Tryptophan Fluorescence Shifts in Proteins from Hybrid Simulations: An Electrostatic Approach

Patrik R. Callis* and Berit K. Burgess

Department of Chemistry and Biochemistry, Montana State University, Bozeman, Bozeman, Montana 59717

Received: July 25, 1997; In Final Form: September 24, 1997[⊗]

We present hybrid quantum chemical−molecular dynamics computations on 15 partially hydrated proteins that support the view that the fluorescence wavelength maximum of a tryptophan (Trp) in a protein is determined almost entirely by the local electric field projection along the long axis of the indole ring, that is to say, an internal Stark effect. The response is 1.3×10^6 V cm^{-1}/nm, with a shift to longer wavelength when the field forces electrons from the pyrrole to the benzene ring. Of 20 Trps simulated, representing a range of exposure to solvent, 15 are predicted within 10 nm of the experimental value. The findings should greatly aid the important task of finding effective methods for computing electrostatic fields and potentials within proteins.

Introduction

The fluorescence of the tryptophan (Trp) amino acid residue in proteins is known to depend on its specific protein/solvent environment, with its spectral maximum (λ_{max}) covering the range from 308 to 355 nm.[1−3] While it is widely agreed that the shortest wavelength case (azurin) involves a Trp buried in a hydrophobic pocket and the long wavelength limit comprises Trps highly exposed to mobile water, the vast intermediate ground remains a matter of speculation. The sensitivity to environment is generally acknowledged to be due to a large dipole change associated with excitation to the ^1L_a state.[4,5] The correlation of λ_{max} with solvent dielectric properties is consistent with this,[5] and recent Stark effect experiments on Trp in rigid polar glasses yield a dipole change of 6 D.[4] Although not generally recognized as such, the large red shift observed in fluid solvents with large dielectric constant is also a Stark effect, with the strong electric field (reaction field) at the solute being created by the partial orientation of the solvent dipoles due to the large solute dipole. Indeed, it is the theme of this Letter to introduce the postulate that the Trp fluorescence λ_{max} in proteins is almost entirely determined by a single parameter: the average electric field projection onto the dipole change direction accompanying excitation to the ^1L_a state. That is, the fluorescence shift is due to what has been termed an internal Stark effect (ISE).[6−8]

The majority of Trp environments (electric field projections) in proteins are probably determined by two sources, a more or less static field produced by the atoms of the protein and the reaction field produced by nearby solvent molecules that are mobile enough to be aligned during the excited state lifetime. It is the contention of this work, however, that the source of the field is irrelevant—only the magnitude and sign of its projection on (as will be shown) the long axis of the indole ring matters.

This almost obvious unification embodied in the ISE, which has been applied to absorption shifts of other chromophores in proteins, e.g., visual pigments,[9] porphyrins,[6] and of a non amino acid probe molecule by an α helix,[7,8] has not previously emerged for Trp primarily because of two obfuscating issues surrounding the nature of Trp fluorescence. One is the near degeneracy of the two lowest singlet states, ^1L_b and ^1L_a, and the other is the view that specifically bonded excited-state complexes (exciplexes) between hydroxyl groups and the Trp π cloud cause significant red shifts, because of the absence of such complexation in the ground state.[10] Recent experimental and theoretical work on model indoles in our laboratory[11] now makes the ISE a reasonable hypothesis, as will be discussed below.

In this paper, we report initial findings from an attempt to predict the tryptophan fluorescence maximum in 15 proteins, using a hybrid semiempirical quantum mechanical−molecular dynamics (MD) method that was found effective for indoles in water.[12−14] The method considers only the variation of electrostatic potential across the indole ring and the extent of electron density transfer during the ^1L_a transition, as determined by the quantum mechanical procedure. The results indicate that the ISE picture is valid and promise to have an impact on the important question of how to effectively predict electrostatic fields and potentials in proteins. Representative works including important references dealing with general issues and techniques involving electrostatics in proteins,[15] and specifically to electron transfer in photosynthesis[16] and to pK_a shifts of amino acids,[7,17] have recently appeared.

