# Modeling Vibrational Spectra of Amino Acid Side Chains in Proteins: The Carbonyl Stretch Frequency of Buried Carboxylic Residues

**Andrei K. Dioumaev and Mark S. Braiman***

*Contribution from the Department of Biochemistry, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908*

*Received June 19, 1995*®

**Abstract:** To improve the use of vibrational spectra for modeling Asp and Glu environments buried in proteins, $\nu_{C=O}$ frequencies of aliphatic carboxylic acids and *N*-acetylaspartic acid α-amide were compared in several different solvents. These data indicate that propionic acid and longer-chain aliphatic carboxylic acids are all quite similar, and serve as better models for Asp and Glu residues buried in proteins than does acetic acid. For propionic acid, $\nu_{C=O}$ frequencies above 1745 $cm^{-1}$ are observed only in non-H-bonding solvents. Furthermore, in such non-H-bonding solvents, the $\nu_{C=O}$ frequency exhibits a linear correlation with Onsager's parameter, $2(\epsilon - 1)/(2\epsilon + 1)$, which is expected to be proportional to the strength of the solute-induced electrostatic ("reaction") field of the solvent. We also measured a $\nu_{C=O}$ frequency of 1742 $cm^{-1}$ for the protonated Asp-26 residue of thioredoxin which is known to be surrounded principally by nonpolar groups. These results are used to model the environment of the Asp-85 residue of bacteriorhodopsin, for which the $\nu_{C=O}$ frequency has been measured previously in several photointermediate states. In the unphotolyzed (bR) state, the Asp-85 residue ($\nu_{C=O} \simeq 1730$ $cm^{-1}$) is in a relatively polar hydrogen-bonding environment, but this environment is drastically changed upon photoconversion to the M state ($\nu_{C=O} = 1762$ $cm^{-1}$). We conclude that in the latter state, the Asp-85 COOH group is in a highly nonpolar environment, characterized by the absence not only of hydrogen bonding but also of other forms of external dielectric stabilization.

## Introduction

Aspartic and glutamic acids buried within the hydrophobic regions of proteins such as thioredoxin, bacteriorhodopsin (bR), and visual rhodopsin have been observed to be protonated at pH 7 and higher. In the retinal proteins, these buried Asp and Glu residues have been associated with elevated $\nu_{C=O}$ stretching frequencies, in some cases even above 1760 $cm^{-1}$.[1,2] There is a voluminous literature on the C=O stretching vibration of carboxylic acids (see for instance refs 3−5 and references therein), but there are few if any attempts to relate these results to the kinds of environment that Asp or Glu groups are exposed to inside proteins.

Here, we compare a number of simple aliphatic carboxylic acids to more complex model compounds such as *N*-acetylaspartic acid α-amide, and an Asp-containing protein, thioredoxin (which has a protonated Asp group and a structure known to atomic resolution). We demonstrate that acetic acid is a substantially poorer model for vibrational studies of Asp and Glu residues in proteins than longer-chain carboxylic acids (such as propionic or butyric). We also conclude that the only way to model $\nu_{C=O}$ frequencies above 1760 $cm^{-1}$ for Asp and Glu residues is by a nearly complete absence of stabilizing external electric field at the carbonyl group, i.e. not only the absence of H bonding but, more surprisingly, an effective dielectric constant of near 2 in the immediate vicinity of the COOH group.

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* October 1, 1995.

(1) Rothschild, K. J.; Zagaeski, M.; Cantore, W. A. *Biophys. Biochem. Res. Commun.* **1981**, *103*, 483−489.

(2) DeGrip, W. J.; Gillespie, J.; Rothschild, K. J. *Biochim. Biophys. Acta* **1985**, *809*, 97−106.

(3) Bellamy, L. J. *Advances in Infrared Group Frequencies*, 2nd ed.; Methuen & Co. Ltd.: London, 1968; Vol. 2, Chapter 5, pp 172−190.

(4) Bellamy, L. J. *Advances in Infrared Group Frequencies.* Chapman & Hall Ltd.: London, 1975; Vol. 2, Chapter 10, pp 183−202.

(5) Nyquist, R. A.; Clark, T. D.; Streck, R. *Vib. Spectrosc.* **1994**, *7*, 275−286.

## Experimental Section

Fourier transform infrared (FTIR) spectra were measured on Nicolet 60SXR and 740 spectrometers equipped with TGS detectors. To improve the signal-to-noise ratio, 200−2000 scans were averaged for each measurement. All measurements (besides those on thioredoxin) were done on samples in solution at room temperature in a demountable cell (Harrick), using two $CaF_2$ windows separated by a 0.2−0.95 mm PTFE spacer. For FTIR measurements on thioredoxin, the protein was dried from pH 6.5 buffer and measured as a film on a $CaF_2$ window at room temperature. All spectra were taken with 1-$cm^{-1}$ resolution over the range 900−4900 $cm^{-1}$.

