

Analytical Harmonic Vibrational Frequencies for the Green Fluorescent Protein Computed with ONIOM: Chromophore Mode Character and Its Response to Environment

Lee M. Thompson,[†] Aurélie Lasoroski,^{†,‡} Paul M. Champion,[‡] J. Timothy Sage,[‡] Michael J. Frisch,[§] Jasper J. van Thor,^{*,||} and Michael J. Bearpark^{*,†}

[†]Department of Chemistry, Imperial College, London SW7 2AZ, United Kingdom

[‡]Department of Physics and Center for Interdisciplinary Research on Complex Systems, Northeastern University, Boston, Massachusetts 02115, United States

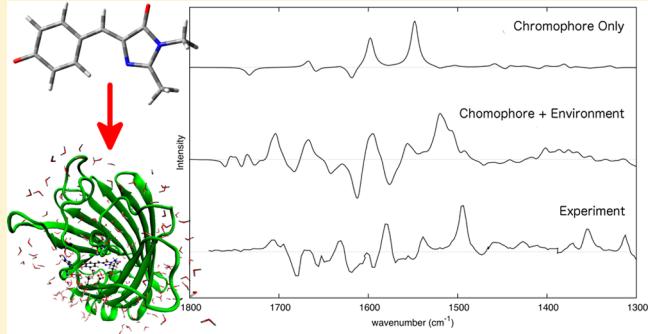
[§]Gaussian, Inc., 340 Quinnipiac Street, Building 40, Wallingford, Connecticut 06492, United States

[¶]Division of Molecular Biosciences, Imperial College, London SW7 2AZ, United Kingdom

^{||}Ecole Normale Supérieure, Département de Chimie, ENS-CNRS-UPMC UMR8640, 75005 Paris, France

S Supporting Information

ABSTRACT: A systematic comparison of different environmental effects on the vibrational modes of the 4-hydroxybenzylidene-2,3-dimethylimidazolinone (HBDI) chromophore using the ONIOM method allows us to model how the molecule's spectroscopic transitions are modified in the Green Fluorescent Protein (GFP). ONIOM(QM:MM) reduces the expense of normal mode calculations when computing the majority of second derivatives only at the MM level. New developments described here for the efficient solution of the CPHF equations, including contributions from electrostatic interactions with environment charges, mean that QM model systems of ~100 atoms can be embedded within a much larger MM environment of ~5000 atoms. The resulting vibrational normal modes, their associated frequencies, and dipole derivative vectors have been used to interpret experimental difference spectra (GFP₁₂-GFP_A), chromophore vibrational Stark shifts, and changes in the difference between electronic and vibrational transition dipoles (mode angles) in the protein environment.



1. INTRODUCTION

Interpretation of vibrational infrared spectra is greatly assisted by computation of the normal modes of the system. Advances in femtosecond coherent infrared spectroscopy experiments have enabled measurement of the vibration-vibration couplings in molecules and provide structural,^{1,2} reactive,^{3,4} and dynamic information.^{5,6} 2D-IR spectroscopy^{5,7} and vibrational coherent spectroscopy (VCS)^{8–10} measure properties of coupled vibrational oscillators and interpreting the resulting spectra requires detailed knowledge of the strength, direction, and coupling of transition dipole vectors. Theoretical models used to interpret this new generation of experiments must therefore include the coupled environment. In this paper we compute the vibrational modes of the Green Fluorescent Protein (GFP) and illustrate how chromophore modes respond to each environmental perturbation.

Vibrational modes may be computed analytically, usually within the harmonic approximation, to enable a mapping of spectroscopic signals to associated normal coordinates. Assembly and diagonalization of the Hessian (second

derivative) matrix is required for this, and for large molecules or systems where accurate electronic structure methods are required, this can be prohibitively expensive computationally. One workaround is partial Hessian diagonalization, often using the QM region only.^{11–13} An alternative approach is to compute only modes of interest through algorithms such as vibrational mode-tracking, which has the advantage of directly incorporating low-level effects.^{14,15} It is also possible to obtain normal modes from molecular dynamics simulations, where oscillations around a minimum may be used to extract second derivatives: this will include anharmonic effects, but the simulation time is required to tend to infinity.^{16,17}

Vacuum calculations of isolated chromophores provide a guide to aid mode assignment for chromo-proteins. However, neglect of the protein environment means that not all experimental features can be realistically described this way.^{18,19} While mechanical coupling will increase mode

Received: July 26, 2013

Published: January 31, 2014

delocalization through covalent and hydrogen bonds, the electrostatic environment results in transition dipole coupling and Stark tuning.^{20–24} Thus, the structural environment of a biological chromophore will modify each of its modes differently as a result of the dipole–dipole interactions. In addition to transition dipole coupling, vibrational Stark tuning results from the directional electric fields experienced by the chromophore.

The chromophore is a form of cyanine dye known as an oxonol, in which charge-transfer transitions occur via an inversion of the bond-length alternation over the carbon bridge.^{25,26} Vibrational modes involving the bridging atoms connect the resonance forms, and so it is expected that there will be strong coupling between the protein electronic environment and the vibrational state of the chromophore.²⁷ This suggests that including the protein electronic environment will be essential to correctly reproduce experimental vibrational spectroscopy results as the Stark tuning effect is likely to be significant.²⁸

ONIOM(QM:MM) can be used to include both mechanical and electrostatic effects in calculations, reducing the size of the QM region while treating the remainder of the molecule with MM, so that the cost associated with assembling the Hessian is reduced. Despite this, the dimensions of the region requiring a QM description often remain sizable (100+ atoms). A further complication is the use of electronic embedding where the number of geometrical degrees of freedom can be much larger than the number of AO basis functions. Modification of the way in which the CPHF equations are solved, as detailed here, give ~20% improvement in disk, I/O and CPU time for QM systems with 100+ atoms. Further developments reduce the expense of computing frequencies with electronic embedding substantially, such that these calculations are of the same order of magnitude as mechanically embedded calculations. This allows full harmonic vibrational frequency analysis of large molecules with both mechanical and electronic embedding.

Normal mode analysis of flexible proteins gives a large number of modes that are delocalized over the entire system, resulting in difficulties in interpretation.²⁹ Comparison with measured experimental parameters, however, requires a global picture of vibrational motion. Here, we present harmonic frequency calculations of the entire Green Fluorescent Protein using the ONIOM scheme and compare calculated frequency, intensity, and vibrational transition mode angle (VTMA) with experiment as follows:

- Computationally constructed difference spectra show that the experimental spectrum can be reproduced even within the harmonic approximation and indicates the importance of electronic embedding in correctly describing certain vibrational modes. This has allowed direct assignment of chromophore modes by matching the computed and experimental spectra (Section 4A).

- The change in chromophore mode frequency between mechanically and electronically embedded calculations allows calculation of the vibrational Stark shift explicitly. This has been used (Section 4B) as a new tool for analyzing frequency shifts of local dielectric probes such as C=O bond stretches in protein environments.

- Analyzing mode angles (VTMAs) indicates how the vibrational normal coordinate is modified in the protein environment. This shows that chromophore modes respond to both structural and electronic perturbations and that the coupling between the chromophore and its environment often

generates several modes for the protein matching one for the isolated chromophore (Section 4C).

This article is organized as follows: we first summarize the theoretical model used in this study. In this section we detail improvements made to the ONIOM frequency code in Gaussian 09 required to carry out the calculations described. After the practical computational details, we present results showing how the frequency, intensity, and VTMA of chromophore modes in GFP change as a function of the protein environment. ONIOM allows the separation of electronic and mechanical effects of the environment and we illustrate that both are required to reproduce the experimental I₂-A difference spectrum. The role of the protein electrostatic field is then related to frequency changes through the vibrational Stark effect. Finally, we examine the sensitivity of the vibrational transition dipole to environment and computational procedure. Acronyms used in the text are defined in the Nomenclature section.

2. THEORETICAL BACKGROUND

2A. ONIOM Methodology. The effect of the environment on chromophore vibrational modes can be explicitly determined by employing a two-layer ONIOM hybrid method.^{30–35} This has the advantage of including the environment with a less computationally expensive force-field calculation, while allowing vibrational analysis on the full system with no extra degrees of freedom, thus avoiding any link-atom artifacts in the normal-mode analysis.³⁶

The energy of a two-layer ONIOM calculation is obtained through an extrapolation

$$E^{\text{ONIOM}} = E_{\text{real}}^{\text{low}} + E_{\text{model}}^{\text{high}} - E_{\text{model}}^{\text{low}} \quad (1)$$

where each term represents the energy of an individual subcalculation. The model system is a fragment of the whole (real) system, which is calculated at both levels of theory (termed high and low). The ONIOM scheme can, therefore, be viewed as the addition of an environment at a lower level of theory to a model system, such as an isolated chromophore.

Dangling bonds in the model system that result from covalent interactions at the boundary are satisfied using hydrogen link atoms, the position being dependent on the positions of the corresponding atom in the real system. This allows the link atoms to be projected onto atoms within the real system so that there are no extra degrees of freedom. The ONIOM gradient can, therefore, be written as

$$\frac{\partial E^{\text{ONIOM}}}{\partial \mathbf{q}} = \frac{\partial E_{\text{real}}^{\text{low}}}{\partial \mathbf{q}} + \frac{\partial E_{\text{model}}^{\text{high}}}{\partial \mathbf{q}_m} \mathbf{J} - \frac{\partial E_{\text{model}}^{\text{low}}}{\partial \mathbf{q}_m} \mathbf{J} \quad (2)$$

where \mathbf{J} is the Jacobian matrix that projects the link atoms in the model system with coordinates \mathbf{q}_m onto atoms in the real system \mathbf{q} . The fact that the potential energy surface is well-defined means that minima, transition states, and reaction paths connecting them can be determined using standard optimization procedures.

ONIOM is able to treat the interaction of the model with the environment in several different ways. In this study, two different methods are employed. Mechanical embedding (ME), where the interaction between layers is treated at the low level of theory, is implicitly incorporated into the ONIOM scheme through the $E_{\text{real}}^{\text{low}} - E_{\text{model}}^{\text{low}}$ term. This means that the electronic structure of the model does not directly interact with the environment, but the nuclear geometry does. Electronic

embedding (EE) allows the wave function of the model region to interact with the electric field created by the environment.³⁵ In the implementation used here we allow the chromophore wave function to respond to the partial charge distribution of the environment but do not account for polarization of the environment by the model or for the individual environment partial charges to change.

