Calculating the Fluorescence of 5-Hydroxytryptophan in Proteins

David Robinson,[†] Nicholas A. Besley,[†] Paul O'Shea,[‡] and Jonathan D. Hirst*[,†]

School of Chemistry, University Park, The University of Nottingham, Nottingham, NG7 2RD, United Kingdom, and Cell Biophysics Group, School of Biology, University Park, The University of Nottingham, Nottingham, NG7 2RD, United Kingdom

Received: July 26, 2009; Revised Manuscript Received: August 27, 2009

5-Hydroxytryptophan is a non-natural amino acid that has attracted a lot of recent interest as a fluorescent probe of protein structure, dynamics, and function. We have investigated its fluorescence in various protein environments, using a decoupled quantum mechanics/molecular mechanics approach. Classical, all-atom molecular dynamics simulations of several proteins containing single tryptophans were performed for both the wild-type and the 5-hydroxy derivatives. The excited state of the fluorophore was described using parameters from complete active space self-consistent field calculations. Time-dependent density functional theory calculations on 5-hydroxytryptophan and a significant portion of its explicit immediate surrounding environment, sampled by the simulations, show that the emission energies of 5-hydroxytryptophan shift, depending on the strength of hydrogen bonding and $\pi-\pi$ stacking interactions. This quantitative description of how the fluorescence responds to different protein environments should enhance the insight that fluorescence studies using 5-hydroxytryptophan can provide at a molecular level.

Introduction

A large body of literature now exists on the extensive applications of molecular probes utilized to study the properties of many types of biological molecule. These approaches typically rely on the use of extrinsic labeling with spectroscopic probes, as biomolecules possess few natural moieties with useful optical properties.[1−9] In membranes and cells, for example, it is necessary to utilize spectroscopic labels directed at useful properties.[10] These often include measurement of shifts of the emission or excitation spectra that yield desirable information about some cellular property such as membrane fluidity, ion concentration, etc.

Fluorescent probes, in particular, have found widespread use; changes in their physical environment lead to characteristic shifts in their spectral characteristics. Aminonaphthylethenylpyridinium (ANEP), for example, yields a direct response to membrane-potential changes in the submillisecond regime. The ANEP class of dyes contains many structural variations, making some more suited to specialized tasks. One highly successful 'fluorescent label' is the green fluorescent protein (GFP),[1−5] consisting of 238 amino acids. The fluorophore of GFP arises from a post-translationally modified sequence of serine, tyrosine, and glycine but only exhibits fluorescence within the extended environment within the GFP structure, where a charge transfer mechanism can operate. The active fluorophore is thus rather small but only becomes active within the much larger environment. Theoretical studies on the nature of spectral properties arising from GFP are very challenging.[3] GFP has been well characterized, with an extended mutation (EGFP) finding much use in biological imaging applications. The rather large size of the GFP structure, however, may interfere with normal biological function and so the search for smaller, less invasive probes is an active area of research.

Of the intrinsic biological fluorophores that are available, tryptophan can offer a powerful tool to identify subtle structural changes of the protein during its biological action.[11−16] In some cases a protein may contain a single tryptophan. This information can be readily exploited to report on the environment.[17] The folding and unfolding of proteins, substrate binding, and external quencher accessibility can all be monitored using tryptophan probes. Gruebele et al.,[18] for example, studied the folding of a λ-repressor by measuring the decay of fluorescence intensity from tryptophan as the protein refolded, having been exposed to a very rapid pressure change.[18] Substrate binding can be studied if the tryptophan is directly involved, since the shifts in fluorescence wavelengths are large.[19]

These foregoing examples illustrate the power of using tryptophan in proteins to illuminate an important biomolecular process. Most proteins, however, contain many tryptophans, and deconvolving their specific contribution to complex spectra can sometimes be intractable with present methods.[9] There are some schemes that deconvolve such spectra, but this is not straightforward. Additionally, tryptophan, along with the other aromatic residues, participates in extensive $\pi-\pi$ interactions, which are important in proteins.[20−23] The folding and thermal stability of proteins is known to be affected by these interactions. Protein−ligand recognition relies on noncovalent aromatic interactions, e.g., the recognition of mRNA by proteins. These types of interactions may be responsible for shifts in both the absorption and emission energy.[23] It is important to understand the effect of such interactions on emission energies, since these interactions are common within protein environments.[20]

Theoretical models have been developed to determine shifts in absorption and emission spectra. Callis et al. have proposed a hybrid quantum mechanics/molecular mechanics (QM/MM) model based on semiempirical configuration−interaction calculations for the tryptophan residue,[14] using ground and excited-state geometries optimized with high-level multiconfigurational ab initio excited-state methods. The results for the proteins studied agreed well with experiment, thus giving an insight into

* Corresponding author. E-mail: Jonathan.Hirst@nottingham.ac.uk.
[†] School of Chemistry.
[‡] Cell Biophysics Group, School of Biology.

10.1021/jp9071108 CCC: $40.75
Published on Web 10/01/2009

TABLE 1: Calculated and Experimental Emission Energies (in eV) for the Proteins Studied, With Standard Deviation in Parentheses

			tryptophan		5-hydroxytryptophan
protein	PDB code	no. waters[a]	calcd	expt[b]	calcd
cobra toxin	1CTX	5642	3.69 (0.13)	3.65	3.68 (0.13)
glucagon	1GCN	10910	3.68 (0.06)	3.53	3.57 (0.08)
subtilisin Carlsberg	1SBC	7096	3.90 (0.10)	3.86	3.73 (0.07)
phospholipase A2	2BPP	6026	3.67 (0.08)	3.65	3.54 (0.04)
metmyoglobin	1MYT	6580	3.80 (0.07)	3.87	3.80 (0.13)
ribonuclease T1	9RNT	3420	3.70 (0.09)	3.86	3.66 (0.16)

[a] Number of explicit water molecules included in the MD simulations. [b] Experimental results from ref 14.

the shifts caused by the local environment, i.e., the role of surrounding residues and the effect of the solvent. Rogers et al. incorporated a larger number of residues into the environment calculated explicitly using time-dependent density functional theory (TDDFT).[11] The absorption energies did not converge until all the surrounding residues within 4.0 Å of the ring-fusing bond of indole were explicitly included within the TDDFT calculations. Thus, the incorporation of explicit molecular interactions appears to be necessary for qualitative and quantitative accuracy in such analysis. This, of course, also has broader implications for the development of more stringent molecular models of the complex environments represented by proteins.

