

Metal Coordination, H-Bond Network Formation, and Protein-Solvent Interactions in Native and Complexed Human Carbonic Anhydrase I: A Molecular Mechanics Study

Angelo Vedani,* David W. Huhta, and Stephen P. Jacober

Contribution from the Biographics Laboratory, Departments of Chemistry and Biochemistry, University of Kansas, Lawrence, Kansas 66045. Received August 15, 1988

Abstract: Details of metal coordination, H-bond network formation, and protein-solvent interactions in native and complexed human carbonic anhydrase I have been investigated using molecular mechanics techniques. The studies are based on the 2.0-Å resolution X-ray crystal structure of the native enzyme with the addition of 503 computer-generated water molecules, representing an explicit solvent structure. Our results suggest that both the native and the complexed enzyme find their highest stabilization with a slightly (4 + 1) distorted tetrahedral coordination at the catalytic zinc. Two proton-relay networks have been studied in detail: one (already detected in the crystal structure) linking the proximal zinc-bound water molecule over three amino-acid residues and two mediating water molecules to His 64, located on the surface of the protein. In our study, we found evidence for a second H-relay network, providing a more direct connection from a distal zinc-bound water molecule through the active site cleft to the surrounding solvent. Both proton relay networks remain intact when the natural substrate bicarbonate is coordinated to the zinc, but are disrupted upon sulfonamide inhibitor binding.

Carbonic anhydrase, an extremely efficient catalyst of the reversible hydration of carbon dioxide, consists of 260 amino acids and a zinc cofactor.¹ It is almost universally assumed that a zinc-bound hydroxide ion is the nucleophile in the catalytic reaction.²⁻⁵ Aromatic and heterocyclic sulfonamides are competitive inhibitors of carbonic anhydrase with respect to the dehydration of bicarbonate when the sulfonamide N atom is unsubstituted.⁶ The inhibition of carbonic anhydrase by sulfonamide drugs finds clinical application in the treatment of glaucoma, epilepsy, and acute mountain sickness.⁷

The X-ray crystal structure of native human carbonic anhydrase I (HCA I; formerly HCAB) has been determined to a resolution of 2.0 Å and refined to a final *R* value of 19% by Kannan, Ramanadham, and Jones.⁸ As shown in their study, the active site of carbonic anhydrase, a 12-Å deep conical cavity, is part of the central twisted pleated β-sheet. The catalytic zinc is located at the apex of the cone, coordinated to three histidine N atoms (His 94, His 96, and His 119) and to the O atom of a water molecule. About 300 solvent molecules have been located bound to the protein molecule; however, only the coordinates of the protein portion have been deposited in the Brookhaven Protein Data Bank (BPDB⁹). As shown in Figure 3 of ref 8, three of these water molecules engage in an H-bond network that might be important for the catalytic mechanism, involving the catalytic zinc, the zinc-bound water molecule, and the amino-acid side chains of the residues Thr 199, Glu 106, Tyr 7, and His 64.

Detailed structural data on small-molecule binding to human carbonic anhydrase are not presently available. For the natural substrate bicarbonate, the lack of information is due to weak binding and limited solubility, which has hitherto prevented the

study of tightly formed 1:1 complexes of the enzyme with hydration substrates.¹⁰ For the sulfonamide inhibitor acetazolamide (2-acetamido-1,3,4-thiadiazole-5-sulfonamide), an X-ray crystal structure has been determined.^{11,12} At 3.0-Å resolution, the electron-density map clearly revealed the position of the two S atoms of the inhibitor molecule (cf. ref 11, Figure 4.3), but only few further details of enzyme-inhibitor interactions were visible.

For modeling studies on biological macromolecules, it is of greatest importance to consider interactions with the solvent. Of particular interest are internal water molecules that form water-mediated hydrogen bonds, cross-linking individual residues. If such water molecules are not included in the calculations, macromolecular cavities may expand or collapse during the refinement. In addition, incomplete allowance for solvation can lead to unreasonably large atomic shifts that poorly correlate with atomic thermal displacement factors as obtained from the X-ray crystal structure determination (cf. below and Figure 3). Unfortunately, X-ray diffraction experiments do not always reveal all the positions of protein-bound water molecules or at the desired level of accuracy. Here, the solvent structure must be generated or completed by computational techniques. For the purpose of generating internal and surface solvent molecules of HCA I, we have used the program SOLVGEN, developed at this laboratory.¹³

YETI,¹⁴ one of the two molecular mechanics programs used for the refinement of the various complexes, includes a special set of directional potential functions for H bonds, salt linkages, and metal-ligand interactions in its force-field energy expression. Directional functions can be calibrated in a subtle manner against experimental data and allow a more realistic modeling of biological systems where these interactions play a critical role. The directional potential function for H bonds and salt linkages in YETI¹⁵ is based on studies of H-bonding patterns in small-molecule¹⁵⁻¹⁷ and accurate protein crystal structures,¹⁸ which showed that donor H atoms are concentrated in directions commonly ascribed to the lone-pair orbitals of the acceptor atom. The H-bond potential

(1) Notstrand, B.; Vaara, I.; Kannan, K. K. In *The Isozymes*; Markers, C. L., Ed.; Academic Press: New York, 1975; pp 575-599.

(2) Lindskog, S.; Engberg, P.; Forsman, C.; Ibrahim, S. A.; Jonsson, B.-H.; Simonsson, I.; Tibell, L. In *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 61-75.

(3) Pocker, Y.; Deits, T. L. In *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 76-83.

(4) Silverman, D. N.; Vincent, S. H. *CRC Crit. Rev. Biochem.* **1983**, *14*, 207-255.

(5) Pocker, Y.; Sarkanen, S. *Adv. Enzymol.* **1978**, *47*, 149-274.

(6) Keilin, D.; Mann, T. *Nature (London)*, **1940**, *146*, 164.

(7) Maren, T. H. In *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 10-17.

(8) Kannan, K. K.; Ramanadham, M.; Jones, T. A. In *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 49-60.

(9) Bernstein, F.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.

(10) Mukherjee, J.; Rogers, J. I.; Khalifah, R. G.; Everett, G. W., Jr. *J. Am. Chem. Soc.* **1987**, *109*, 7232-7233.

(11) Kannan, K. K.; Vaara, I.; Notstrand, B.; Lovgren, S.; Borell, A.; Fridborg, K.; Petef, M. In *Proceedings on Drug Action at the Molecular Level*; Roberts, G. C. K., Ed.; McMillan: London, 1977; pp 73-91.