Having determined a stationary point on the PES (Potential Energy Surface), a Hessian is required to carry out normal-mode analysis within the harmonic approximation. As with other derivatives with respect to nuclear coordinates, this is straightforward to obtain within the ONIOM scheme

$$\frac{\partial^2 E_{\text{ONIOM}}}{\partial \mathbf{q}^2} = \frac{\partial^2 E_{\text{real}}^{\text{low}}}{\partial \mathbf{q}^2} + \mathbf{J}^T \frac{\partial^2 E_{\text{model}}^{\text{high}}}{\partial q_m^2} \mathbf{J} - \mathbf{J}^T \frac{\partial^2 E_{\text{model}}^{\text{low}}}{\partial \mathbf{q}_m^2} \mathbf{J} \quad (3)$$

where, as previously, the Jacobian is used to transform coordinate systems. Further to this, molecular properties dependent on the electric field \mathbf{E} are also defined in a similar manner. The dipole moment μ is thus given as

$$\mu = \frac{\partial E_{\text{real}}^{\text{low}}}{\partial \mathbf{E}} + \frac{\partial E_{\text{model}}^{\text{high}}}{\partial \mathbf{E}} - \frac{\partial E_{\text{model}}^{\text{low}}}{\partial \mathbf{E}} \quad (4)$$

and the intensities are determined from the vibrational transition dipole:

$$\langle \nu | \mu | \nu' \rangle = \left(\frac{\partial^2 E_{\text{real}}^{\text{low}}}{\partial \mathbf{q} \partial \mathbf{E}} + \frac{\partial^2 E_{\text{model}}^{\text{high}}}{\partial \mathbf{q}_m \partial \mathbf{E}} \mathbf{J} - \frac{\partial^2 E_{\text{model}}^{\text{low}}}{\partial \mathbf{q}_m \partial \mathbf{E}} \mathbf{J} \right) \langle \nu | \mathbf{q} | \nu' \rangle \quad (5)$$

The theoretical infrared spectra of large biomolecules such as GFP can, therefore, be obtained from the ONIOM methodology in the harmonic approximation using the standard Hessian diagonalization procedures used in single layer systems. Given the large number of atoms in systems that ONIOM is commonly used for, some extra considerations are required for assembly and diagonalization of very large Hessians (Section 2B).

2B. ONIOM Frequency Analysis for Large Molecules.

Computation of the force constants (nuclear coordinate second derivatives) for the models considered here raise two classes of problems:

1. Issues with the efficient evaluation of force constants for sizable QM systems ($N_A > 100$ atoms, as in the model systems used here), which are not particular to QM/MM models.

2. Additional issues specific to the case of QM/MM with electronic embedding, in which the number of geometrical degrees of freedom can be much larger than the number of AO basis functions. We consider each problem in turn.

2B.I. SCF Second Derivatives for Large Molecules.

Straightforward implementation of the second derivatives of an SCF energy (as in ref 37) leads to terms which involve second derivatives of one- and two-electron integrals, which can be summed directly into the force constants and whose computational cost is asymptotically quadratic in the size of the system, and two terms which require the density derivatives from CPHF (Coupled Perturbed Hartree–Fock). In the notation of ref 37 these second derivative terms are

$$\langle \mathbf{P}^x \mathbf{F}^{(y)} \rangle - \langle \mathbf{W}^x \mathbf{S}^y \rangle \quad (6)$$

where x and y are nuclear coordinates and

$$\mathbf{P}^x = \bar{\mathbf{U}} - \mathbf{S}_{\text{oo}}^x = \mathbf{U}^x + \mathbf{U}^{x\dagger} - \mathbf{S}_{\text{oo}}^x \quad (7)$$

and \mathbf{U}^x is computed by solving the CPHF equations:

$$\mathbf{AU}_{ai}^x = \mathbf{DU}_{ai}^x - \mathbf{G}(\bar{\mathbf{U}}^x) = \mathbf{B}_{ai}^x \quad (8)$$

$$\mathbf{D}_{ai} = \boldsymbol{\epsilon}_a - \boldsymbol{\epsilon}_i \quad (9)$$

$$\mathbf{B}^x = \mathbf{F}^{(x)} - \mathbf{G}(\mathbf{S}_{\text{oo}}^x) - \mathbf{FPS}^x - \mathbf{S}^x \mathbf{PF} \quad (10)$$

Each iteration during the solution of eq 8 requires forming a Fock-like matrix $\mathbf{G}(\mathbf{X})$ for some trial \mathbf{X} for each perturbation, along with transforming \mathbf{X} to the AO basis and then $\mathbf{G}(\mathbf{X})$ to the MO basis. If M is the number of perturbations ($M = 3N_A$) and N is the number of basis functions ($N = 9N_A$ for 6-31G*; larger for better basis sets), then the Fock formation has a formal computational cost of $O(N^4)$ for each perturbation or $O(MN^4)$ overall. Standard techniques for computation of Fock matrices in linear time^{38–42} reduce this to $O(MN)$ when N_A is roughly 100, depending on the basis set. The transformations between the AO and MO basis sets are $O(N^3)$ for each perturbation, hence $O(MN^3)$ overall, but are simple matrix multiplications and hence only become significant for larger systems. [For $N_A < 1000$, the matrices involved are not sufficiently sparse to make sparse matrix methods competitive for this problem.]

Solving for \mathbf{U}^x requires storage and I/O for $\mathbf{F}^{(x)}$ $O(MN^2)$, \mathbf{B}^x $O(MOV)$ and $O(MOV)$ intermediate quantities during solution, where O = number of occupied orbitals, $V=N-O$ is the number of virtual orbitals.

Assembling the derivative matrices required in eq 6 requires $O(MN^2)$ storage and I/O, and forming the traces of the matrices involves $O(M^2N^2)$ computations. For larger systems, these costs can be significant.

For larger ($N_A > 100$) molecules it is important to avoid as much of the storage and I/O as possible and to assemble the terms as efficiently as possible. The second derivative term eq 6 can be rearranged to

$$\langle \mathbf{P}^x \mathbf{F}^{(y)} \rangle - \langle \mathbf{W}^x \mathbf{S}^y \rangle = \langle \mathbf{P}^x \mathbf{B}^y \rangle = \langle \bar{\mathbf{U}}^x \mathbf{B}^y \rangle - \langle \mathbf{S}_{\text{oo}}^x \mathbf{B}^y \rangle \quad (11)$$

The second term can be computed when \mathbf{B} is formed in the AO basis prior to transforming to the occupied-virtual block to form the right-hand side of eq 8. The first term can be computed directly from the CPHF solutions without assembling quantities such as \mathbf{P}^x , \mathbf{F}^x , or \mathbf{W}^x .

The CPHF solutions are typically expanded in a common expansion space.⁴³ The M initial \mathbf{B} vectors are orthogonalized, leaving L_1 linearly independent right-hand sides, and then the space is expanded until $L_2 > L_1$ vectors solve the CPHF problem for the given perturbations.

In the pure QM case, $L_1 = M$ or $L_1 = M-3$ and typically $L_2 \approx SL_1$ [If electric field perturbations are included, their \mathbf{B} 's are linearly dependent on the atomic displacement \mathbf{B} 's by translational invariance]. So forming \mathbf{U}^x from the solutions in the reduced set has $O(ML_2OV)$ cost and then evaluating the second derivative term has $O(M^2OV)$ cost. A better alternative is to form \mathbf{B} in the reduced space at $O(ML_1OV)$ cost followed by evaluating the second derivative term in the reduced space at $O(M^2L_1)$ cost, since $L_1 < L_2 < OV$.

The effect of these rearrangements is about a 20% savings in disk, I/O, and CPU time for 100+ atoms systems.

2B.II. Second Derivatives of ONIOM(SCF-MM) or QM(SCF)/MM for Large Molecules. When electronic embedding is included, the QM wave function depends on the coordinates of every atom. For a system of the type studied here, a typical protein with a QM region containing >100 atoms ($N_A = 4367$,

Table 1. Parameters and Elapsed Times in Minutes for Mechanical Embedding and Electronic Embedding of Sample ONIOM Calculations with Two Different Model Sizes and Two Basis Sets Computed Using a 32-Processor SGI Altix with IA64 Processors

model	N_A	M	N_Q	N	ME			EE			EE/ME
					L_1	L_2	time ^a	L_1	L_2	time ^a	
small 6-31G*	4367	13104	107	864	321	1845	165	1559	3369	234	1.4
small 6-31+G*	4367	13104	107	1064	321	1894	617	2244	4790	13687	2.2
large 6-31G*	4367	13104	167	1244	501	2763	231	2637	5394	401	1.7

^aTotal elapsed times for the ONIOM frequency calculation in minutes, including 62 min outside of the QM calculation.

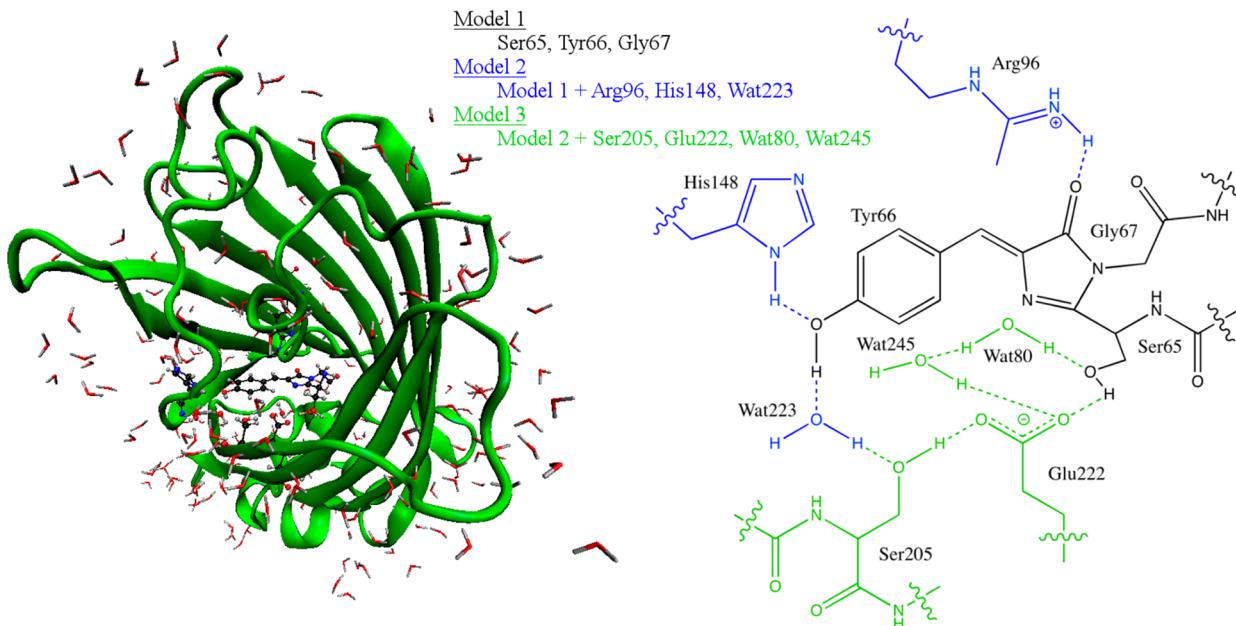


Figure 1. GFP_A showing the full real (left) and chromophore model (right) systems used in the ONIOM calculations presented here. The Model 3 system is highlighted within the real system on the left using a ball and stick representation.

