is the conversion of CO_2 into bicarbonate or vice versa.

The catalytic mechanism of HCAII has been studied in detail.[6-9] The following observations were crucial in developing the catalytic cycle given in Scheme I: (1) The catalysis is dependent on a group whose pK_a is around 7.[6-9] After much debate it was decided that a zinc-bound water satisfied this criteria.[6-9] (2) The rate-limiting step for maximal velocity at high buffer concentration is an intramolecular proton transfer between a zinc-bound water and a residue in the active site. The identity of the residue that serves this function is thought to be His 64.[6-9] (3) The intermolecular proton transfer between protonated His 64 and the surrounding milieu depends on the concentration of buffers in the surrounding medium. At low buffer concentrations this step is rate limiting, while at high buffer concentrations this step is not and (2) above is. These observations suggest that buffers play a crucial role in CA activity.[6-9] (4) Finally, the mechanism follows ping-pong kinetics where the hydration of CO_2 (**1** → **4** below) and proton transfer steps (**4** → **1**) are kinetically distinct.[6-9] The first step in the catalytic mechanism is the introduction of CO_2 into the active site of the zinc-hydroxide form of the enzyme (**1** → **2**) followed by the hydration of CO_2 via an innersphere mechanism[6] to form the zinc-bicarbonate intermediate (**2** → **3**). Water then displaces the bicarbonate ion to form the zinc-water form of the enzyme (**3** → **4**), which then undergoes an intramolecular proton transfer to His 64 which forms **5**. This intermediate next transfers its proton to a buffer or water molecule in the surrounding milieu. Numerous other mechanisms have been proposed,[6] but all remaining mechanisms that fit the experimental evidence are variations on this theme, bar one. One mechanism that is significantly different than that given above, but cannot be ruled out, is the proton shuttle mechanism of Kannan et al.[7,8] The zinc ion still has a similar role, but the proton is not shuttled out of the active site via His 64, Instead, this mechanism involves the active site residues Thr 199 and Glu 106 in a proton relay. The one drawback of this mechanism, though, is that it requires the activity linked group whose pK_a is around 7 to be Glu 106. The pK_a of a glutamic acid is normally around 4.5, but highly perturbed pK_a's for glutamic acids have been observed.[1] The present paper will focus on **1** → **2**, which is specifically the recognition and binding of the substrate molecule CO_2 for subsequent conversion into bicarbonate.

Substrate interactions with the HCAII active site are still not well-understood.[10-14] On the basis of NMR studies[10-13] it is thought that binding of bicarbonate occurs at the zinc ion, while on the basis of NMR[10-12] studies the CO_2 binding site has been found anywhere from 3.6 to 9.7 Å (distance for Zn to CO_2 carbon) away from the zinc ion and IR[14] studies were interpreted as indicating that CO_2 is not directly coordinated to the zinc ion. These experiments[10-11,13] suggest that CO_2 is not covalently bound to the zinc ion. The tertiary structure of HCAII with water molecules in its active site has been solved to 2.0-Å resolution.[7-9] The crystallographic results indicate that there are eight water molecules in the HCAII active site, which are ordered. One water molecule was found to be directly bound to the zinc ion, another, which is referred to as the "deep" water, is at the bottom of the active site in a region that is "possibly involved in substrate and inhibitor binding",[9] the rest of the water molecules are hydrogen bonded to various other residues and water molecules present in the active site.[9] The active site region of HCAII is unique in that it is very deep (~15 Å) and wide (~15 Å); furthermore, the active site is divided into a hydrophobic half and a hydrophilic half.[6-9,15]

Scheme II



The goal of the present research was to study CO_2 binding in the zinc–hydroxide (high pH) form of HCAII,[6] which is the form of the enzyme that reacts with CO_2.

Computational Procedure

The MD simulations were carried out with use of the MD module from the AMBER suite of programs. The AMBER[16] potential function was used throughout and its form is given below in eq 1. The first three

$$E_{\text{Total}} = \sum_{\text{bonds}} \frac{K_r}{2}(r - r_{\text{eq}})^2 + \sum_{\text{angles}} \frac{K_\theta}{2}(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \sum_{n} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \epsilon_{ij}\left[\left(\frac{R_{ij}^*}{R_{ij}}\right)^{12} - \left(\frac{R_{ij}^*}{R_{ij}}\right)^{6}\right] \times \frac{1}{\text{VDW}_{\text{scale}}} \sum_{\substack{i<j \\ \text{1-4 terms}}} \epsilon_{ij}\left[\left(\frac{R_{ij}^*}{R_{ij}}\right)^{12} - \left(\frac{R_{ij}^*}{R_{ij}}\right)^{6}\right] + \sum_{\text{H bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right] + \sum_{i<j} \frac{q_i q_j}{\epsilon R_{ij}} + \frac{1}{\text{EE}_{\text{scale}}} \sum_{\substack{i<j \\ \text{1,4 terms}}} \frac{q_i q_j}{\epsilon R_{ij}} \quad (1)$$