(12) Kannan, K. K. In *Biophysics and Physiology of Carbon Dioxide*; Gros, H.; Bartels, H., Eds.; Springer: Berlin, 1979; pp 184-205.

(13) Jacober, S. P. SOLVGEN: An Approach to Protein Hydration; M.S. Thesis, Department of Computer Sciences, University of Kansas, 1988.

(14) Vedani, A. *J. Comput. Chem.* **1988**, *9*, 269-280.

(15) Vedani, A.; Dunitz, J. D. *J. Am. Chem. Soc.* **1985**, *107*, 7653-7658.

(16) Murray-Rust, P.; Glusker, J. P. *J. Am. Chem. Soc.* **1984**, *106*, 1018-1025.

(17) Taylor, R.; Kennard, O. *Acc. Chem. Res.* **1984**, *17*, 320-326.

(18) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97-179.

function in YETI (see below) therefore includes as an additional variable, the deviation of the H bond from the closest lone-pair direction at the acceptor atom (angle $H\cdots Acc-LP$).¹⁴ The metal-center potential function in YETI includes as variables the various metal-ligand distances and the angles subtended at the metal.¹⁹ This function allows specifically for distortions from and transitions between frequently occurring types of coordination geometries (e.g., tetrahedron, square plane, square pyramid, trigonal bipyramid, and octahedron). Because of differences in distribution of metal-ligand distances between axial and equatorial ligands,¹⁹ YETI calculates the interaction energy of a metal with axial ligands (broad experimental distribution) using a 6/12 Lennard-Jones function plus an electrostatic term, whereas equatorial ligands (narrow experimental distribution) are treated by a 10/12 function plus an electrostatic term.

Methods

The coordinates of native human carbonic anhydrase I (HCA I, 2.0-Å resolution; cf. ref 8) were obtained from the Brookhaven Protein Data Bank.⁹ Of the 300 solvent molecules bound to HCA I, only five have been described in the literature (cf. ref 8, Figures 2 and 3): a zinc-bound water molecule (Wat 262p, p = proximal; see below) and four water molecules forming water-mediated H bonds between the residues Glu 106 and Tyr 7 (Wat 263), Tyr 7 and His 64 (Wat 264), His 67 and His 200 (Wat 265), as well as Ser 65 and Wat 264 (Wat 266).

Because the coordinates of the 300 protein-bound water molecules have not been deposited in the BPDB, we have generated the solvent structure of HCA I using the program SOLVGEN. This program generates internal and oriented surface water molecules of proteins based on the directionality of H bonds.¹³ The algorithm of SOLVGEN includes the following main steps.

(1) The first step is H-bond pattern recognition. This is performed by generating lone-pair vectors (LPV) originating at H-bond acceptors and H-extension vectors (HEV) originating at H-bond donors. The endpoint of a LPV marks the ideal position for a H-bond donor atom relative to the H-bond acceptor fragment; the endpoint of a HEV marks the ideal position of H-bond acceptor atom relative to the H-bond donor fragment. Both LPV and HEV are derived from analyses of small-molecule and accurate protein crystal structures (cf. ref 15-18).

(2) To identify "free" vectors able to engage in H-bond interactions with coordinating water molecules, it is necessary to delete LPV and HEV associated with already existing hydrogen bonds. Ideally, the HEV and the LPV of a specific hydrogen bond would overlap entirely, but this is rarely observed in protein structures. LPV and HEV are therefore extended to a cone, representing the experimental distribution width.

(3) In a next step, the remaining vectors are analyzed for clustering. Probable sites for water molecules (i.e., the position of their O atom) can be assigned to points where at least two vector tips (LPV or HEV) intersect or meet within a short distance at an appropriate angle. If the surrounding cavity or crevice is large enough to accommodate a water molecule (van der Waals contact scan), the water H atoms are generated and oriented along the LPV's; water lone pairs are oriented along the HEV's using interactive computer graphics techniques.

(4) This procedure is repeated in combination with molecular mechanics optimizations of the solvent until the desired level of solvation is obtained but at least until there are no more unoccupied hydrophilic cavities within the protein, and its surface is "solvated" by one layer of water.

The structure of the generated solvent shell is not very sensitive to changes in geometrical constraints (distances, linearity, directionality of H bonds) used to generate it. Stringent determinants are the size of cavities and crevices of the protein (van der Waals contact scan) and the subsequent molecular mechanics refinements (i.e., the particular force field used). Presently, the real limitation of SOLVGEN is the solvation of larger hydrophobic regions exposed to the solvent, since only few HEV and LPV originate at side chains of such residues (Phe, Trp, Tyr, Ala, Val, Leu, Ile, and Pro).

To test SOLVGEN, we have performed a simulation aimed to reproduce the 82 experimental water molecules found in the high-resolution X-ray crystal structure of native bovine trypsin (1.7-Å resolution, 19.8% R factor; cf. ref 20). Of the 82 water molecules found in the X-ray crystal structure, 23 engage in two or more H bonds with the protein and thus represent primary targets for SOLVGEN. Without (!) coupled molecular

mechanics optimizations, SOLVGEN reproduced 16 of these 23 within a reasonable distance: 2 within 1 esd (estimated uncertainty in atomic position due to thermal motion, calculated from the isotropic temperature factor of the O atom), 8 within 1.5 esd, 10 within 2.0 esd, and 16 within 3.0 esd. Subsequent runs of SOLVGEN yielded a total of 374 water molecules, reproducing most experimental water molecules within reasonable distance from the experimental position. At this point, it would not seem possible to test the method further, since most of the available high-resolution X-ray crystal structures have been used indirectly (data used from ref 18) to calibrate SOLVGEN. Moreover, one should be careful in interpreting such agreements because even high-resolution X-ray crystal structures already include some regularization.

For modeling complexed HCA I, the X-ray crystal structures of the bicarbonate ion (as found in $KHCO_3$)²¹ and the acetazolamide molecule²² have been retrieved from the Cambridge Structural Database (CSD²³) and docked to the active site of the enzyme using interactive computer graphics techniques (software FRODO,²⁴ V6.4 running on an Evans & Sutherland PS390 color display).

Energy calculations have been performed using the molecular mechanics programs YETI v3.8¹⁴ and AMBER v3.0²⁵ on μ -VAX II computers of the KU Biophysics Laboratory and the VAX 8650 computer of the University of Kansas. The refinements included all 2008 heavy atoms (C, N, O, S) and 450 polar H atoms (i.e., O-H, N-H, S-H) of the protein, the metal-ion cofactor, the zinc-bound water molecules, the substrate or inhibitor molecule, as well as 501 water molecules representing an explicit solvent structure. For the protein portion, aliphatic and aromatic H atoms were included within united atoms.