$M = 13104$ [Including electric field perturbations so that IR intensities can be predicted.] along with $N_Q = 107$ or 167 , $N = 864$ or 1244), new performance issues arise:

1. Storage and I/O for the remaining $O(MN^2)$ quantity, $F^{(x)}$, is substantial.

2. The cost of the $O(M^2OV)$ steps is much larger relative to the other steps than for pure QM on the model system.

There is also the compensating simplification that, since the basis functions in the QM calculation do not depend on the positions of the non-QM atoms, $F^{(x_a)}$ for the displacement of MM atom a in the x direction is given by a simple one-electron operator

$$F_{\mu\nu}^{x_a} = h_{\mu\nu}^{x_a} = \frac{\partial}{\partial x_a} \int \chi_\mu(\vec{r}) \frac{1}{|r - R_a|} \chi_\nu(\vec{r}) dr \quad (12)$$

which can be computed with $O(N)$ cost for each x and hence can be recomputed as needed. Similarly, $S^x = 0$ if x is a displacement of an MM atom.

Hence the second derivative procedure can be modified as follows:

1. $F^{(x)}$ is computed and stored on disk only when x involves a QM atom.

2. When $B^{(x)}$ is formed, $F^{(x)} = h^x$ is computed if x is an MM atom.

3. The terms involving $\langle S_{oo}^x B^y \rangle$, computed using the AO B^x , involve only $O(N_Q N^2 O) + O(N_Q N O^2)$ cost.

4. When a large fraction of the perturbations involve MM atoms, the full set of M right-hand sides are highly linearly dependent, and when the set is orthonormalized to produce L_1 linearly independent vectors, L_1/M is in the range 0.1–0.2, depending on the size of the model system and the spatial extent (diffuseness) of the AO basis. Similarly, the dimension of the reduced space which solves for all the perturbations, L_2 , ranges from 0.36 M to 0.41 M . So the increase in the number of Fock formations and matrix transformations during the CPHF solution for EE compared to ME or pure QM on the model system is much less than the formal factor of N_A/N_Q .

5. The improvement from evaluating $\langle U^x B^y \rangle$ in the reduced space is much greater than for the pure QM case, because the difference between $O(M^2 OV)$ and $O(M^2 L_1)$ is much larger when M is big.

Table 1 gives parameters and elapsed times (on a 32-processor SGI Altix with IA64 processors) for mechanical embedding (essentially the time for a bare GFP model system QM calculation) and electronic embedding, using two sizes of model system and two basis sets. For either size of model system using a basis without diffuse functions, the cost of frequencies with EE is less than twice that of the pure model system or ME. Even with a more diffuse basis, the extra cost is only a factor of 2.2. The calculations include about 62 min of non-QM steps, including the real system MM force constants, diagonalizing the force constant matrix, and computing intensities, which is a significant part of the total time for the

calculations using the valence basis set. The QM portion of the calculations scales as $O(N_A^{1.2})$ with ME and $O(N_A^{1.5})$ with EE, respectively.

These new developments mean that, even for systems with large numbers of atoms, frequency calculations using EE are of comparable expense to those using ME. This is despite the extra work required to compute the contribution of the low-level point charges in the CPHF equations required for EE. We can now routinely compute the polarization effect of the low-level partial charge distribution on the vibrational modes of large molecules.

3. COMPUTATIONAL DETAILS

Initial atomic coordinates were taken from the A-chain of the 1W7S.pdb GFP_A X-ray structure, including crystallographic water, determined at 1.8 Å resolution.⁴⁴ Standard residues were protonated using PDB2PQR,⁴⁵ where the protonation states of residues were determined at pH 7.8 with PROPKA.⁴⁶

The Amber Cornell 96 force-field and TIP3P water model⁴⁷ was used as the low-level MM (Molecular Mechanics) method. The chromophore which forms autocatalytically from the cyclization of Ser65, Tyr66, and Gly67 residues required further parametrization: Leap (part of the AmberTools package⁴⁸) was employed to produce an AMBER library file for this residue. Stretching, bending, and torsional parameters not present in the force field were obtained from the GAFF force-field,⁴⁹ while RESP⁵⁰ partial charges calculated with HF/6-31G(d) were assigned using R.E.D. Tools.⁵¹ The protein was then optimized using MM alone in Gaussian 09⁵² with all force field parameters defined, and these coordinates were subsequently used as the initial structure for ONIOM geometry optimizations. This approach approximates that the protein crystal structure represents a low-temperature average^{53,54} and that the nearest minimum (required in order to make the transformation to decoupled normal coordinates within the harmonic approximation) is close by. In the case of GFP this is less of an approximation, as the protein is known to have a relatively rigid structure.⁵⁵ An alternative approach would be to average over several minimized snapshots from an MD calculation after equilibration, which may be a requirement for other proteins.

The B3LYP/6-31+G(d) level of theory was used to describe the model system as it had previously been shown to give reasonable agreement with the ground state vibrational spectra of the chromophore (HBDI).⁵⁶ In order to examine the effect of increasing ONIOM model size, three different-sized model systems were selected (Figure 1). The partitioning of each model was chosen so as to conform as best as possible to the various guidelines indicated in the literature.⁵⁷ Model 1 includes the three residues comprising the chromophore only (Ser65, Tyr66, and Gly67), giving 38 atoms modeled at the high level of theory (plus two link atoms to replace covalent bonds to the environment). Model 2 expanded this to include the first hydrogen bonding shell, namely residues His148, Arg96, and Wat223, giving 71 atoms in the model system (including 4 link atoms). The largest model, Model 3, additionally includes the chemically important proton transfer chain Glu222, Ser205 plus hydrogen bonded Wat80, Wat245, giving 103 atoms in the model system out of a total of 4366 atoms (including 6 link atoms). This allowed the proton transfer network to be included at the high level of theory, so that a continuous PES exists between anionic (I_2) and neutral (A) forms of GFP. In the crystal structure from which the initial coordinates were obtained, Thr203 was found not to hydrogen bond to the

chromophore (based on donor–acceptor bond distance) and so was not included in the model region.

Each model partition (Figure 1) was optimized using ONIOM-ME as implemented in Gaussian 09⁵² with micro-iterations.³⁴ These structures were subsequently used as the initial coordinates for electronically embedded (EE) calculations, in which the environment was electronically coupled to the model. Thus the steric and electronic effects of the protein environment could be studied independently. The initial coordinates of the GFP₁₂⁵⁸ structure were obtained from the electronically embedded GFP_A Model 3 structure by manually transferring the hydrogen atoms along the proton wire formed by Tyr66, Wat223, Ser205, and Glu222. This was then optimized using ME and EE schemes.

After optimization, the nature of the critical point was determined through normal-mode analysis. This both confirmed that a minimum energy structure had been obtained and produced the vibrational dipole derivative vectors (μ_2) of each mode for VTMA analysis.

Chromophore modes were selected by comparison with chromophore-only calculations. This was achieved by analyzing normal coordinates in internal coordinates, where modes were selected based on large displacements along chromophore-only internal coordinates. For the Phenol 2 mode, for example, this was determined to be bending of the phenol H–O–C angle. A minimum displacement limit was set to reduce the number of modes to select from, until the number of modes was less than 50. From these a mode could be selected by eye or if required, a histogram could be used to locate clusters of important modes (Figure S1) from which the mode was selected. The construction of difference spectra then provided an additional check, as this highlighted chromophore modes by canceling the low-level effect (Figure S6). This provided a method of rapidly selecting modes of the correct character which was sufficient for the analysis given here; however, other schemes have been proposed.⁵⁹

Frequent reference is given in the discussion below to the extent to which a mode is localized within the model region. This can be quantified as the fraction of the normal mode vector belonging to atoms within the model.