terms represent the "bonded" interactions present in a molecule, namely the bond, angle, and torsion interactions. The bond and angle interactions are represented by a quadratic potential, while the torsional interactions are represented by a truncated Fourier series. K_X (where $X = r, \theta$) is the force constant for the bond or angle, while X_{eq} is the experimentally observed equilibrium bond length or angle associated with force constant K_X. X is the calculated value for the bond or angle. V_n, n, ϕ, and γ represent the torsional barrier, the periodicity, the calculated dihedral angle, and, finally, the phase. The next five terms represent the "nonbonded" interactions in a molecule and they are the Lennard–Jones (the $R_{ij}^{-12} - R_{ij}^{-6}$ or 6–12 terms), the hydrogen bond (the 10–12 term), and the electrostatic interactions. R_{ij} is the distance between atoms i and j, R_{ij}^* and ϵ_{ij} are parameters that define the shape of the Lennard–Jones potential for the interaction between atoms i and j, C_{ij} and D_{ij} define the shape of the hydrogen bond potential, q_i and q_j are the atomic point charges for atoms i and j, and ϵ is the dielectric constant. $1/\text{VDW}_{\text{scale}}$ and $1/\text{EEL}_{\text{scale}}$ are scaling factors for the 1–4 electrostatic and van der Waals terms. In the next paragraph we discuss how the necessary parameters were determined.

The three histidines around the zinc ion used the all-atom representation, while the rest of the protein used the united-atom representation.[16] The zinc ion was incorporated into the force field by first carrying out MNDO[17] ESP[18] calculations of $(\text{imidazole})_3\text{Zn–OH}^{+1}$.[17] This resulted in a charge on zinc of +0.5 with the remainder of the net +1 charge being dispersed on the zinc ligands. The point charges used are given in Scheme II. The resulting point charges were then incorporated into the

(10) Led, J. J.; Neesgaard, E.; Johansen, J. T. *FEBS Lett.* **1982**, *147*, 74. Led, J. J.; Neesgaard, E. *Biochemistry* **1987**, *26*, 183.
(11) Stein, P. J.; Merrill, S. P.; Henkens, R. W. *J. Am. Chem. Soc.* **1977**, *99*, 3194.
(12) Bertini, I.; Borghi, E.; Luchinat, C. *J. Am. Chem. Soc.* **1979**, *101*, 7069. Bertini, I.; Canti, G.; Luchinat, C.; Borghi, E. *J. Inorg. Biochem.* **1983**, *18*, 221. Also see ref 32.
(13) Yeagle, P. L.; LochMüller, C. H.; Henkens, R. W. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 454.
(14) Riepe, M. E.; Wang, J. H. *J. Biol. Chem.* **1968**, *243*, 2779.
(15) Hansch, C.; McClarin, J.; Klein, T.; Langridge, R. *Mol. Pharmacol.* **1985**, *27*, 493. Hydrophobic half: Ile 91, Val 121, Phe 131, Leu 141, Val 143, Gly 145, Leu 198, Pro 201, Pro 202, Val 207, Trp 209, Val 211. Hydrophilic half: Tyr 7, Asn 61, His 64, Asn 67, Glu 69, Gln 92, His 94, His 96, Glu 106, His 119, Thr 199, Thr 200.
(16) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Chem. Soc.* **1986**, *7*, 230.
(17) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899. For the MNDO zinc parameters see: Dewar, M. J. S.; Merz, K. M., Jr. *Organometallics* **1986**, *5*, 1494.
(18) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comp. Chem.* **1990**, *11*, 431.

Table I. Parameters Derived for This Study[a]

Nonbonded Parameters		
atom type	R^*	ϵ
Zn	1.1	0.0125
OZ	1.768	0.152
HZ	1.0	0.0
CO	1.4	0.18
OC	1.9	0.13

Bond Parameters		
bond type	r_{eq}, Å	K_r, kcal/(mol Å^2)
Zn–NB	2.05 Å	40.0
Zn–OZ	1.80	94.0
OZ–HZ	0.96	553.0
CO–OC	1.162	1116.0

Angle Parameters		
angle type	θ_{eq}, deg	$K\theta$, kcal/(mol rad^2)
NB–Zn–NB	109.5	20.0
CR–NB–Zn	126.0	20.0
CC–NB–Zn	126.0	20.0
CV–NB–Zn	126.0	20.0
Zn–OZ–HZ	126.0	100.0
NB–Zn–OZ	109.5	20.0
OC–CO–OC	180.0	41.0

Torsion Parameters		
torsion type	V_n, kcal/mol	γ, deg
X–Zn–OZ–X	0.0	0.0
X–Zn–NB–X	0.0	0.0

[a] The atoms labels are as follows: Zn, zinc ion; OZ, oxygen bound to zinc; HZ, hydrogen bond to OZ; NB, nitrogen atom of a histidine bound to zinc; CV, CR, CC, carbon atoms in the five-membered histidine ring; OC, carbon dioxide oxygen; CO, carbon dioxide carbon.