With this latter comment in mind, a solution to the problem of spectroscopic analysis of multiple tryptophans in proteins may reside in replacing a specific tryptophan in the protein with the spectrally distinct non-natural amino acid, 5-hydroxytryptophan.[6−9] The first absorption band of 5-hydroxytryptophan is significantly longer than that of tryptophan, allowing easy discrimination of the signal. There have been some qualitative experimental studies using 5-hydroxytryptophan, such as with bacteriophage λ-cI repressor, in which 5-hydroxytryptophan replaced the three wild-type tryptophans.[8] The absorption and emission spectra of the 5-hydroxy derivative had an absorption shoulder at a wavelength 30 nm longer than found for native tryptophan within the protein environment. Another study[9] looked at phosphoglycerate kinase (PGK). The protein was labeled by inducing expression in the presence of 5-hydroxytryptophan at a site where conventional fluorescent labels could not be used, because of their size. The insights that such studies could provide would benefit from a more quantitative understanding of both the electronic structure of 5-hydroxytryptophan and the effect of the surrounding environment.

Many studies have focused on the electronic structure and spectroscopy of indole or 3-methylindole.[15,16,24−30] In a previous paper, we explored the electronic structure of 5-hydroxyindole, the fluorescent moiety of 5-hydroxytryptophan.[31] Geometry optimizations in the gas phase were performed for the ground state and the first two singlet excited states, allowing calculation of the emission energies as well as the band origin of absorption. These calculations were extended to include into the quantum mechanical (QM) calculations implicit solvation and finally explicit solvent, where snapshots from classical molecular dynamics (MD) simulations were used. Inclusion of explicit solvation parameters was necessary to describe the effects of hydrogen bonding, which were essential to confirm that the ^1L_a state is responsible for emission.

In recent years, TDDFT has been frequently utilized for the calculation of electronic spectroscopy, because of its speed and simplicity in comparison with more traditional techniques, such as EOM-CCSD[32] and multideterminant approaches.[33] However, a problem arises in TDDFT from the approximate nature of the exchange-correlation functional, where charge-transfer excitations are not accounted for correctly and have a spuriously low energy of transition.[34] Two approaches have been suggested to remedy this problem. The first is a gradual switchover of the exchange integrals from that calculated by the density functional to 100% Hartree−Fock exchange.[35,36] This is applicable to many existing functionals and gives good results, where standard functionals fail dramatically, such as for the π-stacked adenine dimer.[37] The second approach is to have a corrected functional, whereby the correction is part of the definition of that functional.[38−40] This approach has given the correct ordering of states in indole systems where previous attempts with TDDFT have failed.

In the present paper, we report calculations on the emission spectroscopy of 5-hydroxytryptophan within six proteins that have both experimental and theoretical results readily available. For comparison, we also look at the same proteins with unmodified tryptophan. We use a decoupled QM/MM scheme, where snapshots from a MD simulation are used to generate emission spectra. Large numbers of atoms are treated explicitly in a truncated subspace TDDFT excited-state calculation, giving detailed information on the role of the molecular environment.

Computational Methods

For the calculation of fluorescence, the excited-state geometries of indole and 5-hydroxyindole were optimized using the complete active space self-consistent field (CASSCF) level of theory,[33] employing an atomic natural orbital (ANO) basis set[41] contracted to 4s3p2d for carbon, nitrogen, and oxygen, and 2s1p for hydrogen. Both indole and 5-hydroxyindole had CASSCF active spaces comprising the π-system, with the nitrogen lone pair (and oxygen lone pair for 5-hydroxyindole) included. This led to active spaces of (10,9) and (12,10) for indole and 5-hydroxyindole, respectively. These calculations were performed with Molcas 7.0.[42]

MD simulations were performed, using the CHARMM program[43] and the CHARMM22 all-hydrogen parameters,[44] on six proteins (Table 1), surrounded by water molecules described by the TIP3P potential,[45] where the geometries of the indole or 5-hydroxyindole side chains (of tryptophan and 5-hydroxytryptophan, respectively) were constrained to the excited-state geometries, providing sampling for subsequent QM calculations of fluorescence. The excited-state charges calculated from CASSCF calculations using the LoProp[46] analysis were used (see Supporting Information). The LoProp partial atomic charges, in the ground state, agree closely with those used in the CHARMM topology files. The LoProp scheme is capable of providing partial atomic charges that can accurately model the system regardless of the electronic

TABLE 2: Commonly Occurring Clusters Found from MD Simulations, Given in Order of Relative Population for Each Protein[a]

	tryptophan		5-hydroxytryptophan	
protein	rel. pop.	residues in cluster	rel. pop	residues in cluster
cobra toxin	2	{^{25}Trp,^{26}Cys}, {^{25}Trp,^{26}Cys,^{27}Asp}	2	{^{25}FHA,^{26}Cys}, {^{25}FHA,^{26}Cys,^{52}Val}
glucagon	1	{^{25}Trp}	1	{^{25}FHA}
subtilisin Carlsberg	1	{^{48}Ala,^{50}Phe,^{112}Trp,^{116}Asn}	1	{^{48}Ala,^{50}Phe, ^{112}FHA,^{116}Asn}
phospholipase A2	1	{^{3}Trp}	1	{^{3}FHA}
metmyoglobin	2	{^{10}Trp,^{13}Val,^{68}Gly,^{71}Leu}, {^{10}Trp,^{13}Val,^{17}Tyr,^{68}Gly,^{71}Leu}	5	{^{10}FHA,^{13}Val,^{68}Gly,^{71}Leu}, {^{10}FHA,^{13}Val,^{17}Tyr,^{68}Gly,^{71}Leu}, {^{10}FHA,^{13}Val,^{68}Gly,^{71}Leu,^{106}Leu}, {^{10}FHA,^{13}Val,^{17}Tyr,^{68}Gly,^{71}Leu,^{106}Leu}, {^{10}FHA,^{13}Val,^{14}Glu,^{68}Gly,^{71}Leu}
ribonuclease T1	5	{^{19}Ala,^{22}Ala,^{23}Gly,^{59}Trp,^{60}Pro,^{67}Val,^{68}Tyr}, {^{19}Ala,^{22}Ala,^{23}Gly,^{38}Tyr,^{39}Pro,^{59}Trp, ^{60}Pro,^{67}Val,^{68}Tyr}, {^{19}Ala,^{23}Gly,^{38}Tyr,^{59}Trp,^{60}Pro, ^{67}Val,^{68}Tyr}, {^{19}Ala,^{22}Ala,^{23}Gly,^{38}Tyr,^{59}Trp,^{60}Pro, ^{67}Val,^{68}Tyr}, {^{23}Gly,^{38}Tyr,^{59}Trp,^{60}Pro,^{67}Val,^{68}Tyr}	4	{^{19}Ala,^{23}Gly,^{59}FHA,^{60}Pro,^{68}Tyr}, {^{19}Ala,^{22}Ala,^{23}Gly,^{59}FHA,^{60}Pro,^{68}Tyr}, {^{19}Ala,^{22}Ala,^{23}Gly,^{59}FHA,^{60}Pro,^{67}Val, ^{68}Tyr}, {^{19}Ala,^{23}Gly,^{59}FHA,^{60}Pro,^{67}Val, ^{68}Tyr}