Single point ONIOM calculations were carried out on the ground state minimum structures using TD-B3LYP as the high level theory to give the $S_0 \rightarrow S_1$ electronic transition dipole (μ_1). It has been demonstrated that the direction of μ_1 is not very sensitive to the level of theory used, including long-range corrected functionals.⁶⁰ This was verified for the ONIOM calculation by comparison with TD-CAM-B3LYP in the M1 ONIOM-ME system, which revealed $<1^\circ$ difference in the direction of the electronic transition dipole. This assumes that the excitation is localized within the model region and that the interlayer interaction is based on the ground-state electronic structure, allowing a force-field to be used as the low level of theory.^{61–63} Experimentally determined VTMA_s (ζ) can be compared with theoretically calculated VTMA_s from the scalar product of the electronic transition dipole μ_1 and the infrared transition dipole of the vibrational normal mode μ_2

$$\zeta = \cos \frac{\mu_1 \cdot \mu_2}{|\mu_1| |\mu_2|} \quad (13)$$

As well as VTMA_s, normal-mode analysis also gives calculated frequencies and intensities for comparison with experiment. In order to account for differences arising between theory and

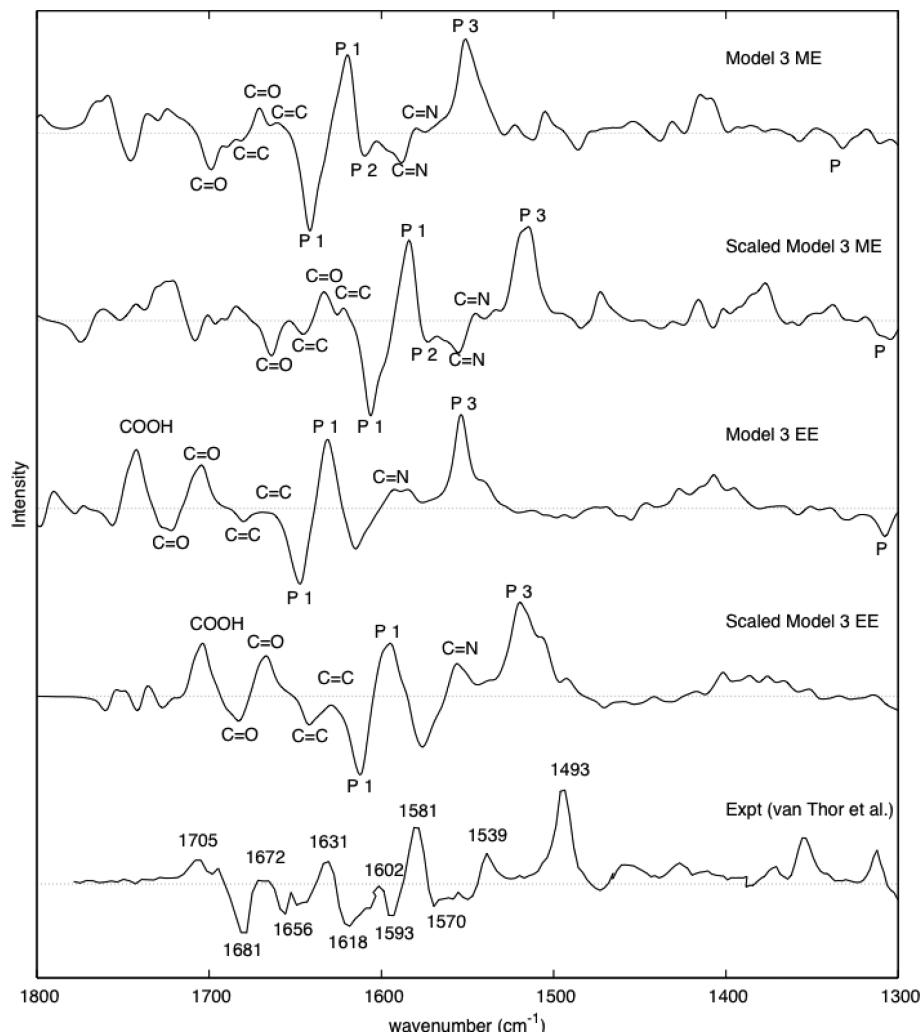


Figure 2. Computed GFP I₂-A difference spectra obtained using mechanically (ME) and electronically (EE) coupled environments using unscaled and scaled Hessian components, compared with the time-independent experimental difference spectrum from van Thor et al.⁷⁰ Gaussian line widths of 5 cm⁻¹ were applied to the computed spectra. In this difference spectrum, 'up' corresponds to the I₂ form, while 'down' corresponds to the A form.

experiment due to the harmonic approximation and finite basis set it is possible to scale the frequencies.⁶⁴ In ONIOM, however, the Hessian is constructed from two different levels of theory, so in order to correct the frequencies, the Hessian of each component can be scaled by a factor:

$$\mathbf{H}^{\text{ONIOM}} = c_{\text{low}}^2 \mathbf{H}_{\text{real}}^{\text{low}} + c_{\text{high}}^2 \mathbf{J}^T \mathbf{H}_{\text{model}}^{\text{high}} \mathbf{J} - c_{\text{low}}^2 \mathbf{J}^T \mathbf{H}_{\text{model}}^{\text{low}} \mathbf{J} \quad (14)$$

As it is the Hessian that is scaled in ONIOM rather than the eigenvalues this may result in a different set of normal modes being obtained. A comparison of VTMAs, frequencies, and intensities of chromophore modes using both scaled and unscaled Hessians was carried out for both the A and I₂ structures with the Model 3 partition (Figure 1), in which c_{low} was set to 1.000⁴⁷ (i.e., unscaled) and c_{high} was set to 0.9784.⁶⁵ This allowed us to observe how sensitive the computed properties were to shifts in the frequency of the chromophore oscillator that would result from going beyond the harmonic approximation. Furthermore, we were able to determine if simple scaling of the Hessian elements could reproduce the frequencies of features in the difference spectrum, without having to resort to expensive computation of higher order derivatives.

The effect on vibrational frequencies of the electrostatic potential resulting from the low level charge distribution was determined from electronic embedding calculations based on a static partial charge model. This neglects higher order multipoles and polarization of the environment by the chromophore and so may not exactly mimic the protein electronic distribution and underestimate transition dipole coupling. Despite this it gives a well-defined first approximation that we show improves the description of the environment effect. The connection to the local electrostatic field was determined by solution of the linearized Poisson–Boltzmann equation using the APBS software (v0.5.1) as described in the Supporting Information.⁶⁶

Difference spectra were computed by taking the computed spectra of GFP_A and GFP_{I2} and adding Gaussian line widths of 5 cm⁻¹ to each line. The resulting line shapes were saved at 0.5 cm⁻¹ resolution and subtracted from each other before plotting the resulting value against frequency. Subtraction of the two spectra means that modes which are unaffected by proton transfer cancel, so that the difference spectrum gives an indication of chromophore modes. Despite this, computation of the whole spectrum is required in order to match theory with experiment, as it is not possible to know a priori the extent to

which the environment modes cancel. The new developments listed in Section 2B now allow difference spectra of molecules the size of GFP to be routinely computed.

4. RESULTS AND DISCUSSION

UV/vis spectroscopy of GFP reveals two absorbing species.⁶⁷ The neutral A state/structure of GFP corresponds to the species which absorbs at $\lambda_{\text{max}} = 398 \text{ nm}$, with protonation of the Tyr66 phenol group in the chromophore HBDI residue. Within the proposed GFP photocycle,^{68,69} this A state (GFP_A) corresponds to the global minimum on the ground state potential energy surface (PES), and we calculate its harmonic vibrational frequencies using a sequence of ONIOM partitions (Models 1 to 3, Figure 1) as described above. The three properties obtained from each vibrational mode are as follows: the frequency, which corresponds to the curvature of the PES; intensity, which corresponds to the magnitude of the vibrational transition dipole moment; and VTMA/mode angle, which corresponds to the direction of the vibrational transition dipole moment.

We first determine in Section 4A how GFP frequencies and intensities correlate with those obtained experimentally. Additional computation of the Model 3 (M3) ground state anionic species - formed from proton transfer between nearby protein residues (GFP_{12}^-), from which recovery in the A state is obtained experimentally in 400 ps⁵⁸ - allows us to construct the I_2 -A difference spectrum. The electronic environmental effect on both frequency and intensity of chromophore modes can then be determined by comparing mechanical and electronic embedding calculations, and compared directly with experiment, enabling mode assignments to be made from a calculation on the full molecule. The results we obtain agree with alternative techniques such as isotopic labeling.

The effects of the electronic environment on the frequencies of important chromophore modes - the vibrational Stark effect (Section 4B) - can also be determined by comparing mechanical and electronic embedding calculations and by evaluating the electric field produced by the surrounding low-level partial charge distribution directly. While these results depend upon the choice of model partition, the correlations we obtain between frequency shifts and computed fields are reliable for modes that are highly localized in the chromophore/model region.

We finally discuss (Section 4C) the changes in the VTMA's as a result of the protein environment. As VTMA's are an indicator of the direction of the vibrational transition dipole, they are sensitive to any interaction between the chromophore and the environment that modifies the normal coordinate. The complications of coupling to the environment are discussed here.

4A. $\text{GFP}_{12}-\text{GFP}_A$ Difference Spectra for Chromophore Mode Assignment. Experimentally, in order to overcome limitations such as spectral congestion and signal-to-noise error, reaction pathways are studied relative to an initial state (nuclear configuration) through difference spectra. From computations, the GFP I_2 -A difference spectrum can be constructed by subtracting the computed Model 3 A-state spectrum from that of the I_2 state, using a Gaussian line width of 5 cm^{-1} (Figure 2).

Figure 2 shows that these computational and experimental difference spectra⁷⁰ are in general agreement. The main difference - that highlights the need for electronic embedding in calculations - is the location of the C=O modes around 1700 cm^{-1} , which are $\sim 20 \text{ cm}^{-1}$ higher in frequency with EE

than with ME. This reveals that the C=O modes are highly sensitive to the low-level region's electrostatic field acting on the chromophore and that neutral and anionic states experience similar fields as their C=O modes are shifted by similar amounts.

The use of Hessian scaling (as described in the computational details, section 3) has a small effect on the shape of the spectrum due to modification of the calculated intensities: the same features are present but shifted to lower wavenumbers. Using 0.9784 as the high-level scaling factor resulted in an overall $\sim 30 \text{ cm}^{-1}$ change in wavenumber relative to the unscaled calculation. This appears to have suitably matched the C=O and C=C frequencies (Figure 2), while lower frequency peaks such as Phenol (3)/P3 remain slightly shifted to higher frequencies by up to 10 cm^{-1} .

The agreement between experimental and theoretical spectra demonstrates that not only are frequencies well reproduced but so too are the intensities of each mode. This can be inferred from the fact that if the mode intensities do not cancel in the correct way, then even if the frequencies are computed correctly, it would not be possible to reproduce the spectrum. The use of the difference spectrum to provide mode assignments therefore means that we are not just trying to match a particular frequency but also the shape of the spectrum.

Based on our computational results, the experimental spectrum can be assigned by matching the line shape of experimental and theoretical spectra, as shown in Table 2. The

Table 2. Assignments of the GFP A and GFP I_2 State Chromophore Normal Mode Frequencies, Derived from the Computed GFP I_2 -A Difference Spectra

mode	A	I_2
Glu222 COOH	-	1705
C=O	1681	1672
C=C	1656	1631
Phenol (1)	1593	1581
Phenol (2)	1570	-
C=N	-	1539
Phenol (3)	-	1493

computed normal modes that lie under the peak maxima can then be examined and the character of the modes determined. We are thus able to connect the experimentally computed spectrum with the computed normal modes of the entire protein. The assignments made in this manner agree with previous assignments based on isotopic labeling and normal mode calculations of vacuum HBDI.^{56,70,71} In this case, however, a direct mapping of the peaks shown in Figure 2 to normal coordinates of the entire protein (rather than just the chromophore) has been achieved.

