Ab Initio Quantum Chemical Study of the Cobalt d–d Spectroscopy of Several Substituted Zinc Enzymes

David R. Garmer[†] and Morris Krauss[*,‡]

Contribution from the Department of Physiology and Biophysics, Mt. Sinai Medical Center, New York, New York 10029, and Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received March 26, 1993[⊗]

Abstract: The visible spectroscopy of cobalt-substituted zinc enzymes provides insight into the electronic and structural environment of the active site. Ab initio calculations of the d–d spectra of model complexes of cobalt are applied to an analysis of the following enzymes: liver alcohol dehydrogenase (LADH), carboxypeptidase A (CPA), thermolysin (TLN), and carbonic anhydrase (CA). For the well-characterized active site of the LADH resting state, the calculated spectra agrees well with experiment both in the infrared and visible. The spectra in the presence of metal-bound deprotonated water or alcohol is predicted and can be used to characterize this stage of the reaction mechanism. Similarities in the active sites and resulting spectra of CPA and TLN are examined with models that explore the rotational flexibility of ligands implied by experimental differences in crystal and solution structures. Insight into the mechanism of anion inhibitor binding is obtained by comparing the spectra of CA with thiocyanate and cyanate. Experimental and theoretical spectra are in good agreement for thiocyanate, but the theoretical cyanate spectra does not fit experiment. The spectra is found to resemble more closely the high pH form of the uninhibited CA which is in accord with the recent X-ray crystallographic observation that the cyanate binds in the second shell of the zinc active site. However, the present suggestion that the water bound to the cobalt is deprotonated leads to a different interpretation of the mechanism of binding of inhibitors like cyanate than proposed earlier. Although the cyanate in the second shell has a small effect on the d–d spectra of cobalt, the deprotonated water suggests that desolvated cyanate abstracts the proton from water.

Introduction[1]

Cobalt substitution is a frequently used experimental tool for probing the nature of zinc binding sites in enzymes. Co^{2+} is the transition metal ion with useful spectroscopic properties that is least expected to perturb the enzyme structure and function.[1] Cobalt carbonic anhydrase (CoCA), for example, has substantially the same enzymatic activity as does the native zinc. The cobalt ion electronic absorption spectra has been used to suggest the coordination number of the first shell of ligands and the residue types that are present. The spectral perturbations are used to analyze the binding properties of inhibitors and substrates and the effects of temperature and pH on active sites in relation to activity.[2] Analysis of the visible spectra of CoCA is judged to be capable of discriminating between tetra- and pentacoordination based on analysis of characterized protein binding sites and inorganic complexes.[1,3] The coordination number at the active site will determine the number of waters that can bind and solvate reactants, transition states, and products and may, therefore, be important to the elucidation of the mechanism.

Ligand-field theory can predict the pattern of energy splittings for high symmetry and known geometry but can be ambiguous for the low-symmetry structures found in proteins.[4] In limited studies on the active site of carbonic anhydrase we have recently found that the ab initio configuration interaction method can approximately reproduce details of the d–d spectra of cobalt in different environments.[5] In this work an ab initio quantum chemical configuration interaction procedure is applied to the calculation of the excited state energies of several cobalt substituted enzymes. The goals are to check the spectral assignments of the experimentalists from consideration of appropriate active site models and to make predictions about the spectra of important geometrical configurations and chemical states at the metal binding sites which are suggested but not well-characterized by experimental probing.

The key issue for a successful theoretical application is the use of a configuration interaction approximation treating the excited states at the same level of accuracy as the ground state. This is relatively easy to accomplish as described in Methods because the d–d excitations are the important spectral features. These occur for the cobalt dication in a lower range of energies than any other states involving the binding site such as the charge transfer excitations. Therefore, the relevant excited states can be described properly using the same occupied orbitals as the ground state, but with different electron occupation patterns.

The present work extends the study of carbonic anhydrase to a further analysis of anion inhibitor binding and to other zinc enzymes whose active site is not completely characterized. Since information about the binding site first-shell structure obtained from crystallographic reports[6] has been used to refine the structural models as described below, there is hope of predicting the d–d pattern accurately for specific binding site models. In addition to testing this proposition, we have constructed models and simulated spectra for chemical states of the binding sites which have not been well-characterized. For carbonic anhydrase inhibitor binding new interpretations of the experimental information of inhibitor binding provide consistency between the simulated spectra and observed experimental properties. In addition to carbonic anhydrase, the active sites of the following enzymes have been considered: liver alcohol dehydrogenase

[†] Mt. Sinai Medical Center.
[‡] Center for Advanced Research in Biotechnology.
[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1993.

(1) Bertini, I.; Luchinat, C. *Acc. Chem. Res.* **1983**, *16*, 272.
(2) Bertini, I.; Luchinat, C. *Met. Ions Biol. Syst.* **1983**, *15*, 101.
(3) Bertini, I.; Luchinat, C.; Scozzafava, A. *Struct. Bond.* **1982**, *48*, 46.
(4) Bertini, I.; Luchinat, C. *Adv. Inorg. Biochem.* **1984**, *6*, 71.
(5) Garmer, D. R.; Krauss, M. *Int. J. Quantum Chem.* **1992**, *42*, 1469.
(6) (a) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535. (b) Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J. Protein Data Bank. In *Crystallographic Databases-Information Content*; Software Systems, Scientific Applications; Allen, F. H., Bergerhoff, G., Sievers, R., Eds.; Data Commission of the International Union of Crystallography; Bonn/Cambridge/Chester, 1987; pp 107–132.

0002-7863/93/1515-10247$04.00/0

(LADH), carboxypeptidase A (CPA), and thermolysin (TLN). The spectra from structural sites in LADH and aspartate transcarbamylase will also be analyzed.

Specific tests of the effect of the protein environment are included in this work. The d–d spectroscopy is found to be relatively insensitive to this electrostatic perturbation. Therefore, the electronic spectra should be predominately determined by the structure of the first shell of ligands. Variations in structure within the first shell, such as changing bidentate carboxyl to unidentate, in contrast, have a very significant effect on the predicted spectroscopic properties.

Comparisons were made between the energetics of structural models for zinc and cobalt ions. Such comparisons may be necessary to reconcile some aspects of the experimental situations in which cobalt and zinc enzyme structures apparently differ. The modest differences in energetic properties suggest that comparative experimental analysis of zinc and cobalt forms of these enzymes would possibly be probing multiple chemical and structural species present in a reactive sequence.

Biological crystal structure data almost universally indicate coordination structures for zinc in which the first-shell ligands are distinct from the rest of the protein environment. Atomic contact distances and coordination polyhedra are similar to those found for model small-ligand clusters. These constraints reflect the balance of the very strong electrostatic, polarization, and charge transfer attractive interactions against the short-ranged repulsions due to exchange forces. Therefore, realistic models of the coordination suitable for spectroscopic evaluation may include only the functional groups acting as ligands for a first approximation.

Irregularities in coordination polyhedra are common for the enzyme binding sites in contrast to small-ligand clusters where high-symmetry forms predominate. The theoretical examination of the spectroscopy of the irregular, asymmetric binding sites is most easily carried out by direct solution of the Schrodinger equation to give energies and other properties of the important electronic states. The fundamental ab initio approach is chosen instead of the semiempirical approximations employed by other workers[3,4] because of the need to treat low-symmetry sites and the extensibility of the ab initio calculations to the inclusion of protein environment.

In an enzyme there is usually some hydrogen bonding of ligand functional groups to the rest of the peptide structure. The complex interactions of the metal-bound functional groups to the remainder of the protein may therefore cause first-shell energetics involving low-energy modes such as torsion angles not to be dominant in determining the detailed coordination structure. The approach used in this work to overcome this complication is to identify important low-energy modes in possible coordination structures and then determine the correct structural parameters from crystallographic information if possible. In a few cases, more than one geometry for a specific binding site model was generated and tested for spectroscopic and energetic properties. The details of the binding site models are discussed in the Methods section and with the results for the various proteins.

Methods

1. Generating Structural Models. Ab initio configuration interaction (CI) calculations were carried out to estimate the pattern of spectroscopic excited states observable with cobalt substitution of several zinc enzymes. Only the functional groups belonging to first-shell ligands were included in the quantum calculations. The molecular structures were energy minimized but with significant constraints derived from crystallographic geometries in the Protein Data Bank (PDB).[6a] The partial structure optimizations were carried out because the crystallography on proteins has limited resolution. The essentially random errors in such features as the metal–ligand bond lengths are usually of a magnitude which could introduce significant errors into the spectrum calculations. We also believe that more consistent energetics is obtained if the internal structures of the ligands are optimized with the basis set to be used for the CI calculations rather than using standard geometries. The relative orientations of various groups are less in doubt and the constraints in optimizations were applied primarily to maintain these orientations.

Table I. Zinc Representation in Structure Optimizations with a 28-Electron Core Compact Effective Potential[a]

CEP	A_k	n_k	B_k
V_f	−2.95556	1	0.73655
V_{s-f}	−0.81189	2	0.42740
	5.98306	0	0.55756
V_{p-f}	−0.55769	2	0.32955
	4.64785	0	0.40095
V_{d-f}	1.79370	2	0.50931
	2.22461	0	1.05839

Basis Set Energy Optimized for a 4s4p Atomic Configuration

shell	type	exponent	C_s	C_p
1	sp	0.93380	−0.240332	−0.037389
		0.15590	0.715374	0.489346
		0.04865	0.442231	0.614419

[a] The CEP functional form is given in ref 7.