AMBER force field. This model of the point charges for the zinc ion and its associated ligands works significantly better than models which treat the zinc ion and the hydroxide ion as having net charges of +2.0 and −1.0, respectively. Short MD simulations (18 psc) using our charge model and the latter charge model (in conjunction with the parameters listed in Table I) demonstrated that the former charge model retained the structure of the active site, while the latter predicted that the coordination of the zinc ion was octahedral and not tetrahedral as has been determined experimentally.[6-9] Thus, our approach to the incorporation of metal ions is capable of reproducing the structure of the active site in HCAII, while a model in which charges are assigned by using a more formalistic approach are not. The zinc Lennard-Jones parameters were obtained by varying them until $Zn^{2+}\cdots CO_2$ and $Zn^{2+}\cdots H_2O$ interaction energies and geometries computed by using ab initio techniques were reproduced.[19] The Lennard-Jones parameters for the hydroxide ion bound to zinc were taken directly from the T1P3P water model.[20] The bond and angle parameters between the zinc ion and its ligands were obtained from experimental force constants and from MNDO geometry optimizations of the $(imidazole)_3Zn–OH^+$ fragment.[21] For example, the zinc to nitrogen bond force constant in $Zn(NH_3)_4^{2+}$ has been determined to be 0.54 mdyn/Å which corresponds to about 40 kcal/(mol Å^2), which is the value we used in our force field. For the N–Zn–N force constant the experimental value is 0.018 mdyn/Å which corresponds to 5 kcal/(mol rad^2); however, we used a slightly larger value (20 kcal/(mol rad^2)) in our force field in order to bring it into line with the values used to successfully reproduce the structure of other transition-metal complexes using molecular mechanics.[21b] The barrier height for the torsions around the Zn–O and Zn–N bonds were all chosen to be 0.0 kcal/mol as has been used in previous work on transition-metal complexes.[21b] The CO_2 bond and angle parameters were taken from the literature,[22] and the point charges were taken from 6-31G*[23] ESP calculations (q_c = 0.908, q_o = −0.454).[24] The Lennard-Jones parameters were chosen such that they reproduced the solvation free energy of CO_2[25] and 6-31G* CO_2–water interaction energies. The parameters derived for the present work are given in Table I.

In order to identify CO_2 binding sites in HCAII using MD simulations we need to generate guess starting structures. These were generated by model building CO_2 into locations that were less than 10 Å away from the zinc ion. In order to cover this region as best as possible we generated ten starting configurations that covered the wide range of environments presents within 10 Å of the zinc ion.[15] The structures were generated in the following way: For the first structure we placed the CO_2 near a pocket that we identified from van der Waals surfaces overlaid on the surface of the HCAII active site. This pocket consisting of His 64,94,96, Ala 65, Asn 244, Tyr 7, Phe 93, and Thr 200 and was later found to be a CO_2 binding site. From this site we model built four more sites by simply moving the CO_2 molecule around the surface of the active site in a circular fashion. In this way we "ringed" the active site at a distance of about 10 Å from the zinc ion. The next starting structure placed a CO_2 molecule near the zinc ion (about 2.0 Å) and in between His 119 and His 94, which are two of the zinc-bound histidines. Moving the CO_2 molecule slightly put it into the "deep-water" pocket (Val 121,143,207, Trp 209, Leu 198, His 119, Thr 199 and the residue consisting of Zn–OH), which generated the next structure. Moving along the "lip" (Val 121, 143, and 207) of the deep-water pocket toward the exit of the active site we generated the eighth starting structure. The final two guess structures were generated by placing the CO_2 molecules in sites that were in the center of the active site. While ten starting sites might not be a "statistical" representation of the surface of the active site we expect that we are sampling enough of the region of interest to have confidence in the binding sites we might observe. Furthermore, we have carried out MD simulations of 9 ps on each structure, as described below, and we expect that a total of 90 ps of sampling around the zinc ion in the active site will sample a large, statistically significant, portion of this region. An alternative method by which small molecule binding sites can be located in enzymes has been described by Elber and Karplus,[26] but we have not used that approach here.

The starting structures for the MD simulations were first minimized fully to remove any bad intermolecular contacts. These structures were then solvated with a 22-Å sphere of T1P3P[20] water molecules, which was centered at the zinc ion in the HCAII active site. Any water molecules that came within 2.4 Å of a protein atom were removed. This resulted in about 300 water molecules solvating the active site of the protein. The water molecules were kept in this sphere with the use of harmonic restraining forces (0.5 kcal/(mol Å)), which are applied to any water molecule that strays out of the sphere. The simulations were kept at 298 K by coupling to a temperature bath.[27] All residues within 15 Å of the zinc ion were permitted to move during the course of the simulation as were all water molecules, while residues lying outside this region were held fixed. SHAKE[28] was used to constrain bond lengths at their equilibrium value and a timestep of 1.5 fs was employed. The nonbonded pairlist had a cutoff of 10 Å, was updated every 50 timesteps, and a constant dielectric of one was used throughout. The equilibration period was 9 ps. After this period of time the final structure was examined graphically to determine whether or not the simulation should be continued. Examination of all ten structures that were generated after 9 ps of simulation time clearly indicated that there were only two regions that one could describe as CO_2 binding sites. These were the deep-water site (eight of the initial coordinates led to this site) and the pocket consisting of His 64,94,96, Ala 65, Asn 244, Tyr 7, Phe 93, and Thr 200 (two sets of initial coordinates gave this site). No other unique binding pockets were located with this procedure. Two of the initial simulations were then carried out for a total of 126 pc each and these were started from the first set of coordinates we generated (as described in the paragraph above) and the structure in which the CO_2 was initially bound in the deep-water pocket. Since all other simulations had resulted in situations where CO_2

