THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

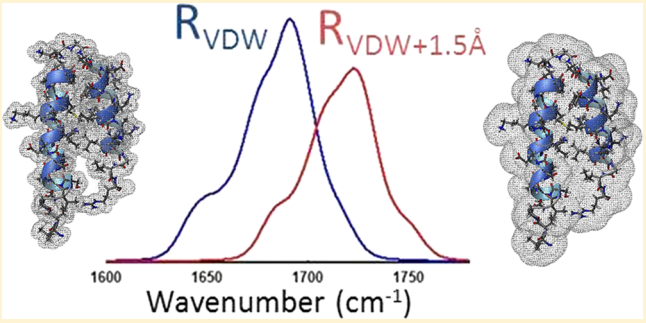
pubs.acs.org/JPCB

DFT-Based Simulations of Amide I′ IR Spectra of a Small Protein in Solution Using Empirical Electrostatic Map with a Continuum Solvent Model

William R. W. Welch and Jan Kubelka*

Department of Chemistry, University of Wyoming, Laramie, Wyoming 82071, United States

ABSTRACT: A continuum solvent model was tested for simulations of amide I′ IR spectra for a 40-residue subdomain of P22 viral coat protein in aqueous solution. Spectra obtained using DFT (BPW91/6-31G**) parameters for a reduced all-Ala representation of the protein were corrected by an electrostatic potential map obtained from the solvent cavity surface and AMBER99 side-chain atom partial charges. Various cavity sizes derived from van der Waals atomic radii with an added effective solvent radius up to 2.0 Å were tested. The interplay of the side-chain and solvent electrostatic effects was investigated by considering the side chains and solvent separately as well as together. The sensitivity to side-chain conformational fluctuations and to the parametrization of C_β group partial charges was also tested. Simulation results were compared to the experimental amide I′ spectra of P22 subdomain, including two ^{13}C isotopically edited variants, as well as to the previous simulations based on the molecular dynamics trajectory in explicit solvent. For small cavity sizes, between van der Waals and that with added solvent radius of 0.5 Å, better qualitative agreement with experiment was obtained than with the explicit solvent representation, in particular for the ^{13}C-labeled spectra. Larger protein cavities led to progressively worse predictions due to increasingly stronger electrostatic effects of side chains, which could no longer be well compensated for by the solvent potential. Balance between side-chain and solvent electrostatic effects is important in determining the width and shape of the simulated amide I′, which is also virtually unaffected by side-chain-geometry fluctuations. The continuum solvent model combined with the electrostatic map is a computationally efficient and potentially robust approach for the simulations of IR spectra of proteins in solution.



INTRODUCTION

The sensitivity of the amide I/I′ infrared (IR) absorption to the physical environment has proven to be very valuable for probing secondary structure and dynamics of proteins.[1] There are numerous factors that contribute to spectral characteristics, including mode coupling, hydrogen bonding, and solvation.[1−3] In order to translate IR spectra into structural information, a detailed understanding of how all of these processes affect the amide I signal is necessary. In this regard, theoretical simulations of the IR spectra for solvated proteins should be very informative.

One of the most outstanding challenges in simulating protein spectra is accurately accounting for the effects of the solvent (water), which is known to dramatically influence amide I absorption frequencies as well as intensities.[3−15] In many proteins, amides in different parts of the molecule experience different degrees of solvation ranging from total exposure to solvent to complete burial in hydrophobic regions.[12,16,17] Due to the strong effects of the solvent on the amide vibrations, this differential solvation of the protein amide groups may lead to additional broadening and band shape changes that can significantly complicate the spectral analyses.[12]

Electrostatic maps have been developed as an efficient means for approximating the effects on frequency and intensity that arise from the aqueous environment. The electrostatic map approach empirically relates the vibrational frequency and intensity shifts of the solute vibrations with respect to their gas-phase values to the electrostatic potential, field, and/or field gradient of the surrounding medium.[3,9,11,18−20] Recently, we have applied this methodology to simulation of the amide I′ IR spectrum for a small 40-residue subdomain of a P22 viral coat protein,[21] which has also been experimentally studied in our laboratory.[22] The electrostatic map, obtained from the explicit water molecules along a molecular dynamics (MD) trajectory, was used to correct a DFT-level vibrational force field and intensity parameters[20] calculated for a reduced, all-Ala analogue of the protein in gas phase.

It was shown that the predominant characteristics of the amide I′ spectrum of the P22 subdomain arise from the protein backbone vibrational parameters, with solvent playing only a secondary role.[21] In fact, the electrostatic map solvent

Received: June 1, 2012
Revised: August 1, 2012
Published: August 14, 2012

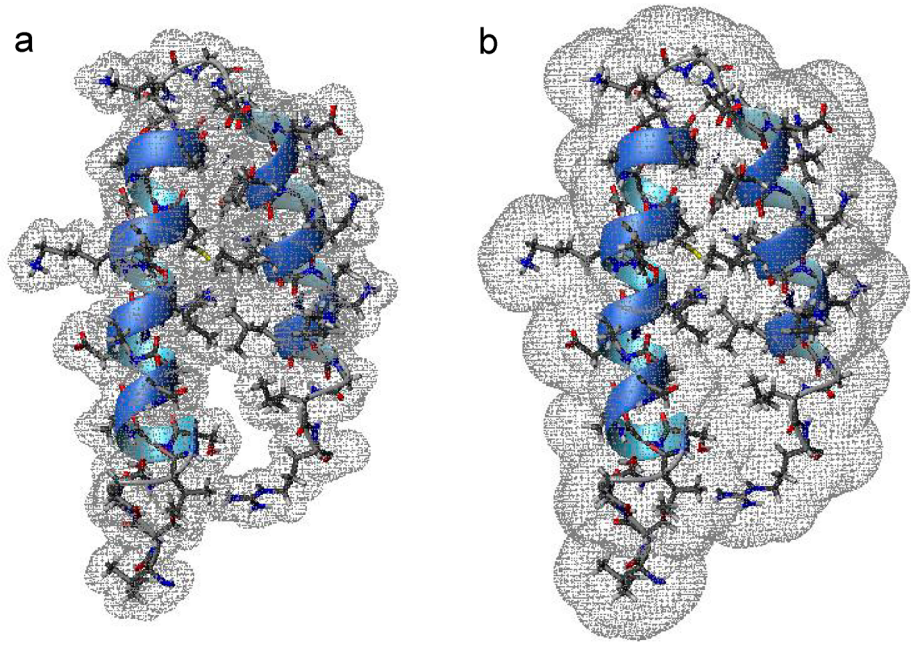


Figure 1. Molecular surface of the P22 subdomain used for parametrization of the continuum solvent model: (a) the van der Waals surface created using van der Waals atomic radii R_{VdW} and (b) the solvent-accessible surface created with $R = R_{VdW} + 1.5$ Å. The images were generated using MOLMOL.[29]

correction, rather than improving the agreement of the simulation with experiment, made it worse. In particular, the amide I′ contours were predicted to be too broad and nonuniform, with partially resolved shoulders arising predominantly from the most solvent exposed parts. Even more telling was the comparison of the amide I′ profiles for ^{13}C isotopically labeled P22: although the gas-phase all-Ala spectra qualitatively reproduced experiment, adding the solvent correction failed to give even qualitatively correct patterns.[21] The explicit solvent model therefore introduced unwanted artifacts, while the remarkably good performance of the gas-phase calculations could be rationalized by an apparent cancellation between the effects of side chains and water.