Partial optimizations were performed using the CEP-31G basis set[7] for ligand atoms and a zinc core effective potential (CEP) and basis set listed in Table I derived using the same methods. The constrained optimizations at the Hartree–Fock level were performed conveniently using the Gaussian program package.[8] Optimizations with cobalt were not used because the near degeneracy occurring with the open-shell ground state causes the convergence of the Hartree–Fock methods to be very undependable. Often only a particular choice of initial orbital guess and convergence methods succeeds for a particular geometry which makes automatic geometry optimization impractical. The characteristics of the structures calculated with these small basis sets are qualitatively the same in most respects as if larger, polarized sets are used. The latter are impractical to apply to many of the large coordination structures used for this work for optimizations and CI calculations. Significant differences in structure occur only along low-energy modes of the systems and we have in any case made an effort to identify these and examine their effect on the spectra. A similar statement can be made concerning the effect of differences between optimized structures using zinc or cobalt.

The structural parameters taken from the crystallography are those that determine relative orientation of ligands in the first shell. This constrained set generally included the angles between the liganded atoms. For example, the optimization to give the structure illustrated in Figure 1a had the O–Zn–S, O–Zn–N, and S–Zn–N angles fixed and all other independent angles optimized.

Constraints were applied forcing an imidazole ring-plane orientation to match its particular crystal structure. The two independent C–N–Zn–X torsions involving a path through chemical bonds were frozen, where X represents one other ligand atom. In the structures having several imidazole ligands the number of basis functions usually exceeds the practical limits of the present computer code for the CI calculations. In these cases we substituted ammonia ligands for imidazoles and used the STO-3G basis set for the ammonia atoms. Our earlier study of carbonic anhydrase[9] had indicated that STO-3G ammonia made a good electronic mimic of the CEP-31G imidazole ligands. Direct comparisons between the imidazole and ammonia models in spectroscopic calculations are described below in the first section on LADH, showing that realistic results are still obtained with the ammonia model. The torsional angle of one proton in each ammonia was fixed with respect to another ligand so that obvious hydrogen bond contacts with other ligands would be disfavored.

Cysteinate ligands were represented either by HS^- or CH_3S^- groups, with direct comparisons also included below in the first results section. The sulfur ligation always has a sharp angle about the sulfur atom due to the lone pair directionality. Therefore, the ligand field presented by a cysteinate ligand should be quite asymmetric about the Co–S axis. The torsion angle C_B,H–S–Zn–X was constrained at the crystal structure value for each ligand to produce a qualitatively correct ligand field. The C_B,H–S–Zn angle was free to optimize, producing values similar to the crystallography. When the ligand model was a CH_3S^- group, an H–C_B–

(7) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026.
(8) Frisch, M. J.; Head-Gordon, M.; Trucks, G. W.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M. A.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R. A.; Melius, C. F.; Baker, J.; Martin, R. L.; Kahn, L. R.; Stewart, J. J. P.; Topiol, S.; Pople, J. GAUSSIAN 90, Gaussian, Inc., Pittsburgh, 1990.
(9) Krauss, M.; Garmer, D. R. *J. Am. Chem. Soc.* **1991**, *113*, 6426.

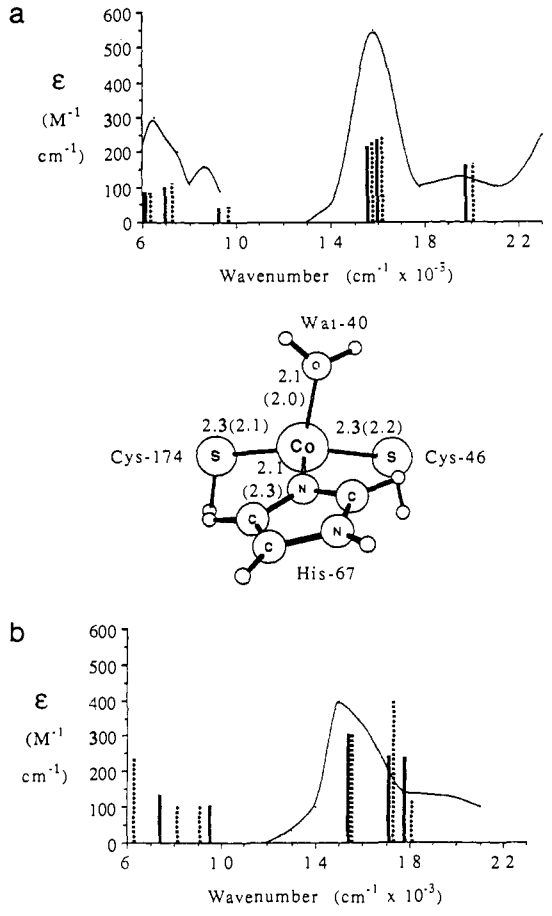

Figure 1. (a) Simulated spectrum for liver alcohol dehydrogenase with cobalt substitution at the active site using the structural model shown (solid bars) or using two methiolates and ammonia (dotted bars). The ^4P bands are shifted down by 4500 cm^{-1}. The experimental visible spectrum region is from ref 14 with the near-infrared spectrum for the completely substituted form from ref 15. The optimized metal–ligand contact distances are indicated along with the crystallographic estimates in parentheses. (b) Spectrum simulations with deprotonated water from the structure in part a with hydroxyl proton orientation optimized (dotted bars) or directed toward Ser-48 (solid bars). The ^4P bands are shifted down by 4500 cm^{-1}. The experimental spectrum is from ref 17 with coenzyme H_2NADH bound.

S–Zn torsion was constrained to the crystallographic C_A–C_B–S–Zn value thus positioning the C_B methylene hydrogens in correct locations relative to the ion. This torsion angle was always greater than 120° in absolute magnitude so that C_A was farther from the ion than the C_B methylene hydrogens.

Another group requiring special treatment is the carboxylate from Glu residues. The carboxylate has a variable degree of unidentate to bidentate binding in the sites examined here. Matching this crystal structure characteristic required freezing the Zn–O_A–O_B angle plus a O_B–O_A–Zn–X torsion to avoid rotations of the group during optimization. Here O_A designates the carboxylate oxygen more closely coordinated to zinc. Metal–ligand distances from the optimization and crystallography are drawn in the figures to show that these constraints produced an appropriate level of agreement.

With coordinated water or hydroxide, the positions of the protons are freely determined by the energy minimization step unless specifically noted in a results section. The internal contacts which can be made with lone pairs of other ligands are dominant in a first-shell optimization when a constraint is not applied. This effect is in reality altered by contacts with water molecules which are always present near these binding sites, and many of the protein structures indicate clear hydrogen bond preferences to second-shell residues. This indeterminacy in proton positions is probably the most significant source of error in reproducing the true structural characteristics. However, each active site structure we have examined where this might be a problem appears to have a clearly preferred contact for the protons from first-shell forces. For example, in the carboxypeptidase A and thermolysin structures the coordinated water O–H bond is oriented toward the coordinated carboxylate, as produced by the energy minimization, in preference to orientations specifically toward one of the two imidazole ligands. In electrostatic terms, each of these structures has one or two protein ligands producing significant dipoles along their dication–ligand contacts. These dipoles attract the water molecule O–H bond dipoles. In like manner, these dipoles will tend to repel the lone pairs of coordinated hydroxyl and attract the hydroxyl proton. These internal orientation preferences do not represent optimal hydrogen bond-like interactions, but they would have to be overcome by external contacts in order to produce orientations differing from those we generate.

The frozen internal coordinates also represent low-energy modes within the first shell and are therefore determined partly by external forces such as second-shell interactions. Constraining the angles between ligands particularly aids in examining the five-coordinate structures which have very low energy modes representing variations between square-pyramidal and trigonal-bipyramidal extremes. The four-coordinate structures are all approximately tetrahedral. This feature would be reproduced in our energy optimization if the constraints on ligand angles were released.

2. Spectrum Simulations. The configuration interaction method was the same as in our previous study on carbonic anhydrase.[5] Molecular orbitals for a CI calculation were generated from a restricted open-shell Hartree–Fock calculation (ROHF). The cobalt CEP and basis set are from ref 10 with the ligand data from ref 7. Trials showed that the spectrum simulations were insensitive to the ligand basis set used but significantly sensitive to the ion–ligand distances.[5] This was a primary motivation for using an optimization step before the spectroscopic calculations to cancel out possible random uncertainties in distances obtained from the crystal structures.

All occupied orbitals were considered as active orbitals in the CI stage. The orbital excitation level was set as sufficient to generate all possible configurations from within this active orbital set. This is a minimal level of configuration interaction for treating all of the cobalt d-shell derived states in a balanced fashion. The Gamess code system was used[11] because it contained all of the necessary features such as convergence accelerators for the ROHF calculations, a flexible CI method, and transition moment calculations. The latter module was modified to operate with all of the excited states simultaneously in order to save computer time.

The coordination structures around high-spin Co^{II} produce seven low-lying electronic states connecting to the ^4F ion term. Previous results[5] and trials for this work showed that the minimal CI level produced ^4F energy states in good agreement with more extensive CI treatments. The stronger and more significant absorptions are to three high-energy states connecting to the ^4P ion term. Trials showed that the energy difference between the ground state and ^4P states decreases by a few thousand wavenumbers if extensions to the CI treatment such as first-order or second-order configuration interaction[5] are used. The results are then in better agreement with experiment. However, the experimentally significant internal pattern and intensities of the ^4P levels are not much affected. These results are to be expected since the primary effect of more extensive correlation should be to change the relative energies of different spin couplings of the electrons. Therefore, we can use the minimal CI level for calculations on the large, asymmetrical shell models generated from crystal structures. Our graphical presentation of the results includes indicated shifts of the ^4P energy levels to overlay in the best agreement with the experimental spectra. The comparison with better CI methods could be carried out only for the isolated cobalt ion and for high-symmetry coordination structures[5] because a more extensive treatment consumes much greater disk space. This is compounded by increased system size or lack of symmetry.