[a] For clarity, the 5-hydroxytryptophan residue constrained at its emitting state (^1L_a) is referred to as FHA.

state. The simulations were carried out at constant volume and temperature (300 K). The histidine residue was rendered electrically neutral for all simulations, with default values used for all other titratable groups. The SHAKE constraint[47] was applied to all bonds to hydrogen (excluding those on the 5-hydroxyindole ring). A time step of 1 fs was used. Metals, cofactors, and waters of crystallization were included in the simulations, where relevant. Equilibration lasted for 140 ps, followed by production dynamics for 1 ns. Cubic periodic boundary conditions were imposed. A 10 Å cutoff for nonbonded interactions was used for all six proteins. Long range electrostatics were accounted for by the PME method, using default values for the parameters. Tryptophan fluorescence lifetimes in proteins can range from a few hundred picoseconds to ~8 ns.[17] It is not our aim to predict fluorescence lifetimes; rather, we are trying to sample the conformational space in order to predict fluorescence within tryptophan. However, to model protein dynamics occurring on a time scale longer than 1 ns and their influence on fluorescence, one would need longer MD simulation times than performed here.

In order to calculate the emission spectrum, snapshots were taken from the MD simulations and any residues within 4.5 Å of the ring-fusing carbon atoms of the indole side chain were included in a TDDFT excited-state calculation, which employed the Tamm−Dancoff approximation.[48] For each cluster, 20 snapshots were used. The distance cutoff was based on the convergence in exploratory calculations of the emission energy with respect to distance, as well as the conclusions of Rogers et al. on similar systems.[11] Performing a TDDFT calculation on a system of this size would be computationally demanding and so a scheme is utilized involving a truncated set of single excitations associated with the atoms of the fluorophore.[49] In this scheme, occupied orbitals are chosen based upon their Mulliken populations. If $\{\lambda\}$ is the subset of basis functions centered on the fluorophore atoms, a parameter κ_i^{occ} is defined such that

$$\kappa_i^{\mathrm{occ}} = \sum_{\lambda} M_{\lambda i}$$

where $M_{\lambda i}$ is the contribution to the Mulliken population of orbital i from basis function λ. κ_i^{occ} is, therefore, a measure of the atoms on which orbital i is localized. Similarly, a parameter κ_a^{vir} can be defined for virtual orbitals, based upon molecular orbital coefficients c,

$$\kappa_a^{\mathrm{vir}} = \sum_{\lambda} c_{\lambda a}^2$$

For this study, an occupied orbital i was included if $\kappa_i^{\mathrm{occ}} \geq 0.4$ au and a virtual orbital a was included if $\kappa_a^{\mathrm{vir}} \geq 0.5$ au. In previous studies of indole systems, TDDFT has not predicted the ordering of the ^1L states correctly.[11] The ^1L_a state is due to a transfer of charge density across the two rings,[14] which TDDFT fails to describe properly. Therefore, in the current study, the asymptotically corrected PBEOP functional was employed,[38,39] based on our work on 5-hydroxyindole.[31] The 6-311G(d) basis set was used on all atoms. If a group was hydrogen-bonded to the fluorophore, the orbitals of this group were also included, with the 6-311G(d,p) basis[50] on the hydrogen atoms in question. These calculations were performed with Q-Chem.[51]

From the MD simulation, $\pi-\pi$ stacking was thought to contribute to the stability of the excited state, as discussed in more detail below. For these calculations, the interaction energies were calculated for 5-hydroxytryptophan and tyrosine in the gas phase, at the geometry obtained from the classical MD simulations. Second-order perturbation theory (MP2) with the 6-311++G(d,p) basis set[50,52] was employed, since coupled-cluster calculations for this system would have been beyond the available computational resources. Interaction energies calculated using MP2 and a medium-sized basis set are likely to be overestimated.[53] They do, however, provide a useful insight into the relative strengths of interaction energies in $\pi-\pi$ stacking interactions. Basis-set superposition error (BSSE) was accounted for using the counterpoise correction.[54]

Results and Discussion

First we consider the MD simulations for the native proteins. The RMSDs of the protein backbones with respect to the experimental structures (with variance in Å^2 in parentheses) are 3.00 Å (0.44), 2.65 Å (0.33), 1.27 Å (0.14), 1.01 Å (0.06), 1.88 Å (0.21), and 1.13 Å (0.17) for cobra toxin, glucagon, metmyoglobin, subtilisin Carlsberg, phospholipase A2, and ribonuclease T1, respectively. For the same proteins modified by exchanging tryptophan with 5-hydroxytryptophan, the values

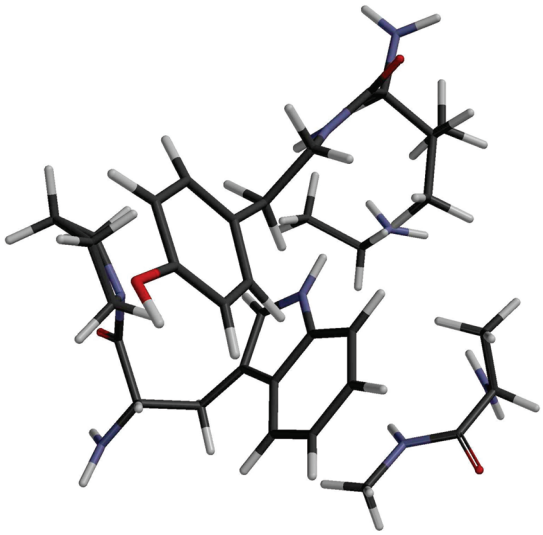

Figure 1. Molecular composition of the shared cluster of ribonuclease T1 (shown with natural tryptophan).