Another important result was that details of the solvent dynamics had qualitatively small effects on the simulated amide I′ band shapes. Motional narrowing, modeled by the time-averaging approximation (TAA),[23] led to narrower mode dispersion compared to the inhomogeneous limit as expected. However, the change in the amide I′ bandwidth was relatively small and the shape remained qualitatively the same. While this may seem at odds with the importance of motional narrowing of amide I in *N*-methylacetamide (NMA),[11,18,24,25] it is primarily due to the spread of the multiple protein amide modes, which dominates the bandwidth. This is further corroborated by the explicit calculation of the normal modes under average potential from water configurations over the entire trajectory, which yielded only very small additional narrowing and the same band shape.[21] The latter result suggests that MD simulations in explicit water might be unnecessary and, since only the average potential from the solvent may be a good approximation, a continuum model might be appropriate which requires much less computational effort.

In the present study, we test this hypothesis by using the same model protein (the P22 subdomain) and DFT/electrostatic map approach, but considering a continuum solvent model instead of the explicit water.

METHODS

Parameterization of the Continuum Solvent Model. The structure of the P22 subdomain was taken from its NMR solution structure (PDB code 2GP8).[21,26] Molecular cavities were generated using the program GEPOL93.[27] The van der Waals surface is obtained from spheres with a van der Waals radius,[28] R_{VdW}, generated around each atom, approximated by pentakis dodecahedra, and points belonging to the intersection of any two atomic tessellated spheres deleted (Figure 1a). Solvent-accessible surfaces (SAS) usually consider some effective radius R_{solv} of the solvent molecule that is added to the solute van der Waals radii. The cavities are computed in the same way, but by using radius $R = R_{VdW} + R_{solv}$ (Figure 1b). Since there is no universal consensus about the most appropriate radii, we tested surfaces constructed using the following values: R_{VdW} (van der Waals surface), R_{VdW} + 0.25 Å, R_{VdW} + 0.5 Å, R_{VdW} + 1 Å, R_{VdW} + 1.5 Å, and R_{VdW} + 2.0 Å.

Using the tessellated molecular surface, the implicit solvent model is constructed assuming the continuum outside the molecular cavity is a conductor, following the COSMO model of Klamt.[30] The electrostatic response of the solvent is expressed in terms of surface charges q induced by the solute on the individual tessera of the cavity, which are obtained as follows

$$\vec{q} = -\mathbf{A}^{-1}\mathbf{B}\vec{Q} \quad (1)$$

where Q are the partial charges on all protein atoms (including the peptide backbone). We use atomic partial charges as defined in AMBER99 force field.[31] Matrices **A** and **B** contain reciprocal distances between different cavity points and between cavity points and atoms, respectively:

dx.doi.org/10.1021/jp305387x | J. Phys. Chem. B 2012, 116, 10739−10747

$$A_{ik} = \frac{1}{r_{ik}} \quad \text{for } i \neq k$$
$$A_{ii} = 3.8\sqrt{\text{area}_i}$$
$$B_{ij} = \frac{1}{r_{ij}} \tag{2}$$

In (2), r_{ik} is the distance between the ith and kth tesserae centers, area_i is the area of the ith tessara, and r_{ij} is the distance between the ith tessera center and the jth protein atom. The self-interaction term A_{ii}, which corrects for the finite area of the tesserae, is due to Klamt.[30]

Simulations of the Amide I′ Spectra. The amide I′ vibrations for the peptide backbone of the P22 subdomain were modeled by transferring force constants and dipole derivatives (atomic polar tensors) onto an all-Ala analogue of the protein from DFT BPW91/6-31G** calculations on smaller fragments as previously described.[21] The DFT force field and intensity parameters were then corrected using the electrostatic potential at backbone atoms using an electrostatic map developed by Bour[20] based on the following empirical relationships

$$\Delta\omega(i) = \sum_j l(i)_j \varphi(i)_j \quad \Delta D(i) = \sum_j m(i)_j \varphi(i)_j \tag{3}$$

where the shift in frequency, $\Delta\omega(i)$, or dipole strength $\Delta D(i)$ for a mode i from a reference state (the gas phase) to that in an electrostatic medium is proportional to electrostatic potentials (φ) on atoms j via the empirical parameters $l(i)_j$, $m(i)_j$. Vibrational force fields and dipole derivatives are used to compute the vibrational frequencies and intensities using standard methods. Spectral band shapes were calculated by assigning a pseudo-Voight function (70% Gaussian and 30% Lorentzian) to each transition with the full width at half-maximum (fwhm) of 14 cm^{-1}, except for the ^{13}C isotopically labeled difference spectra, where a fwhm of 24 cm^{-1} was used. (The same width was used in the preceding study,[21] although it was wrongly specified as 7 cm^{-1}, which is half-width (hwhm), rather than full width (fwhm) of the band.) Since oscillators with different degrees of solvent exposure may exhibit different line widths,[32,33] the uniform fwhm values can be regarded as an approximation, representing an effective average value.

The electrostatic potential φ in (3) is calculated from the cavity surface charges q, representing the solvent obtained from (1) together with protein side-chain partial charges Q (AMBER99). The amide backbone atoms are excluded, since they are already contained in the DFT vibrational parameters. However, since Ala side chains ($-C_\beta H_3$ groups) are also explicitly accounted for by the DFT calculations, it is not completely clear how to treat the side-chain β groups. As a default, we adopted the most direct scheme that uses the full AMBER99 partial charges of these groups. Several other charge assignments were also tested: (i) H_β partial charges set to zero, (ii) correcting the C_β partial charge by subtracting that of Ala C_β, (iii) half of this corrected value, and (iv) zero partial charge for the β groups.

Finally, to account for the side-chain structural fluctuations, which may directly influence the solute cavity, we made use of an equilibrated 100 ps molecular dynamics trajectory of the P22 subdomain in explicit water, described previously.[21] Since the DFT-based approach limits the simulation of the spectra to a single protein geometry, the peptide backbone was constrained in the same conformation used for the DFT all-Ala calculations (see above). The trajectory therefore samples only the solvent and side-chain fluctuations, which were left unconstrained. The molecular cavities were computed for 166 peptide geometries separated by 600 fs: this value was chosen since it approximately corresponds to the averaging time proposed to take into account the motional narrowing.[21,23] To explicitly test the effect of the variations in side-chain geometries on the simulated spectra, the amide I′ band shapes were calculated both for individual MD snapshots selected for maximum differences in the side-chain arrangements, as well as for averages over the individual configurations, and compared.