Unless otherwise stated, all values for COOH stretch frequencies discussed below are for the highest frequency band in the carbonyl region, i.e. corresponding to the monomer state. Concentrations of model compounds were ~3 mM. Residual water vapor bands were routinely subtracted from calculated absorbance spectra.

All solvents were of reagent grade or better and were used without further purification. Acetic and propionic acids were from Aldrich Chemical Co. (St. Louis, MO). *N*-Acetylaspartic acid α-amide (Ac-Asp-$NH_2$) was from Bachem Bioscience Inc. (King of Prussia, PA). Thioredoxin (recombinant form from *Escherichia coli*) was from Calbiochem-Novabiochem Corp. (La Jolla, CA).

## Results and Discussion

Previous studies of solvent-induced shifts in $\nu_{C=O}$ in carboxylic acids (for review see refs 3 and 4) have not included any peptide derivatives of Asp or Glu. To determine the extent to which peptide linkages affect $\nu_{C=O}$, we have obtained IR spectra of *N*-acetylaspartic acid α-amide (Ac-Asp-$NH_2$) in several solvents and compared its $\nu_{C=O}$ frequency to those obtained with simpler carboxylic acids in the same solvents (Table 1). Our data indicate that the carbonyl frequency of Ac-Asp-$NH_2$ never exceeds that of propionic acid and is also always markedly lower than that of acetic acid.

A more instructive "model compound" is provided by Asp-26 of thioredoxin (Figure 1), which has perhaps the most nonpolar environment of any COOH group in a protein with a

0002-7863/95/1517-10572$09.00/0 

**Table 1.** Carbonyl Stretching Frequency ($cm^{-1}$) of Carboxylic Acids in Hydrogen Bonding Solvents[a]

| | solvent | | | |
|---|---|---|---|---|
| acid | acetonitrile | ethanol/chloroform, 1:100 | dioxane | trifluoroethanol |
| $CH_3COOH$ | 1755 | 1755 | 1754 | 1726 |
| $CH_3CH_2COOH$ | 1743 | 1740 | 1740 | 1723 |
| Ac-Asp-$NH_2$ | 1740 | 1738 | 1736 | 1723 |

[a] Only the highest frequency band, corresponding to the monomer state, is included.

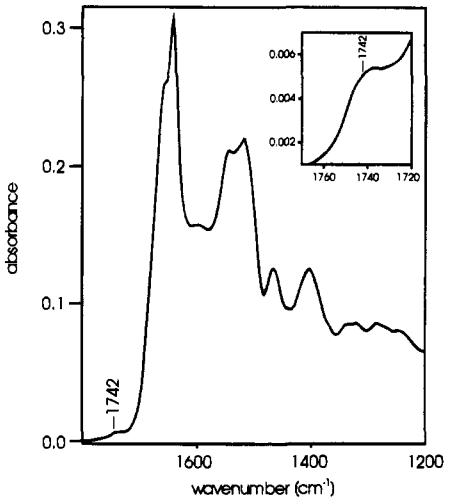


**Figure 1.** FTIR spectrum of thioredoxin (Calbiochem, La Jolla, CA) at pH 6.5 and room temperature. Inset, $\nu_{C=O}$ region of the same spectrum. The exact position of $\nu_{C=O}$ was determined by the curve-fitting procedure of LabCalc software (Galactic Industries Corp., Salem, NH), assuming a linear baseline in the range 1800−1700 $cm^{-1}$ and a single band. The band center was unchanged regardless of whether a Gaussian or Lorentzian bandshape was assumed.

published atomic-resolution X-ray structure, even though it appears to be H bonded to a tightly-bound water molecule.[6] Observation of its COOH $\nu_{C=O}$ band position at 1742 $cm^{-1}$ (Figure 1) supports the proposal[7] that this Asp is protonated inside the folded protein. This frequency is in excellent agreement with $\nu_{C=O}$ of propionic acid and Ac-Asp-$NH_2$ in ethanol/$CHCl_3$ (1:100), where these model compounds would be expected also to be singly H bonded, but is considerably lower than the frequency for acetic acid in the same solvent (Table 1).