Methods

For this initial survey, the method has been kept almost naively simple, with the underlying philosophy being to capture the electrostatic effects. The potentials at each atom of the "3-methylindole" (3MI) portion of the Trp were computed from the Coulomb's law sum over all other atomic point charges assigned to individual atoms of the protein and solvent by the CVFF[18] force field using the expression

$$\phi_i = \sum_k q_k / r_{ik}$$

where ϕ_i is the potential at the ith 3MI atom, q_k is the partial charge on the kth protein atom, and r_{ik} is the distance between atoms i and k. Likewise, the electric field at the ith 3MI atom is given by $\boldsymbol{\xi}_i = \sum_k (q_k / r_{ik}^3)\boldsymbol{r}_{ik}$. A dielectric constant of 1 was assumed throughout, and polarization effects are included only for the tryptophan implicitly through the quantum mechanics. No attempt was made to include the relatively small effects due to differences in dispersion forces upon excitation.

Protein structures were obtained from Brookhaven Protein Data Bank files. Hydrogens were added using Insight (Molecular Simulations Inc.) with the ionization state generally at pH 7, assuming pK_a values in aqueous solution. Waters, which are crucial to obtaining agreement with experiment, were typically added only within 15 Å of the Trp whose fluorescence

[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1997.

S1089-5647(97)02436-X CCC: $14.00

was being computed, using the "soak" option of Insight. The number of waters varied from 40 to 200, depending on the degree of solvent accessibility. Because of the short trajectories and partial solvation, no constraints such as periodic boundary conditions were imposed on the waters; no cutoffs were used.

The largest obstacle in extending our earlier studies on 3MI in polar solvents[12−14] and on crystal fields[19,20] was finding a way to apply quantum mechanics to a Trp residue that is covalently connected to a large protein. The present method involves computing the Coulombic potential and electric field at the atoms constituting the 3MI part of the tryptophyl group and applying these to a reference 3MI ground- or excited-state structure and orientation. The electric fields were transformed according to the transformation required to rotate the principal moments of inertia of the 3MI residue from the protein orientation to the reference orientation. Small differences between relative locations of the indole ring atoms in the protein and the reference geometry are ignored.

Once the potentials and fields are established at a given time in a trajectory, an INDO/S-CIS[21,22] calculation is made, using a modification allowing input of electrostatic potentials and fields, previously applied to crystals of nucleic acid bases with success.[19,20] This procedure uses 196 singly excited configurations, Mataga-Nishamoto electron repulsion screening, and the original CNDO/S overlap factors. The trajectory is interrupted every 10 fs, and the transition energy and charge distribution are computed for the 3MI using the coordinates of the surrounding protein and water atoms. The new indole charges are incorporated into the molecular dynamics, and the trajectory is continued. Because of the partial solvation and the short trajectories, only 100 fs of minimization and 100 fs of initialization was done prior to the first data collection run. Nevertheless, the first trajectory was not significantly different from the subsequent six or seven included in this study.

A key to predicting the correct fluorescence wavelength is incorporating a realistic geometry change upon excitation. Unlike our treatment of 3MI in water,[14] here the geometry change is taken from recent ab initio results that gave accurate detailed fluorescence band shapes at the vibronic level.[23] To track the ^1L_a state with confidence, a transition density projection technique was developed.[22,24]

Results

The inset of Figure 1 defines the molecular *x*, *y*, and *z* axes for the indole ring and shows the relative INDO/S-CIS computed dipoles for the ground and ^1L_a states. The magnitudes are 3 and 9 D when the *x* field is 4×10^7 V/cm, and the arrow indicates the positive end. Although the magnitudes and directions depend somewhat on the field, the difference of about +6 D is reasonably independent of field and is in reasonable agreement with that found experimentally.[4] The sign of the change reflects electron transfer from the pyrrole to the benzene ring, meaning that an electric field pointing in the $+x$ direction will create a red shift of the fluorescence (such as would be created by a positive charge near the benzene ring).

Figure 1 shows the correlation of the average computed *x* (solid squares), *y* (open squares), and *z* (open circles) electric field with experimentally observed fluorescence λ_{max} for 19 tryptophans in 16 proteins and for 3MI in water. Each point represents the average of 200−300 points during MD trajectories sampled between 0.6 and 1.0 ps following simulated excitation. Longer trajectories would be of questionable significance, given the partial solvation, but several trajectories were carried to 10 ps for selected proteins, with the finding that no significant relaxation occurs after 1 ps. It was found that those Trps highly exposed to water showed a time evolution of λ_{max} similar to