Comparison Experimental Data. The experimental spectra of the P22 subdomain, along with its ^{13}C isotopically labeled variant, were collected in D_2O phosphate buffer, pH 7 at 0.7 °C, as previously described.[22] The amide I′ band was extracted from the experimental spectrum by subtracting the amino acid side-chain signals, which were constructed using published parameters.[21,34]

RESULTS

Optimization of Protein Molecular Cavities. The molecular cavity is a critical parameter of any continuum solvent model. Simulated amide I′ IR spectra for the P22 subdomain using different cavity sizes, obtained using radii ranging from van der Waals (R_{VdW}) to R_{VdW} + 2.0 Å are compared in Figure 2. Previous results from an averaged 100 ps MD trajectory in explicit water[21] are also shown, along with the gas phase all-Ala amide I′ (Figure 2c), and an experimental spectrum (Figure 2d).

The smallest cavity (van der Waals) produces amide I′ at significantly lower frequency than any of the larger cavities. Even with just 0.25 Å larger radius, the spectrum shifts by ~20 cm^{-1} higher. Another 0.25 Å increase (van der Waals +0.5 Å radius cavity) results in further ~10 cm^{-1} shift of the maximum. The band shapes, however, remain effectively unchanged—the partially resolved features only become less pronounced with larger cavities and the width (fwhm) is nearly the same (~40 cm^{-1}). Increasing the cavity radius further (Figure 2b) does not significantly shift the band position. For the cavity with 1 Å larger than R_{VdW} the band shape is narrower than the R_{VdW} + 0.5 Å, with less pronounced features to the low-frequency side, and on the high-frequency side the R_{VdW} + 0.1 Å spectrum perfectly lines up with the all-Ala spectrum (Figure 2c). Further increase of the atomic radii (R_{VdW} + 1.5 Å) starts to produce an additional high-frequency shoulder which shifts even higher for the largest cavity (R_{VdW} + 2.0 Å). For the latter the main maximum also starts to split into two peaks and the low-frequency feature is more distinct.

It is difficult to find a physical explanation for the amide I′ band features above the gas-phase frequencies, as the solvent is expected to shift the amide frequencies significantly lower.[6,10] On the other hand, it is readily explained by the protein molecular cavity becoming too large: for large cavities the water electrostatic potential becomes too weak to properly balance that of the side chains. This was suggested in the preceding study using the explicit representation of water[21] and is explored more in detail in the following section. Therefore, from Figure 2a,b, it can be concluded that the optimum cavity size is obtained with the radii between van der Waals and R_{VdW} + 1 Å. We note that the latter was also used by Klamt in the original COSMO paper.[30]

How do the implicit solvent amide I′ simulations compare to those with the same electrostatic map but explicit water and to

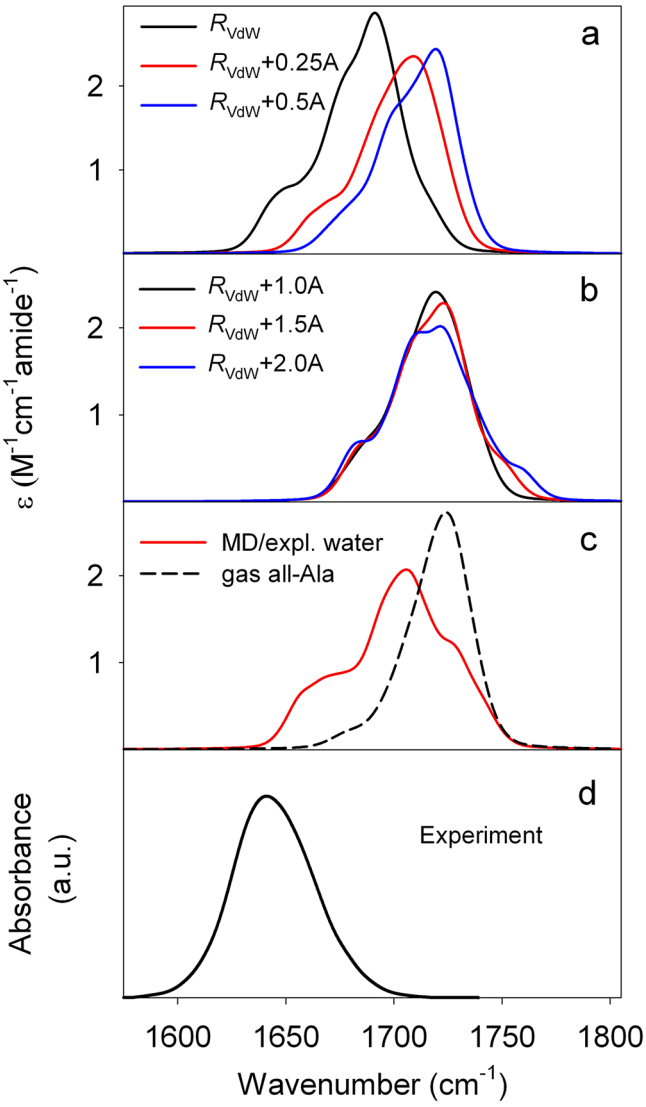


Figure 2. Simulations of amide I′ IR spectra for P22 subdomain using the electrostatic map with the continuum solvent model and different size molecular cavities: (a) van der Waals surface (black), surface constructed using R_{VdW} + 0.25 Å (red) and R_{VdW} + 0.5 Å (blue); (b) R_{VdW} + 1 Å (black), R_{VdW} + 1.5 Å (red), and R_{VdW} + 2.0 Å (blue). (c) Simulation of amide I′ in solution using the same electrostatic map from an averaged 100 ps MD trajectory in explicit solvent (red, ref 21) and gas-phase all-Ala representation of the protein (black), used as an input to all electrostatic map solvent corrections. (d) Experimental amide I′ spectrum of the P22 subdomain corrected for the side-chain absorptions.

experiment? The continuum solvent spectra are uniformly much narrower (40 cm^{-1} versus 53 cm^{-1}) than that from the explicit solvent simulation (Figure 2c). Nevertheless, the band shapes have the same qualitative features: partially resolved shoulders on both sides of the main maximum, but especially to the lower frequency side. In the continuum solvent calculations these features are less pronounced resulting in more uniform spectral contours. The van der Waals cavity spectrum is lower in frequency (by 10 cm^{-1} at the maximum), while the R_{VdW} + 0.25 Å maximum is 4 cm^{-1} higher than the amide I′ peak in explicit water. The significant distinction, however, is that for these two smallest cavities, the whole band is shifted lower from the gas-phase amide I′, while the explicit water contour stretches to the same high-frequency limit. In comparison to experiment (Figure 2d), aside from the much higher frequencies, the shapes of the all simulated amide I′ contours have the opposite asymmetry (more gradual on the low-frequency side of the maximum). The simulated bands also have partially resolved features, indicating less uniform normal-mode distribution. Since the continuum solvent model simulations are narrower than the experimental band (fwhm ~45 cm^{-1}), an increase in the width of the band-shape function would partially offset this, while the band shape from the explicit solvent simulation is much broader.