Additional evidence that acetic acid is a poor model for Asp and Glu side chains is shown in Figure 2, where it is shown that in nonpolar solvents as well, acetic acid has a markedly higher $\nu_{C=O}$ frequency than longer aliphatic carboxylic acids. The solubility of Ac-Asp-$NH_2$ in these less-polar solvents, except for the weak H bond acceptor acetonitrile, was too low to measure an IR spectrum. However, acetonitrile, in which solvent $\nu_{C=O}$ frequencies for five different carboxylic acids were obtained (Table 1 and Figure 2), demonstrates continuity of the

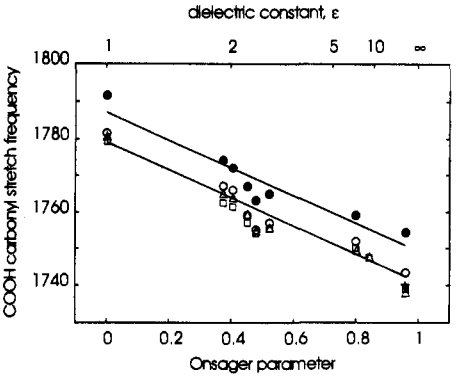


**Figure 2.** COOH carbonyl stretch frequency of acetic (●), propionic (○), butyric (Δ), and caproic (□) acids, as a function of external dielectric constant $\epsilon$. Values are given for all acids in the vapor phase ($\epsilon = 1$) and in hexane ($\epsilon = 1.89$), cyclohexane ($\epsilon = 2.02$), $CCl_4$ ($\epsilon = 2.24$), toluene ($\epsilon = 2.38$), $CS_2$ ($\epsilon = 2.63$), $CH_3I$ ($\epsilon = 6.97$), and acetonitrile ($\epsilon = 36.6$). Values in $CH_2Cl_2$ ($\epsilon = 8.93$) are given for caproic and butyric acids only. The COOH carbonyl stretch frequency of Ac-Asp-$NH_2$ (★) in acetonitrile is also included for comparison (see also Table 1); solubility of this model compound in the other solvents was too low to measure $\nu_{C=O}$. Data on caproic and butyric acids, as well as vapor phase data for acetic and propionic acids, are from ref 8. Dielectric constants of solvents are from refs 9−11. Each plotted solid line is a least-squares fit of $\nu_{C=O}$ for a particular carboxylic acid to a linear function of Onsager's parameter, $2(\epsilon - 1)/(2\epsilon + 1)$; the horizontal axis scale has also been transformed to be linear with this parameter.

behavior between nonpolar and H-bonding solvents. Our results are consistent with a recent conclusion that $\nu_{C=O}$ for carboxylic acids is strongly dependent on the number of hydrogens on the α-carbon atom.[5] Therefore, the best models for vibrations of Asp and Glu side chains in proteins must be other aliphatic carboxylic acids with a secondary α-carbon.

Another concern with acetic acid as a model compound is that its carbonyl stretch is affected by a Fermi resonance which is not present in other carboxylic acids.[12] Failure to consider this has affected recent interpretations of acetic acid as a model compound for COOH vibrations in proteins. For example, Maeda et al.[13] attributed the bands at 1754 and 1728 $cm^{-1}$ for acetic acid in dioxane to distinct monomer states differing in the degree of H bonding to solvent. However, as had been shown previously,[12] the two bands actually arise from a Fermi resonance splitting between two close frequencies present in one and the same molecule: $\nu_{C=O}$ and the first overtone of $\nu_{C-C} \simeq 870$ $cm^{-1}$. In other carboxylic acids lacking the Fermi resonance, the unusually high frequency of 1754 $cm^{-1}$ is not seen when dioxane is used as solvent; the $\nu_{C=O}$ frequency for the COOH group is instead typically near 1740 $cm^{-1}$ (see Table 1).

For such reasons, it is appropriate to focus on propionic acid as a model for understanding the environments of protonated Asp and Glu residues in proteins. In H-bonding solvents, $\nu_{C=O}$ of propionic acid, and also of an Asp model compound, is always below 1743 $cm^{-1}$ (Table 1) and is not significantly correlated with solvent dielectric constant. For example, acetonitrile has a much higher dielectric constant ($\epsilon = 36.6$) than dioxane ($\epsilon = 2.21$), but $\nu_{C=O}$ is slightly higher in acetonitrile for all of the compounds shown in Table 1. Moreover, mixing an H-bonding with a non H-bonding solvent, even in a ratio too small to significantly alter the dielectric constant of the latter, is nevertheless enough to downshift the propionic acid C=O band, e.g. from its position at 1767 $cm^{-1}$