In YETI, optimizations are performed in a mixed internal/Cartesian coordinate space using a conjugate-gradient minimizer. Degrees of freedom include the conformation of all protein side chains; position, orientation, and conformation of the substrate or inhibitor molecule; position of the zinc cofactor, as well as position and orientation of all water molecules, yielding 3564 degrees of freedom for the native enzyme, 3565 for the bicarbonate complex, and 3563 for the acetazolamide complex. In the YETI force field, the total energy of the system is represented in terms of torsional, electrostatic, van der Waals, H-bond, and metal-ligand energies. Bond lengths and bond angles are not varied during the refinement (those have been optimized using AMBER; see below); therefore no bond-stretching and angle-bending terms are present in the YETI force-field energy expression (cf. ref 14):

$$E_{\text{total}} = \sum_{\text{dihedrals}} \frac{Y_n}{2} \{1 + \cos(n\phi - \gamma)\} + \sum_{i < j} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{D(r) r_{ij}} + \sum_{i < j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} + \sum_{\text{H bonds}} \left(\frac{A'}{r_{H-Acc}^{12}} - \frac{C'}{r_{H-Acc}^{10}} \right) \cos^2(\theta_{\text{Don-H}\cdots\text{Acc}}) \cos^4(\omega_{\text{H}\cdots\text{Acc-LP}}) + \sum_{\text{ML bonds}} \left(\frac{A''}{r_{M-L}^{12}} - \frac{C''}{r_{M-L}^{10}} \right) \prod_{\text{indep angles}} \cos^2(\Psi_{L-M-L} - \Psi_0)$$

Using smooth cutoff criteria of 9.5/10.0 Å (switch-on, switch-off; cf. ref 26) for electrostatic interactions, 6.5/7.0 Å for van der Waals interactions, and 5.5/6.0 Å for H bonds, the initial list of nonbonded interactions numbered about 580 000. Convergence criteria for the energy refinement have been set to 0.025 kcal/(mol-deg) for torsional, to 0.050 kcal/(mol-deg) for rotational, and to 0.750 kcal/(mol-Å) for translational rms first derivatives, respectively.

Full relaxations (including also the protein backbone) have been performed using the molecular mechanics software AMBER.²⁵ AMBER uses a combined steepest-descent/conjugate-gradient minimizer in Cartesian-coordinate space. Apart from the treatment of H bonds and metal-ligand interactions, the AMBER force field for nonbonded interactions is identical with the YETI force field. In addition, AMBER allows bond

(21) Thomas, J. O.; Tellgren, R.; Olovsson, I. *Acta Crystallogr., Sect. B* 1974, B30, 2540-2549.

(22) Mathew, M.; Palenik, G. J. *J. Chem. Soc., Perkin Trans. 2* 1974, 532-536.

(23) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr., Sect. B* 1979, B35, 2331-2339.

(24) Pflugrath, J. W.; Saper, M. A.; Quioco, F. A. In *Methods and Applications in Crystallographic Computing*; Hall, S., Ashiaka, T., Eds.; Clarendon Press, Oxford, 1984; pp 407.

(25) Singh, U. C.; Weiner, P. K.; Caldwell, J. W.; Kollman, P. A. University of California, San Francisco, 1986.

(26) Brooks, B. R.; Brucolieri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* 1983, 4, 187-217.

(19) Vedani, A.; Dobler, M.; Dunitz, J. D. *J. Comput. Chem.* 1986, 7, 701-710.

(20) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. *Acta Crystallogr., Sect. B* 1983, B39, 480-490.

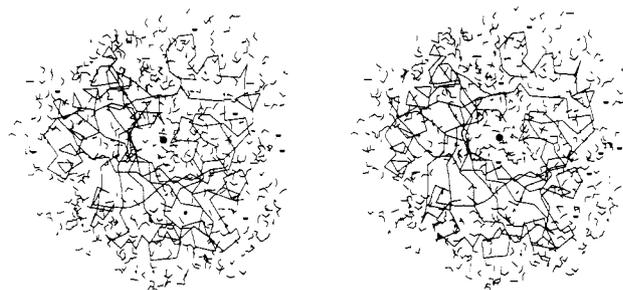


Figure 1. Stereoscopic view of the solvated structure (16 internal and 487 surface water molecules) of native human carbonic anhydrase I. For the protein portion, only the α -carbon atoms are shown.

distances and bond angles to vary and therefore includes terms for bond stretching and angle bending (cf. ref 27):

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} |1 + \cos(n\phi - \gamma)| + \sum_{i < j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right) + \sum_{\text{H bonds}} \left(\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right)$$

Convergence criteria for all AMBER refinements have been set at 0.050 kcal/(mol·Å) for first-derivative rms values; the cutoff value for non-bonded interactions has been set at 10.0 Å. In both the YETI and AMBER refinements, a weight of 0.5 has been assigned to nonbonded 1–4 interactions (electrostatic and van der Waals terms).

Atomic partial charges for amino-acid residues of the protein were taken from ref 27; those for the substrate and the inhibitor molecules were calculated by MNDO methods (program AMPAC²⁸). Electrostatic energies are evaluated using a distance-dependent dielectric parameter of $D(r) = 2r$. Earlier studies justify this value for modeling metallo-enzymes (cf. ref 19, pp 275–6 for details).

Results and Discussion

Native Human Carbonic Anhydrase I. In the X-ray crystal structure of native human carbonic anhydrase (both I and II), the zinc cofactor was found to be tetrahedrally coordinated by three histidine N atoms (His 94, His 96, and His 119) and by the O atom of a water molecule (Wat 262p, p = proximal). For the catalytic reaction, it is generally assumed that the OH⁻ ion acting as the nucleophile is metal bound. However, there seems to be no general agreement about the metal coordination during the catalytic reaction (cf. ref 2, 3, 29, 30). To retain its nucleophilicity, the hydroxide ion should not be bound too tightly to the metal, but rather occupy a distal position (in contrast to His 94, His 96, His 119, and Wat 262p which are proximal ligands; for proximal/distal terminology of zinc ligands in proteins, cf. ref 31, 32). In the native enzyme, this nucleophile is present either as a hydroxide ion or a water molecule, depending on the microscopic pH value.

Two zinc-bound, 16 internal, and 485 surface water molecules representing an explicit solvent structure of HCA I have been generated using the program SOLVGEN.¹³ A first run yielded 128 water molecules that initially engaged in at least two H-bond interactions with the protein. Subsequent runs yielded 196 (including also water molecules that initially engage in one H bond with the protein and one additional interaction with a neighbored

(27) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Chandra Singh, U.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.