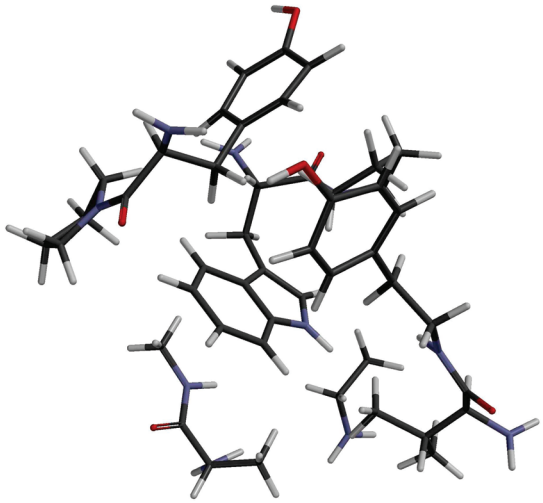

Figure 2. The largest cluster treated explicitly with TDDFT, ribonuclease T1 (cluster 2). It contains 137 atoms (63 heavy atoms), giving 1356 basis functions.

are 2.89 Å (0.21), 2.69 Å (0.36), 1.36 Å (0.12), 1.03 Å (0.06), 1.85 Å (0.19), and 1.19 Å (0.15). In all cases, the energies converged with respect to time, showing that the simulations are stable. Tryptophan has three conformers, trans ($120 \leq \chi_1 \leq -120$), gauche$^-$ ($0 \leq \chi_1 \leq 120$), and gauche$^+$ ($-120 \leq \chi_1 \leq 0$), where χ_1 is the dihedral angle specified by the atoms $N-C_\alpha-C_\beta-C_\gamma$. For all proteins, the trans conformer was found almost exclusively at 300 K; only a temperature greater than 300 K yielded a population of all three conformers. Only trajectories obtained at 300 K have been analyzed further.

Next we compare the theoretical and experimental[55] spectra of the wild-type proteins with tryptophan. The average emission energies, computed over all snapshots from the MD simulations, and the corresponding experimental values are given in Table 1. The largest difference (0.16 eV) occurs for ribonuclease T1, and the average difference for this set of proteins is 0.08 eV, which represents quantitative accuracy. This compares well with the work of Callis et al., who reported an average difference of 0.07 eV and a maximum difference of 0.18 eV for the same proteins using a semiempirical approach with optimized parameters and the surrounding environment described by a series of point charges.[14] TDDFT can give accuracies to within a couple of tenths of an electron volt. In similar systems, such as tryptophan and 5-hydroxytryptophan, the relative difference between excitations of the two molecules should show smaller errors, and so a difference of ~0.05 eV between the two different molecules is a real effect and not an artifact of TDDFT errors. The predicted emission energies for the proteins modified with 5-hydroxytryptophan are also given in Table 1. In most cases, the emission energy occurs below that of the wild-type protein with tryptophan. This allows for the discrimination of the signal from 5-hydroxytryptophan-labeled proteins from those of unmodified proteins.

Table 2 lists the most commonly occurring clusters (collection of residues within 4.5 Å of the center of the ring-fusing bond of the indole side chain) for each protein taken from the MD simulations. The average emission energy for each cluster reveals the solvatochromic effect of specific residues on the emission spectrum. Long cobra toxin is a 71 amino acid venom from the cobra snake. It comprises a β-sheet surrounded by interconnecting loop regions. For unmodified cobra toxin, the shift in the emission spectrum from the first to the second cluster is +0.19 eV, suggesting that the close proximity of the aspartate residue, ^{27}Asp, to the tryptophan causes the emission peak to blue shift. The modified protein (with 5-hydroxytryptophan) also has two clusters, although ^{27}Asp does not feature in either. The difference between the emission energies of the two clusters is +0.06 eV, representing a modest hypsochromic shift when the hydrophobic valine residue is close to the 5-hydroxytryptophan.

Metmyoglobin acts as an oxygen storage protein, providing a diffusion gradient to absorb oxygen from the blood into the cytoplasm. It consists of 146 amino acids, forming 8 helices. The region around the tryptophan in metmyoglobin, which is located on helix A, is largely hydrophobic in all clusters observed from the MD simulations for both the modified and unmodified proteins. The most populated cluster for both proteins contains ^{13}Val, ^{68}Gly, and ^{71}Leu close to the tryptophan (5-hydroxytryptophan). The average emission energy is 3.77 and 3.76 eV, for the modified and unmodified metmyoglobin, respectively. The overall average emission energies are similar as well. In the second cluster, there is an additional tyrosine residue, with the hydroxyl group almost always hydrogen-bonded to the (N)H group of the indole ring. The average emission energy for this cluster is slightly higher for both the modified and unmodified proteins than for the first cluster. However, the effect of the tyrosine is more subtle. At hydrogen bond distances greater than 2.0 Å, the hydrophobic effect dominates and the emission energy is shifted by ~0.1 eV (relative to the first cluster), but, with hydrogen bond distances shorter than 2.0 Å, stabilization of the excited state occurs and the emission energy is red-shifted, by up to ~0.2 eV. The third cluster of the modified metmyoglobin contains an additional leucine residue (with respect to the first cluster), giving an average emission energy of 3.84 eV. Again, the addition of a hydrophobic residue gives a mild red shift in the emission energy. The fourth cluster incorporates both the tyrosine residue (observed in cluster 2) and the leucine residue (from cluster 3). The average emission energy is computed at 3.86 eV and thus follows the trend observed upon adding hydrophobic residues (clusters 2 and 3). The tyrosine again displays a more subtle balance, dependent upon the hydrogen bond distance, although the shifts are not as large as seen in the absence of the additional leucine residue. The final cluster corresponds to the addition of a glutamate residue to cluster 1. The carboxylate group lies reasonably close to the five-membered ring of the indole ring,