An even more stringent test for the implicit solvent model is the simulation of the isotopic effects on the amide I′ in ^{13}C-labeled proteins. Isotopic labeling of specific residues can be used to elucidate information about the local structure and environment of the labeled groups.[32,35,36] Similarly, the uniquely shifted amide I bands of proline can be used as a local probe.[37,38] While good agreement with experiment for isotopically labeled systems has also been demonstrated using coupled oscillator models[39,40] as well as DFT methods,[41−43] correct prediction of frequency distributions of amide I signals for labeled oscillators using electrostatic maps has proven to be elusive,[21,44,45] despite their very good agreement with the unlabeled experimental spectra. The reason is that the isotopic shifts and intensity changes are much more sensitive to the correct underlying normal-mode distribution than the overall band-shape contour. For the same reason, simulations of isotopically edited spectra are also a very useful test for the most realistic molecular cavity.

The amide I′ IR spectra were calculated for two of the experimentally studied[22] isotopic variants of the P22 subdomain: the first with ^{13}C on Ala_{19}, Ala_{20}, and Ala_{21} (denoted h1), and the second on Leu_{32}, Ala_{34}, and Leu_{36} (h2). The results are summarized in Figure 3, in the form of difference spectra (^{13}C-labeled minus unlabeled), which are the most convenient way to highlight the isotopic effects. A broader line-shape function (fwhm of 24 cm^{-1}) was used for the difference spectra; while it does not alter the results, it provides a better comparison of the important, dominant features.

It is clearly seen that the continuum simulations are in best qualitative agreement with experiment for radii R_{VdW} + 0.25 Å and R_{VdW} + 0.5 Å. The van der Waals radius calculation still reproduces the most important features, namely the relative positions and intensities of the h1 and h2 positive differences, but the h1 negative signal has a somewhat different shape. For R_{VdW} + 1 Å the agreement becomes worse as the positive h1 signal broadens and shifts lower with respect to that of h2. This trend of less and less satisfactory agreement continues for R_{VdW} + 1.5 Å and R_{VdW} + 2.0 Å (not shown) consistent with the above, unlabeled amide I′ results (Figure 2). The isotopic difference spectra therefore suggest that the optimal cavity size is between R_{VdW} and R_{VdW} + 0.5 Å.

The most remarkable result is that compared to the explicit water simulation, which obviously fails to even resemble correct patterns, the continuum model gives qualitatively correct predictions, aside from the obvious frequency and width differences. As the gas-phase difference spectra are in qualitative agreement with experiment, the discrepancy is due to including the electrostatic correction with the explicit water. Despite the similar overall amide I′ band-shape patterns (Figure 2), the continuum solvent preserves the correct normal-mode distribution, while the explicit water simulation does not. With the continuum solvent and optimum cavity sizes (~R_{VdW} to R_{VdW} + 0.5 Å) the experimental amide I′ data are therefore

dx.doi.org/10.1021/jp305387x | J. Phys. Chem. B 2012, 116, 10739−10747

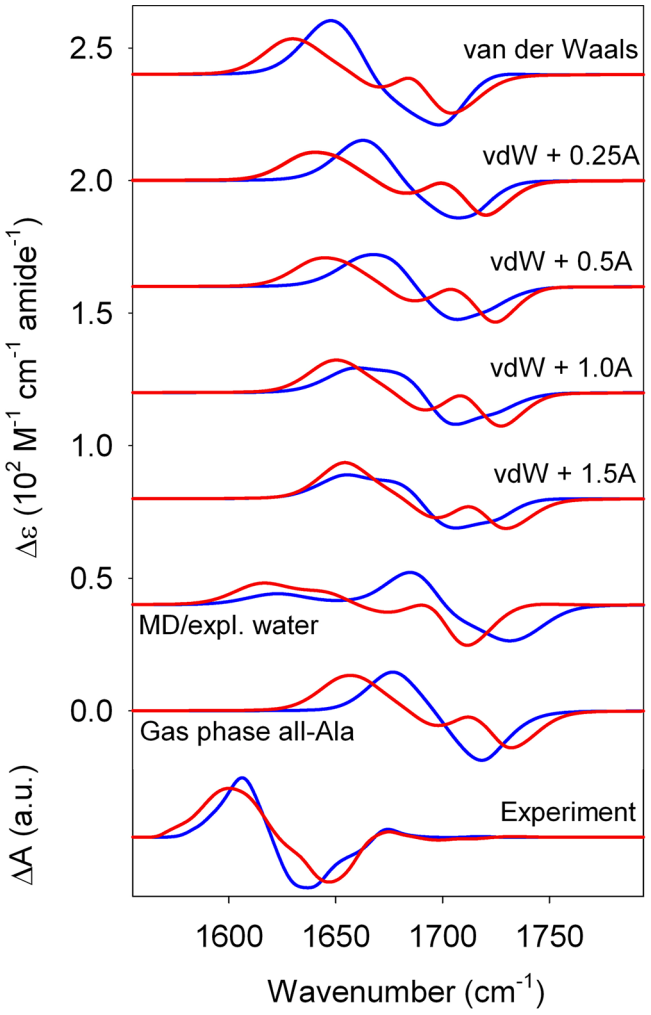


Figure 3. Simulations of amide I′ spectra for ^{13}C isotopically edited P22 subdomain. Difference spectra (^{13}C labeled − unlabeled) for h1 (blue) and h2 (red) isotopic labels. From top to bottom: simulations using the electrostatic map with the continuum solvent model for cavities calculated using van der Waals, R_{VdW} + 0.25 Å, R_{VdW} + 0.5 Å, R_{VdW} + 1.0 Å, and R_{VdW} + 1.5 Å radii; simulation using the electrostatic map and the MD trajectory in explicit water (ref 21); gas-phase all-Ala simulations and experimental data (ref 21). Spectra are offset by 40 M^{-1} cm^{-1} $amide^{-1}$.