Using computationally constructed difference spectra, therefore, we have shown that the experimental spectrum can be reasonably reproduced even within the harmonic approximation. An important observation is that EE is important to describe some modes, such as the frequency of the C=O mode. We now seek to understand this further by explicitly calculating how the frequency shift observed in the C=O mode corresponds to the electrostatic field in which it is embedded.

4B. Chromophore Stark Shifts as Local Dielectric Probes. The field induced frequency shift in the chromophore modes is due to vibrational Stark coupling of the protein

environment with the model's electronic structure. Using the chromophore modes as a probe, the nature of the local electric field they experience may be understood.

In this section we compare the field, as determined from the frequency shift of the imidazolinone C=O bond stretch on going from mechanical to electronic embedding, with an explicit Poisson–Boltzmann calculation of the field arising from the low level partial charge distribution. This is done by comparing the frequency shift with values of the field computed at the C and O atoms and the C=O midpoint. Our aim is to correlate the frequency shifts we observe with computed fields, not to predict the exact field in the protein cavity. However, we show that for delocalized modes, just knowing the field value at the bond midpoint is not enough to predict the frequency shift, as this partly depends on the choice of model size. Thus explicit calculation of the frequency shift presents an alternative method for understanding how a local dielectric probe, such as C=O, responds when embedded in the protein environment.

The chromophore wave function in ONIOM-ME is modified by expanding the model region to include nearby residues (Figure 1). This is one of the reasons that ONIOM-ME frequencies change between model sizes. The response to charges lying outside the model region, however, is only included when ONIOM-EE is used. The change in vibrational frequency when electronic embedding is used, therefore, represents the electric field perturbation caused by the surrounding low-level atoms to the wave function of the model region.

The magnitude of the resulting frequency shift is the vibrational Stark effect. This depends upon the alignment of the vector defining the Stark tuning rate with the electrostatic field.^{72,73} The change in the vibrational frequency of a given mode upon the application of the field is, therefore, an indicator of the size of the component of the field aligned with the vector defining the Stark tuning rate

$$\hbar c \Delta \tilde{\nu} = -\Delta \mu \cdot E \quad (15)$$

where E defines the electric field vector, $\Delta \mu$ is the vector that defines the Stark tuning rate, and $\Delta \tilde{\nu}$ is the change of the frequency in wavenumbers.

Several recent studies have computed the Stark shift resulting from protein mutation by explicit calculation of the protein electrostatic field projected along an oscillator probe.^{74–77} Accurate prediction of Stark shifts in proteins was shown to be complicated by a strong dependence upon conformation and the quality of the computed electrostatic potential. To obtain quantitative agreement with experiment, polarization of the MM partial charges,⁷⁸ through the use of a charge update scheme such as MoD-QM/MM,⁷⁹ as well as some form of conformational sampling was suggested, such as molecular dynamics (MD) or Monte Carlo (MC) simulation. In this section we rationalize the shift in vibrational frequency on going from mechanical to electronic embedding of the model region by examining the local protein electrostatic environment, generated by the low-level partial charges. We do this by focusing on the main C=O mode, as this appears from Figure 2 to have the largest Stark shift of the selected chromophore modes. Here, we do not attempt to compute the field exactly, but we calculate vibrational frequencies explicitly and compare with computed fields.

In order to calibrate the response of the C=O oscillator to an electric field it is necessary to first determine the Stark

tuning rate at the level of theory employed. To do this we first calculate the response of a vacuum CO molecule at B3LYP/6-31+G(d) to a field of 25.8 MV cm⁻¹ coincident with the internuclear axis such that the electric field vector points from O to C (CO-Z+50 in Table 3). Using this data the Stark tuning

Table 3. Frequency and Geometry Response of the CO Oscillator to an Electric Field Perturbation in Several Environments: Vacuum CO and within GFP Models M1, M2, and M3 (Figure 1)^a

system	$\nu(0)/\text{cm}^{-1}$	$\nu(E)/\text{cm}^{-1}$	$\nu(E)-\nu(0)/\text{cm}^{-1}$	$\Delta R(\text{CO})/\text{\AA}$
CO-Z+50	2202.72	2188.88	-13.84	+0.0017
GFP-M1-Z+50 SP	1789.11	1794.34	+5.23	-
GFP-M1-Z+50	1789.11	1734.35	-54.76	+0.0084
GFP-M1	1789.11	1776.53	-11.42	+0.0079
GFP-M2-Z+50	1741.78	1671.39	-70.39	+0.0132
GFP-M2	1741.78	1761.03	+19.25	-0.0086
GFP-M3	1698.87	1721.19	+22.32	-0.0129

^aResponse computed using B3LYP/6-31+G(d). Z+50 indicates a 0.005 au (25.8 MV cm⁻¹) field applied along the CO axis otherwise response due to environment electrostatic field as calculated from difference between mechanical and electronic embedding. SP indicates frequency response at field-free geometry.

rate of this molecule is determined to be +0.54 cm⁻¹ (MV/cm)⁻¹. We use the convention that positive $\Delta \mu$ indicate a vector pointing from O to C. This is in line with previous studies of CO based on single determinant methods,⁸⁰ which typically give 50% greater shifts than those determined by multiconfigurational methods, such as CASSCF (+0.38 cm⁻¹ (MV/cm)⁻¹).⁸¹ As previously noted, calculations of the Stark tuning rate in the free molecule may not be directly comparable with that in a bound species.²³

The frequency response of the C=O probe within the GFP protein environment was carried out by computing normal modes of the mechanically embedded Model 1 system with a field of 25.8 MV cm⁻¹ coincident with the C=O bond, without allowing for geometry optimization (GFP-M1-Z+50 SP in Table 3). The change in vibrational frequency of the C=O mode was found to be +5.23 cm⁻¹ ($\Delta \mu = -0.20 \text{ cm}^{-1} (\text{MV}/\text{cm})^{-1}$), showing significant sensitivity to the applied field, especially considering the absence of any geometry relaxation.

Allowing for geometry optimization of the entire protein in response to the same field results in an increase in C=O bond length of 0.0084 Å (GFP-M1-Z+50 in Table 3). This response is about 5 times greater than that of the vacuum CO molecule and is associated with a frequency change of -54.76 cm⁻¹, suggesting the Stark tuning rate of the C=O oscillator within the protein is +2.12 cm⁻¹ (MV/cm)⁻¹. Using this value to analyze the vibrational frequency perturbation from electronically embedding the Model 1 partition in the ONIOM calculation (GFP-M1 in Table 3), the field component coincident with the C=O axis is 5.38 MV cm⁻¹.

Figure 3 shows the electrostatic potential determined from the low-level charge distribution around the Model 1 system in the plane coincident with the chromophore. The electric field generated is displayed as black field lines, with the direction of the electric field vector along the C=O bond indicated by the three red cones. This indicates that the electric field lies along the imidazolinone C=O bond, with a component that points into the plane at the oxygen atom and changes to point out of

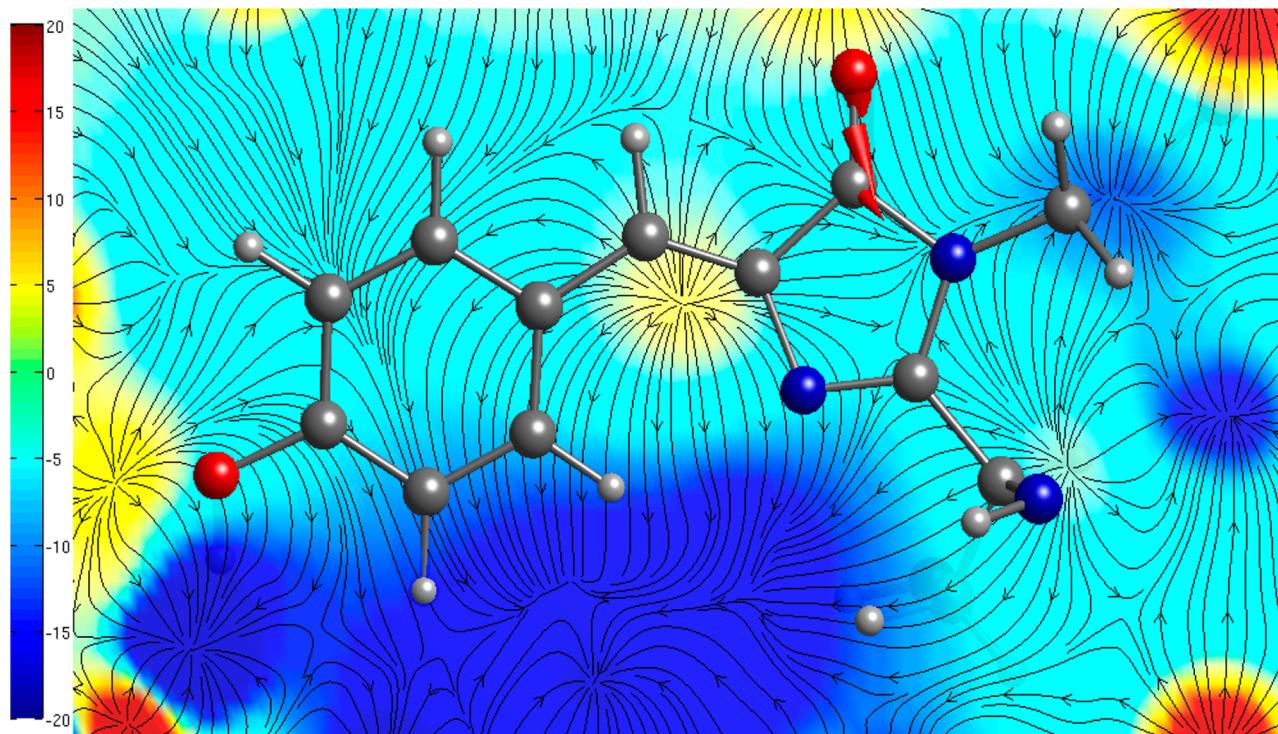


Figure 3. The electrostatic potential of the low level region around the Model 1 partition, ranging from -20 (blue) to 20 (red) kT/e calculated from solution of the Poisson–Boltzmann equation using AMBER partial charges and radii.⁶⁶ Field lines indicating local directionality of the electric field are shown in black. The red cones on C, the C=O midpoint, and O indicate the direction of the electric field vector along the C=O bond. ($1 kT/e = 25.68$ mV at 298 K.)

the plane as it moves along the C=O internuclear axis to the carbon atom. The magnitude of this field is 3.90 MV cm $^{-1}$ at the carbon atom, 5.20 MV cm $^{-1}$ at the midpoint of the C=O bond, and 6.70 MV cm $^{-1}$ at the oxygen atom, demonstrating a significant variation over a short distance. The corresponding component of the field projected along the C=O internuclear axis is 1.96 , 4.56 , and 6.30 MV cm $^{-1}$ for the three positions, respectively. The value obtained from the frequency response of the C=O mode of 5.38 MV cm $^{-1}$ determined previously appears, therefore, to be in good agreement with this electrostatic model.