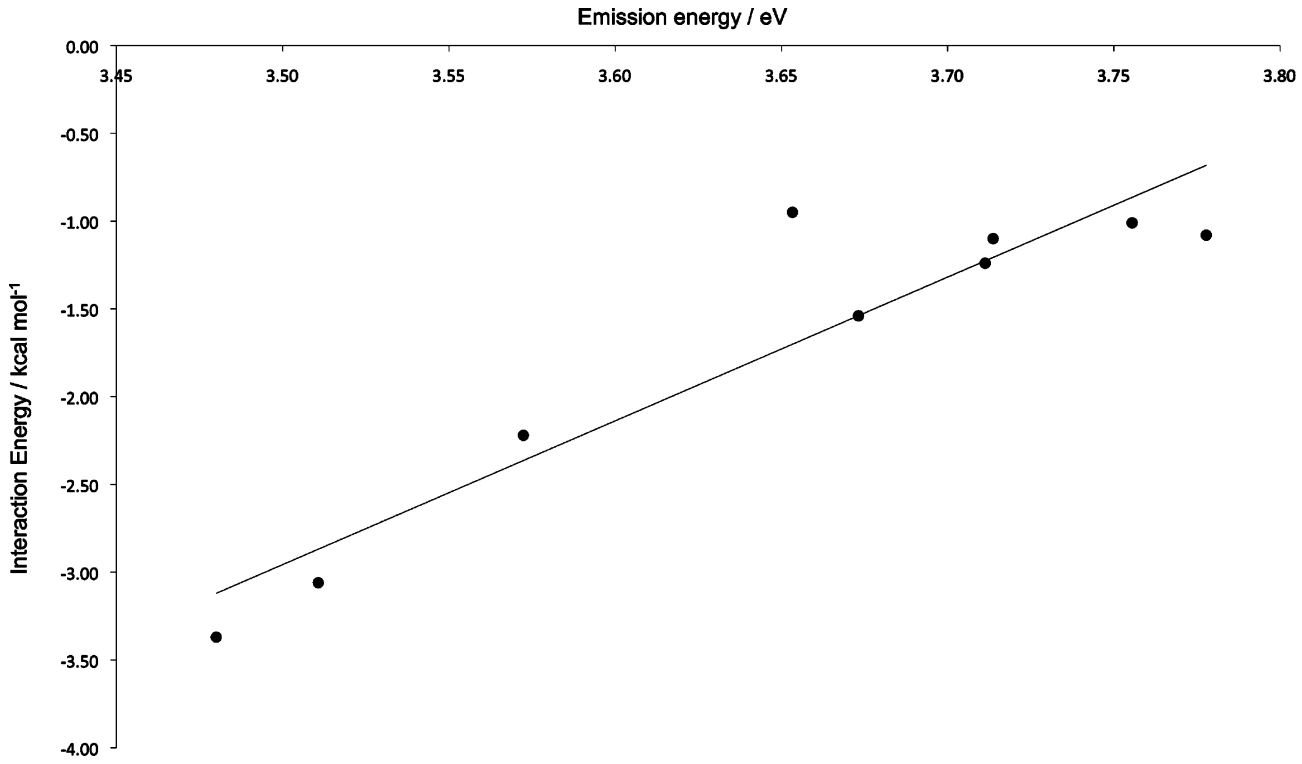

Figure 3. Relationship between the interaction energy and the emission energy for modified ribonuclease T1. Interaction energies were calculated at the MP2/6-311++G(d,p) level of theory; emission energies were calculated with TDDFT (PBEOP/6-311G(d)). $r^2 = 0.87$ (omission of the point at 3.65 eV causes r^2 to rise to 0.96).

although hydrogen bonding to the NH group does not occur. The average emission energy for this cluster shows a hypsochromic shift of 0.02 eV.

Ribonuclease T1 from *Aspergillus oryzae* is a 104 amino acid enzyme with a $\alpha\beta$ structure, responsible for the phosphodiester bond hydrolysis of single-stranded RNA. In this protein, the tryptophan is buried, with no exposure to the solvent. Ribonuclease T1 shows a large number of clusters for both tryptophan and 5-hydroxytryptophan (see Table 2), of which only one is common to both (Figure 1). This shared cluster (cluster 1 for tryptophan; cluster 3 for 5-hydroxytryptophan) gives the only direct comparison in this particular hydrophobic pocket. Cluster 2, comprising 137 atoms, is the largest system we considered in the TDDFT (Figure 2). For both the 5-hydroxytryptophan and the natural tryptophan, the average emission energy is 3.60 eV. There are small shifts dependent upon cluster composition, except when ^{38}Tyr appears in the tryptophan clusters and then shifts of ~0.1 eV are observed (with respect to the overall average for the protein). This is a substantial shift and can only be observed if the explicit environment is included within the calculation.

The individual shifts within each cluster are sensitive to the proximity of ^{68}Tyr to the fluorophore. The phenol side group of tyrosine is arranged in a T-shaped stacking interaction with the six-membered ring of the indole side chain. Figure 3 presents a quantitative relationship between the interaction energy and the excitation energy. This relationship suggests the $\pi-\pi$ interaction stabilizes the excited state rather than the ground state. By comparing the Hartree−Fock orbital energies, we see that the π^* orbitals are lowered in energy upon forming the $\pi-\pi$ interaction with tyrosine. Figure 4 shows a schematic representation of this stabilization. Since standard density functionals do not include dispersion terms, the TDDFT calculations were recalculated using the DFT-D correction of Grimme,[56] yielding emission energies with an overall difference from the standard functional of less than 0.05 eV. These preliminary calculations indicate that dispersion is not the dominant factor for this type of complex, similar to the result found for benzene and indole.[57] In the case of 5-hydroxytryptophan, the shift due to this $\pi-\pi$ interaction energy can be as large as 0.25 eV from the calculated average for the protein. In contrast, the shifts due to $\pi-\pi$ interactions for tryptophan are ~0.1 eV.