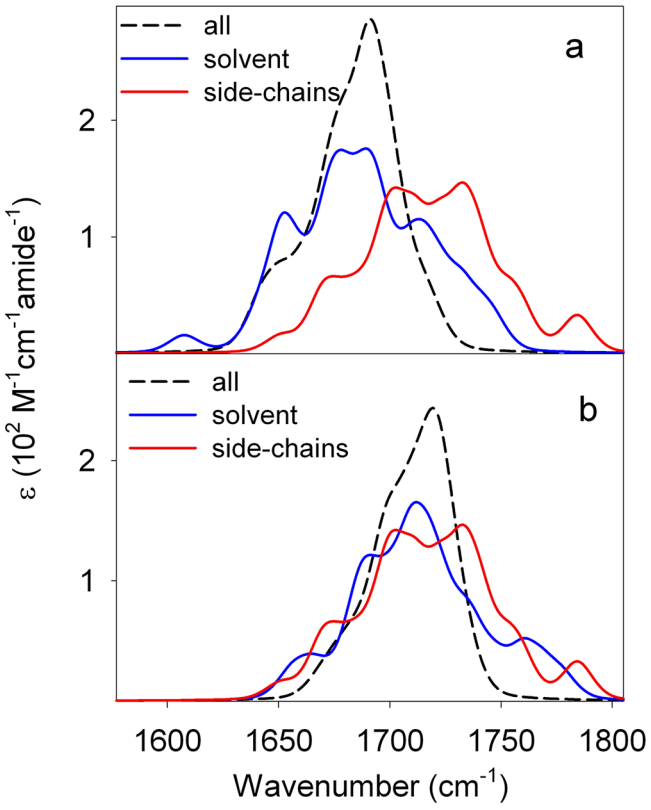


Figure 4. Electrostatic effects of the solvent and amino acid side chains on the amide I′ spectra of P22 subdomain. Simulated spectra with the electrostatic map and continuum solvent model using (a) van der Waals cavity radius and (b) R_{VdW} + 0.5 Å predicted by considering the electrostatic potential from both side chains and solvent (dashed black line), just the solvent (blue). Spectra simulated using just the side-chain partial charges (red) are independent of the cavity.

reproduced much more reliably than with the explicit water using the same electrostatic map.

Electrostatic Effects of the Side Chains. The arrangement of the solvent and its resulting electrostatic potential naturally respond to the charge distribution of the protein, and contributions of both the side chains and the solvent have to be included in the electrostatic map in a balanced fashion.[21] As already mentioned above, this is the reason why increasing the size of the molecular cavity beyond the optimum radius results in unrealistic simulated spectra. It is interesting to explore how the amide I′ spectra are affected by the electrostatic potential of the side-chain charges and the solvent individually.

Simulated amide I′ spectra with the electrostatic map containing potentials from both the side chains and (continuum) solvent, solvent only, and side chains only are overlaid in Figure 4 for two different cavity sizes, van der Waals and R_{VdW} + 0.5 Å. Omitting either the side-chain or the solvent potentials leads to much broader and more structured spectral contours, with multiple partially resolved features. Both are broadened beyond the high-frequency limit of the gas-phase simulations, although the corresponding high-frequency modes differ. In the solvent-only simulation, the high-frequency shoulder is due to amides in a region of high surface positive charge density on the C-terminal helix, which was also observed with the explicit water model when the side-chain partial charges were neglected.[21] For spectra generated considering only the potential from side-chain atoms (obviously, these are independent of the cavity model), the characteristic high-frequency peak at about 1775 cm^{-1} corresponds to the extreme N-terminal amide vibration (isoleucine in the P22 subdomain sequence). Substantial narrowing, and disappearance of the additional high- and low-frequency band components by the combination of both the solvent and side-chains potentials in the electrostatic map, show explicitly the importance of the cancellation of these two contributions.

Since the electrostatic potential depends inversely on distance, side-chain atoms of the C_β groups which are closest to the peptide backbone can be expected to have significant effects on the predicted spectra. As the DFT gas-phase spectra calculations were done on an all-Ala analog of the protein, C_β methyl groups were in principle taken into account before the corrections using the electrostatic map. Because most residues in P22 are not Ala, it is questionable what partial charges should be used for non-Ala C_β. As described in Methods, we adopted the most straightforward scheme, assigning standard AMBER99

dx.doi.org/10.1021/jp305387x | J. Phys. Chem. B 2012, 116, 10739−10747

partial charges for all C_β group atoms, but tested additional possible schemes. The results are summarized in Figure 5 for two different sizes of the solvent cavity.

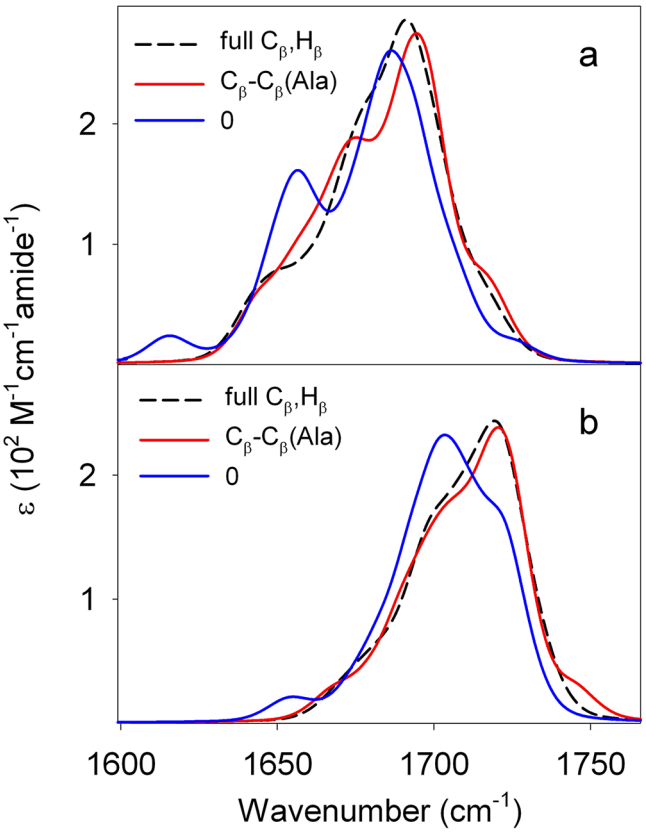


Figure 5. Tests of different partial charge assignments for C_β side-chain groups. Simulated amide I′ spectra for the P22 subdomain with the electrostatic map and continuum solvent model using (a) van der Waals surface and (b) R_{VdW} + 0.5 Å surface with the "default" unaltered AMBER99 partial charges for C_β groups (dashed black), with H_β partial charges set to zero (red) and with all β atom partial charges set to zero (blue).

The exclusion of hydrogen partial charges was tested, because the hydrogens may be positioned even closer to the amide group atoms than the carbons, and potentially introduce artifacts. However, compared to the "default" full C_β group partial charges, excluding the hydrogens seems to broaden the simulated spectra, in particular again introducing the unwanted high-frequency shoulder. On the other hand, correcting the C_β partial charges by subtracting charges for Ala yielded virtually indistinguishable spectra from those using full partial charges, and are not shown. This may be due to the relatively small AMBER99 partial charge for Ala C_β, about $-0.18e$. At the other extreme, assigning zero partial charges to the C_β groups obviously yields unrealistic spectra, additional shoulders now appearing at the low-frequency side. To interpolate between the extremes, we have also tested partial charges that were half of the AMBER99 C_β value. The data are not shown, but as may be expected, the simulated contours fall in between the full C_β and zero partial charges in Figure 5. Simply using the partial charges defined by AMBER99 force field for all atoms except the amide backbone seems to yield the most satisfactory simulated spectra.