(6) Katti, S. K.; LeMasters, D. M.; Eklund, H. *J. Mol. Biol.* **1990**, *212*, 167−184.
(7) Langsetmo, K.; Fuchs, J. A.; Woodward, C. *Biochemistry* **1991**, *30*, 7603−7609.
(8) Lascombe, J.; Haurie, M.; Josien, M.-L. *J. Chim. Phys.* **1962**, *59*, 1233−1246.
(9) Weast, R. C., Ed. *Handbook of Chemistry and Physics*, 76th ed.; CRC Press: Cleveland, 1995; pp 6-159−6-192.
(10) Yukawa, Y., Ed. *Handbook of Organic Structural Analysis*; W. A. Benjamin, Inc.: New York, Amsterdam, 1965; p 353.
(11) Riddick, J. A.; Bunger, W. B.; Sakano, T. K. *Organic Solvents. Physical Properties and Methods of Purification*; Wiley: New York, 1986; Vol. 2.
(12) Haurie, M.; Novak, A. *J. Chim. Phys.* **1967**, *64*, 679−684.
(13) Maeda, A.; Sasaki, J.; Shichida, Y.; Yoshizawa, T.; Chang, M.; Ni, B.; Needleman, R.; Lanyi, J. K. *Biochemistry* **1992**, *31*, 4684−4690.

in 100% hexane (Figure 2) to 1742 $cm^{-1}$ in ethanol/hexane (1:200) (data not shown). The latter frequency is similar to that recorded for both propionic acid and Ac-Asp-$NH_2$ in ethanol/chloroform (1:100) (Table 1). These observations demonstrate that a single hydrogen bond to the C=O group is sufficient to downshift its frequency by 25 $cm^{-1}$, and the effect of this single hydrogen bond dominates over other aspects of the environment, including the amide carbonyl dipoles present in Ac-Asp-$NH_2$ or (presumably) a peptide backbone. This is in accord with the well-established fact that for the strongly hydrogen-bonded cyclic dimers of carboxylic acids, the $\nu_{C=O}$ frequency is practically solvent invariant.[3,8] This lack of correlation between $\nu_{C=O}$ and solvent dielectric constant is due to the dominant electric field associated with hydrogen bonding.

In the absence of hydrogen bonding, however, the COOH carbonyl frequency is much more sensitive to other types of dipolar interactions with its environment, as demonstrated in Figure 2. In the frequency range above 1740 $cm^{-1}$ characteristic of non-H-bonded COOH groups, there is a clear inverse correlation of $\nu_{C=O}$ with dielectric constant $\epsilon$. Previous workers have observed a similar correlation for the $\nu_{C=O}$ of ketones and esters and explained it as resulting from interaction of the carbonyl group with both permanent and induced dipoles of neighboring solvent molecules, which leads to their partial reorientation.[14] This Debye screening creates a stabilizing external electric field corresponding to that of a dipole opposite to the C=O group. The external field polarizes the C=O bond, decreasing its bond strength and therefore $\nu_{C=O}$. Such an induced electric field is expected to increase linearly not with $\epsilon$ itself, but rather with the Onsager parameter, $2(\epsilon - 1)/(2\epsilon + 1)$, reflecting the calculated influence of the reaction field.[15] In the absence of the even-stronger local fields expected from hydrogen bonding, therefore, $\nu_{C=O}$ is expected to approach asymptotically some high-$\epsilon$ "saturation" value, which we observe to be approximately 1740 $cm^{-1}$.

The correlation in Figure 2 is not perfect. Note, for instance, the dip at toluene, $\epsilon = 2.38$. This is because the Onsager theory[15] is an oversimplification, and assumes the absence of all sterically-specific solute−solvent interactions such as H-bonding. To be more precise, $\epsilon$ determines the general trend, while weak sterically-specific solute−solvent interactions can still contribute minor deviations of $\nu_{C=O}$ from that general trend. However, significant deviations from the trend are likely to be negative, indicating specific solute−solvent interactions that would further stabilize the C=O dipole. For example, aromatic rings are known to act as weak H bond acceptors;[16] this probably explains the deviation of toluene below the general trend line in Figure 2.

Positive deviations from the trend line seem a priori to be unlikely. It is possible to imagine positioning a dipole, or even a monopole (point charge), so as to increase $\nu_{C=O}$, e.g. by concentrating negative charge near the carbonyl oxygen so as to *raise* the effective bond order. However, any such simple monopole or dipole field, arranged to oppose the carbonyl dipole so as to raise the C=O bond order, would clearly be energetically unfavorable. While such configurations cannot absolutely be ruled out inside a protein, it is difficult to imagine constraints that could render them stable relative to rapid local reorganizations such as COOH tautomerization.