(28) Distributed by QCPE, University of Indiana, Bloomington, In. Original reference on MNDO method: Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(29) Coleman, E. In *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 26–48.

(30) Cook, C. M.; Allen, L. C. In *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N.Y. Acad. Sci.* **1984**, *429*, 84–88.

(31) Dutler, H.; Ambar, A. In *The Coordination Chemistry of Metallo-enzymes*; Bertini, I., Drago, S. R., Lucchinat, C., Eds. (NATO Advanced Study Institute Series); Reidel: Dordrecht, 1982; pp 135–145.

(32) Brändén, C.-I. In *Metalloproteins*; Weser, U., Ed.; G. Thieme: Stuttgart and New York, 1979; pp 220–245.

Table I. H-Bond Partners of (a) Zinc-Bound and (b) Internal Water Molecules in Human Carbonic Anhydrase I^a

		H bond to protein acceptor	H bond to protein donor	H bonds with other water molecules	total no. of H bonds
(a)	Wat 262p ^b	Thr 199, OH		Wat 404 (Acc)	2 (+Met)
	Wat 681d ^b	none		Wat 402 (Acc) Wat 484 (Acc)	2 (+Met)
(b)	Wat 263 ^b	Glu 106, OE2 {Thr 199, O} ^c	Tyr 7, OH		2 (3)
	Wat 264 ^b	Tyr 7, OH His 64, ND1		Wat 266 (Don)	3
	Wat 265 ^b	His 200, NE2	His 67, NE2	Wat 266 (Acc)	3
	Wat 274	Pro 13, O	Arg 246, NH2		2
	Wat 286	Gly 25, O Pro 250, O	Gln 28, N		3
	Wat 288	Ser 29, OG	Arg 254, NH1		2
	Wat 320	His 64, O His 243, O		Wat 444 (Don) Wat 476 (Don)	4
	Wat 323	Ser 65, OG Phe 66, O	Phe 95, N		3
	Wat 324	Phe 66, O Phe 93, O			2
	Wat 329	Glu 71, O Leu 90, O		Wat 313 (Don) {Wat 503 (Don)} ^c	3 (4)
	Wat 346	Thr 100, O Ser 115, O	Gly 98, N	Wat 345 (Don)	4
	Wat 360	His 107, O	{Ser 105, HOG} ^c		2 (3)
	Wat 371	Val 112, O Ala 138, O Thr 208, OH			2
Wat 373	Gly 145, O Cys 212, O	Leu 147, N		3	
Wat 374	Leu 147, O Glu 214, O			2	
Wat 800	Lys 156, O	Val 160, N		2	

^aOnly those internal water molecules are listed that have at least two H-bond interactions with the protein. ^bThese water molecules have already been described in the literature.⁸ ^cWeak, usually not very directional H bond: E_{HB} smaller than -0.5 kcal/mol (direct H-bond term).

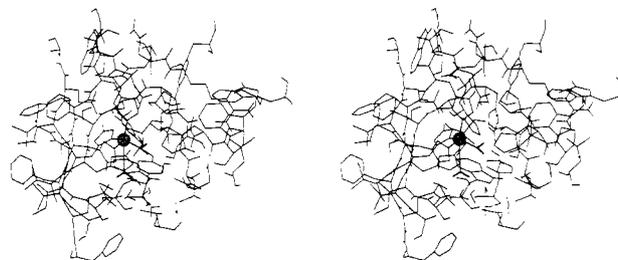


Figure 2. Stereoscopic view of the active site cleft of native human carbonic anhydrase I. Details of zinc binding. The zinc is represented by a sphere; the two zinc-bound water molecules have been drawn enhanced.

water molecule), 439 (at least one H bond; protein or water molecule), and finally (after protein side-chain relaxation) 503 oriented water molecules (cf. Figure 1). Internal water molecules, tightly bound to the protein in cavities, were of special interest for this study. For human carbonic anhydrase I, the program SOLVGEN identified 16 such water molecules (cf. Table I). In addition, nine water molecules bound to more or less well-defined crevices were detected. All five internal water molecules associated with the proton-relay network described in ref 8 could be reproduced.

Subsequent molecular mechanics refinements of the "solvated" native enzyme yielded a slightly distorted tetrahedron at the zinc: Zn...N 1.974 (His 94), 1.936 (His 96), 1.968 Å (His 119); Zn...O 1.923 Å (Wat 262p), with angles subtended at the zinc varying from 103° to 115°. A second water molecule has been generated by SOLVGEN, only weakly coordinated to the metal, opposite of His 96 (Zn...O 3.252 Å after refinement). The arrangement at the zinc can thus be best described as a (4 + 1) distorted tetrahedron with four proximal ligands (N atoms of His 94, 96, 119; O atom of Wat 262p) and one distal ligand (O atom of Wat 262p).

Table II. Details of Two Possible H-Bond Relay Networks in Native HCA I^a

	donor		acceptor	Don...Acc, Å	H...Acc, Å	Don-H...Acc, deg	H...Acc-LP, deg
(a)	Wat 262p	O-H...O	Thr 199	2.64	1.79	147	25
	Thr 199	O-H...O	Glu 106	2.74	1.80	162	25
	Wat 263	O-H...O	Glu 106	2.78	1.82	170	10
	Tyr 7	O-H...O	Wat 263	2.78	1.82	173	14
	Wat 264	O-H...O	Tyr 7	2.77	1.84	162	1
	Wat 264	O-H...N	His 64	2.86	1.90	171	14
(b)	Wat 681d	O-H...O	Wat 484	2.77	1.81	177	9
	Wat 484	O-H...O	Wat 411	2.76	1.80	178	11
	Wat 411	O-H...O	Wat 356	2.75	1.79	174	13
	Wat 356	O-H...O	Wat 425	2.76	1.81	166	14
	Wat 425	O-H...O	Wat 449	2.73	1.79	162	16
	Wat 449	O-H...O	Wat 721	2.77	1.80	177	8
	Wat 721	O-H...O	Wat 705	2.78	1.81	173	4
(c)	Wat 681d	O-H...O	Wat 402	2.74	1.79	167	15
	Wat 402	O-H...O	Wat 356	2.78	1.82	170	11
(d)	Wat 262p	O-H...O	Wat 404	2.62	1.77	145	11
	Wat 404	O-H...O	Wat 484	2.69	1.78	157	7

^a(a) Network originating at the proximal zinc-bound water molecule (first described in ref 8); (b) network originating at the distal zinc-bound water molecule (first described in this study); (c) branch completing the water pentagon (Wat 681d, Wat 484, Wat 411, Wat 356, and Wat 402); joining network b) at Wat 356; (d) branch, linking the two zinc-bound water molecules. To avoid too much complexity, water-protein interactions are not listed for b, c, and d.