Geometry Fluctuations. The simulated spectra presented above are averaged over many configurations of the amino acid side chains, extracted from an MD trajectory (Methods), which was available from the preceding explicit solvent study.[21] However, one of the goals of introducing the continuum solvent is to avoid the lengthy MD simulations, along with even more time-consuming averaging that follows.[21] It is therefore important to examine to what extent the simulated spectra are affected by the fluctuations in the side-chains, which also cause changes in the molecular cavities. For this reason, we selected individual MD snapshots that appeared to produce the largest variability in the simulated amide I′ spectra. Five of these are overlaid in Figure 6.

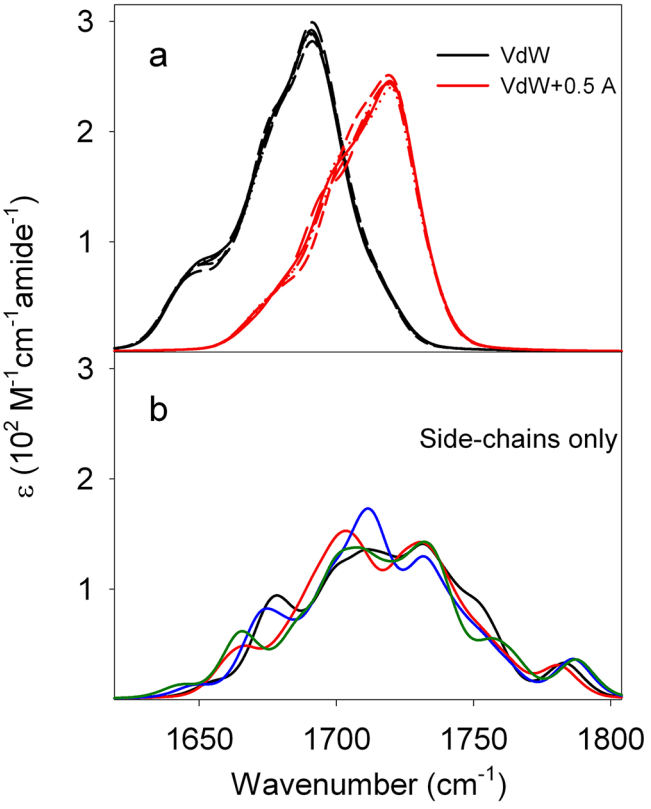


Figure 6. Effects of side-chain-geometry fluctuations. Simulated amide I′ spectra for five structures from a 100 ps MD trajectory of the P22 subdomain in explicit water using the electrostatic map and (a) continuum solvent with van der Waals (black) and R_{VdW} + 0.5 Å (red) surfaces and side-chain partial charges, and (b) side-chain partial charges only. The different colors correspond to the individual side-chain configurations (MD snapshots).

It is obvious that for the two cavity sizes shown the simulated amide I′ spectra are only negligibly influenced by the side-chain-geometry fluctuations. Again, this is due to the cancellation between the electrostatic potential of the solvent and side-chain partial charges, as illustrated by much greater effects when the side chains alone are used in the electrostatic map (Figure 6b). Similar variability is observed in solvent-only simulations (not shown). The results for other cavity sizes are not shown, but as can be expected for larger than ~ R_{VdW} + 1 Å cavities, the fluctuations cause increasingly more spectra variation, as the side-chain potential overwhelms that of the solvent. With the smaller cavities, which are also most consistent with the experimental band shapes (Figure 2) and

dx.doi.org/10.1021/jp305387x | J. Phys. Chem. B 2012, 116, 10739−10747

^{13}C isotopic spectral patterns (Figure 3), the simulated amide I′ spectra are virtually independent of particular side-chain geometries.

DISCUSSION

Our initial objective was to find an implicit solvent representation, which in combination with the electrostatic map allows simulations of the amide I′ spectra comparable to those obtained with the explicit solvent MD simulation.[21] In fact, the continuum representation of the solvent yields substantially improved amide I′ predictions. The first significant improvement is that the amide I′ band shapes simulated with the continuum solvent model are narrower than those calculated with the explicit solvent. Excessively broad simulated bands have been a major problem of this electrostatic map approach.[7,20,21,46,47] As we have shown before,[21] taking into account motional narrowing through TAA[23] did lead to narrower band shapes, but the change in width compared to the inhomogeneous limit spectrum was relatively small. This may be attributed to the limitations of TAA, which was subsequently improved.[48] However, even with the completely averaged water (and side-chain) electrostatic potential, the broad dispersion in the normal modes persists, giving only slightly narrower width than TAA.[21] By contrast, the continuum solvent eliminates the excess broadening, although it corresponds to the same limit in that effective, average potential is considered.

Second, and even more importantly, the ^{13}C isotopically edited spectra which failed to even resemble experiment in the explicit water simulations[21] are calculated qualitatively correctly with the continuum solvent. Reliable simulations of ^{13}C spectral patterns have been difficult to achieve using electrostatic map solvent corrections[21,44,45] even though generally good agreement with overall band shapes could be obtained.[32,35,36] Interestingly, this suggests errors due to the electrostatic map scaling, given that with simple coupled oscillators models, but not accounting for solvent, experimental spectra could be simulated very closely.[39,40] Likewise, gas-phase DFT methods proved very reliable in this respect.[41−43] Although before we attributed the discrepancies of the explicit solvent/MD results to the empirical electrostatic map, it now appears that the map works very well with the continuum solvent. While explicit solvent is generally considered more realistic, its much poorer performance compared to the implicit, reaction field solvent model is perhaps not that surprising. The implicit model is naturally based on electrostatics as is the electrostatic map. On the other hand, the explicit water, which can approximate specific interactions such as hydrogen bonding (whether correctly or incorrectly) may not be entirely compatible with the simplicity of the electrostatic potential map.

The continuum solvent model is much more computationally efficient than the explicit water approach. The latter not only requires the MD trajectory but also subsequent averaging,[21] or even more laborious partial optimization of fast degrees of freedom to account for motional narrowing.[47] Furthermore, as we have shown above (Figure 6), the results are virtually invariant with respect to the small geometry fluctuations in the protein side chains. Although we calculated averages over a number of configurations from the MD trajectory, since we had computed it before, using any single protein structure would have produced the same results, as would averaging over more side-chain configurations from the same or longer MD simulations. However, we again stress that only the protein side chains can vary in these simulations, which, through the corresponding changes in the computed solute cavity, also effectively represents the fluctuations of the solvent. The polypeptide backbone conformation, on the other hand, is held constant: this is necessitated by the transfer of DFT vibrational parameters that can be calculated only for the particular protein conformation, as detailed in Methods.