To generalize, $\nu_{C=O}$ acts quite generally as a sensor for local electric fields that stabilize the carbonyl dipole, whether these fields originate from hydrogen bonding or from other electrostatic interactions. In simple non-H-bonding solvents, dielectric constant differences account for most of the variation in $\nu_{C=O}$. Of course, the possibility of correlating $\nu_{C=O}$ with a single parameter ($\epsilon$) is expected to decrease with increasing size and complexity of the surrounding molecules. The local and directional character of solute−solvent electrostatic interactions is reflected most strongly in hydrogen bonds, which involve a highly localized and sterically specific electric field, i.e. one that most strongly affects a particular bond. In proteins, the polar and polarizable groups nearest to an Asp or Glu residue are also expected to be able to cause local and directional influences on $\nu_{C=O}$. Nevertheless, in most cases Asp or Glu residues and surrounding side chains can be expected to have some thermodynamic freedom to arrange themselves as if they were solute and solvent. Thus, for proteins without known atomic-resolution structures, values of $\nu_{C=O}$ of propionic acid in "isotropic" solvents should be of significant value for modeling environments around protonated Asp and Glu residues.

The most detailed assignments of COOH bands available for any protein are for bacteriorhodopsin (bR) and its photoproducts.[13,17] In unphotolyzed bR, a $\nu_{C=O}$ of ~1730 $cm^{-1}$ for Asp-85[18] probably indicates rather strong hydrogen bonding, in between that measured for propionic acid in a hydrogen-bonded solvent like ethanol (1735−1738 $cm^{-1}$, see also Table 1) and that in the cyclic dimer state (1714−1718 $cm^{-1}$; our spectral data, not shown in Table 1). This degree of hydrogen bonding is characteristic of a relatively polar environment, which has also been attributed to this region of the protein in an atomic-resolution model based on cryo-electron microscopy.[19] However, to upshift the $\nu_{C=O}$ of Asp-85 to 1762 $cm^{-1}$, as occurs in the M state of the bR photocycle,[1,17,18] this residue must be transferred to an environment capable of producing only very weak stabilizing electrical fields. Certainly, as has been recognized previously, this means that its carbonyl group is not H bonded.[1,17,20] However, a more striking conclusion is now possible. A $\nu_{C=O}$ frequency this high for an Asp or Glu side chain would almost certainly require a solvent capable of generating an equivalent frequency for propionic acid, e.g. $CCl_4$ (Figure 2). The environment around Asp-85 in the M state therefore corresponds to an effective dielectric constant of ~2.2, i.e. it must be nearly like an aliphatic hydrocarbon. The only way to model this very high $\nu_{C=O}$ frequency is if Asp-85 in M is sterically constrained from participating in almost all stabilizing interactions that might be present in the unphotolyzed bR state—including not only those with H-bonding groups (the Schiff base on Lys-216, Arg-82, Asp-212, and Thr-55) but also most of those involving nearby polar and polarizable groups, such as Met-56, Tyr-57, Tyr-185, Trp-86, and Phe-208.[19] At present, we have no direct evidence for a particular mechanism, but this change could probably be created by fairly small displacements of Asp-85 away from the dipole-stabilizing residues. Concomitantly, the space left by these "departing" side chains could undoubtedly be filled by nearby nonpolar groups, such as Ile-52, Met-56, Ala-84, Phe-88, Phe-208, Leu-211,[19] or even lipid tails.

**Acknowledgment.** This work was supported by NIH Grant No. GM46854 (M.S.B.).

JA9519905

(14) Nyquist, R. A. *Vib. Spectrosc.* **1994**, *7*, 1−29.
(15) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486−1493.
(16) Makhatadze, G. I.; Privalov, P. L. *Biophys. Chem.* **1994**, *50*, 285−291.
(17) Braiman, M. S.; Mogi, T.; Marti, T.; Stern, L. J.; Khorana, H. G.; Rothschild, K. J. *Biochemistry* **1988**, *27*, 8516−8520.
(18) Braiman, M. S.; Dioumaev, A. K.; Lewis, J. R. *Biophys. J.* **1995**, in press.
(19) Henderson, R.; Baldwin, J. M.; Ceska, T. A.; Zemlin, F.; Beckmann, E.; Downing, K. H. *J. Mol. Biol.* **1990**, *213*, 899−929.
(20) Engelhard, M.; Gerwert, K.; Hess, B.; Kreutz, W.; Siebert, F. *Biochemistry* **1985**, *24*, 400−407.