The intensities of the transitions were computed from a linear combination of the length and velocity forms of the oscillator strength integrals. These alternative algebraic forms are numerically equivalent only in the case of a complete configuration interaction treatment. The length form is considered to be empirically the more accurate for limited CI. However, for this problem we noted that the length form often overestimated the relative strength of the ^4F bands. The linear combination used instead weighted the length form by a factor of 250 over the velocity form. This linear combination was chosen to give the best global agreement between the intensities of the ^4F and ^4P levels. The velocity form has an effect on the results because its oscillator strength predictions are usually 10 to 100 times larger for the stronger transitions.

A constant relationship between the calculated oscillator strengths and the experimental extinction coefficients is used for all binding sites.

(10) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612.

(11) Schmidt, M. W.; Boatz, J. A.; Baldridge, K. K.; Kokeski, S.; Gordon, M. S.; Elbert, S. T.; Lam, B. GAMESS. *QCPE Bull.* **1989**, *7*, 115.

Therefore, we have the ability to judge how well intensities are calculated by comparing between spectral simulations as well as within one spectrum.

Results

Liver Alcohol Dehydrogenase (LADH). 1. Catalytic Site. This enzyme is one of a class of NAD^+-dependent zinc enzymes catalyzing hydride transfer from a substrate to NAD^+. A mechanism for the transfer has been proposed.[12] NAD^+ is first bound close to the active site. This lowers the pK of zinc-bound water sufficiently for deprotonation. A substrate alcohol is then deprotonated and rapidly undergoes hydride transfer to NAD^+. The existence of bound alcoholate is somewhat controversial because of the necessary very substantial pK lowering from solution. Therefore, our simulated spectra included structures representing bound alcoholates and hydroxyl in different conformations which could be compared with experimental spectra.

This enzyme represents the least complex spectroscopic experimental system and was also used by us to test structural approximations. Each enzyme unit has two nonequivalent zinc/cobalt binding sites, one of which is required for catalytic activity and another for structural stability.[13] The cobalt-substituted forms are also enzymatically active. Selective substitution by cobalt[14] has been used to produce distinct catalytic and structural site spectra for the ^{4}P region, but the ^{4}F region spectrum is available only for the totally substituted form.[15] The solution spectra are reported for a dimer; therefore we have divided the reported intensities by two to reproduce the experimental spectra in our figures. The crystal structure designated as 8adh[6,16] in the PDB (2.4-Å resolution) is the source of structural information about the binding sites.

The catalytic site has two coordinated cysteine residues, one histidine imidazole, and a coordinated solvent molecule which is very probably a water. We have prepared three spectral simulations of this site to test structural simplifications which were required for other enzyme simulations due to limitations in the computer resources and code. In each case, the water protons are directed toward sulfur lone pairs by the optimization step while maintaining the ion–oxygen contact as shown in Figure 1a. The crystal structure indicates a distorted tetrahedral binding site, in particular having a large S–Zn–S angle of 126°. We have carried out an optimization on a complex $Zn^{2+}(SH^-)_2$-OH_2imidazole without the usual constraints on angles between ligands. The optimized S–Zn–S angle at 134° is the only one greater than 109.5°, indicating that the repulsion between thiolates is qualitatively stronger than that between the other pairs of ligands and is probably responsible for this distortion in the enzyme binding site.

The first structure and spectral simulation given in Figure 1a used HS^- ligands and imidazole. The second also reported in Figure 1a used CH_3S^- ligands and ammonia with the latter ligand having an STO-3G basis instead of the CEP-31G used for all other ligands. A third calculation used HS^- ligands and STO-3G ammonia but we have not reproduced the spectrum here. All three simulated spectra have line positions which are equivalent to within about 300 cm^{-1} and comparable intensities so that they are nearly indistinguishable. The CH_3S^- and HS^- ligands are therefore spectroscopically equivalent as are imidazole and ammonia with the latter computed using the STO-3G basis. The simulations reproduce the ^{4}P experimental pattern and approximate intensities using an intensity scaling common to all enzyme calculations reported here. The ^{4}F experimental spectrum, which is a summation from cobalt catalytic and structural sites, is seen to be reproducible from the model of the catalytic site spectrum.

The atomic populations in this system are nearly unaffected by the d–d transitions as for the other cobalt systems we have studied.[5] The strong absorption rising at above 21 000 cm^{-1} is typical of charge transfer bands between cysteinate ligands and cobalt. The minimal CI treatment does not satisfactorily reproduce these transitions and efforts to increase the level of treatment were blocked by problems with the code and computer limitations.

In Figure 1b, predicted spectra for conformations in the high-pH form of the catalytic site are also given. The corresponding experimental spectrum has not been reported to our knowledge. The geometries were obtained by deprotonating water in the structure shown in Figure 1a. Our results showing that HS^- and CH_3S^- are spectroscopically equivalent would imply that the predicted hydroxyl spectra also represent possible alcoholate spectra. The dotted bars in Figure 1b indicate the results with no constraint on the hydroxyl proton orientation. In this case, the N–Co–O–H torsion is at approximately 180° with a Co–O–H angle of 133°, an orientation which probably serves to minimize repulsion between the hydroxyl and thiolate lone pairs. The solid bars represent the spectrum with the hydroxyl proton directed toward O_G of Ser-48 from an approximately 90° change in the torsion (proton now roughly toward Cys-46). Directing the proton away from Ser-48 results in intensity shifts but similar energetics because this is approximately a plane reflection in C_s symmetry as far as the first shell is concerned. The results in Figure 1b therefore represent the most likely conformational forms in which Ser-48 might act as hydrogen bond donor or acceptor to zinc-bound oxygen.

For the binding of deprotonated water or alcohol at the catalytic site, the simulations predict significant changes which would be indicative of this particular deprotonation as opposed to other processes not directly involving the coordinated residues. The observed spectrum should consist of two peaks separated by approximately 2000 cm^{-1}. This prediction appears not to be sensitive to the orientation of the proton or methyl group.

The experimental spectrum shown in Figure 1b results from binding H_2NADH to the enzyme at neutral pH. With this neutral coenzyme the experimental spectrum is more consistent with a structurally perturbed version of the resting state as proposed in ref 17. This spectra is similar to that in Figure 1a and does not resemble the predicted spectra for the deprotonated water or alcohol. Binding a positively charged coenzyme is hypothesized[12] to lower the pK of zinc-bound water.

As illustrated and explained below, the spectra are sensitive to changes outside the first shell only if they propagate a significant structural effect into the first shell. The non-coordinated N_{D1} of His-67 is hydrogen bonded to Asp-49 so that proton removal there would lead to a local protein structural disruption. Proton transfer to the Asp would probably not occur at high pH because the His already binds at least two anionic ligands. Therefore, a spectral shift resembling our calculated form would be clearly indicative of water deprotonation and also could be used to identify deprotonation of alcohol substrates.

2. Structural Site. The structural site with four cysteine sulfur atoms binding cobalt has an absorption spectrum as reproduced in Figure 2a. The spectral simulation used four SH^- ligands as a model. The rest of the side chain should be relatively unimportant as in the test comparison noted above. This simulation agrees approximately in pattern and intensity with the experimental spectrum. Only a small fraction of the ^{4}F absorption region at the lower frequencies is predicted to arise from the structural site.

The crystal structure has two of the reported zinc–sulfur distances 0.1 to 0.15-Å longer than the other two in the structural

(12) Tapia, O.; Cardenas, R.; Andres, J.; Krechl, J.; Campillo, M.; Colonna-Cesari, F. *Int. J. Quantum Chem.* **1991**, *39*, 767.
(13) Galdes, A.; Vallee, B. L. *Met. Ions Biol. Syst.* **1983**, *15*, 1.
(14) Maret, W.; Andersson, I.; Dietrich, H.; Schneider-Bernlöhr, H.; Einarsson, R.; Zeppezauer, M. *Eur. J. Biochem.* **1979**, *98*, 501.
(15) Drum, D. E.; Vallee, B. L. *Biochem. Biophys. Res. Commun.* **1970**, *41*, 33.
(16) Eklund, H.; Nordström, B.; Zeppezauer, E.; Söderlund, G.; Ohlsson, I.; Boiwe, T.; Söderberg, B.-O.; Tapia, O.; Brändén, C.-I.; Akeson, A. *J. Mol. Biol.* **1976**, *102*, 27.
(17) Dietrich, H.; Maret, W.; Wallén, L.; Zeppezauer, M. *Eur. J. Biochem.* **1979**, *100*, 267.