(19) Clementi, E.; Corongiu, G.; Jönsson, B.; Romano, S. *J. Phys. Chem.* **1980**, *72*, 260. Clementi, E.; Corongiu, G.; Jönsson, B.; Romano, S. *FEBS Lett.* **1979**, *100*, 313. Clementi, E.; Corongiu, G.; Jönsson, B.; Romano, S. *Gazz. Chim. Ital.* **1979**, *109*, 669. Jönsson, B. In *Catalysis in Chemistry and Biochemistry. Theory and Experiment. 12th Jerusalem Symposium*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1979; p 67.

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(21) (a) Nakamoto, K.; Takemoto, J.; Chow, T. L. *Appl. Spectrosc.* **1971**, *25*, 352. Sacconi, L.; Sabatini, A.; Gans, P. *Inorg. Chem.* **1964**, *3*, 1772. Nakagawa, I.; Shimanouchi, T. *Spectrochim. Acta* **1964**, *20*, 429. Nakamoto, K. *Infrared and Raman Spectra of Inorganic and Coordination Compounds*; Wiley: New York, 1978. Jones, L. H. *Inorganic Vibrational Spectroscopy*; Marcel Dekker: New York, 1971. (b) Hancock, R. D. *Prog. Inorg. Chem.* **1989**, *37*, 187.

(22) Herzberg, G. *Molecular Spectra and Molecular Structure III. Electronic Spectra and Electronic Structure of Polyatomic Molecules*; Van Nostrand: New York, NY, 1966.

(23) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley and Sons: New York, 1986.

(24) Williams, D. E.; Yan, J.-M. *Adv. Atomic Mol. Phys.* **1988**, *23*, 87. Chirlian, L. E.; Franci, M. M. *J. Comput. Chem.* **1987**, *8*, 894. Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(25) Wilhelm, E.; Battino, R.; Wilcock, R. J. *Chem. Rev.* **1977**, *77*, 219.

(26) Elber, R. Personal communication. Elber, R.; Karplus, M. *J. Am. Chem. Soc.* In press. Elber, R.; Karplus, M. *Proteins.* In press.

(27) Berendsen, H. J. C.; Potsma, J. P. M.; van Gunsteren, W. F.; DiNola, A. D.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(28) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.

Table II. ZnOH–CO_2 Distances and a ZnOH–His 64 Distance[a]

	min, Å	max, Å	⟨Q⟩, Å
run 1[b]			
C–OZ	2.52	4.28	3.09 ± 0.88[c]
Zn–C	3.32	5.57	4.15 ± 1.13
Zn–O1	2.82	5.86	3.85 ± 1.52
Zn–O2	3.76	6.20	4.77 ± 1.22
OZ–HND	5.46	11.90	9.21 ± 3.22
run 2[d]			
C–OZ	2.97	6.90	5.55 ± 1.97
Zn–C	3.47	6.76	4.93 ± 1.65
Zn–O1	3.22	7.32	4.99 ± 2.05
Zn–O2	3.32	7.31	5.10 ± 2.0
OZ–HND	8.14	12.78	10.56 ± 2.32

[a] The atom labels are as follows–CO_2: O1, O2, C; R_3Zn–OH: Zn, OZ, HZ; Thr 199, His 64: the hydrogen atom bound to ND1 is HND. [b] Run 1 started from the coordinate set where the CO_2 was in the deep-water pocket and covered 126 ps. [c] The ± values indicate the total range of the fluctuations of these distance (i.e. (max − min)/2). [d] Run 2 started from the first coordinate set described in the simulation protocol section and it covered 126 ps.

was bound in regions very close to either of these two pockets we did not extend these simulations further because we felt that they would give redundant information. Every 100 timesteps energies and coordinates were saved for later study. Every 18 ps a coordinate and velocity set was saved for later use as a starting structure for our free energy perturbation simulations.