The proximal zinc-bound water molecule is also hydrogen bonded to the hydroxyl O atom of Thr 199 (here, the water molecule acts as H-bond donor). A H...O distance of 1.79 Å and a slightly bent O—H...O arrangement (147°) are found. The water molecule in the distal position at the zinc (Wat 681d) is ideally positioned for nucleophilic attack at the CO₂. Apart from the interaction with the metal, it engages in two H bonds with Wat 402 and Wat 484 (cf. Table II). A close-up of the active-site region is shown in Figure 2.

All 503 water molecules engage in at least one H bond. The average water molecule engages in 2.2 H bonds (total average H-bond energy = -7.6 kcal/mol for the direct H-bond term; electrostatic and van der Waals contributions are calculated in addition); 44 water molecules engage in four H bonds (average H-bond energy = -13.5 kcal/mol), 145 in three H bonds (-9.8 kcal/mol), 199 in two H bonds (-6.9 kcal/mol), and 115, mainly located on the surface, in one H bond (-3.9 kcal/mol). Two- and three-dimensional H-bond network formation is observed at the inner (active site) and outer surface of the protein which yielded a complete hydration of charged amino-acids side chains (Asp, Glu, Lys, Arg).

The proton-relay network, detected by Kannan and co-workers in their X-ray study,⁸ originating at the proximal zinc-bound water molecule and including three internal water molecules can be described as follows (see Table II). The proximal zinc-bound water molecule (Wat 262p) is H-bonded to the side-chain hydroxyl O atom of Thr 199. Thr 199 engages in a H bond with the side-chain carboxyl O(E1) atom of Glu 106. The O(E2) atom of Glu 106 forms a water-mediated H bond (Wat 263) with the side-chain hydroxyl O atom of Tyr 7. Wat 264 links the O atom of Tyr 7 and the imidazole N atom of His 64 which is located at the entrance of the active site cavity and provides the contact with the surrounding solvent (Wat 465).

An alternative proton relay network, associated with the distal zinc-bound water molecule, has been found in our study. In contrast to the proton relay network detected by Kannan and co-workers,⁸ this second network would exclusively include water molecules. An interesting feature is the formation of a water pentagon in the active-site funnel, involving the distal zinc-bound water molecule (Wat 681d), Wat 484, Wat 411, Wat 356, and Wat 402 (see Figure 3 and Table II). Additional features include a branch (Wat 402, part of the water pentagon) and a water molecule (Wat 404) linking the two zinc-bound water molecules, therefore also connecting the two proton-relay networks.

To analyze the impact of protein-bound water molecules to the protein conformation during molecular mechanics optimizations, a corresponding refinement without internal and surface water molecules, including only the proximal zinc-bound water molecule (Wat 262p) with both AMBER and YETI, has been performed.

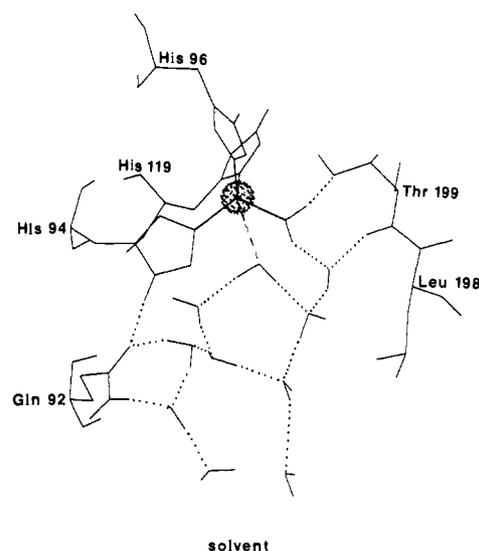


Figure 3. Schematic representation of the proton-relay network associated with the distal zinc-bound water molecule in native human carbonic anhydrase I. The zinc is represented by a sphere; hydrogen bonds are indicated by dotted lines.

Relative to the X-ray crystal structure,⁸ the change in backbone conformation is 50% larger (rms shift for all α -C atoms: 0.367 Å) if protein-bound water molecules are omitted compared to the refinement "in solution" (rms shift: 0.240 Å). Particularly large differences are observed at turns in the secondary structure on the protein surface. Figure 4 compares the changes in position of the α -C atoms of each amino-acid residue with the uncertainty in the atomic position associated with thermal motion (derived from the isotropic temperature factor; from ref 8 and 9). For the refinement including the 503 water molecules (as generated by SOLVGEN), only 21 of 256 residues show a shift larger than $(\bar{U}^2)^{1/2}$, whereas for the refinement without water molecules, 96 residues show a larger shift. This strongly underlines the need for protein-bound water molecules in molecular mechanics studies.

Calculations with a hydroxide ion (in place of Wat 681d) occupying the distal position at the metal were considerably complicated by the high negative partial charge of the hydroxide O atom. The strong electrostatic attraction between the zinc and the hydroxide O atom leads to an unreasonably short Zn...O distance of 1.687 Å which is not in agreement with structural data (Zn...O distances 1.89–2.05 Å, mean 1.98 (5) Å for five-coordinated zinc complexes; cf. ref 19). A reasonable (4 + 1) coordination (with the OH⁻ ion occupying a true distal position)

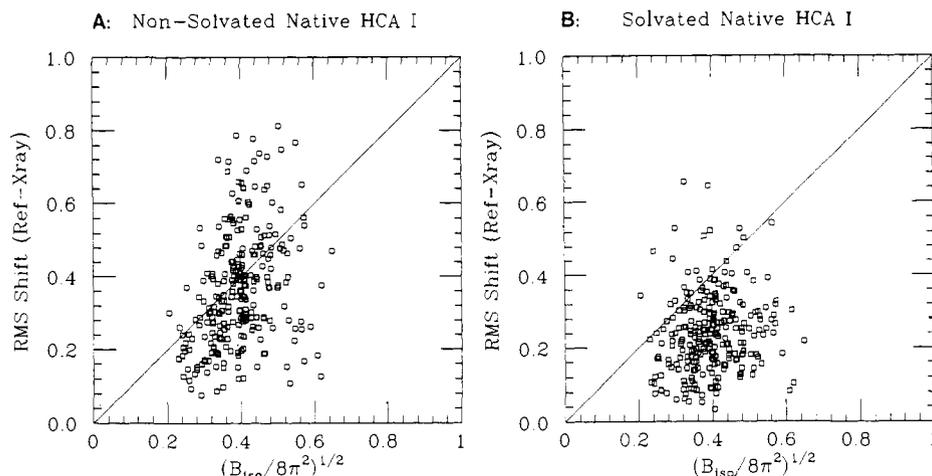


Figure 4. Comparison of C- α atomic shifts from molecular mechanics refinements (AMBER) and uncertainty in atomic position due to thermal motion (derived from the isotropic temperature factor obtained from the x-ray crystal structure determination; cf. ref 8 and 9): (A) refinement without solvent molecules (including only the proximal zinc-bound water molecule: W 262p); (B) refinement including 503 solvent molecules.