The three remaining proteins which we studied each have only one commonly occurring cluster for each of tryptophan and 5-hydroxytryptophan, with the surrounding environment being equivalent for each case. Subtilisin Carlsberg, a 274 amino acid endopeptidase with $\alpha\beta$ structure, is the only one of these proteins that contains other amino acids in the cluster, as well as explicit water molecules. The tryptophan is partially exposed to solvent. The emission energy for the 5-hydroxytryptophan mutant is 0.17 eV lower than the natural protein. Phospholipase A2 is a 123 amino acid α-domain enzyme. For phospholipase A2, the difference is 0.13 eV. Glucagon, a 29 amino acid α-helical pancreatic hormone, is responsible for glycogenolysis and gluconeogenic pathways, helping to increase blood-glucose levels. The tryptophan residue is solvent exposed. The difference for glucagon was calculated to be 0.11 eV. Since in these proteins the tryptophan is close to the interface with water, it is in a relatively polar environment. It is understood that the $^1L_a \leftarrow 1\,^1A'$ transition involves a shift of electron density from the pyrrole ring to the benzene ring of the indole side chain.[14] The presence of the hydroxyl group of 5-hydroxytryptophan would be expected to stabilize this state relative to tryptophan, which is duly reflected in the results for these proteins.

The solvent-accessible surface areas (SASA) of the fluorophores were calculated for each snapshot from the MD

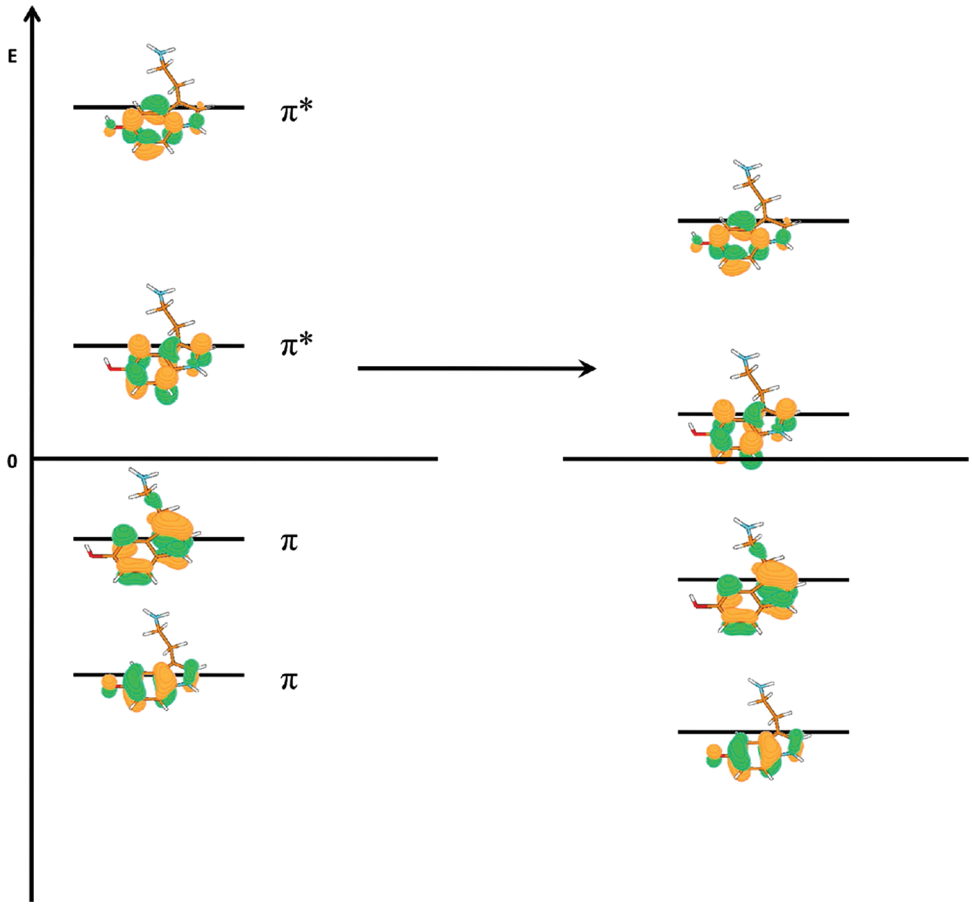

Figure 4. Schematic of the lowering of orbital energies for 5-hydroxytryptophan upon forming a $\pi-\pi$ interaction with tyrosine.

TABLE 3: Solvent-Accessible Surface Area, Calculated Using the VMD Visualization Software[58]

	solvent accessible surface area/Å^2	
protein	tryptophan	5-hydroxytryptophan
cobra toxin	162	114
glucagon	215	220
subtilisin Carlsberg	117	137
phospholipase A2	204	228
metmyoglobin	4	5
ribonuclease T1	<1	<1

simulations, giving average values for each protein (Table 3). In general, the larger the SASA, the lower the emission energy for that protein. Ribonuclease T1 does not quite fit these results; the tryptophan is shielded from the solvent and thus solvation plays no part in the emission spectrum. Where tryptophan residues were hydrogen bonded to water, the emission energy was red-shifted. This lowered the emission by ~0.2 eV for the strongest hydrogen bonds (Figure 5). Figure 5 also shows an apparent linear relationship between hydrogen bonding and emission energy. This appears as a result of the relative stabilization of the virtual orbitals in analogy to the $\pi-\pi$ stabilization discussed above.

Finally, we considered the effect of the extended protein environment, by comparing the emission energies with those calculated with only the model system, i.e., without point charges (Table 4). Also in Table 4 are estimates of the electrostatic potential due to the explicit residues (excluding indole) calculated from the Mulliken charges, and the electrostatic potential due to the extended environment modeled by point charges. Where the electrostatic potential of the extended environment is large, a shift is observed in the emission energy compared to the gas phase model. The average shift for the tryptophan with the extended environment is 0.10 eV, with four of the six proteins showing a blue shift. The 5-hydroxytryptophan emission energies are less affected, with an average shift of 0.06 eV, but all results are red-shifted on inclusion of the point charges. The extended environment also amplifies the more subtle effects of the $\pi-\pi$ stacking interaction in ribonuclease T1; without this environment, shifts due to this interaction are ~0.10 eV from the average, whereas with it, the same emission energies shift by up to 0.25 eV.