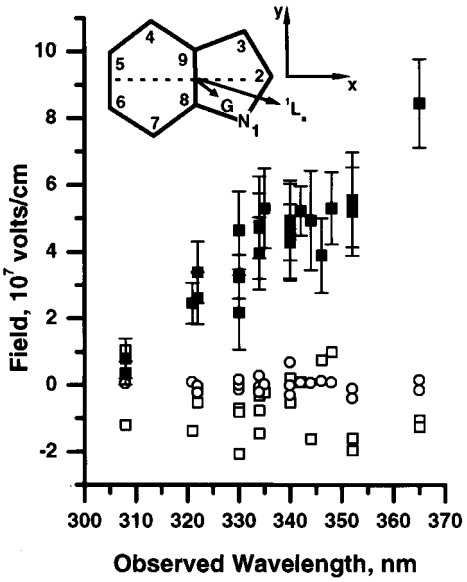

Figure 1. Correlation of experimental tryptophan fluorescence maxima in proteins and 3-methylindole in water with *x* (solid squares), *y* (open squares), and *z* (open circles) electric field components computed from hybrid quantum mechanics−molecular dynamics generated from relaxed, partially solvated crystal structures using the CVFF force field. Each point is the average of 200−300 computations sampled 0.6−1.0 ps following simulated excitation. The vertical bars give the standard deviation. The field for the longest wavelength cases is mainly the solvent reaction field and is therefore proportional to the quantum mechanically determined indole ring dipole in the ^1L_a state, but in the short wavelength cases the field is mainly from the protein.

that for 3MI in water[12−14] and effectively complete after 0.6 ps. Buried Trps showed virtually no time dependence of λ_{max} following excitation. As in our previous study,[14] the excited-state dipole was closely correlated with the transition energy and was seen to evolve in time in proportion to the transition energy. For exposed Trps, the dipole evolved from a value near 9 D at the point of excitation to about 11 D, with fluctuations of ±2 D. For the buried Trps, the dipole stayed near 9 D, with fluctuations of ±0.5 D. Fluctuations of individual atomic charges on π atoms were on the order of ±0.02 electrons during typical trajectories.

The *x* and *y* fields were generated from potential differences, but the average of the *x* electric fields over ring atoms and trajectories gives a very similar plot. A strong correlation with the *x* field is seen, with a slope of 1.3×10^6 V cm^{-1}/nm. In contrast, the plots of the *y* and *z* (out-of-plane) electric field components show virtually no correlation with observed wavelength.

The "error bars" on the *x* field points are the standard deviation of the field, indicating that fluctuations in the field range from 7×10^6 V cm^{-1} for azurin at pH 5 to 3×10^7 V cm^{-1} when the Trp is fully exposed to water. Standard deviations for the *y* and *z* fields (not shown) are similar to those shown for the *x* field. The computed average fields span the range 5×10^6 to 8.5×10^7 V/cm. Such large fluctuations in average field require a large degree of field inhomogeneity. Accordingly, the average standard deviation of the *x* field over the nine ring atoms was found to range from 0.5×10^7 V cm^{-1} for Trp48 of azurin to 2×10^7 V cm^{-1} for exposed Trps. As found in our previous study of 3MI in water,[14] there is a strong correlation between dipole change and the transition energy (and therefore with the internal field).

Figure 2 shows the results of pure applied fields on the INDO/S-CIS computed vertical transition energies (converted to λ_{max}) for the lowest excited singlet state (S_1) of 3MI in vacuum at the model ^1L_a geometry. It is seen that the effect of *y* and *z*

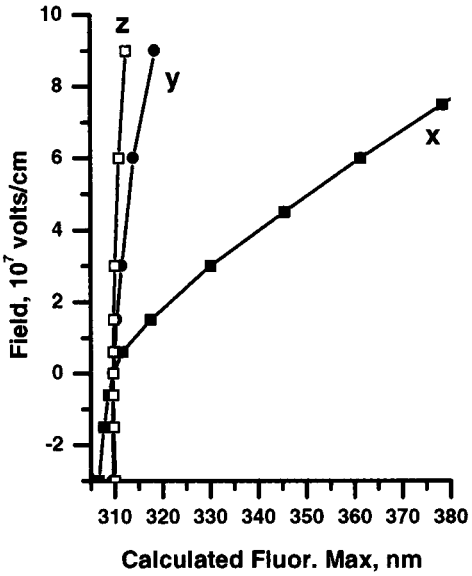

Figure 2. INDO/S-CIS computed vertical fluorescence transition wavelengths as a function of *x*, *y*, and *z* electric fields. The electronic state is ^1L_a for all wavelengths longer than 312 nm and is ^1L_b for shorter wavelengths.

components of the field have little effect compared to the *x* field. With the present geometry and parameters, S_1 is ^1L_a for $\lambda_{max} > 312$ nm. This figure is consistent with an earlier, less extensive INDO/S study on indole in the context of electrostatic perturbations of Trp59 of RNase T1.[25]

Figure 3 shows the fluorescence maxima predicted with the INDO/S-CIS scheme demonstrated in Figures 1 and 2, plotted against the experimentally observed maxima for the same cases shown in Figure 1. Each point is the average of 200−300 INDO/S-CIS computations of the ^1L_a transition energy in the model ^1L_a excited-state geometry with the fields and potentials at each atom included in the Fock matrix. The average includes only configurations from 0.6 to 1.0 ps following simulated excitation. The good correlation follows from the results of Figures 1 and 2.