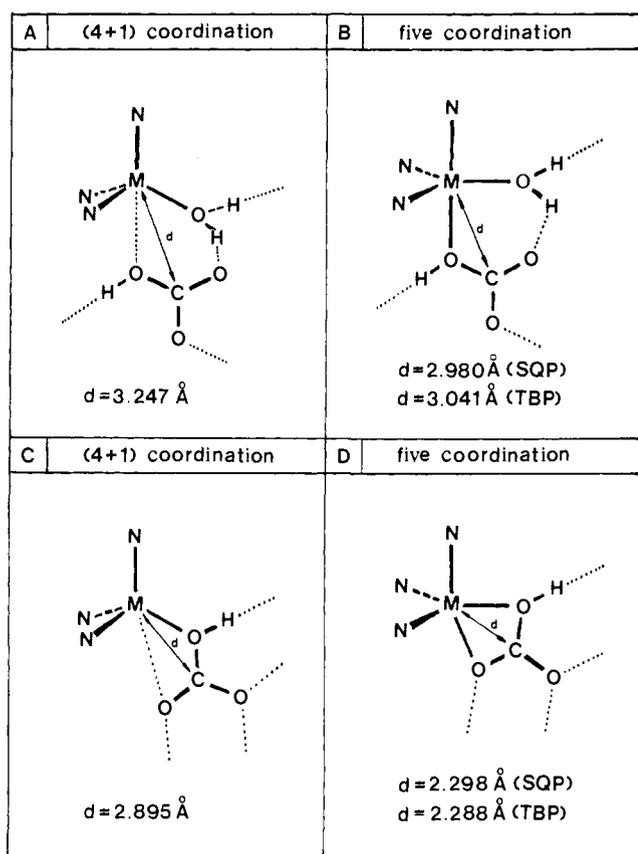


Figure 5. Comparison of the various analyzed binding modes of the bicarbonate ion to human carbonic anhydrase I. The indicated Zn...C distances are compared with the experimental value of 3.2 Å obtained by a ^{13}C NMR study on Co(II) carbonic anhydrase.³³ H bonds are indicated by dotted lines.

could only be obtained using a distant-dependent dielectric parameter of $D(r) = 6r$ or higher which does not seem appropriate to model the rest of the protein. For $D(r) = 2r, 4r, 6r, 8r$, Zn...O distances of 1.69, 1.86, 2.66, and 2.75 Å were obtained, respectively. Presently, an unconstrained refinement does not seem possible. To define appropriate geometrical parameters, we plan to search the CSD²³ for four- and five-coordinate zinc fragments with at least one hydroxide ligand.

The HCAI-Bicarbonate Complex. Because of its molecular topology, the natural substrate bicarbonate can bind mono- or bidentate to the zinc in carbonic anhydrase. Recent experimental³ and ab initio studies³⁰ suggested that HCO_3^- does not displace

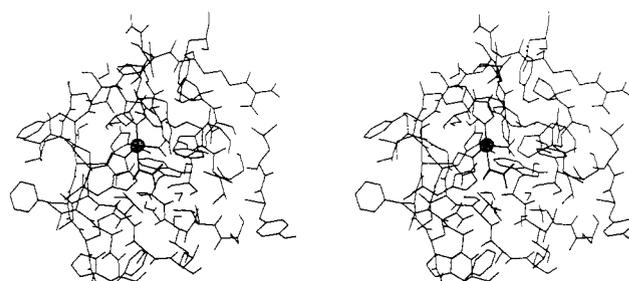


Figure 6. Stereoscopic view of the complex of HCA I and the natural substrate bicarbonate (details of the active-site region). The zinc is represented by a sphere; the bicarbonate molecule has been drawn enhanced.

the proximal zinc-bound water molecule (Wat 262p) and that a monodentate binding mode is more likely. In addition, a second bicarbonate O atom might engage in a H bond with the proximal water molecule (Wat 262p; cf. ref 3, Scheme 3).

For our study, the refined structure of the native enzyme has been used as starting model for the bicarbonate complex. The distal zinc-bound water molecule (Wat 681d) has been replaced by the substrate molecule, assuming that its deprotonated form attacks the CO_2 molecule, yielding bicarbonate. Two water molecules (Wat 484 and Wat 411) in the active site region were reoriented (rotation about their O atom) to include the bicarbonate molecule into the existing H-bond network.

Several possible binding modes of the bicarbonate ion to the catalytic zinc have been analyzed (cf. Figure 5), including four- and five-coordination combined with mono- or bidentate binding. Our calculations suggest that a monodentate (4 + 1) coordination represents the most stable arrangement. Here, the N atoms of His 94, His 96, and His 119 and the O atom of Wat 262p are defined as proximal ligands. The fifth ligand (i.e., the hydroxyl O atom of the bicarbonate ion) is only subject to electrostatic and van der Waals forces but is not treated by the directional metal-center function which allows its unbiased positioning. Table III gives some geometric details of this most stable arrangement; Figure 6 shows a close-up of the active site region.

For this arrangement (monodentate substrate binding and (4 + 1) metal coordination), the computed distance between the zinc and the bicarbonate C atom of 3.247 Å is in excellent agreement with ^{13}C NMR relaxation studies on Co(II) carbonic anhydrase which yielded an average Co...C distance of 3.2 Å.³³ This compares with 3.222 Å in this study (after correction for different

(33) Henkens, R. W.; Merrill, S. P.; Williams, T. J. In *Biology and Chemistry of the Carbonic Anhydrases*; Tashian, R. E., Hewett-Emmett, D., Eds. *Ann. N.Y. Acad. Sci.* 1984, 429, 143-145.