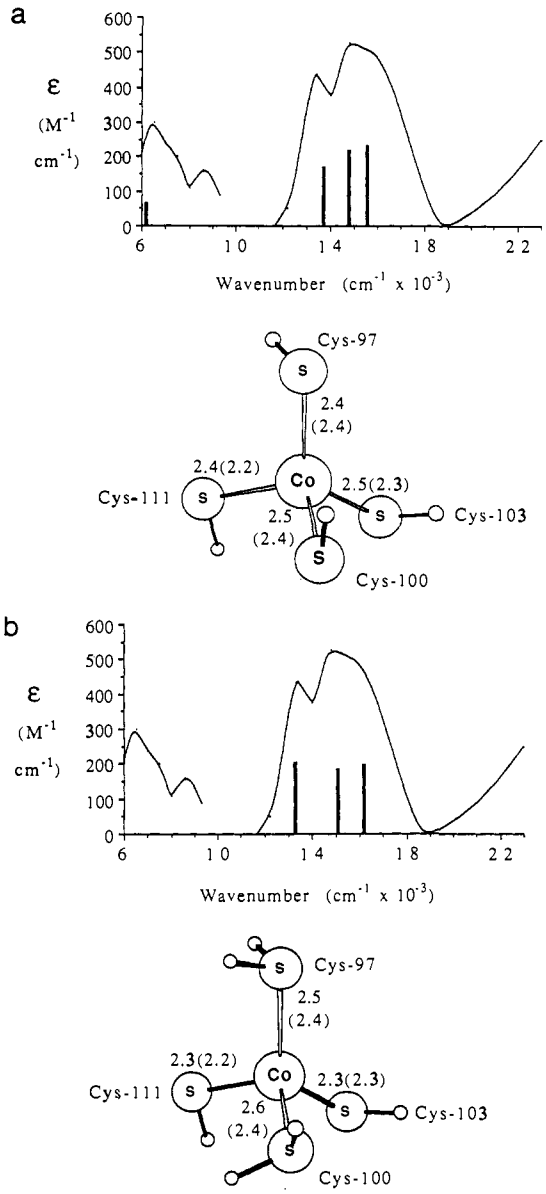

Figure 2. (a) Spectrum simulation for liver alcohol dehydrogenase with cobalt at the structural site using the model shown. The ^4P bands are shifted down by 5000 cm^{-1}. Experimental spectrum sources as in Figure 1a. (b) Spectrum as indicated in Figure 2a simulated using the diprotonated model shown. The ^4P bands are shifted down by 4500 cm^{-1}.

site. This is similar to the result obtained in *ab initio* structural optimizations using $Zn^{2+}(SH_2)_2(SH^-)_2$ as a model system. In this case, the protonated ligands have zinc–sulfur distances larger by approximately 0.25 Å. The ligands with the longer distances in the crystallography are exposed to solvent and therefore might be protonated to produce a structural effect as in the model system. We were also led to consider this possibility by the almost exact superimposibility of the catalytic and structural site charge transfer band structures.[14] Chemical intuition leads one to expect that this result is most likely if both sites have a comparable charge distribution with two anionic and two neutral ligands. However, as noted above, we have not been able to calculate the charge transfer band positions accurately to support this conjecture.

The d–d spectrum for this model has been calculated and is reported in Figure 2b. The structure was prepared as outlined in the Methods section. The additional protons were started off in positions directed toward solvent and remained in the same general orientation after optimization. The spectrum for this model is similar to that of the fully deprotonated structure with increased separation of the two higher energy bands from the lower energy transition at 13 000 cm^{-1}. Within the error limits

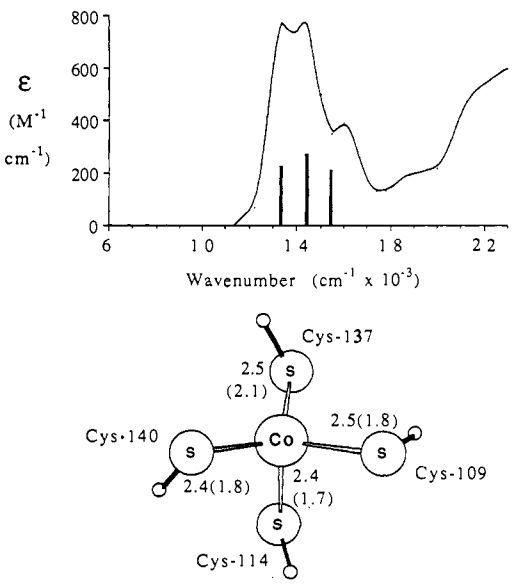

Figure 3. Simulated spectrum for cobalt aspartate transcarbamylase using the indicated structure. The ^4P bands are shifted down by 4500 cm^{-1}. The experimental spectrum is from ref 19.

of this type of calculation both spectra are in reasonable agreement with the experimental observation but the protonated model in Figure 2b has a pattern moving toward better agreement with experiment. The theoretical evidence is therefore suggestive of this alternative protonation state to what is usually assumed but does not give enough information to choose unambiguously.

Aspartate Transcarbamylase (Carbamoyl Transferase). This enzyme is considered because its metal binding site is stoichiometrically identical with the LADH structural site. Aspartate transcarbamylase is an oligomeric enzyme in which six equivalent zinc cations stabilize the assembly of subunits to form a fully active enzyme. A crystal structure was available in the PDB entry 2atc[6] for a subunit of an assembled enzyme at 3.0-Å resolution.[18] The low resolution was evident at the zinc binding site where three of the four contacts with cysteine sulfur atoms have unusual bond lengths of 1.7–1.8 Å. The other deposited structure at 2.5-Å resolution is for an allosteric form of the enzyme. It was decided not to use this form but rather to constrain values of only the H–S–Zn–X torsions from the native structure. The partial optimization is relied on to produce a reasonably strain-free and representative model with good S–Zn–S angles and Zn–S distances.

The geometry and spectrum are reproduced in Figure 3 for comparison with the experimental observation. Both indicate a closely spaced pattern of three ^4P bands. The spectral features above 18 000 cm^{-1} are probably due to charge-transfer excitations which the limited configuration interaction treatment cannot reproduce accurately. Near-infrared bands were reported for the nickel-substituted enzyme in the same study by Johnson and Schachman[19] but not for cobalt. However, it is not clear whether these workers were able to see such weak features as typical for cobalt ^4F bands or if these are absent at observable energies as predicted by the simulation.

These experimentalists noted for the nickel form that the charge-transfer bands were modified by particular allosteric shifts while the d–d bands were practically unaffected. The results we have obtained with modeling the environmental affects, discussed below for carbonic anhydrase, suggest that this result implies small or no change in the first-shell geometry. An external electrostatic shift, possibly from a change in the hydrogen bonding of the cysteine sulfurs, is indicated by the shift in the charge-

(18) Honzatko, R. B.; Crawford, J. L.; Monaco, H. L.; Ladner, J. E.; Ewards, B. F. P.; Evans, D. R.; Warren, S. G.; Wiley, D. C.; Ladner, R. C.; Lipscomb, W. N. *J. Mol. Biol.* **1982**, *160*, 219.
(19) Johnson, R. S.; Schachman, H. K. *J. Biol. Chem.* **1983**, *258*, 3528.

transfer bands. Only charge-transfer bands involve a substantial dipolar flow of charge from ligand to metal allowing for an outer-shell electrostatic influence on transitions.

The spectral simulations involving cysteine residues in this study (Figures 1–3) have apparently produced consistent underestimates of the ^4P intensities although the band patterns are reasonably accurate. This may be a consequence of the inadequacy of the configuration interaction treatment in representing the charge-transfer excitations. The d–d bands could conceivably borrow intensity from these very intense transitions when they are relatively close in energy to the upper ^4P states. Our treatment places the charge-transfer states too high in energy and so any mixing with the d–d excitations may be reduced. This may also be the physical origin for the consistent red shift of ^4P bands in the structures with cysteine ligands from the usual wavelengths which has been remarked on by experimentalists.[2] We have been unable to produce reasonable charge-transfer band positions or ^4P intensity changes by increasing basis set size or level of configuration interaction within the computational resources available. Charge-transfer excitations require treatment of the dynamical correlation as well as the degeneracies among the d electrons. The scale of the ab initio oscillator strengths to the experimental extinction coefficients was adjusted for the best average agreement throughout all spectra simulated.

We have noted that the experimental structures for the ZnS_4 binding sites used for this work are closely superimposable on the local minimum structures obtained in unrestrained optimizations of $Zn^{2+}(SH^-)_4$ or $Zn^{2+}(SCH_3^-)_4$ structures. The aspartate transcarbamylase structure has approximate S_4 symmetry at the $Zn(SC_B)_4$ group level. This is the global minimum conformation for the two types of model complexes. The liver alcohol dehydrogenase structural site is irregular and is superimposable on a local minimum energy structure for these models approximately 1–2 kcal/mol above the global minimum. No other local minima were observed and the internal energies for the $Zn^{2+}(SH^-)_4$ model can vary up to about 8 kcal/mol as a function of the torsional angles. Therefore, the enzyme binding sites appear to be arranged to minimize unfavorable interactions of the sulfur lone pairs leading to low internal energies. These forms are also among those that present a large surface area with lone pair density for external hydrogen bonding.

Carboxypeptidase A (CPA) and Thermolysin. These enzymes are discussed together because both hydrolyze peptides using a zinc binding site and a common set of first-shell ligands. The active sites of both have been characterized by X-ray crystallographic studies reported in PDB entries 5cpa at 1.54-Å resolution and 3tln for thermolysin at 1.6-Å resolution.[6] Zinc is bound by two histidine imidazoles and a glutamate carboxyl group with a fourth site occupied by a probable water molecule. The first-shell carboxyl group was indicated to bind in a unidentate fashion in TLN[20] and close to ideal bidentate configuration in CPA.[21,22] The observed spectra in solution with cobalt substitution reproduced in Figure 4 for CPA and Figure 5 for TLN are significantly different. The near-infrared energy range spectrum of TLN was not reported in ref 23.