It was found that the explicit electric field at the individual atoms, which serves only to hybridize atomic orbitals on that atom, was much less important than the implicit electric field asserted by the potential differences between the 3MI atoms. Accordingly, the quantum mechanically computed shift could be well approximated by

$$\Delta E = \sum_i \phi_i \Delta q_i \qquad (1)$$

where Δq_i is the change in charge on the *i*th atom of the 3MI caused by transition to the ^1L_a state.

Discussion

The reasonable correlation seen in Figure 3 gives credence to the ISE hypothesis. The agreement with experiment, however, is better than expected, considering that no scaling has been attempted to create a good fit and that many questionable assumptions have been made. Critical variables that impact the fit are the effective dielectric constant (assumed to be 1), the ability of the INDO/S−CI method to accurately predict the dipole change, the extent of solvation, and the parameters of the force field. There is an obvious systematic deviation from the slope = 1 line wherein the predicted λ_{max} exceeds the experimental value by about 10 nm. However, two of the crucial approximations made in this study are the short trajectories and the incomplete solvation—both of which should reduce the predicted red shift. Indeed, preliminary runs with

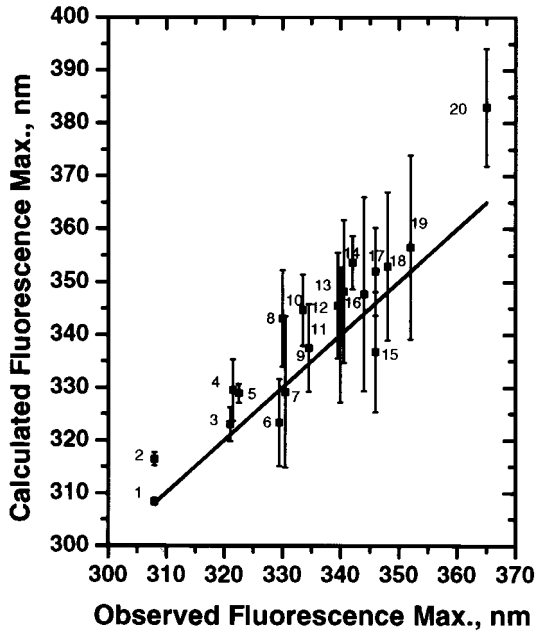

Figure 3. Comparison of computed and observed fluorescence maxima for 19 tryptophans in 16 proteins of known structure and for 3-methylindole in water. The calculated transition energies are from INDO/S-CIS computations on ^1L_a for 3-methylindole with added electrostatic potentials and fields computed from the structures with partial solvation. Each point is the average of 200−300 protein/solvent configurations taken from dynamics trajectories 0.6−1.0 ps following simulated excitation. The error bars are the standard deviations, and they give an indication of the extent of predicted inhomogeneous broadening. The solid straight line is not a regression line but merely a line with slope = 1. The numbers map to proteins as follows: (1) azurin, Trp48 at pH 5; (2) azurin Trp48 at pH 7; (3) myoglobin; (4) ribonuclease T1; (5) subtilisin Carlsberg; (6) T4 lysozyme Trp138; (7) T4 lysozyme Trp158; (8) T4 lysozyme Trp126; (9) staph. nuclease; (10) azurin Trp118; (11) α-cobra toxin; (12) cholera toxin; (13) prophospholipase A2; (14) monellin monomer; (15) che-Y; (16) phospholipase A2; (17) melittin monomer; (18) calmodulin; (19) glucagon; (20) 3-methylindole.

full solvation to 10 Å about the entire protein do increase the predicted red shift by 5−10 nm, but extending the simulation time to 10 ps following excitation in four representative cases did not lead to a significantly increased red shift. At present the two most likely candidates for the overestimation of the shifts are the force field partial charge scheme and the INDO/S results. We are actively investigating the force field dependence, which has been found to affect calculations such as these.[7] Preliminary results using the AMBER force field give equally well-correlated results with somewhat better agreement with experiment. However, the high value found for 3MI in water indicates that the INDO/S dipole is computed to be too large.