Table III. Details of Metal Coordination in Various HCA I-Bicarbonate (BIC) Complexes^a

(a) Zinc-Ligand Interactions: (4 + 1) Coordination, Monodentate						
N (His 94), Å	N (His 96), Å	N (His 119), Å	O (Wat 262p) proximal, Å	O (Bic 900) distal, Å	C (Bic 900), Å	
1.969	1.946	2.026	1.962	2.408	3.247	
(b) Zinc-Ligand Interactions: 5-Coordination, Monodentate						
N (His 94), Å	N (His 96), Å	N (His 119), Å	O (Wat 262p) proximal, Å	O (Bic 900) proximal, Å	C (Bic 900), Å	
SQP:	2.049	2.094	2.007	2.029	2.209	2.980
TBP:	2.021	2.083	2.018	2.028	2.248	3.041
(c) Zinc-Ligand Interactions: (4 + 1) Coordination, pseudo-bidentate						
N (His 94), Å	N (His 96), Å	N (His 119), Å	OH (Bic 900) proximal, Å	O (Bic 900) distal, Å	C (Bic 900), Å	
1.978	1.952	1.965	1.911	3.060	2.895	
(d) Zinc-Ligand Interactions: 5-Coordination, Bidentate						
N (His 94), Å	N (His 96), Å	N (His 119), Å	OH (Bic 900) proximal, Å	O (Bic 900) proximal, Å	C (Bic 900), Å	
SQP:	2.073	2.050	2.056	2.011	1.976	2.298
TBP:	2.052	2.041	2.067	1.993	1.948	2.288
(e) H-Bond Interactions: (4 + 1) Coordination, Monodentate						
donor	acceptor	Don...Acc, Å	H...Acc, Å	Don-H...Acc, deg	H...Acc-LP, deg	
Bic 900	O-H...O	Wat 402	2.80	1.80	170	10
Wat 484	O-H...N	Bic 900	2.65	1.82	143	34

^a Interactions of proximal and distal ligands with the zinc for the (4 + 1) coordination (monodentate binding); (b) interactions of ligands with the zinc for a pure five-coordination (monodentate binding); (c) interactions of proximal and distal ligands with the zinc for the (4 + 1) coordination (pseudo-bidentate binding); (d) interactions of ligands with the zinc for a pure five-coordination (bidentate binding); (e) H-bond interactions with water molecules for the (4 + 1) coordination (monodentate binding).

effective ionic radii of Zn(II) and Co(II); from ref 34). Alternative binding modes (see below) all resulted in shorter Zn...C distances (cf. Table III and Figure 5) and are therefore in less agreement with the NMR experiment. A ¹³C NMR study on Mn(II) carbonic anhydrase yielded a Mn...¹³C distance of 2.71 ± 0.03 Å.³⁵ Since Mn(II) carbonic anhydrase displays only about 4% of the activity of the Zn(II) enzyme, and Mn(II) in contrast to Zn(II) and Co(II) does not stabilize five-coordinated complexes particularly well (cf. ref 36), Mn(II) data should not be used to discuss the zinc enzyme.

In the monodentate (4 + 1) coordination, the bicarbonate ion also engages in H bonds with Wat 402 (H...O, 1.80 Å; O-H...O, 170°; H...O-LP, 10°) and Wat 484 (H...O, 1.82 Å; O-H...O, 143°; H...O-LP, 34°; both water molecules are part of a proton relay network (cf. Table II)); however, a strong interaction with the protein other than with the zinc (e.g., Thr 199) could not be found. The proximal zinc-bound water molecule (Zn...O, 1.962 Å) remains H-bonded to the side-chain hydroxyl O atom of Thr 199 (H...O, 1.76 Å; O-H...O, 156°; H...O-LP, 34°). According to our calculations, the H bond between the bicarbonate molecule and the proximal zinc-bound water molecule (Wat 262p) is very weak (H...O, 2.36 Å; O-H...O, 94°; H...O-LP, 63°).

For a pure five-coordination (with the bicarbonate still binding monodentate but its hydroxyl O atom also occupying a proximal position at the metal), the Zn...C distances are calculated to be 2.980 and 3.041 Å (for square-pyramidal and trigonal-bipyramidal arrangements, respectively), which are somewhat shorter than the experimental value (cf. Table IIIB).

We have also analyzed arrangements assuming only one zinc-bound water molecule at the beginning of the catalytic cycle. Again, the bicarbonate ion could bind mono- or bidentate to the zinc, leading to a four- or five-coordinated species. For the four-coordination, a Zn...C distance of 2.895 Å was obtained; for

a five-coordination, distances of 2.298 Å (square-pyramidal arrangement) and 2.288 Å (trigonal-bipyramidal arrangement), respectively. In all three arrangements, the bicarbonate ion would engage in two H bonds with Thr 199: one with the side-chain hydroxyl O atom, the other with the main-chain amide N atom.

HCAI-Acetazolamide Complex. Calculations have also been made for the complex of HCA I with acetazolamide (2-acetamido-1,3,4-thiadiazole-5-sulfonamide). ¹⁵N NMR experiments of complexed carbonic anhydrase have shown that (sulfon)amides bind via a deprotonated N atom to the zinc.³⁷ However, there is no direct experimental evidence whether the sulfonamide group binds monodentate (through the N atom only) or bidentate (through the N and one O atom) to the zinc.

Again, the refined structure of the native enzyme has been used as starting model for the acetazolamide complex. The two zinc-bound water molecules are displaced by the inhibitor molecule, leaving 501 protein-bound water molecules. Eight water molecules in the active site were reoriented to embed the inhibitor molecule in the existing H-bond network. Subsequent molecular mechanics refinements with both AMBER and YETI showed that a (4 + 1) coordinated zinc represents a more stable arrangement than any five-coordinated complex (trigonal bipyramid or square pyramid) which could only be obtained using a strongly damped dielectric forces, i.e., $D(r) > 4r$.

In agreement with experimental data (cf. ref 11; see also ref 14, 37), three key interactions of the sulfonamide group with the enzyme are observed in our calculations. The deprotonated sulfonamide N atom coordinates to the fourth coordination site of the zinc ($d = 1.904$ Å) and replaces the proximal zinc-bound water molecule (Wat 262p) from the native enzyme, thus interrupting one possible proton-relay network. One O atom (the second "tooth" of the sulfonamide group) binds at a fifth coordination site ($d = 3.131$ Å) and replaces the distal zinc-bound water molecule (Wat 681d), also disabling the second H-relay network. The sulfonamide moiety engages in two H bonds with Thr 199: one from the sulfonamide N atom to the side-chain

(34) Huhey, J. E. *Inorganic Chemistry - Principles of Structure and Reactivity*; Harper & Row: London, 1975; pp 74, 75.

(35) Led, J. J.; Neesgaard, E. *Biochemistry* **1987**, *183*, 183-192.