The optimized structures shown are superimposable on the crystal structures due to the constraints applied as described above in Methods. The CPA structures were generated with a water proton oriented to contact Glu-270 oxygens as suggested by the apparent hydrogen bond available in the crystal structure. The TLN structure has no such clear hydrogen bonding pattern and therefore no constraint was applied to a water proton. Both crystallographic reports indicate a shorter distance between the active site water and a ligand carboxyl oxygen compared with the distances of the water from the imidazole ligands. This may indicate that an electrostatic attraction between a water proton and carboxyl in the tripod is structurally significant.

The calculated spectra (solid bars in Figures 4a and 5) predict

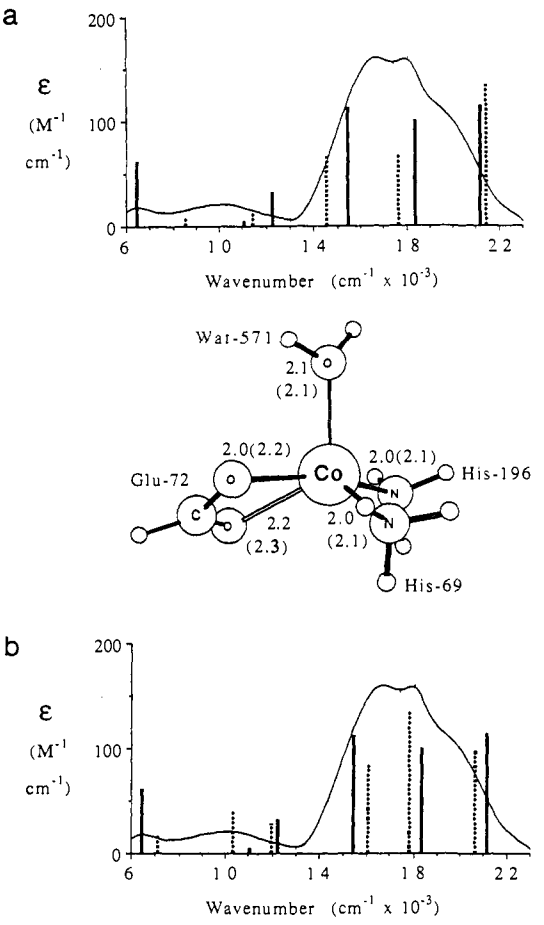

Figure 4. (a) Simulated spectrum for cobalt carboxypeptidase A using the indicated structure with bidentate formate (solid bars) or with formate rotated perpendicular to its plane to make unidentate binding (dotted bars). The ^4P bands are shifted down by 4500 cm^{-1}. The experimental spectrum is from ref 26. (b) Experimental spectrum as in part a with the crystal structure simulation using water (solid bars) or hydroxyl (dotted bars). The ^4P bands are shifted down by 4500 cm^{-1}.

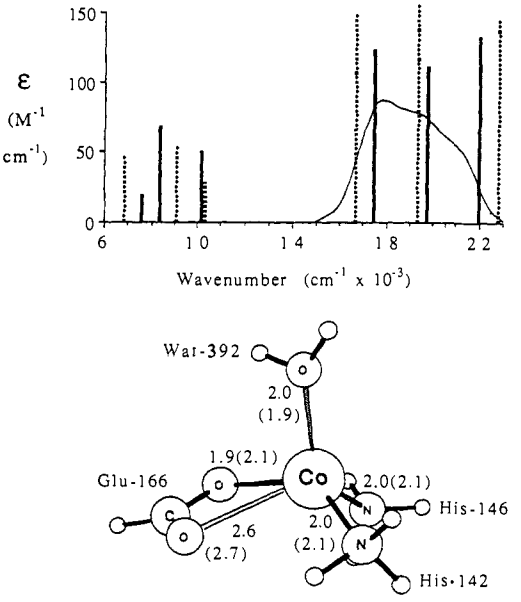

Figure 5. Simulated spectrum for cobalt thermolysin using the indicated structure with water (solid bars) or hydroxyl (dotted bars). The ^4P bands are shifted down by 3500 cm^{-1}. The experimental spectrum is from ref 23.

(20) Holmes, M. A.; Matthews, B. W. *J. Mol. Biol.* **1982**, *160*, 623.
(21) Rees, D. C.; Lewis, M.; Lipscomb, W. N. *J. Mol. Biol.* **1983**, *168*, 367.
(22) Rees, D. C.; Lewis, M.; Honzatko, R. B.; Lipscomb, W. N.; Hardman, K. D. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 3408.
(23) Holmquist, B.; Vallee, B. L. *J. Biol. Chem.* **1974**, *249*, 4601.

a spread of ^4P peaks consistent with the observed spectra but the intensities between and within each simulation do not agree well with experiment. The spectrum simulations are also more similar than the experimental spectra are to each other. The possible reason for these discrepancies is indicated in the results of a recent XAFS study of CFA[24] in which the solution structure was deduced as a unidentate binding configuration of Glu-72 instead of the bidentate contact in the crystal structure. The solution structure was also indicated to have a somewhat disordered active site which would give rise to a broadened absorption spectrum. The characteristics of the zinc and cobalt enzymes from the XAFS study were very similar in solution or in the crystal.

The model structure for CPA could be adjusted to be in better accord with the XAFS prediction. The $Zn–O_A–O_B$ angle constraint applied to the carboxyl group in generating the CPA active site model was modified to lengthen the longer $Zn–O_B$ contact distance. The unidentate contact so produced had dication–oxygen distances approximately as reported in the XAFS study. The predicted spectrum for this structure is also indicated in Figure 4a (dotted bars) and it is in poorer agreement with the experimental spectrum. Therefore, the structural change detected by XAFS may also involve a rotation of the carboxyl plane as well as the indicated change in the carboxyl oxygen distances. The spectral simulations indicate that the observed structural difference could produce a noticeable effect on the cobalt spectroscopy in crystal versus solution. A similar effect is worth looking for with TLN since the simulation is only in very rough agreement with the observation for this case also.

In examining this problem we noted that there is an alternative model of the active site structural behavior which could be consistent with the crystallographic and XAFS data. The crystallographic report[21] indicated a possible second site or disordered structure for the electron density associated with Wat-571. The second site could be interpreted as having a very high B value and roughly 3.2-Å distance from zinc at the center of density. The XAFS result for a solution structure shows four atoms at a distance of 2.1 Å and another at 2.5 Å. Instead of the latter peak representing a Glu-72 oxygen, it could be a result of Wat-571 located at unusually large distances at least with partial occupancy as in the crystallographic interpretation. With the effective loss of one ligand, the dication might also attract the glutamate into a tighter bidentate configuration to fill up its coordination sphere. The solution structure in the XAFS study could therefore be interpreted as predominately of the site 2 type (ref 21, p 385) with only three tightly bound ligands and the crystal structure as of the site 1 type with the more usual four ligands.

There is currently some interest in the effect of pH on the active site of CPA due to observed dependencies of binding constants and spectroscopies on pH with various species.[25] This is an issue that is directly relevant to competing mechanistic proposals for the enzymatic hydrolysis function of both CPA and TLN. The zinc hydroxide mechanism requires that a water proton is removed before or after substrate binding so that the hydroxyl anion can attack a substrate carbonyl. Figures 4b and 5 show also the predicted spectra for the inactive high-pH form of each enzyme in which the active site water is probably deprotonated. The comparable CPA spectrum was described by Latt and Vallee[26] as having a central peak with surrounding shoulders. The predicted spectrum is reasonably consistent with this kind of band structure.

The fact that the high-pH forms are inactive might not necessarily rule out a variant of the zinc hydroxide mechanism in which the water is deprotonated simultaneously with the attack of the incipient hydroxyl on substrate. The high-pH form with zinc hydroxide may simply experience an interaction with peptides similar to that of carbonic anhydrase in which an amide can be deprotonated and bind through nitrogen as an anion. It was recently proposed that the zinc hydroxide of carbonic anhydrase acts to deprotonate an amide, followed by a rearrangement to expel the water molecule produced.[27] The peptide-hydrolysis enzymes may therefore be required to have a higher pK_a for bound water without substrate to avoid this dead-end pathway with peptides.

Carbonic Anhydrase. 1. Active Site Inhibitors. This is a hydrolyzing enzyme particularly used to interconvert carbon dioxide with bicarbonate rapidly using a zinc-binding active site. Carbonic anhydrase is known to be active at slightly lower velocity with cobalt substituted for zinc. The experimental situation is complex for this enzyme active site because there is a known pentacoordinated structure along with the tetracoordination for the high-pH resting state. It is well-known that pentacoordinated spherical dications can adopt square pyramidal, trigonal pyramidal, and intermediate structures with comparable energetics while tetrahedral structures predominate in tetracoordination. Features determining these behaviors have only partially been illuminated. Hydrogen bonding with the side-chain oxygen of Thr-199 and its backbone amide proton have been emphasized,[28,29] but this alone is not sufficient to explain some of the crystallographic results such as the different behavior of cyanate and thiocyanate inhibitors. The behavior of inhibitors is an active research area for this enzyme because a complete picture of the interactions in the active site region might lead to better drug treatment for carbonic anhydrase-related disease such as glaucoma.[30] The protonation state of inhibitors and the structural characteristics of resting states of this enzyme can be examined using experimental and simulated cobalt spectra.