The simulation protocol used for the MDFEP simulations on the enzyme was identical with that used for the MD simulations. The free energy evaluations were done using the slow growth procedure.[5] The simulations were started from a total of six different starting coordinate sets. For the case where the CO_2 is bound close to the zinc ion (4 Å) three coordinate and velocity sets from the MD simulation were used. These sets were saved after 90, 108, and 126 ps of total MD simulation time (times do not include 9 ps equilibration time). For the second pocket the same approach was adopted. Each of the MDFEP simulations were only done in the forward direction (disappearing CO_2) and covered a total of 45 ps each. The reason for carrying out the simulations only in the forward direction was due to the observation that at the last step when CO_2 was turned into a set of three "dummy atoms" it tended to become embedded in the protein matrix, which when it reappeared resulted in free energies of binding that were not indicative of the binding pockets that we were studying. In our opinion running the simulation from widely separated starting structures is the best way to proceed in this case. The determination of the free energy of solvation for CO_2 was done in an analogous fashion. Constant temperature (298 K) and pressure (1 atm) conditions were employed with periodic boundary conditions.[27] SHAKE[28] was used in conjunction with a timestep of 1.5 fs. The nonbonded pairlist had a cutoff of 8 Å and was updated every 50 timesteps, and a constant dielectric of one was used throughout. The equilibration period was 27 ps. A MD simulation of 36 ps more was carried out to generate two more sets of coordinates and velocities, which were used as starting points for our MDFEP simulations. Thus, three separate MDFEP simulations were carried out for the determination of the free energy of solvation for CO_2. The three separate simulations started from the equilibrated structure and the 18- and 36-ps structures. The simulation time period was 90 ps, and the simulations were carried out in the forward ($\lambda = 1 \rightarrow 0$) and backwards ($\lambda = 0 \rightarrow 1$) directions for a total of 180 ps of simulation time.

Results and Discussion

In order to identify the binding loci we made use of MD simulations. We started our MD simulations by placing the CO_2 molecule in a number of different positions within the HCAII active site (see computational procedure section). Within the equilibration period (9 ps) we found that only two sites were ever occupied. Thus, only two MD simulations were carried out for 126 ps. These two MD simulations indicate that there are two binding sites for CO_2 near the zinc ion, each of which have different affinities for CO_2 (see below). The first binding pocket is about 4 Å away from the zinc ion (Zn···C) and is surrounded by Val 121,143,207, Trp 209, Leu 198, His 119, Thr 199, and the residue consisting of Zn–OH. This pocket is surrounded by hydrophobic residues, and from the crystallographic studies it has been identified as the deep-water site.[9] The second pocket is about 5 Å away from zinc and is formed by His 64,94,96, Ala 65, Asn

Table III. Further Analysis of the Zn–C Distance[a]

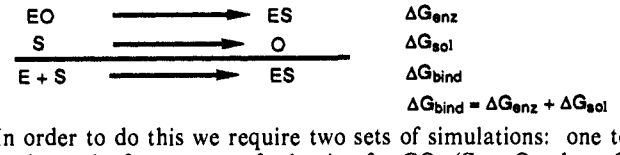
	Zn–C distance		
	min, Å	max, Å	⟨Q⟩, Å
run 1[b]			
0–18 ps	3.45	4.84	4.17
18–36 ps	3.49	5.57	4.14
36–54 ps	3.44	5.11	4.04
54–72 ps	3.41	5.48	4.20
72–90 ps	3.32	5.30	4.04
90–108 ps	3.37	5.27	4.16
108–126 ps	3.43	5.47	4.27
run 2[c]			
0–1 ps	4.55	6.43	5.29
18–36 ps	3.90	6.66	5.07
36–54 ps	3.47	5.95	4.53
54–72 ps	3.91	6.14	4.89
72–90 ps	3.76	6.30	4.73
90–108 ps	3.66	6.76	4.91
108–126 ps	4.12	6.40	5.08

[a] The atom labels are as follows: CO_2: C; Zn–OH: Zn. [b] Run 1 started from the coordinate set where the CO_2 was in the deep-water pocket and covered 126 ps. [c] Run 2 started from the first coordinate set described in the computational procedure section and it covered 126 ps.

244, Tyr 7, Phe 93, and Thr 200. This pocket is about 5 Å at the mouth, is 5 Å deep, and is surrounded by mostly hydrophilic residues, in contrast to the previous pocket. We find that the CO_2 molecules stay in these locations for the whole course of our simulations.

Table II gives the average, maximum, and minimum Zn–C, Zn–O1, Zn–OZ, and Zn–O2 distances for both of the simulations. The nonbonded distances from run 1 indicate that the CO_2 molecule is more tightly constrained than the CO_2 of run 2. Interestingly, nonbonded distances involving the CO_2 carbon have smaller fluctuations than do the distances involving O1 and O2, which indicates, perhaps, that the motion parallel to the O–C–O axis is less constrained than is its motion perpendicular to this axis. In Table III we break down the fluctuations in the Zn–C nonbonded distance into 18-ps sections. These numbers indicate that the fluctuation in this value for run 1 is relatively constant in each of the 18-ps windows, while for run 2 there is significantly more variability in the average value in each 18-ps time period. This further indicates that CO_2 is more constrained in the deep-water pocket than it is in the more remote site. In conclusion, these numbers along with visual inspection of the trajectories (MD movies) indicate that the CO_2 molecule remains in these regions of the active site throughout the MD simulation. This is understandable because CO_2 is a nonpolar molecule (ΔG_{sol}(exp) = 0.0 kcal/mol at 298 K[25]) and would prefer to interact with a less polar environment than water. Another interesting feature of these simulations is the motion of the catalytically important residue His 64. The nonbonded distance between OZ and HND of His 64 is given in Table II. For both simulations His 64 undergoes significant fluctuations, but in run I His 64 undergoes a rotational transition after about 90 ps of simulation time, while no such transition occurs in run 2 possibly due to the CO_2 bound in the second site interfering with this process. While the observation of one such transition can hardly be deemed statistically significant, it does indicate that this sort of motion is not substantially hindered at 298 K, which is in agreement with the assessment that this group is involved in shuttling a proton from the active site to the outside environment.[6]