In addition, individual deviations from the average fit could arise from discrepancies between the crystal/molecular dynamics structure and the actual solution structure, "anomalous" pK_a values caused by the nonaqueous protein environment, and structural relaxations caused by the dipole change that may occur on a time scale >10 ps.[26] This is not a complete list, nor are these reasons independent of one another.

An earlier study by Levy et al. concluded that higher multipoles than dipole must be important in the solvent-induced fluorescence shift, because increasing the magnitude of the ground state partial charges for 3MI in water until the excited-state dipole magnitude was reached gave a 7-fold greater free energy of stabilization than using the excited-state partial charges. However, the charge distribution difference in that case apparently involved large changes on *hydrogens* (a cumulative increase of over +2), which would understandably have a potent effect on solvation but would be a poor

representation of the charge change pattern for any $\pi\pi^*$ transition. We find that INDO/S-CI computations with large electric field gradients (0.5×10^{16} V cm^{-2}) that would project well onto quadrupole changes upon excitation create shifts an order of magnitude smaller than constant fields of the same magnitude.

Much as Burstein et al.[1] concluded over 20 years ago, we can see in the simulations three broad classes of environment, wherein: (1) The Trp is sufficiently isolated from solvent that on the 1 ps time scale the field is due purely to the protein. These have the shortest wavelength λ_{max} values and include azurin, ribonuclease T1 (RNase T1), subtilisin Carlsberg, and myoglobin. (2) The Trp is highly exposed to water. These show a large red shift due almost entirely to the water reaction field. Examples are glucagon, melittin, and calmodulin. (3) The Trp is partially exposed to solvent and shows an intermediate red shift. For these, the model offers a variety of detailed suggestions as to the reason for the intermediate λ_{max} values. For monellin monomer, staph nuclease, che-Y, and cholera toxin, which have λ_{max} values between 334 and 345 nm, water contributes a small fraction of the red shift. In fact, for staph nuclease and che-Y, the water actually contributes a net blue shift. The models predict that the λ_{max} for these will be little changed upon cooling. On the other hand, phospholipase A_2, prophospholipase A_2, and cobra toxin have wavelengths in the same range but owe their red shifts largely to water relaxation, so that their λ_{max} values are predicted to shift dramatically to the blue upon cooling. Whether these model predictions are valid must be established through experiment.

It is now fairly evident that Trp fluorescence in all proteins (with the probable exception of azurin) is from the ^1L_a state—including the vibrationally structured fluorescence from RNase T1.[11,23] Therefore, the issue of ^1L_b emission affects only the limited wavelength range below 310 nm. However, the fluorescence spectrum of Trp is seen from Figure 2 to be blind to fields more negative than about 5×10^6 V/cm because S_1 is then expected to be ^1L_b, which has a small dipole change accompanying excitation. It is now possible, in principle, to distinguish ^1L_a and ^1L_b fluorescence,[23] and future refinements should allow limits to be placed on the minimum field that supports ^1L_a emission.

The Trps in most cases emit with $\lambda_{max} > 320$ nm, and it is the properties of the ^1L_a state that determine λ_{max}. In this region, Figure 2 implies a fairly linear correlation of λ_{max} with the x field, with a slope of about 1×10^6 V cm^{-1}/nm. Much of the red shift is from relaxation of nearby mobile water, and, accordingly, we find a rough correlation with shift and the number of waters found within 5 Å of the indole ring center. As noted previously,[14] in some cases single waters transiently produce red shifts of 10−15 nm. These are due to unspecific dipole−dipole interactions and offer an alternative to the commonly invoked exciplex language.[10]

Since these red shifts are caused by the solvent reacting to the excited-state dipole, they mask the underlying field due to the protein. The protein field can in principle be extracted, because the solvent relaxation component depends on the proximity and mobility of the solvent. If the fluorescence spectrum is measured prior to the response of mobile polar groups, the λ_{max} should be indicative of only the field existing prior to excitation. This can be accomplished either by ultrashort measurements (<50 fs)[27] at room temperature or with steady state measurements at low temperature. When the Trp is essentially buried, e.g., Trp48 of azurin and Trp59 of RNase T1, the wavelength may well be dictated solely by the protein field. For example, the present model attributes the 15 nm difference in λ_{max} observed for these two Trps mainly to the nitrogen atom of the adjacent proline (Pro60) in RNase T1, which is assigned a large negative charge and lies directly over the pyrrole ring of Trp59.