The replacement of natural tryptophan with 5-hydroxytryptophan provides spectroscopists with a useful intrinsic probe, introducing only a minor perturbation in the structure local to the tryptophan site. In like-for-like environments, this study shows that the 5-hydroxytryptophan variant has an emission signal about 0.2 eV below that of natural tryptophan, allowing for easy discrimination from natural tryptophan. In polar environments, sensitivity to hydrogen-bonding results in shifts of nearly 0.3 eV. More subtle environmental effects, such as $\pi-\pi$ interactions, can be detected by a 5-hydroxytryptophan, especially in nonpolar environments.

Conclusions

The general scheme we introduced for the calculation of fluorescence has been extended to compare 5-hydroxytryptophan to natural tryptophan in several protein environments. The level of molecular detail that can be assigned to a particular shift has been revealed for both the natural and mutant proteins. In

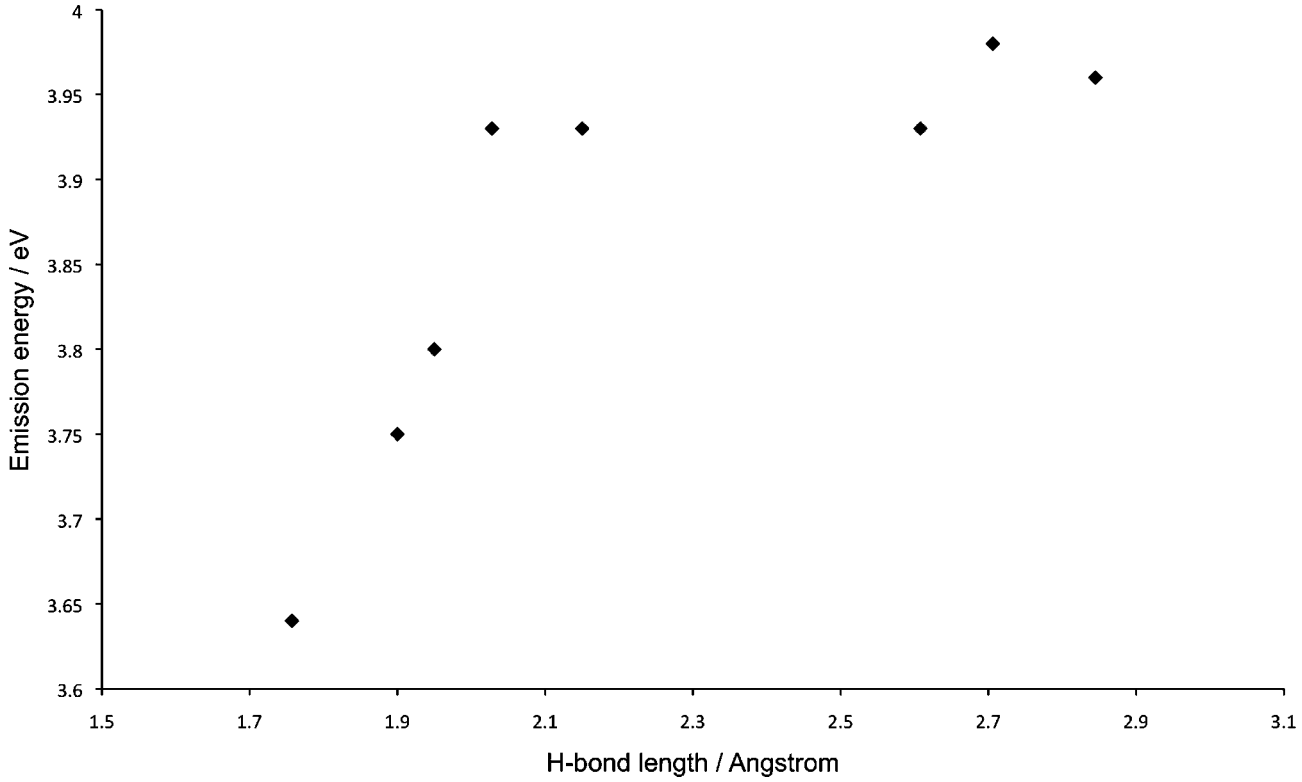

Figure 5. Relationship between hydrogen bond length and emission energy for 5-hydroxytryptophan in metmyoglobin.

TABLE 4: Effect of the Protein Environment Modeled by Point Charges on the Calculated Spectral Shifts (in eV) from the Gas-Phase Model

	tryptophan				5-hydroxytryptophan			
	with environ[a]		model[b]		with environ[a]		model[b]	
protein	E_{emis}	V	E_{emis}	V	E_{emis}	V	E_{emis}	V
cobra toxin	3.69	0.074	3.62	0.104	3.68	0.062	3.69	0.125
glucagon	3.68	−0.006	3.74	0.098	3.57	−0.012	3.62	0.111
subtilisin Carlsberg	3.90	0.016	3.76	0.090	3.73	0.012	3.77	0.093
phospholipase A2	3.67	−0.038	3.61	0.126	3.54	−0.028	3.55	0.109
metmyoglobin	3.80	0.052	3.61	0.092	3.80	0.049	3.87	0.062
ribonuclease T1	3.70	−0.115	3.77	0.093	3.66	−0.113	3.84	0.114

[a] V is the electrostatic potential (in au) due to the point charges. [b] V is the electrostatic potential (in au) due to the explicitly calculated residues.

general, the 5-hydroxytryptophan emission signal is lower in energy than for the tryptophan signal for any given protein. The notable exception for this is where the environment immediately around the fluorophore site is very hydrophobic, in which case the emission energies become very close in energy. In ribonuclease T1, tyrosine is found in a T-shaped $\pi-\pi$ stacking arrangement with the indole side chain. The results presented here show that 5-hydroxytryptophan can be used as a sensitive probe to these $\pi-\pi$ interactions, opening a new line of possibilities with this probe. In calculating such effects, it appears that the extended protein environment must be included for quantitative accuracy.

Acknowledgment. This work was supported through the Engineering and Physical Sciences Research Council through the award of a grant (EP/F006780/1). We thank the University of Nottingham for time on the high-performance computing service.

Supporting Information Available: Excited-state CHARMM parameters for tryptophan and 5-hydroxytryptophan. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Heim, R.; Prasher, D. C.; Tsien, R. Y. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12501–12504.