We have evaluated the absolute free energy of CO_2 binding with the aid of the following relationship,[29] which can alternatively be derived from thermodynamic cycles:[29]



In order to do this we require two sets of simulations: one to evaluate the free energy of solvation for CO_2 (S → O, where O

Table IV. Computed Free Energies for E-CO_2 → E·O[a]

run no.	90 ps	108 ps	126 ps	ΔG_{enz}[b]
1	−4.79, 4.81[c]	−2.93, 2.94	−3.98, 3.99	−3.91 ± 0.83
2	−3.05, 3.05	−2.68, 2.62	−2.44, 2.40	−2.71 ± 0.29

[a] Coordinate and velocity sets from runs 1 and 2 were saved at 18-ps intervals past the 9-ps equilibration period. Run 1 corresponds to CO_2 binding to the more hydrophobic deep-water pocket near the zinc ion, whereas run 2 corresponds to the CO_2 binding to the more hydrophilic pocket. For the free energy evaluations we used the 90-, 108-, and 126-ps coordinate and velocity sets. The simulation protocol for the MDFEP calculations is given in the computational procedure section. All values reported in kcal/mol. [b] The average values are given for EO → ES. [c] The first number correspond to EO → ES, while the second number corresponds to going in the ES → EO direction.

Table V. Computed Free Energies for CO_2 → O[a]

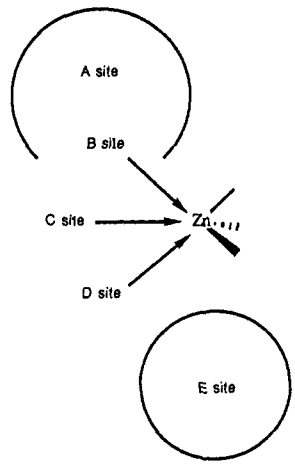
direction[b]	0 ps	18 ps	36 ps	ΔG_{sol}[c]
forward	−0.56, 0.57	−1.51, 1.52	−1.07, 1.08	
backward	0.86, −0.85	−0.70, 0.71	−0.21, 0.22	0.54 ± 0.78[d]

[a] Coordinate and velocity sets were saved at 18-ps intervals past the 27-ps equilibration period. For the free energy evaluations we used the 0- (27-ps equilibrated starting structure), 18-, and 36-ps coordinate and velocity sets. The details of the simulations are given in the computational procedure section. All values reported in kcal/mol. [b] The forward direction indicates that the coupling parameter λ is going from 1 to 0 and backwards indicates that it is going from 0 to 1. Each cell contains two numbers, which indicates that for each individual run we have sampled in both the backwards (O → S) and forward (S → O) directions. See refs 4 and 5. [c] Total average value for both forward and backwards runs in the S → O direction. [d] The experimental value for ΔG_{sol}(O → S) is 0.0 kcal/mol at 298 K.[25]

implies that the substrate exists as a set of non-interacting dummy atoms, ΔG_{sol}) and the second to evaluate the free energy of binding of CO_2 to the protein (EO → ES, ΔG_{enz}). This approach has been applied successfully to the evaluation of the free energy of association of two methane molecules.[29] Our MDFEP simulations indicate that the binding pocket closest to the zinc ion binds CO_2 more tightly (−3.37 ± 1.14 kcal/mol) than the more distant one (−2.17 ± 0.83 kcal/mol). The results are summarized in Table IV and V. An experimental number for the binding of CO_2 ($K_{eff}^{CO_2}$) to the zinc–hydroxide form of HCAII at near neutral pH (7.5) is −2.2 kcal/mol (25 mmol).[30] This number represents the maximum for this binding constant. However, when used in computer simulations of a proposed kinetic scheme for HCAII it is able to reproduce the experimental steady-state kinetic patterns of HCAII.[30] Hence, we have chosen this value to compare our results to, but we note that no one has unambiguously identified the true affinity constant for CO_2 binding to the active site of HCAII.[6,30,31] With this in mind we find that one of our computed values is too high, while the other is in excellent agreement. However, if we assume, for example, that two CO_2 molecules are binding at any one time to the enzyme we know that the true affinity constant for CO_2 binding to HCAII would be dominated by the larger binding constant. Thus, the affinity constant that we would obtain from our theoretically determined individual binding free energies would be on the order of −3.37 kcal/mol (or 3.4 mmol) or tighter. Thus, our computed value is larger than expected, but it is reasonably close to the experimental value. These calculations suggest that the deep-water pocket is the high affinity pocket while the second site has a lower affinity for CO_2, which suggests that the deep-water site is the important catalytic site while the more remote site serves another function.