Comparing the simulations with the experimental spectrum, perhaps the most obvious discrepancy is that amide I frequencies are calculated significantly higher. This is often attributed to a systematic shift in DFT-predicted vibrational frequencies. However, it has been demonstrated that DFT calculations for gas-phase frequencies may not be far off from experiment.[6] For example, the amide I frequency calculated for NMA at the same DFT level used here (BPW91/6-31G**) is 1735 cm^{-1}, compared to ∼1725 cm^{-1} experimental gas-phase frequency. The solvent (water) is responsible for the shift of the NMA amide I by about a 100 cm^{-1} to ∼1625 cm^{-1}. It was also shown that, with the solvent taken into account, the DFT frequency is again much closer to experiment.[6] While theoretical works often pick the most convenient "unperturbed frequency" or a scaling factor, they may in fact completely bypass an important effect. For this reason, we do not use scaling, other than the electrostatic map, which models only the electrostatic environment. The gas-phase or "unperturbed" frequencies are unscaled DFT values. As seen above, this can be useful in judging the validity of different approximations, such as parametrizations of the molecular cavities or even the electrostatic map model.

Even though traditionally the vibrational frequencies are regarded as the IR "signatures" of different structures, it is now clear that they are highly variable.[1] From DFT calculations with explicit water, it was shown that solvent can significantly shift the amide bands in frequency without essentially changing the normal-mode distributions.[10,12] This is also apparent from the above results (Figure 2) for different size cavities (e.g., van der Waals, versus van der Waals +0.5 Å), which produced amide I′ contours shifted by ∼30 cm^{-1}, but similar shapes and isotopic patterns.

The discrepancy in frequencies, though perhaps visually unpleasant, is therefore not the most important criterion for judging the quality of the simulations. More important is the band shape, which, unfortunately, is also not well reproduced by our simulations. The qualitatively correct amide I′ shape was obtained only from calculations using DFT for a cluster of the all-Ala model protein with explicit water molecules hydrogen bonded to amide groups.[21] It was also shown that the electrostatic map gives a very different result for this model, implying, again, that it cannot reliably describe specific, hydrogen-bonding interactions with the solvent.[21] Neglect of these effects is one possible explanation. Another is that the protein is not fully folded even at 0.7 °C where the experimental spectra were measured: in fact, our unfolding ^{13}C isotopically edited experiments[22] suggest that the N-terminus is significantly disordered. This can obviously be addressed only by considering an ensemble of different backbone conformations. By contrast, our ^{13}C edited simulation results presented here are consistent with experiments, which show that the labeled regions are fully folded at cold temperatures.

Concerning the simulated spectral shapes, as pointed out in Methods, another approximation used here is that for simulating the spectral contours the same line-shape function

is assigned to all transitions. It was shown[32,33] that the line widths associated with particular amide transitions may vary depending on the environment, in particular the solvent exposure. In this approach, however, the modes are largely delocalized, with varying contributions from amide groups that may be considered exposed or buried from the solvent.[21] Furthermore, as discussed in more detail below, the electrostatic effects of amino acid side chains complicate this distinction even further. The uniform widths are therefore best regarded as an effective average. Using average widths is justified on one hand as naturally fitting our simplified approach, which takes into account the external environment (solvent) in an average fashion, and on the other as the simplest solution and one most commonly done in the absence of any definite indication of the line width or line shape distribution.

The major assumption of continuum solvent models is that of a well-defined "solute cavity". There is no consensus in the literature about the optimal cavity sizes. For example, Negre et al.[49] simply took the van der Waals radii of solute atoms to construct the cavities, while Tomasi and co-workers[50] added 20%. When calculating the zwitterion of glycine, however, it was found that reduced cavity sizes were needed for N and O of the charged amino and carboxylate groups. The Gaussian suite of programs[51] has van der Waals radii +10% as the default. Klamt, whose COSMO model we adopted, has taken van der Waals radii plus 1 Å for all atoms.[30] For proteins, Lee and Richards' radius R_{solv} = 1.4 Å, which corresponds to the "spherical" water molecule, is often used to define the SAS.[52] According to our results, the optimal size cavity is smaller, between those constructed with the van der Waals radius (R_{solv} = 0 Å) to R_{solv} = 0.5 Å. Within this range, the frequencies of the simulated amide bands change significantly, but the band shapes and ^{13}C isotopically edited patterns change very little. This reinforces the above-discussed point that the dominant solvent effect on the amide I′ spectrum is the frequency shift of the band as a whole, but very little perturbation to the underlying normal-mode structure. Within these limits, the particular choice of the radii for cavity construction is therefore not critical.

By contrast, when the radii are increased slightly beyond the optimum range ($R_{solv} \gtrsim 0.5$ Å), unwanted features start to appear in the simulated spectra, as the contribution of the side-chain electrostatic potential begins to overwhelm that of the solvent. The opposite, i.e., distortion of the simulated spectra by too strongly weighted solvent potential, as might be expected for the smallest cavity sizes, did not occur. Perhaps distortions toward the spectral shapes resembling predictions with the solvent only electrostatic map might be seen for even smaller cavity sizes; however, it did not seem to make much physical sense to use smaller radii than the van der Waals. This underlines the role of the side chains, within the electrostatic map model, as a dominant factor in shaping the protein amide I′ spectra. The importance of the side chains is easy to rationalize by their proximity to the amide groups. The solvent, on the other hand, mainly counteracts the electrostatic effects of the side chains.

This mutual electrostatic shielding of the side chains and solvent shows that reasonable approximations to the amide I′ spectra can be obtained from considering just the polypeptide backbone, as exemplified here by the gas-phase all-Ala protein model. As also noted previously,[21] this is consistent with the success of the very simple, coupled oscillator models of amide I in proteins, in which both the solvent and side-chain effects are completely neglected.[4,39,40,53] Nevertheless, given the aforementioned importance of the side chains' effects, it is obviously desirable to model them at higher level than simply a collection of atomic partial charges, derived from the molecular mechanics force field, as done here. Taking the side chains into account at the DFT level, consistently with the backbone, will also resolve the problems with parametrization of C_β side-chain groups, which had to be resolved empirically. Such simulations are the objective of our future studies.

CONCLUSIONS

We have combined an empirical electrostatic map with an implicit solvent model and DFT calculations to simulate amide I′ spectra for a small protein in aqueous solution. Continuum solvent is not only a more computationally efficient alternative to explicit water/electrostatic map approaches but also eliminates some of their unwanted artifacts. In particular, much better agreement with experiment was obtained for ^{13}C isotopically labeled spectra simulations, which are more sensitive to the normal-mode structure than the overall band contours. We have also shown that the most important role of the solvent, at least within the electrostatic map approximation, is the cancellation of the side-chain electrostatic potential. The effective cancellation of the side-chain effects by solvent requires relatively small molecular cavities: the best results were obtained with the surfaces constructed with the van der Waals atomic radii up to surface the solvent probe radius of 0.5 Å. Within this range, the simulated amide I′ band shapes and isotopic patterns remain relatively unchanged and qualitatively correct, despite substantial frequency shifts of the amide I′ band as a whole. Moreover, for these molecular cavities the simulated amide I′ spectra are essentially insensitive to the side-chain-geometry fluctuations. The continuum solvent model therefore allows spectral simulations for a given protein conformation without the need for molecular dynamics trajectories and extensive conformational averaging over side-chain and solvent configurations.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jkubelka@uwyo.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation, Grant CAREER 0846140.