Structural information is available for several of the chemical states considered to be of interest. We first discuss the case of the active site with bound inhibitors at physiological pH. The zinc or cobalt ion is bound by three histidine imidazole ligands. PDB entry 2ca2 at 1.9-Å resolution of the human isozyme II with thiocyanate inhibitor (6) also shows this anion bound through nitrogen in the first shell along with a probable water molecule.[31] The coordination pattern is a distorted square-pyramidal structure. Our modeling of this site utilized ammonia ligands (STO-3G basis) instead of the imidazole tripod. Unrestrained optimization of this model structure using either the zinc or cobalt ion (unrestricted Hartree–Fock level) produced a trigonal-bipyramidal optimum. Using angle constraints between the ligands to force the geometry into the crystal structure conformation raised the internal energy by 3 kcal/mol for zinc and 1 kcal/mol for cobalt. It is therefore likely that the enzyme structure is forcing a slightly non-optimal internal geometry for the first shell. The greater ease of this deformation for cobalt is consistent with experimental results[32] and ligand field arguments, indicating that cobalt ion is more likely to prefer square-pyramidal conformations than zinc.

Spectral simulations for the restrained and unrestrained structures are shown together in Figure 6 with the experimental spectrum for the bovine enzyme. Only the pentacoordinate crystal structure conformation produces a ^4P band pattern in good agreement with the observation reproduced in Figure 6. It is important to note that the other pentacoordinate structure gives significantly lower predicted intensities. The ^4P band intensity has often been used to judge between 4-, 5-, and 6-coordination

(24) Zhang, K.; Chance, B.; Auld, D. S.; Larsen, K. S.; Vallee, B. L. *Biochemistry* **1992**, *31*, 1159.
(25) Auld, D. S.; Bertini, I.; Donaire, A.; Messori, L.; Moratal, J. M. *Biochemistry* **1992**, *31*, 3840.
(26) Latt, S. A.; Vallee, B. L. *Biochemistry* **1971**, *10*, 4263.
(27) Liang, J.-Y.; Lipscomb, W. N. *Biochemistry* **1989**, *28*, 9724.
(28) Merz, K. M. *J. Mol. Biol.* **1990**, *214*, 799.
(29) Hakansson, K.; Carlsson, M.; Svensson, L. A.; Liljas, A. *J. Mol. Biol.* **1992**, *227*, 1192.
(30) Dodgson, S. J.; Tashian, R. E.; Gros, G.; Carter, N. D., Eds. *The Carbonic Anhydrases. Cellular Physiology and Molecular Genetics*; Plenum Press: New York, 1991; Part III.
(31) Eriksson, A. E.; Kylsten, P. M.; Jones, T. A.; Liljas, A. *Proteins* **1988**, *4*, 283.
(32) Vedani, A.; Huhta, D. W. *J. Am. Chem. Soc.* **1990**, *112*, 4759.

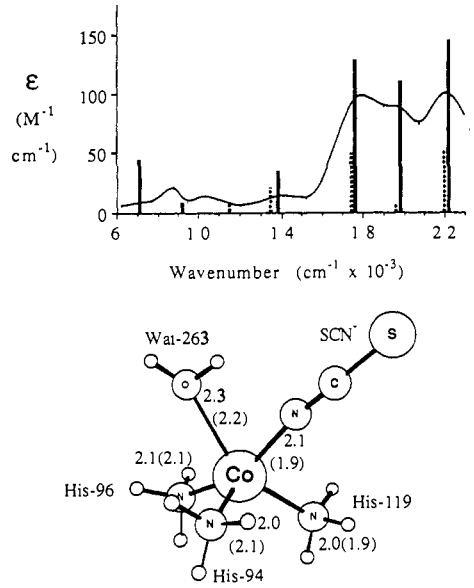

Figure 6. Simulated spectrum for cobalt carbonic anhydrase with thiocyanate inhibitor bound. Structural models also have three ammonias and water in a crystal structure conformation as shown (solid bars) or trigonal bipyramidal with apical NCS^- (dotted bars). The ^4P bands are shifted down by 2500 cm^{-1}. The experimental spectrum is from ref 33.

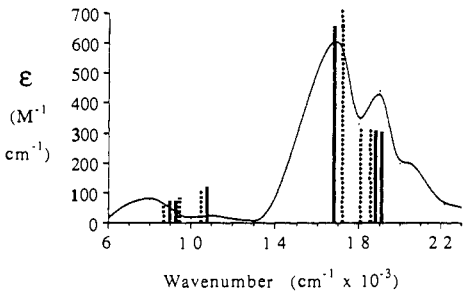

Figure 7. Simulated spectrum for cobalt carbonic anhydrase with cyanate inhibitor bound using three ammonia ligands in structural modes. Solid bars are from nearly collinear Co–N–C–O and dotted bars result from a Co–N–C angle of 120° (with 11 kcal/mol energy cost). The ^4P bands are shifted down by 4000 cm^{-1}. The experimental spectrum is from ref 33.

for bound cobalt. The calculations predict that there is a significant degree of uncertainty in this due to the effect of geometric variation for pentacoordination. If these intensity predictions are reasonably accurate, occurrence of the trigonal bipyramidal form for this inhibitor might have led to the conclusion that the site was actually octahedral in structure. We note that for the zinc enzymes alkaline phosphatase, B-lactamase II, and yeast aldolase octahedral coordination has been proposed for specific zinc/cobalt binding sites on the basis of weak or absent visible absorption for the cobalt forms.[2]

The simulated spectrum was also generated for the tetrahedral form of the active site with thiocyanate or cyanate bound without the water molecule. The cyanate spectrum is reproduced in Figure 7 along with its experimental spectrum from the bovine enzyme. The experimental pattern of peaks is not represented well by the model spectra. The simulated spectrum for thiocyanate is similar in appearance but with an 800 cm^{-1} smaller separation of the ^4P groupings. The binding energies for both anions to the tripod are comparable.[9]

The simulated spectra for the tetrahedral structures have higher ^4P intensities and a narrower pattern than the thiocyanate crystal structure simulation. In addition, the highest energy ^4F levels are lowered in energy to approximately 11 000 cm^{-1}. Similar characteristics in the experimental spectrum for cyanate have been used as the basis of previous suggestions that the cyanate structure was tetrahedral.[1,33]

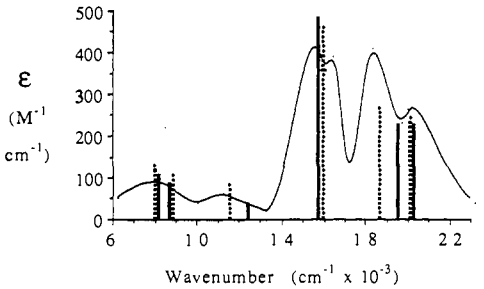

Figure 8. Simulated spectrum for the cobalt carbonic anhydrase high-pH resting state using hydroxyl and three ammonias. Solid bars are from nearby collinear Co–O–H and dotted bars result with O–H directed at Thr-199 O_{G1}. The ^4P bands are shifted down by 4000 cm^{-1}. The experimental spectrum is from ref 36.

Recent crystallographic evidence now shows that the spectra has no bearing on the binding of cyanate to the zinc in the first shell. Hakansson et al. have recently determined the inhibitor occupies a position outside of the first shell of zinc dication but with contact to a bound ligand.[29] Therefore, the observed spectrum could be a perturbed version of low-pH or high-pH resting state spectra of native carbonic anhydrase. Examination of the reproduced experimental spectra in Figures 8 and 9 shows a much greater similarity of the high-pH resting state with the cyanate-bound spectrum shown in Figure 7. If the first-shell ligand is hydroxyl instead of the water molecule that Hakansson et al. proposed, there would be two anions in close contact. Therefore, we believe that the possibility has to be considered that cyanate is protonated by active site water and is therefore a mimic of carbon dioxide as a neutral inhibitor. This could explain how this species resides in the hydrophobic environment of the carbon dioxide binding site and also how it can reside in close proximity to the dication charge without replacing water ligand and entering the first shell.

2. High-pH Resting State. The spectrum of the resting state of CoCA was examined in our previous work.[5] However, we did not use the crystallographic constraints or examine conformational variants. The results of these trials will be briefly discussed here. Figure 8 shows a spectral simulation from a model with bound hydroxyl anion and three ammonia ligands with applied angular constraints from the reported crystal structure of high-pH human isozyme II[34] in PDB entry 1ca2 at 2.0-Å resolution.[6] The splitting of the upper two ^4P states comes in as the Co–O–H angle is decreased from 180°. The approximately linear arrangement is favored with the quantum chemical level used in the optimization step. However, all crystal structures reported for carbonic anhydrase isozymes have the side chain oxygen of a threonine residue in position to hydrogen bond with active site hydroxyl and with the carboxylate of a nearby glutamate. This suggests that the threonine acts as a proton donor to the glutamate and as an acceptor to the hydroxyl proton. From the resting state crystal structure of isozyme II, an optimal linear hydrogen bond requires a Zn–O–H angle of 103°. We have carried out a spectral simulation reported in Figure 8 on a partially optimized structure with this angle constrained. The widely split upper ^4P states are in better agreement with the experimental spectrum in this case. The energy cost of this angle bending varies with the calculation method used (basis set and correlation), but all levels indicate that the energy required is only a few kilocalories per mole in contrast to the much greater difficulty of bending the NCX^- inhibitors. The difference probably arises because the latter species only have one axially directed lone pair to interact with the dication while the hydroxyl has three lone pairs.