Conclusions

There is reason to believe that the primary determinant of the Trp fluorescence maximum is the local electric field projection along the long axis of the indole ring, regardless of whether the field is generated from charged and polar groups of the protein or from the reaction field of nearby oriented water. Positive charges at the benzene end or negative charges at the pyrrole end cause maximal red shifts. Viewed in this way, variations of fluorescence wavelength maximum from tryptophan in proteins from 308 to 355 nm, a range of 4300 cm^{-1}, correspond to a variation of electric field projection on the indole long axis of about 4×10^7 V/cm. If subsequent studies verify that the fluorescence wavelength of Trp in proteins is a faithful indicator of local electric field (internal Stark effect), a convenient and powerful guide for assessing predictive methods for the crucially important electrostatic component of protein energetics will have been born.

Acknowledgment. We thank Professor Edward Burstein for helpful discussions. This work was supported by NIH Grant GM31824.

References and Notes

(1) Burstein, E. A.; Bedenkina, N. S.; Ivkova, M. N. *Photochem. Photobiol.* **1973,** *18*, 263.
(2) Eftink, M. R. *Methods Biochem. Anal.* **1991,** *35*, 127.
(3) Longworth, J. W. In *Excited States of Proteins and Nucleic Acids;* Steiner, R. F., Weinryb, I. Eds.; Plenum: New York, 1971; p 319.
(4) Pierce, D. W.; Boxer, S. G. *Biophys. J.* **1995,** *68*, 1583.
(5) Lami, H. Glasser, N. *J. Chem. Phys.* **1986,** *84*, 597.
(6) Varadarajan, R.; Lambright, D. G.; Boxer, S. G. *Biochemistry* **1989**, *28*, 3771.
(7) Sitkoff, D.; Lockhart, D. J.; Sharp, K. A.; Honig, B. *Biophys. J.* **1994,** *67*, 2251.
(8) Lockhart, D. J.; Kim, P. S. *Science* **1992,** *257*, 947.
(9) Honig, B.; Dinur, U.; Nakanishi, K.; Balogh-Nair, V.; Gawinowicz, M. A.; Arnaboldi, M.; Motto, M. G. *J. Am. Chem. Soc.* **1979**, *101*, 7084.
(10) Hershberger, M., V.; Lumry, R.; Verrall, R. *Photochem. Photobiol.* **1981,** *33*, 609.
(11) Callis, P. R. *Methods Enzymol.* **1997,** *278*, 113.
(12) Muino, P. L.; Harris, D.; Berryhill, J.; Hudson, B.; Callis, P. R. *Proc. SPIE-Int. Soc. Opt. Eng.* **1992,** *1640*, 240.
(13) Muino, P. L.; Callis, P. R. *Proc. SPIE-Int. Soc. Opt. Eng.* **1994,** *2137*, 362.
(14) Muino, P. L.; Callis, P. R. *J. Chem. Phys.* **1994,** *100*, 4093.
(15) Honig, B.; Nicholls, A. *Science* **1995,** *268*, 1144.
(16) Ichiye, T. *Structure* **1996,** *4*, 1009.
(17) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem.* **1997**, *101*, 4458.
(18) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111.
(19) Theiste, D.; Callis, P. R.; Woody, R. W. *J. Am. Chem. Soc.* **1991**, *113*, 3260.
(20) Sreerama, N.; Woody, R. W.; Callis, P. R. *J. Phys. Chem.* **1994**, *98*, 10397.
(21) Ridley, J.; Zerner, M. *Theor. Chim. Acta* **1973,** *32*, 111.
(22) Callis, P. R. *J. Chem. Phys.* **1991,** *95*, 4230.
(23) Callis, P. R.; Vivian, J. T.; Slater, L. S. *Chem. Phys. Lett.* **1995,** *244*, 53.
(24) Callis, P. R. *Int. J. Quantum Chem.* **1984,** *18s*, 579.
(25) Ilich, P.; Axelsen, P. H.; Prendergast, F. G. *Biophys. Chem.* **1988,** *29*, 341.
(26) Pierce, D. W.; Boxer, S. G. *J. Phys. Chem.* **1992**, *96*, 5560.
(27) Jimenez, R.; Fleming, G. R.; Kumar, P. V.; Maroncelli, M. *Nature* **1994,** *369*, 471.