Figure 3 shows that two positively charged regions govern the directionality of the electric field in the plane for Model 1. These result from the hydrogen of the Arg96 residue, which hydrogen bonds to the C=O, and a hydrogen of Wat245, which is held in place by hydrogen bonding with nearby waters and Glu222. However, as we now show, the field changes direction if we change the model system size.

Expansion of the model region to include the first hydrogen bonding shell causes the observed frequency shift upon electronic embedding to change sign and almost double in magnitude (GFP-M2 in Table 3). Arg96 is now included within the model, such that the chromophore is already polarized by this residue in the mechanically embedded calculation, and the low level electrostatic potential does not have the positively charged region above the C=O (shown in Figure S7). As a result, the electric field in the plane now lies along the C=O axis in the opposite direction, although this component of the field is quite small and the overall direction of the field is out of the plane. The magnitude of the electric field at the three points along the C=O axis is greater than in the Model 1 partition: 14.88 MV cm $^{-1}$ at the carbon, 13.75 MV cm $^{-1}$ at the midpoint,

and 12.10 MV cm $^{-1}$ at the oxygen atom. The component directed along the C=O internuclear axis is 4.80 , 1.27 , and -1.14 MV cm $^{-1}$ for the three positions, respectively, showing that the orientation of the field has indeed switched direction, in agreement with the vibrational response. Assuming the Stark tuning rate of $+2.12$ cm $^{-1}$ (MV/cm) $^{-1}$ is still valid for the larger model size, the expected field component along the C=O axis from the vibrational response is 9.08 MV cm $^{-1}$. This represents an overestimation of the size of the electric field. However, calibration of the frequency response with the Model 2 partition in the same manner as before (GFP-M2-Z+50 in Table 3) reveals that the C=O mode is even more sensitive to an applied external field, with a Stark tuning rate of $+2.73$ cm $^{-1}$ (MV/cm) $^{-1}$. Using this new value, a field strength of 7.05 MV cm $^{-1}$ is obtained, in closer agreement with the electrostatics calculations, although still larger than the values at any point along the bond.

The cause of this disagreement is that, by increasing model size to include the Arg96 residue, coupling via the hydrogen bond to the C=O oscillation occurs. The C=O normal coordinate is no longer highly localized to the C–O stretching coordinate but now involves significant NH₂ bending. Measurements along the C=O bond are no longer sufficient to understand the change in electric field gradient along the Stark tuning vector, owing to the highly local nature of the protein electrostatic field and delocalization of the normal coordinate.

The developments described in Section 2B allow us to compute the frequency shift of a chromophore mode in the protein field explicitly, rather than computing the electrostatic field and trying to relate this to observed frequency shifts. This has the advantage of not relying on the assumptions that i) the

Table 4. VTMA_s (ζ) in Degrees between μ_1 and μ_2 for Vibrational Modes Belonging to the Chromophore of GFP_A Calculated Using ONIOM(B3LYP/6-31+G(d):AMBER) with Different Model Partitions^{a,d}

assignment	expt. ζ^{18}	HBDI ζ^b	Model 1 ζ		Model 2 ζ		Model 3 ζ		scaled Model 3 ζ	
			ME	EE	ME	EE	ME	EE	ME	EE
C=O	62/64 ^c	56	60	64	30	36	79	63	81	74 (3 σ = 8)
C=C	13	4	9	21	51	15	51	28 (2 σ = 19)	50	39 (2 σ = 35)
Phenol 1	13	5	15	9 (2 σ = 3)	8 (2 σ = 7)	21 (2 σ = 4)	16	14	18	17
Phenol 2	22/32 ^c	19	40	30 (3 σ = 21)	33 (2 σ = 19)	13 (2 σ = 2)	32	24	28	21 (2 σ = 4)
C=N	60	80	42	44	75 (3 σ = 16)	44 (2 σ = 16)	27 (2 σ = 21)	11	6	15
Phenol 3	46	4	10	14	47 (2 σ = 23)	19 (2 σ = 4)	3	14	10	32 (2 σ = 4)
Phenol 7	35/19 ^c	7	10	25	4	8	2	6	3	26 (2 σ = 1)

^aMode angles of 1550–1800 cm⁻¹ modes in the GFP₁₂ spectrum are reported in Table S3. ^bHBDI VTMA_s differ from that of ref 18, as these were computed using extra polarization and diffuse functions within an implicit water solvent model, whereas we have computed vacuum B3LYP/6-31+G(d) VTMA_s in order to compare directly with ONIOM calculations. ^cIn ref 18 two values for the A structure are given, corresponding to A₁ and A₂. ^dValues in brackets indicate the number of modes that match the assignment and the average deviation (σ).

Table 5. Frequencies of Chromophore Vibrational Modes of HBDI Calculated Using B3LYP/6-31+G(d) and GFP_A Calculated Using ONIOM(B3LYP/6-31+G(d):AMBER) with ME and EE and Different ONIOM Partition Sizes, Compared with Experiment^a

assignment	expt. freq (cm ⁻¹) ¹⁸	HBDI freq (cm ⁻¹)	Model 1 freq (cm ⁻¹)		Model 2 freq (cm ⁻¹)		Model 3 freq (cm ⁻¹)		scaled Model 3 freq (cm ⁻¹)	
			ME	EE	ME	EE	ME	EE	ME	EE
C=O	1681/1678 ^b	1771	1789	1777	1742	1761	1699	1721	1663	1682 (3 σ = 0.70)
C=C	1646/1643 ^b	1696	1697	1689	1685	1688	1676	1680 (2 σ = 0.18)	1640	1642 (2 σ = 0.32)
Phenol 1	1596	1654	1654	1649 (2 σ = 0.02)	1645 (2 σ = 1.75)	1645 (2 σ = 1.30)	1641	1647	1606	1612
Phenol 2	1572	1616	1630	1632 (3 σ = 0.93)	1615 (2 σ = 0.02)	1609 (2 σ = 0.10)	1612	1606	1577	1582 (2 σ = 0.08)
C=N	1553	1614	1585	1604	1595 (3 σ = 4.81)	1599 (2 σ = 0.03)	1588 (2 σ = 1.11)	1601	1555	1565
Phenol 3	1512/1496 ^b	1554	1557	1552	1561 (2 σ = 0.05)	1561 (2 σ = 0.09)	1555	1557	1522	1527 (2 σ = 0.04)
Phenol 7	1275	1300	1307	1298	1330	1315	1332	1333	1304	1305 (2 σ = 0.14)

^aMode angles of 1550–1800 cm⁻¹ modes in the GFP₁₂ spectrum are reported in Table S3. ^bIn ref 18 two values for the A structure is given, corresponding to A₁ and A₂.

vector that defines the Stark tuning rate lies in any particular direction and ii) that the electrostatic field at any one particular point in space defines the magnitude of E. Explicit calculation of the frequency shift can be used as a new tool for analysis of experimental frequency shifts of local dielectric probes.

In this section we highlighted how mode delocalization and coupling complicated correlation of electrostatic field and frequency shift. We now seek to understand and examine the nature of this coupling and how it modifies the character of the normal coordinate by examining changes in mode angles (VTMA_s) through a series of ONIOM partitions in the GFP_A state (Figure 1).

4C. GFP_A Chromophore Vibrational Modes in the Protein Environment. In this section we examine how mode angle (VTMA) is affected by how the interaction between low level and high level (vacuum vs ME vs EE) is computed and model size (Figure 1). The mode angle is computed from the inner product of the vibrational transition dipole (μ_2) and the S_1/S_0 electronic transition dipole (μ_1). This can be compared to experimentally resolved VTMA_s of GFP. The VTMA_s indicate the direction of the vibrational transition dipole and so are highly sensitive to mode composition. The effect of the environment on the chromophore modes can thus be

understood by examining changes in VTMA_s of chromophore modes.