(36) Cotton, F. A.; Wilkinson, G. *Inorganic Chemistry*, 2nd ed.; Wiley: New York and London, 1972; p 781.

(37) Rogers, J. T.; Mukherjee, J.; Khalifah, R. G. *Biochemistry* **1987**, *26*, 5672-5679.

Table IV. Details of Acetazolamide (AAA) Binding to HCA I^a

(a) Zinc-Ligand Interactions						
N (His 94), Å	N (His 96), Å	N (His 119), Å	N (AAA 900) proximal, Å	O Å 900) distal, Å		
1.988	1.975	1.991	1.904	3.131		
(b) H-Bond Interactions						
donor	acceptor	Don...Acc, Å	H...Acc, Å	Don-H...Acc, deg	H...Acc-LP, deg	
AAA 900	N-H...O	Thr 199	2.80	1.86	154	21
Thr 199	N-H...O	AAA 900	2.91	1.97	152	118 ^b
Wat 404	O-H...N	AAA 900	2.87	1.92	172	27
AAA 900	N-H...O	Wat 707	2.87	1.86	173	6
(c) Hydrophobic Interactions						
inhibitor atom	residue, atom	distance, Å	inhibitor atom	residue, atom	distance, Å	
thiadiazole S atom	His 94, CD2	3.422	thiadiazole 4-N atom	Leu 198, CG	3.542	
	Ala 121, CB	4.344		Leu 198, CD1	3.428	
	Leu 198, CD2	3.704		His 200, CD2	3.348	
thiadiazole 3-N atom	Leu 198, CG1	3.503	acetamido carbonyl C atom	Phe 91, CZ	3.675	
	His 200, CD2	3.161	acetamido methyl C atom	Phe 91, CZ	3.740	
(d) Intramolecular Interactions						
inhibitor atom	inhibitor atom	distance, Å				
thiadiazole S atom	acetamido O	2.811				
	sulfonamide O	3.111				

^a(a) Interactions of proximal and distal ligands with the zinc; (b) H-bond interactions with protein and water molecules; (c) hydrophobic interactions; (d) intramolecular interactions of the acetazolamide molecule. ^bIf the sulfonamide O engages only in one H-bond interaction, no definite lone-pair direction can be assumed at the H-bond acceptor atom (cf. ref 13). In those cases, YETI optimizes the angle at the acceptor atom (angle S=O...H) instead.

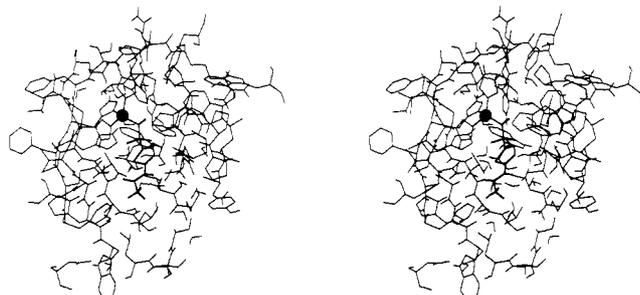


Figure 7. Stereoscopic view of the complex of HCA I and acetazolamide (details of the active-site region). The zinc is represented by a sphere; the inhibitor molecule has been drawn enhanced.

hydroxyl O atom (N-H...O), the other from the second sulfonamide O atom (not coordinated to the zinc) to the main chain amide N atom (S=O...H). A close-up of the active-site region is shown in Figure 7; geometrical details are given in Table IV.

The heterocyclic ring of acetazolamide interacts with residues Leu 198 and His 200, the 3-N atom of the thiadiazole ring engages in a water mediated H bond (Wat 404) with the amide O atom of His 200. The thiadiazole S atom seems not to interact significantly with the protein as found for the acetazolamide complex with HCA II (cf. ref 12, Figure 12) where residue 121 is a valine in HCA II compared to the smaller alanine in HCA I. The main interaction of the thiadiazole S atom is of intramolecular type: a very strong intramolecular interaction with the acetamido O atom ($d = 2.811$ Å) which is also found in the X-ray crystal structure of molecular acetazolamide ($d = 2.751$ Å).²² The acetamido side chain is almost coplanar with the thiadiazole ring (torsion angles S-C-N-C of 8.1° and C-N-C-C of 14.2°, respectively) and engages in H-bond interactions with surrounding water molecules. For example, the amide N atom is linked with the amide O atom of His 200 over two mediating water molecules (Wat 707 and Wat 625).

Conclusions

Molecular mechanics calculations on native and complexed human carbonic anhydrase I based on the 2.0-Å resolution X-ray

crystal structure and including 2 zinc-bound, 16 internal, and 485 surface water molecules lead to a deeper insight into possible mechanisms of catalysis and inhibition of the enzyme with respect to H-bond network formation and metal-ion coordination.

For both native and complexed human carbonic anhydrase I, our calculations suggest a slightly (4 + 1) distorted tetrahedral arrangement at the catalytic zinc, with a fifth ligand binding to a distal position. From the structural point of view, a pure five-coordination could also be considered but it is associated with a less favorable force-field energy, and the resulting geometries are in less agreement with experimental data. In the native enzyme, two proton-relay networks linking the catalytic zinc to the protein surface could be identified: one, first described in this study, originating at the zinc-bound water molecule in the distal position. Both proton relay networks remain intact when the natural substrate bicarbonate is coordinated to the metal, but are disrupted upon sulfonamide inhibitor binding.

Directionality has proven to be a valuable concept for both the subtle evaluation and comparison of metal-ligand and H-bond energies (program YETI) as well as for the generation of oriented internal and surface water molecules of proteins (program SOLVGEN). Protein-bound water molecules are doubtless of greatest importance for modeling studies on biological systems; of particular interest are internal water molecules that form water-mediated hydrogen bonds, cross-linking individual residues. A comparison of the results from molecular mechanics optimizations with the X-ray crystal structure showed that a significantly better agreement could be obtained when protein-bound water molecules were included in the calculations.

Acknowledgment. The authors wish to express their gratitude to Professor Hans Dutler (ETH Zürich, Switzerland), Dr. Radja G. Khalifah (University of Kansas School of Medicine, Kansas City, Missouri), and Dr. Jim P. Snyder (Searle, Research & Development, Skokie, Ill.) for stimulating discussions. We gratefully acknowledge financial support from the University of Kansas and the Swiss "Foundation for the Replacement of Animals in Medical Research" (FFVFF, Zürich).

Registry No. HCAI, 9001-03-0; HiS, 71-00-1; Zn, 7440-66-6; H₂O, 7732-18-5.