REFERENCES

(1) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2002**, *35*, 369−430.
(2) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181−364.
(3) Ham, S.; Cho, M. *J. Chem. Phys.* **2003**, *118*, 6915−6922.
(4) Torii, H.; Tatsumi, T.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 537−546.
(5) Torii, H.; Tatsumi, T.; Kanazawa, T.; Tasumi, M. *J. Phys. Chem. B* **1998**, *102*, 309−314.
(6) Kubelka, J.; Keiderling, T. A. *J. Phys. Chem. A* **2001**, *105*, 10922−10928.
(7) Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2003**, *119*, 11253−11262.
(8) Besley, N. A. *J. Phys. Chem. A* **2004**, *108*, 10794−10800.
(9) Lee, H.; Kim, S. S.; Choi, J. H.; Cho, M. *J. Phys. Chem. B* **2005**, *109*, 5331−5340.

(10) Kubelka, J.; Huang, R.; Keiderling, T. A. *J. Phys. Chem. B* **2005**, *109*, 8231−8243.
(11) Jansen, T. L.; Knoester, J. *J. Chem. Phys.* **2006**, *124*, 004502.
(12) Turner, D. R.; Kubelka, J. *J. Phys. Chem. B* **2007**, *111*, 1834−1845.
(13) Gorbunov, R. D.; Nguyen, P. H.; Kobus, M.; Stock, G. *J. Chem. Phys.* **2007**, *126*, 054509.
(14) Bagchi, S.; Falvo, C.; Mukamel, S.; Hochstrasser, R. M. *J. Phys. Chem. B* **2009**, *113*, 11260−11273.
(15) Cho, M. *J. Chem. Phys.* **2009**, *130*, 094505.
(16) Manas, E. S.; Getahun, Z.; Wright, W. W.; DeGrado, W. F.; Vanderkooi, J. M. *J. Am. Chem. Soc.* **2000**, *122*, 9883−9890.
(17) Walsh, S. T. R.; Cheng, R. P.; Wright, W. W.; Alonso, D. O. V.; Daggett, V.; Vanderkooi, J. M.; DeGrado, W. F. *Protein Sci.* **2003**, *12*, 520−531.
(18) Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. *J. Chem. Phys.* **2004**, *121*, 8887−8896.
(19) Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. *J. Chem. Phys.* **2004**, *120*, 8107−8117.
(20) Bouř, P.; Michalík, D.; Kapitán, J. *J. Chem. Phys.* **2005**, *122*, 144501.
(21) Grahnen, J. A.; Amunson, K. E.; Kubelka, J. *J. Phys. Chem. B* **2010**, *114*, 13011−13020.
(22) Amunson, K. E.; Ackels, L.; Kubelka, J. *J. Am. Chem. Soc.* **2008**, *130*, 8146−8147.
(23) Auer, B. M.; Skinner, J. L. *J. Chem. Phys.* **2007**, *127*, 104105.
(24) DeCamp, M. F.; DeFlores, L.; McCracken, J. M.; Tokmakoff, A.; Kwac, K.; Cho, M. *J. Phys. Chem. B* **2005**, *109*, 11016−11026.
(25) Kwac, K.; Cho, M. H. *J. Chem. Phys.* **2003**, *119*, 2247−2255.
(26) Sun, Y. H.; Parker, M. H.; Weigele, P.; Casjens, S.; Prevelige, P. E.; Krishna, N. R. *J. Mol. Biol.* **2000**, *297*, 1195−1202.
(27) Pascualahuir, J. L.; Silla, E.; Tunon, I. *J. Comput. Chem.* **1994**, *15*, 1127−1138.
(28) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.
(29) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph.* **1996**, *14*, 51−55.
(30) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799−805.
(31) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(32) Woys, A. M.; Lin, Y.-S.; Reddy, A. S.; Xiong, W.; de Pablo, J. J.; Skinner, J. L.; Zanni, M. T. *J. Am. Chem. Soc.* **2010**, *132*, 2832−2838.
(33) Roy, S.; Jansen, T. L. C.; Knoester, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9347−9357.
(34) Chirgadze, Y. N.; Fedorov, O. V.; Trushina, N. P. *Biopolymers* **1975**, *14*, 679−694.
(35) Lin, Y. S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. *J. Phys. Chem. B* **2009**, *113*, 592−602.
(36) Manor, J.; Mukherjee, P.; Lin, Y.-S.; Leonov, H.; Skinner, J. L.; Zanni, M. T.; Arkin, I. T. *Structure* **2009**, *17*, 247−254.
(37) Lessing, J.; Roy, S.; Reppert, M.; Baer, M.; Marx, D.; Jansen, T. L. C.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2012**, *134*, 5032−5035.
(38) Roy, S.; Lessing, J.; Meisl, G.; Ganim, Z.; Tokmakoff, A.; Knoester, J.; Jansen, T. L. C. *J. Chem. Phys.* **2011**, *135*, 234507.
(39) Paul, C.; Wang, J. P.; Wimley, W. C.; Hochstrasser, R. M.; Axelsen, P. H. *J. Am. Chem. Soc.* **2004**, *126*, 5843−5850.
(40) Brauner, J. W.; Flach, C. R.; Mendelsohn, R. *J. Am. Chem. Soc.* **2005**, *127*, 100−109.
(41) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 8163−8163.
(42) Silva, R.; Kubelka, J.; Bouř, P.; Decatur, S. M.; Keiderling, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8318−8323.
(43) Huang, R.; Kubelka, J.; Barber-Armstrong, W.; Silva, R.; Decatur, S. M.; Keiderling, T. A. *J. Am. Chem. Soc.* **2004**, *126*, 2346−2354.
(44) Wang, J.; Zhuang, W.; Mukamel, S.; Hochstrasser, R. *J. Phys. Chem. B* **2008**, *112*, 5930−5937.
(45) Smith, A. W.; Lessing, J.; Ganim, Z.; Peng, C. S.; Tokmakoff, A.; Roy, S.; Jansen, T. L. C.; Knoester, J. *J. Phys. Chem. B* **2010**, *114*, 10913−10924.
(46) Bouř, P.; Keiderling, T. A. *J. Phys. Chem. B* **2005**, *109*, 5348−5357.
(47) Hudecová, J.; Hopmann, K. H.; Bouř, P. *J. Phys. Chem. B* **2012**, *116*, 336−342.
(48) Yang, M.; Skinner, J. L. *J. Chem. Phys.* **2011**, *135*, 154114.
(49) Negre, M.; Orozco, M.; Luque, F. J. *Chem. Phys. Lett.* **1992**, *196*, 27−36.
(50) Bonaccorsi, R.; Palla, P.; Tomasi, J. *J. Am. Chem. Soc.* **1984**, *106*, 1945−1950.
(51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(52) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.
(53) Watson, T. M.; Hirst, J. D. *J. Phys. Chem. A* **2003**, *107*, 6843−6849.