(2) (a) Paulick, M. G.; Forstner, M. B.; Groves, J. T.; Bertozzi, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20332–20337. (b) Limon, A.; Reyes-Ruiz, J. M.; Eusebi, F.; Miledi, R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15526–15530.

(3) Timerghazin, Q. K.; Carlson, H. J.; Liang, C.; Campbell, R. E.; Brown, A. *J. Phys. Chem. B* **2008**, *112*, 2533–2541.

(4) Leake, M. C.; Chandler, J. H.; Wadhams, G. H.; Bai, F.; Berry, R. M.; Armitage, J. P. *Nature* **2006**, *443*, 355–358.

(5) See, e.g.: Prescott, M.; Battad, J.; Wilmann, P.; Rossjohn, J.; Devenish, R. *Biotechnol. Annu. Rev.* **2006**, *12*, 31–66, and references therein.

(6) Zhang, Z.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8882–8887.

(7) Hogue, C. W. V.; Rashquinha, I.; Szabo, A. G.; MacManus, J. P. *FEBS Lett.* **1992**, *310*, 269–272.

(8) Ross, J. B. A.; Senear, D. F.; Waxman, E.; Kombo, B. B.; Rusinova, E.; Huang, Y. T.; Laws, W. R.; Hasselbacher, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 12023–12027.

(9) Botchway, S. W.; Barba, I.; Jordan, R.; Harmston, R.; Haggie, P. M.; William, S. P.; Fulton, A. M.; Parker, A. W.; Brindle, K. M. *Biochem. J.* **2005**, *390*, 787–790.
(10) Loew, L. M. *Pure Appl. Chem.* **1996**, *68*, 1405–1409.
(11) Rogers, D. M.; Besley, N. A.; O'Shea, P.; Hirst, J. D. *J. Phys. Chem. B* **2005**, *109*, 23061–23069.
(12) Callis, P. R.; Burgess, B. K. *J. Phys. Chem. B* **1997**, *101*, 9429–9432.
(13) Callis, P. R.; Parinello, M. *J. Phys. Chem. B* **2004**, *108*, 4248–4259.
(14) Vivian, T.; Callis, P. R. *Biophys. J.* **2001**, *80*, 2093–2109.
(15) Serrano-Andrés, L.; Roos, B. O. *J. Am. Chem. Soc.* **1996**, *118*, 185–195.
(16) Öhrn, A.; Karlström, G. *J. Phys. Chem. A* **2007**, *111*, 10468–10477.
(17) Chadborn, N.; Bryant, J.; Bain, A.; O'Shea, P. *Biophys. J.* **1999**, *76*, 2198–2207.
(18) Dumont, C.; Emilsson, T.; Gruebele, M. *Nat. Methods* **2009**, *6*, 515–519.
(19) Elbaz, Y.; Tayer, N.; Steinfels, E.; Steiner-Mordoch, S.; Schuldiner, S. *Biochemistry* **2005**, *44*, 7369–7377.
(20) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210–1249.
(21) Ringer, A. L.; Sinnokrot, M. O.; Lively, R. P.; Sherill, C. D. *Chem.—Eur. J.* **2006**, *12*, 3821–3828.
(22) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *J. Chem. Phys.* **2005**, *122*, 144323.
(23) Rocha-Rinza, T.; De Vico, L.; Veryazov, V.; Roos, B. O. *Chem. Phys. Lett.* **2006**, *426*, 268–272.
(24) Catalan, J.; Pereze, P.; Acuna, A. U. *J. Mol. Struct.* **1986**, *142*, 179–182.
(25) Slater, L. S.; Callis, P. R. *J. Phys. Chem.* **1995**, *99*, 8572–8581.
(26) Tine, A.; Aaron, J.-J. *Spectrochim. Acta A* **1996**, *52*, 69–70.
(27) Rogers, D. M.; Hirst, J. D. *J. Phys. Chem. A* **2003**, *107*, 11191–11200.
(28) Sharma, N.; Jain, S. K.; Rastogi, R. C. *Bull. Chem. Soc. Jpn.* **2003**, *76*, 1741–1746.
(29) Somers, K. R. F.; Ceulemans, A. *J. Phys. Chem. A* **2004**, *108*, 7577–7583.
(30) Sharma, N.; Jain, S. K.; Rastogi, R. C. *Spectrochim. Acta A* **2007**, *66*, 171–176.
(31) Robinson, D.; Besley, N. A.; Lunt, E. A. M.; O'Shea, P.; Hirst, J. D. *J. Phys. Chem. B* **2009**, *113*, 2535–2541.
(32) See, for example, Krylov, A. I. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.
(33) See, for example, Roos, B. O. *Adv. Chem. Phys.* **1987**, *69*, 399–442.
(34) Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004**, *126*, 4007–4016.
(35) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. *J. Chem. Phys.* **2001**, *115*, 3540–3544.
(36) Song, J.-W.; Hirosawa, T.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2007**, *126*, 154105.
(37) Lange, A. W.; Rohrdanz, M. A.; Herbert, J. M. *J. Phys. Chem. B* **2008**, *112*, 6304–6308.
(38) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
(39) Tsuneda, T.; Suzumura, T.; Hirao, K. *J. Chem. Phys.* **1999**, *110*, 10664–10678.
(40) van Leeuwen, R.; Baerends, E. *J. Phys. Rev. A* **1994**, *49*, 2421–2431.
(41) Widmark, P.-O.; Malmqvist, P.-A.; Roos, B. O. *Theor. Chim. Acta* **1990**, *77*, 291–306.
(42) Karlström, G.; Lindh, R.; Malmqvist, P.-A.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222–239.
(43) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(44) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evenseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(46) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494–4500.
(47) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(48) Hirata, S.; Head-Gordon, M. *Chem. Phys. Lett.* **1999**, *17*, 291–299.
(49) Besley, N. A. *Chem. Phys. Lett.* **2004**, *390*, 124–129.
(50) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.
(51) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, Frerich, J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F., III.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.
(52) Clark, T.; Chandrasekhar, J.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294–301.
(53) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656–10668.
(54) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–556.
(55) Eftink, M. R. *Methods Biochem. Anal.* **1991**, *35*, 127–205.
(56) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
(57) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
(58) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

JP9071108