The covalent binding sites on the zinc ion can be labeled as follows:[9] In the case of four-coordination the binding site that retains the distorted tetrahedral arrangement around the zinc ion is designed **A**. The hydroxide ion in the zinc–hydroxide form of this enzyme binds in site **A**.[9] In the case of five-coordination, two sites have been distinguished; the **B** site is pointing toward the deep-water pocket,[9] which we have described as being the "high

Scheme III



CO_2 affinity" pocket. Site **C** is directly opposite site **B** and is the site at which a water molecule is bound in the HCAII·SCN structure.[9] All of these sites are covalent binding sites. Thus, we need a way in which to describe the CO_2 binding sites we have observed, which are noncovalent. We propose Scheme III. If we consider our high CO_2 affinity site as the **A** site, then we can call the low CO_2 affinity site the **E** binding site in the active site. We next reorder the covalent sites around the zinc ion and label them as **C** (old **A**), **B** (old **B**), and **D** (old **C**). This model has the advantage that the labeling is sequential, starting with the base of the active site, and any new sites that are identified further out can be simply labeled **F**, **G**. ....

HCAII is considered a "perfectly evolved enzyme" because of its very high turnover number of 1×10^6 s^{-1}.[1,6] The rate-limiting step in the catalytic mechanism under physiological conditions is an intramolecular proton transfer from a zinc-bound water to His 64.[6] Thus, the enzyme is not limited by the rate at which CO_2 is bound to the enzyme. This implies that the flux of CO_2 molecules into the HCAII active site has to be very high. Thus, we can ask the question of what significance might the observation of multiple binding sites be to HCAII catalysis given that the enzyme is not diffusion limited at high substrate concentrations?[6] The function of multiple CO_2 binding sites would be to first increase the capability of the enzyme to recognize and bind CO_2 molecules (i.e. increasing the effective "target area"). Thus, this would enhance the likelihood of a CO_2 molecule diffusing toward the active site of HCAII being recognized and bound. Furthermore, the **E** binding site can serve as a reservoir for CO_2 molecule(s), which can be transferred from the **E** site when the **A** site is empty. The **A** site then places the CO_2 molecule in an orientation that is favorable for CO_2 to react with the zinc–hydroxide form of the enzyme. Hence, the second function of multiple binding sites would be to enhance the rate at which CO_2 molecules diffuse within the active site.

Is there any evidence in the carbonic anhydrase literature that suggests that two (or more) CO_2 molecules may be bound at any one time? Bertini et al.[12,31] in light of their work on inactive Cu–CA have indeed suggested this as being one of two possible explanations for their results (the other being one CO_2 directly bound to zinc).[31] The experiments of Bertini et al.[12,31] and Led et al.[10] (Mn–CA), which used inactive and slightly active enzymes, respectively, both have bicarbonate as a ligand to the metal ion. Thus, these studies are actually investigating the binding of CO_2 to the CA–bicarbonate adduct. From a simple computer graphics analysis we find that if we bind bicarbonate to the active site metal ion in either a four- or five-coordinated configuration, we effectively "block off" the **A** binding pocket, thereby interfering with CO_2 binding to this site. This suggests that these experiments were examining the binding of CO_2 to other sites in the active site cavity (possibly our **E** site). The hallmark of these experiments

(29) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.
(30) Lindskog, S. *J. Mol. Catal.* **1984**, *23*, 357. Silverman, D. A. Private communication.
(31) Bertini, I.; Luchinat, C.; Monnanni, R.; Roelens, S.; Moratal, J. M. *J. Am. Chem. Soc.* **1987**, *109*, 7855.

were low binding affinities[12,31] and long metal ion to CO_2 carbon distances,[10,12,31] which would be more consistent with binding to a site other than the A site.

Further analysis of the MD trajectories has provided us with some qualitative insights that may have important implications for the function of the enzyme.[32] (1) The A site places the CO_2 molecule in such a way that it is "perfectly" aligned to react with the hydroxyl oxygen bound to the zinc ion at the active site (the CO_2 carbon to hydroxyl oxygen distance is 3.0 Å). Thus, given our observed structures, it appears likely that once CO_2 binds to this site it is nicely positioned to subsequently react to form the zinc-bicarbonate form of the enzyme.[32] (2) The Thr 199···HO-Zn hydrogen bond appears to be very strong and not fluxional. Thus, we have suggested that the function of this interaction is to "lock" the hydroxyl hydrogen in the zinc-hydroxide form of the enzyme in what we will call the down position. In this orientation the lone pairs on the hydroxyl bound to the zinc ion are pointing toward the CO_2 molecule bound in the A binding site. If this hydrogen-bonding interaction was not present one might imagine that the zinc bound hydroxyl would be able to freely rotate, thereby potentially interfering with CO_2 binding and subsequent reaction.[32]