VTMA_s have been reported for several smaller systems in the literature that we can use to benchmark the extent to which experiment and theory agree. Usman et al.⁸² compared theoretical and experimental VTMA_s of the C=O mode in various charged states of ground state HBDI using B3LYP/6-31+G(d,p) for the vibrational transition dipole moments and TD-B3LYP/6-31+G(d,p) for the electronic transition dipole. Discrepancies of around 20 degrees were observed for all three systems. Zhang et al.⁸³ analyzed VTMA_s of the symmetric and antisymmetric C≡N stretches of *para*-julolidine malononitrile in several electronic states with CAM-B3LYP/6-31G(d,p). Two calculated VTMA_s gave agreement within two degrees of experiment, while another two modes were in error by 10 and 15 degrees. Luber et al.⁸⁴ studied VTMA_s of ground state 2-(2'-hydroxyphenyl)benzothiazole. Tetrachloroethene solvent effects were included in the calculations using a continuum solvent model. Vibrational transition dipole moments were computed using BP86/TZVP, while TD-B3LYP/TZVP was used to obtain the S_0/S_1 electronic transition dipole. In this case, of the four modes presented three gave agreement within

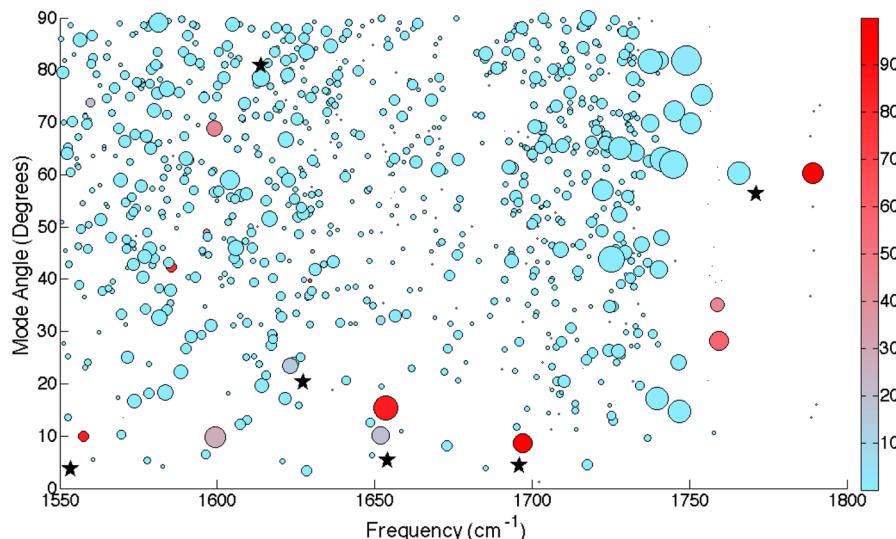


Figure 4. Plot of mode angles (VTMAs) against frequency for computed normal modes in the range 1550–1800 cm^{-1} computed using GFP Model 1 ONIOM-ME(B3LYP/6-31+G(d):AMBER). Size of point indicates intensity and color indicates percentage of mode within model region. Black stars show vacuum HBDI B3LYP/6-31+G(d) modes.

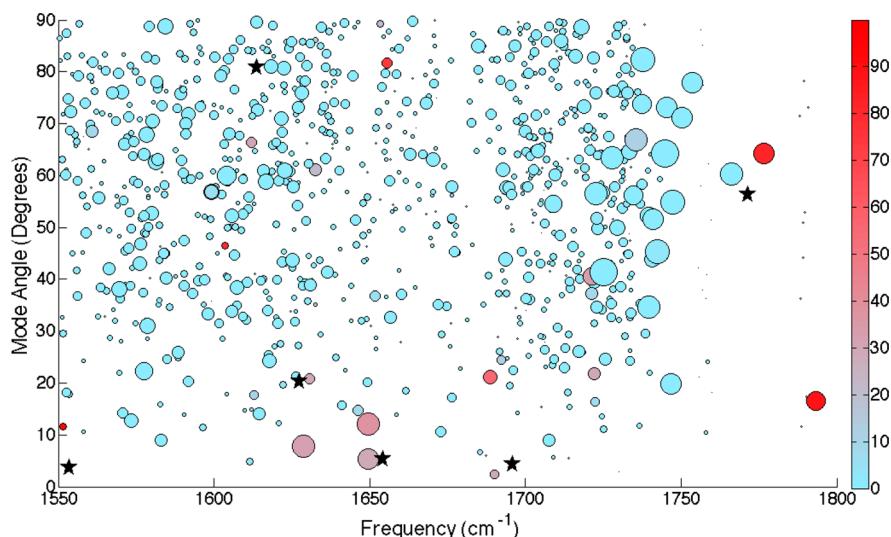


Figure 5. Plot of mode angles (VTMAs) against frequency for computed normal modes in the range 1550–1800 cm^{-1} computed using GFP Model 1 ONIOM-EE(B3LYP/6-31+G(d):AMBER). Size of point indicates intensity and color indicates percentage of mode within model region. Black stars show vacuum HBDI B3LYP/6-31+G(d) modes.

2 degrees, while the remaining mode was over twenty degrees away.

Table 4 shows the calculated mode angles/VTMAs for chromophore modes identified from vacuum HBDI frequency calculations to be in the 1550–1800 cm^{-1} range⁵⁶ (For associated normal coordinates refer to Esposito et al.⁵⁶). Table 5 shows the frequencies in cm^{-1} for the same set of modes.

The Model 1 partition (Figure 1) represents the addition of the protein environment entirely at the low level to the chromophore residue. Figure 4 indicates how the mechanically coupled environment shifts the frequencies and VTMAs of chromophore modes. The C=O, C=C, Phenol 1, and Phenol 3 modes show a small VDMA change (<10 degrees). However, Phenol 1 and Phenol 3 show a second mode that is much more coupled to the environment with the same frequency as the localized mode. While both Phenol 1 modes have similar VDMA, the Phenol 3 VMAs are very different (10° vs 74°). In

Table 4 we report data for the localized mode. The frequency of the C=O mode changes by 18 cm^{-1} (Table 5) due to hydrogen bonding from the Arg96 residue, which increases the stiffness of this vibration.

Table 4 shows that the only VTMAs that are significantly affected by ME in the M1 calculation are the Phenol 2 and C=N modes, which change by 27° and 38°, respectively, from the vacuum HBDI calculation. Figure 4 shows that as with C=C and Phenol 3, there is a localized Phenol 2 mode at 40° and a less local (24%) mode at 46°. The delocalization into the low level is due to coupling with a series of amide I modes that greatly affect the VDMA. The C=N mode is even more complicated as at 1599 cm^{-1} there are two modes which are 28% and 43% within the model region, while at 1585 cm^{-1} there is a less intense but more localized C=N mode (77%). The two higher frequency modes result from amide II vibrations in the backbone to which the chromophore is

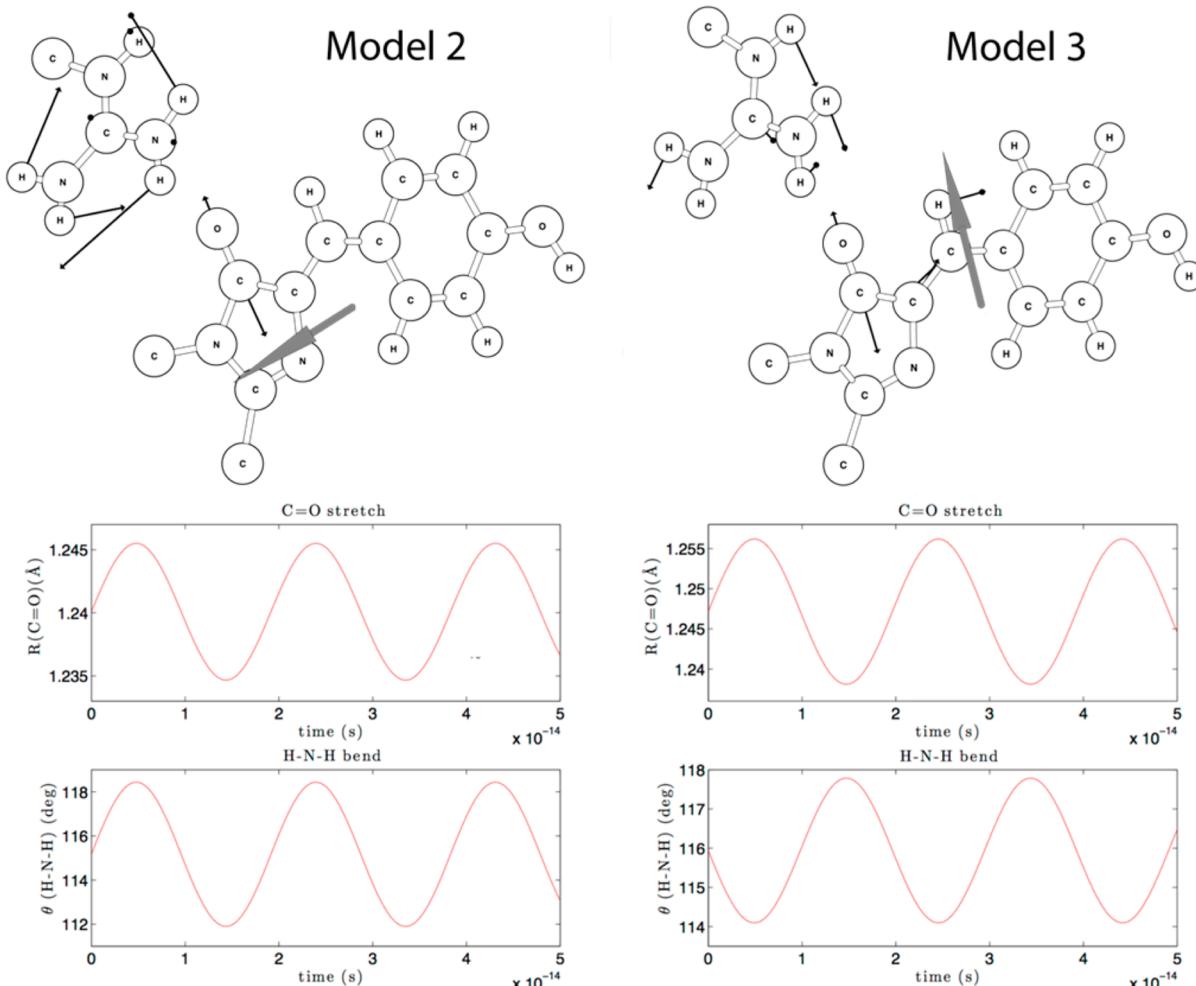


Figure 6. Normal mode vectors (black) of the C=O mode calculated using ME ONIOM(B3LYP/6-31+G(d):AMBER) with Model 2 (left panel) and Model 3 (right panel) partitions. The direction of the associated dipole derivative vector (gray) is also plotted. The phases of the chromophore C=O bond stretch and neighboring Arg96 H–N–H bend in each mode are plotted below with displacement amplitudes scaled to $0.1 k_B T$. This depends on the partition, giving an in-phase oscillation for the Model 2 C=O mode and an out-of-phase oscillation for the Model 3 C=O mode. The change of phase has a large effect on calculated VTMA: 30° in Model 2 and 79° in Model 3.

attached, and so we report the lower frequency local mode in Table 4. This has a smaller VTMA than in HBDI as it is even more localized to the C=N bond. The resulting transition dipole moment is more aligned with the C=N bond and the associated VTMA decreases as the C=N bond happens to be almost parallel with the electronic transition dipole moment. Other modes that show model character are due to the amide I and amide II modes of the backbone attached to the chromophore.