The low-frequency region of the ^4P bands also shows a small splitting from a source which these simulations do not identify. A minority of published Co^{II} spectra have four significant features

(33) Bertini, I.; Canti, G.; Luchinat, C.; Scozzafava, A. *J. Am. Chem. Soc.* **1978**, *100*, 4873.
(34) Eriksson, A. E.; Jones, T. A.; Liljas, A. *Proteins* **1988**, *4*, 274.

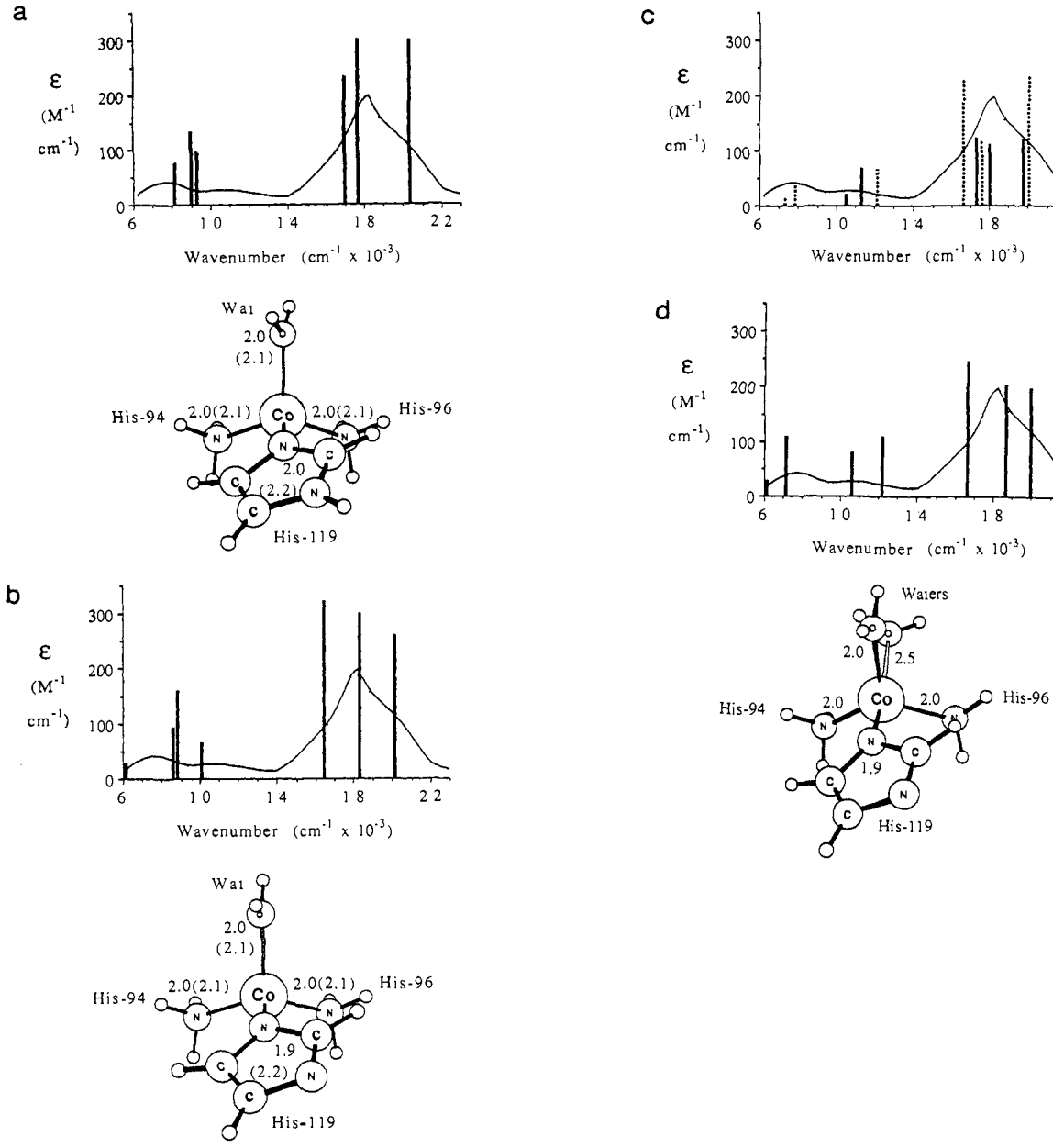

Figure 9. (a) Simulated spectrum for a possible cobalt carbonic anhydrase low-pH native state using the indicated structural model. The ^4P bands are shifted down by 4000 cm^{-1}. The experimental spectrum is from ref 36. (b) Experimental and simulated spectra as in part a but using deprotonated imidazole. The ^4P bands are shifted down by 4000 cm^{-1}. (c) Experimental spectrum as in part a with simulations from two waters and three ammonias in square-pyramidal (solid bars) and trigonal-bipyramidal (dotted bars) structures. The ^4P bands are shifted down by 4000 cm^{-1}. (d) Experimental spectrum as in part a simulated using the pentacoordinate structure shown. The ^4P bands are shifted down by 4500 cm^{-1}.

in the ^4P absorption region while the majority have three in accord with the expectation of observing three excited states with high transition intensities. It is therefore likely that the minority of spectra such as for the carbonic anhydrase resting state are showing effects from the vibrational envelope of one state or from distinct conformational populations. The latter could possibly even result from different proton orientations not producing observable effects in the crystallography.

3. Low-pH Resting State. The number and position of first-shell water molecules in the low-pH structure of carbonic anhydrase is apparently an unresolved point in spite of a recent successful crystal structure determination for human isozyme II at a pH of approximately 6.[35] The active site density excepting the tripod and dication was diffuse and centered at an unusual distance of 2.7 Å from the dication. This suggests that the active site has multiple occupancy to give an indefinite electron density feature or may have another species bound instead of water. A more recent determination by Hakansson et al.[29] shows a four-coordinate structure with a more normal Zn–O distance of 2.05 Å. The cobalt spectra are variable in intensity depending on which isozyme is used. The ^4P peak structures are commonly broad and barely distinguishable and therefore we have reproduced the spectrum of the human isozyme II only in Figure 9a–d. The peak intensities for all isozymes are somewhat lower than those for the high-pH forms. The low-pH resting state for bovine isozyme is reported to have temperature-dependent spectral intensity which has been suggested to signify a variable coordination number.[1] Human isozyme II was proposed to be pentacoordinate as signified by particularly low intensity. Therefore, the spectroscopic conclusion for this isozyme substituted with cobalt is in disagreement with the crystallographic conclusion for the zinc enzyme.

We have attempted to calculate structures, simulated spectra, and energetics for various active site water occupancies and conformations to identify any common features and examine

(35) Nair, S. K.; Christianson, D. W. *J. Am. Chem. Soc.* **1991**, *113*, 9455.
(36) Bertini, I.; Luchinat, C.; Scozzafava, A. *Inorg. Chim. Acta* **1980**, *46*, 85.

agreement with the reported low-pH spectra. A reinterpretation of the NMR spectra of the tripod imidazoles[9] indicated the strong possibility that the His-119 has donated a proton to Glu-117 in the low-pH form and is therefore anionic. Our spectral simulations thus used an imidazole or an imidazolate ring as a model for this ligand in either protonation state. Specific partially optimized structures and simulated spectra are given in Figure 9a–d. Several general results emerged from these and other simulations not reproduced here.

The four-coordinate structures with either an all ammonia tripod or one replacement by imidazole (Figure 9a) have excited states close to the pattern of a C_{3v} symmetry structure. In the tetracoordinate structures generated for this work one of the water protons was constrained to be oriented in the general direction required to hydrogen bond with Thr-199 in addition to the usual constraints. The low-energy modes for the water molecule, a rotation around the Co–O axis or a rolling of the protons away from this axis, have only a small effect on the spectral simulation.[5] The imidazolate ligand field breaks this pattern (Figure 9b) and gives roughly evenly spaced ^4P states of similar intensity. However, given the broad structure of the experimental ^4P region we cannot effectively distinguish between these model structures.

The pentacoordinate structures represented in Figure 9c,d generated lower intensity than tetracoordinate structures to varying degrees in all cases. The pentacoordinate structures tended to have higher energies for the upper states of ^4F character but by only a few thousand wavenumbers. In Figure 9c the effects of different types of five-coordinate structures with two water molecules and an ammonia tripod are illustrated. Again, the detailed structure has a significant effect on the predicted intensities. Similar energies for these conformations are obtained for either zinc or cobalt ions with again a slightly larger preference for the trigonal bipyramidal structure for zinc (2 versus 1 kcal/mol preference with cobalt). Following the reaction coordinate connecting these structures showed that there is no energy barrier. Fluctuations between these types could give rise to a significant vibrational or conformational broadening of a pentacoordinate spectrum. Different proton orientations for water molecules were also tried for these forms having similar effects on the predicted spectra. With imidazolate as a ligand, the trigonal-bipyramidal form illustrated in Figure 9d is more strongly favored with a significant rocking of one water molecule to direct a proton at the π system. More evenly spaced ^4P bands are obtained with imidazolate as shown.