Conclusions

With the aid of two 126-ps MD simulations we have identified two CO_2 binding sites in HCAII, which we have labeled as the A and E binding sites. Analysis of these trajectories demonstrates that the A site provides a rather restrictive environment for the CO_2 molecule while the E site is less so. Using free energy perturbation simulations we have evaluated the absolute free energy of binding for CO_2 interacting with these sites. We find that the A site has a greater free energy of binding (-3.37 kcal/mol) than the E site (-2.17 kcal/mol). While these binding free energies are not quantitatively correct when compared to experiment they make sense from a functional perspective. Thus, the A site one would expect to be the reactive pocket due to its proximity to the zinc ion, and therefore one might expect that it should have a greater binding affinity than does a more remote site like the E site.

Liang and Lipscomb[33] have recently studied this problem using MD simulations and they have identified a total of three binding sites in the HCAII active site, two of which are similar, but not identical with our A and E sites. The third site, which is further out in the active site cavity, we will label as the F site. The addition of other binding sites fits in with our proposal that more binding sites increase the ability of HCAII to bind and to diffuse CO_2 molecules into the A binding site.

Our results provide suggestions for site-directed mutagenesis experiments that could be performed in order to further our understanding of the catalytic mechanism of HCAII and, in particular, the step involving the association of CO_2 with the protein. For the E site one could mutate Ala 65 into a larger group that would fill up this pocket and inhibit CO_2 binding, while for the A site mutating Val 143 into a larger group would block off this site. We look forward to these experiments being done along with high-pressure X-ray studies aimed at addressing this issue.

Acknowledgment. We thank the Center for Academic Computing at the Pennsylvania State University for generous allocations of IBM 3090-600S computer time. Helpful discussions with U. C. Singh and B. Brooks are also acknowledged. We thank J.-Y. Liang and W. N. Lipscomb for providing a copy of their manuscript prior to publication. D. N. Silverman, S. Lindskog, and C. Fierke have provided numerous very helpful and insightful comments on this present work and their help and support is gratefully acknowledged. Helpful comments by one of the referees are also acknowledged. Coordinates for the A and E binding sites can be obtained via an anonymous ftp download. The internet address is 128.118.30.113 and the PDB format coordinates are in the pubs/HCAII directory.

(32) Merz, K. M., Jr. *J. Mol. Biol.* **1990**, *214*, 799.

(33) Liang, J.-Y.; Lipscomb, W. N. *Proc. Natl. Acad. Sci.* Submitted for publication.

Ab Initio Search for the Structure of the 4-Protoadamantyl Cation

R. Dutler, A. Rauk,* S. M. Whitworth, and T. S. Sorensen*

Contribution from the Department of Chemistry, University of Calgary, Calgary, Alberta T2N 1N4, Canada. *Received March 23, 1990*

Abstract: The 4-protoadamantyl cation region of the $C_{10}H_{15}^+$ potential energy hypersurface is searched by means of RHF-SCF calculations and the STO-3G, 6-31G, and (in part) 6-31G* basis sets. Two structures are found. The structure with C_s symmetry is postulated to be the primary intermediate in the solvolysis of 4-*endo*-protoadamantyl derivatives. It is calculated to be 8 kJ/mol less stable than the second structure which has no symmetry. The asymmetric cation is postulated to be the primary intermediate in the solvolysis of 4-*exo*-protoadamantyl derivatives. It is 7 kJ/mol less stable than 2-adamantyl cation to which rearrangement is hindered by a very small barrier. Rearrangement of the C_s structure is expected to be hindered by relatively large barriers. It may be generated directly from bicyclo[4.3.1]dec-3-en-8-yl derivatives. The C_s cation is characterized by IGLO calculations of ^{13}C and ^1H NMR chemical shifts.

Introduction

Solvolysis of 4-*exo*- and 4-*endo*-protoadamantyl derivatives proceeds[1] with an *exo*/*endo* reactivity ratio of $\sim 10^4$. The products of solvolysis of 4-*exo*-protoadamantyl 3,5-dinitrobenzoate were exclusively 2-adamantyl derivatives.[1] The rearrangement has been shown to proceed with retention of configuration of the oxygen moiety.[2] On the other hand, solvolysis of the 4-*endo* tosylate (*endo*-**1**-OTs) under some conditions gave predominantly 2-adamantyl derived products but also some 4-*endo*-protoadamantanol. No 4-*exo*-protoadamantanol was detected.[1] Acetolysis of 2-adamantyl tosylate (**2**-OTs) under buffered con-

(1) Lenoir, D.; Hall, R. E.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1974**, *96*, 2138-2147.

(2) Nordlander, J. E.; Haky, J. E. *J. Org. Chem.* **1980**, *45*, 4780.

0002-7863/91/1513-411$02.50/0