The energetics of these pentacoordinate structures is significantly different from that of the pentacoordinated structures with a bound anionic inhibitor discussed above. The binding energy of the second water molecule is approximately 20 kcal/mol regardless of which tripod model is used. The barrier to moving a water molecule out into a second-shell position of similar binding energy is only 1 or 2 kcal/mol. These calculations are in no sense quantitative so that we interpret a small barrier to also be consistent with any shallow potential surface format, including a square well. In the case of binding a water molecule to a model complex with the thiocyanate inhibitor, a second-shell binding energy of 19.8 kcal/mol and a first-shell binding energy of 12.5 kcal/mol were obtained in a previous study using the same quantum chemical treatment.[9] The comparative first-shell energetics could be modified by the possibility of hydrogen bonding of first-shell water molecules to nearby Thr-199. The energy barrier to moving from the first-shell position into the second-shell position was estimated by constrained optimization steps to be higher at approximately 5 kcal/mol in spite of the conformational change being more downhill. The origin of a lower barrier for the low-pH models may be due to a wider distribution of the dication positive charge into the neutral ligands for this case. The water oxygen is thus able to move out along ligand bonds while maintaining attractive electrostatic contacts more easily. The five-coordinate low-pH forms could therefore be potentially very fluctional for motions within the first shell and for moving water molecules between shells. The broad crystallographic electron density centered at a large distance from the zinc in the Nair and Christianson study[35] could therefore result from stable water molecules in partial first-shell positions effectively intermediate between tetracoordination and pentacoordination, with large fluctuations so that first-shell water was not clearly observable. The broad band shapes for this chemical state support the idea that the structures are unusually fluctional in conformation.

An alternative interpretation of the experimental information is to accept that the Hakansson et al.[29] crystallographic study is correctly describing the low-pH form as tetrahedral at the active site while the cobalt-substituted enzyme is pentacoordinate to display the properties of the experimental and simulated spectra and structures noted above. We have tested this possibility by calculating the binding energy of a second water molecule to zinc in an active site model using an ammonia tripod. Equivalent large basis sets for zinc and cobalt[10] were used with complete structural optimization. The binding energy was 3 kcal/mol larger for the cobalt active site model which we interpret as a significant preference for pentacoordination with two waters for cobalt compared to zinc. Therefore, it is very likely that this crystallographic study and the best interpretation of the cobalt spectroscopy are both accurate, but the metal substitution changes the active site conformation.

4. Calculations with Enzyme Environment Represented. Spectral simulations were carried out for two carbonic anhydrase structural models with partial charges inserted for the purpose of simulating the electrostatic environment of the protein. It is expected that such electrostatics should have little effect on d–d spectra, but this has not even been roughly verified by experiments or theory. The partial charges used were taken from the CHARMM program files at standard protonation states. All residues except those for the tripod histidine side chains were included. Charge positions were taken from the relevant Protein Data Bank structures with polar hydrogens added by the Quanta program. The hydrogen positions close to the active site were examined to make sure that they were consistent with possible hydrogen bond patterns. Charges were inserted into the Hamiltonian for the quantum calculations at the orbital generation and CI stages but after structural optimization of the active site model. The crystal structure constraints applied in generating the optimized active site geometries then allowed for the structures to be overlaid onto the protein coordinates with reasonable accuracy.

The first structure used for this purpose was the pentacoordinate thiocyanate inhibitor with crystal structure constraints yielding a vacuum spectrum as shown in Figure 6. The line positions in the simulation are affected by 150 cm^{-1} at most after inserting the partial charge set. Intensity variations are at approximately 10% level so that there is no significant change in appearance. The protonation states of residues at varying distances from the active site were then systematically varied, producing no effect on the simulation. This indicates insensitivity to "electrostatic" mutants and also that no accidental cancellation of external fields is occurring to minimize spectral perturbations. Systematic deletion of the charge sets showed that the primary perturbation came from the Thr-199 side chain charge set which was in close contact with the first-shell water molecule. Another test was then performed using a tetracoordinate model for the low-pH enzyme. The coordination shell has a water molecule, two ammonia ligands, and an imidazole representing His-119 for which the vacuum spectrum was given in Figure 9a. The imidazole ligand then adds an extra close contact with the partial charge set of Glu-117. In this case the largest state energy perturbation was approximately 200 cm^{-1} and intensity variations were at the 10% level.

These quantum chemical first-shell models interact with the partial charge environment by 21 and 49 kcal/mol, respectively. These large magnitudes are due to the lack of any dielectric

screening in the interactions. In spite of these apparently large field interactions which are probably overestimated, the d–d transitions are insignificantly perturbed. Atomic populations were reported in our previous study[5] showing that the flow of charge between atoms accompanying the transitions is very small. The dominant field change to interact with species outside of the first shell is a quadrupolar field shift which inherently has a short range, while the longer range dipolar shift is very small in magnitude. This characteristic could allow for the use of these spectroscopies to probe the effect of mutations on important properties of carbonic anhydrase and other enzymes. Mutations which do not affect the first-shell structure should not perturb the spectra. It might be possible also to see an effect from remote mutations in the proton transfer pathways if they produce a propagated effect through hydrogen bonds into the active site.

Conclusions

The spectrum simulations on the enzyme liver alcohol dehydrogenase first served to show us that the methodology produces an accurate representation of an experimental cobalt d–d spectrum from an accurate first-shell structure model. In Figure 1b we predict spectra representing the high-pH form of the LADH catalytic site with deprotonated water. This also should be an accurate prediction for the proposed chemical state of the enzyme in which an alcohol substrate is bound as an alcoholate anion. The various reasonable conformational forms were examined and all predict a comparable pattern of two observable ^4P peaks separated much less than the two peaks noted in the resting state of the enzyme and with a somewhat different intensity pattern.

For the structural site with four cysteine residues we propose an alternative model with the two cysteine sulfur atoms exposed to solvent being protonated. This possibility was suggested by the longer Zn–S distances for these cysteines in the crystallography. This model and the usually assumed form with all residues deprotonated produce comparable absorption spectra which agree with the experimental observation fairly closely. We were not able to reproduce the experimental thiolate–cobalt charge-transfer spectrum due to computer limitations preventing us from running the much larger configuration interaction wave functions required. However, the almost exact similarity of the structural and catalytic site charge-transfer spectra from the experimental literature suggested also that both sites have two thiolates only and the same overall charge.

Comparing the structural binding sites for aspartate transcarbamylase and LADH showed two distinct patterns of internal torsional angles. Each corresponds to a local minimum with comparable energies in the structures obtained with zinc and the reduced ligand models SCH_3^- or SH^-. Changing these torsion angles probably would require substantial local reorganization of the protein. The binding sites therefore may have evolved to have particularly stable internal conformations with bound zinc.

The spectrum simulations for carboxypeptidase A and thermolysin resting states are only in fair agreement with the experimental spectra. This may be partly a consequence of the substantial mobility of active site residues, reported from several sources, having the effect of broadening the spectra. An effort was made to adjust the CPA active site model from the crystallography to accommodate the suggestion from XAFS work that the Glu-72 becomes unidentate in solution compared with the bidentate crystal structure orientation.[24] However, the results are not any more satisfactory than the original simulated spectrum. We suggest that a possible partial occupancy form discussed by the crystallographers in which the active site water is unusually far from the metal dication[21] could alternatively be responsible for the solution XAFS characteristics.

High-pH spectra with deprotonated active site water were produced for CPA and TLN. The CPA spectrum resembles the pattern described for CPA in words by Latt and Valee[26] while the TLN spectrum has not been reported to our knowledge. We also discuss in the Results section the comparative chemistry of these enzymes and carbonic anhydrase relating to peptide hydrolysis activity as a function of pH.

The carbonic anhydrase spectra for the resting state and with thiocyanate inhibitor bound are in satisfactory agreement with the experimental spectra. The resting state spectrum is improved by forcing the orientation of the hydroxyl proton to make an optimal hydrogen bonding contact with the nearby Thr-199. The simulations with cyanate inhibitor in the first shell are not satisfactory which would be predicted given the recent observation by crystallographers that cyanate occupies a second-shell position.[29] The cyanate experimental spectrum is probably best interpreted as a perturbed version of the resting state spectrum. This suggests that the cyanate is protonated and occupies a binding site specific for carbon dioxide in a similar mode and in contact with first-shell hydroxyl anion. This overcomes energetic objections to the crystallographers suggestion that an anion can occupy a second-shell position intended for carbon dioxide with only a water molecule in the first shell. The desolvated cyanate inhibitor apparently has a larger effective proton affinity than the water bound to zinc and abstracts a proton from the water. This will probably be true for a variety of anionic inhibitors with large gas-phase proton affinities.

The low-pH spectrum simulations for carbonic anhydrase are only in fair agreement with the very broad experimental spectrum which again could be due to the broadening itself. We show that alternative conformations for pentacoordinate structures having two bound water molecules can be comparable in energy while giving distinct spectra. There is also predicted to be a negligible barrier connecting four- and five-coordinate structures. These low-energy modes having spectrum consequences could induce the observed broadening.

Energy calculations for low-pH models comparing zinc and cobalt showed that cobalt more strongly attracts a second water molecule into the first shell. A recent crystal structure determination showed a tetracoordinate structure at pH 6.0 for the zinc enzyme.[29] Previous evidence from the cobalt absorption spectrum and NMR relaxation times were interpreted as suggesting pentacoordination.[1,2] The suggestion of different energetics producing the distinct structural types could reconcile this point.

Comparative spectrum simulations with partial charges representing the protein environment of carbonic anhydrase showed that the d–d spectra are essentially insensitive to electrostatic perturbations. Experimentally observed spectrum changes thus represent structural perturbations of the first shell.

Our calculations with explicit pentacoordinate conformations of low energy predicted exceptionally weak ^4P absorption for particular structural forms. Experimental occurrences of these could therefore be mistakenly interpreted as having octahedral coordination. Except for this result, our calculations with explicit structural models produced general agreement with trends noted by experimentalists from examination of series of spectra.