The effect of electronic embedding is indicated in Figure 5. The C=C mode is less localized (55%) as it couples to NH₂ bends in Arg96 which causes the appearance of a second even more delocalized (29%) C=C mode. Phenol 1 character is still present in two modes although the level of delocalization is similar in each (39% and 29%) such that it was not possible to single out a particular mode and the average ζ and average distribution is given (Table 4). Similarly, three computed modes were given the Phenol 2 assignment, as it was not possible to single out a particular mode. These had frequencies of 1632.46, 1630.77, and 1628.78 cm⁻¹ (Table 5) with varying degrees of environment delocalization (21%, 29%, and 31% of vector magnitude in the model region, respectively) and as a result gave rise to VTMA of 61, 21, and 8 degrees, depending on the

low-level residues to which they coupled. The coupling to the environment also greatly increased the intensity of modes of this character. The Phenol 3, C=N, and C=O modes show a small change in VTMA (Table 4), but the C=O and C=N mode change in frequency significantly more than the other modes (Table 5).

The Phenol 2 mode in the Model 1 ONIOM-EE calculation demonstrates that the protein environment can split a single vacuum mode into several protein modes. These modes may all have significantly different VTMA. To test this, we froze part of the protein backbone to which coupling was observed in Phenol 2 (backbone amide II), particularly in the mode at 1630.77 cm⁻¹. This resulted in three modes with similar frequencies (shown as unfrozen→frozen: 1632.46→1632.37 cm⁻¹, 1630.77→1630.46 cm⁻¹, and 1628.78→1628.75 cm⁻¹) such that the average frequency was 1630.53 cm⁻¹, $\Delta\tilde{\nu} = -0.14$ cm⁻¹. Although all modes were more localized to the model region (21→46%, 29→64%, 31→33%), the 1630.46 cm⁻¹ mode was significantly more localized than the corresponding mode in the unconstrained calculation (1630.77 cm⁻¹). This suggests that there are two environment residues with quasi-degenerate resonant frequencies to the Phenol 2 mode that couple together to produce 3 modes. By freezing part of one of

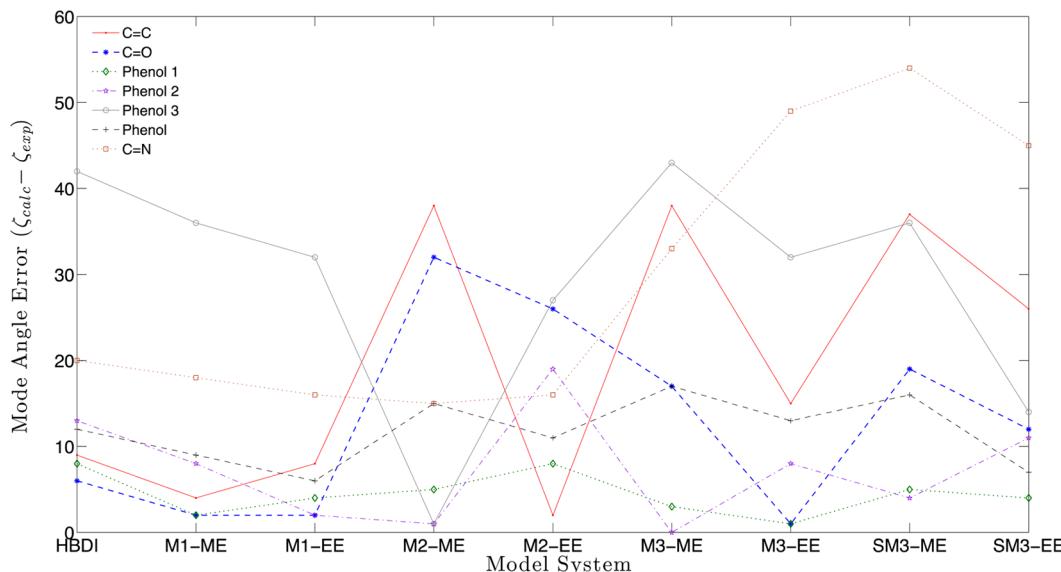


Figure 7. Absolute error in average VTMA of selected chromophore modes for HBDI calculated using B3LYP/6-31+G(d) and GFP calculated using ONIOM(B3LYP/6-31+G(d):AMBER) with three different partitions (M1, M2, M3), each of which treated using mechanical embedding (ME) and electronic embedding (EE). Results of the largest partition calculated using scaled Hessian components (eq 14) is also shown (SM3).

these residues we have increased the contributions of the other two residues to the vibrational mode, and this has resulted in the increased localization to the model. The associated VTMA were very different ($61 \rightarrow 40.4^\circ$, $21 \rightarrow 71.9^\circ$, and $8 \rightarrow 89.2^\circ$) with an average of 67.2° , $\Delta\zeta = +37.3^\circ$ indicating the sensitivity of the VTMA to environment coupling.

The interlayer coupling of normal modes is a concern for a number of reasons: QM/MM methodologies assume that the property of interest is localized well within the model and that the environment is a small perturbation of the model region; a mechanical force-field may not model coupling between layers correctly; and oscillators from different levels of theory that are degenerate in the harmonic approximation may not be if higher order terms are included. In order to address these concerns, extra environment residues were incorporated into the model region (Figure 1).

The VTMA of the C=O mode was found to be very sensitive to the partition size (Table 4). This is caused by the way in which the C=O oscillator interacts with the NH₂ bending displacements in the Arg96 residue. The phase of this oscillation relative to the C=O stretch has a large effect on the dipole derivative, as Arg96 lies perpendicular to the electronic dipole moment (Figure 6). This coupling occurs through hydrogen bonds and demonstrates that the VTMA is very sensitive to this coupling. The C=C mode (Table 4) shows a similar phase difference, but in this case between ME and EE calculations, showing that the way in which the protein environment is modeled can have a large effect. We believe these are a result of small geometry differences between calculations and indicates that statistical averaging from several snapshots is required to reproduce experiment.

The Phenol 1 and Phenol 7 modes were much less sensitive to partition size (Table 4) as they were highly localized to the model region for all model systems. Coupling through hydrogen bonds of Phenol 1 to Arg96 and His148 was found to be dependent on the hydrogen-bonding distance. For example, two modes were assigned to Phenol 1 Model 2 ME calculations, which corresponded to the two phases relative to the coupled Arg96 oscillation. By contrast in the Model 2 EE

calculation, the coupling to His148 was much stronger as the hydrogen bond to this residue was 0.09 \AA shorter while the hydrogen bond to Arg96 increased by 0.19 \AA . The Phenol 7 mode tended to couple with the His148 due to the hydrogen bonding interaction with the frequency being determined by the C–O bond length (Figure S2).

The Phenol 2, C≡N, and Phenol 3 modes were more delocalized into the low-level region, as is indicated by the large σ values in Table 4. This is because, as with the Model 1 Phenol 2 mode, the region of the spectrum in which these modes are found is particularly congested, with extensively delocalized amide backbone modes. An example of the environment's role is seen in the Model 2 Phenol 3 mode (Table 4). In both the ME and EE calculations, two normal modes are given this assignment. Although the environment residue to which the oscillation couples is completely different for each embedding scheme, in both cases the two oscillations have opposite phases. Depending on the position of the environment residue with respect to the electronic transition dipole, the in- and out-of-phase modes may have similar mode angle, such as in the EE case, or very different mode angles, as in the ME case.

Using a scaled Hessian, in which c_{low} was set to 1.000^{47} and c_{high} was set to 0.9784^{65} results in a new set of frequencies, as well as modified normal modes in cases of delocalization into the low level (Table 4). In the majority of cases, the VTMA in the Model 3 scaled Hessian calculation are similar to those computed in the unscaled Model 3 calculation, changing by up to 10° . The EE Phenol 3 and Phenol 7 are the only two modes that give closer correspondence to experiment when Hessian scaling is used. This is caused by slightly greater low-level contributions to the vibrational mode which couple to give an in-phase and an out-of phase mode, the average of which is closer to experiment.

Figure 7 shows the error between the calculated VTMA and the experimental values according to Table 4. All modes in HBDI were within the 20° error indicated by benchmarks on small systems except for Phenol 3. Adding the mechanical and electrostatic effects from the environment causes convergence of the VTMA toward the experimental value, with the Phenol

3 error decreasing by 10° . Expansion of the model region to the Model 2 and Model 3 partitions does not improve the VTMA; in fact the VTMA are sensitive to partition size. This is because a slight change in mode frequency changes the environment residue to which it is coupled, and the VTMA is sensitive to this.

We comment now on the nature of the long-range coupling observed in the mode angle analysis. In the dipole–dipole approximation, the coupling constant between two coupled oscillators scales as $1/r^3$ and depends on the orientations of the transition dipole vectors. If this were as a result of transition-dipole or through-bond coupling, then it would be expected that the coupled residue, and thus the VTMA, would be the same. However, it appears that the coupling is due to the presence of oscillators of the same resonant frequency in the environment, and so small changes in the fundamental frequency could significantly alter the VTMA. A coupled protein environment introduces the full 3N-6 dimensionality to the vibrational problem, resulting in a high density of oscillators that display accidental near-degeneracy in their fundamental frequencies. This resonance effect results in modes of mixed character, complicating the identification and assignment of a particular normal mode. The Phenol 2 and C=N modes seem to be particularly susceptible to this effect as their fundamental frequencies lie in a particularly crowded region of the spectrum. A result of this coupling is that transition dipoles for nominal chromophore modes have significant components normal to the chromophore plane, and so the assumption that the transition dipoles lie in the plane of the chromophore may not be valid.

Our initial hypothesis was that we could assign a single computed chromophore vibrational mode to those measured experimentally for a protein, and that by including the chromophore's environment with ONIOM we would achieve better agreement between computation and experiment. We have seen that coupling between the chromophore and its environment often generates several modes for the protein matching one for the isolated chromophore: there is no simple one to one mapping between chromophore only and chromophore in protein vibrational modes, because of mechanical and electronic coupling, and near degeneracy mixing. For matching computed mode angles with experiment, this proves to be a particular problem for two reasons: first, it is not clear if we should be averaging over the coupled modes or comparing with the most localized – if these modes really are degenerate then the normal modes can be arbitrarily mixed which would modify the VTMA of a single mode but the average would be invariant; second, the residue to which the chromophore couples changes between calculations, suggesting that VTMA calculations are highly sensitive to the protein conformation.

5. CONCLUSIONS

A systematic comparison of different environmental effects on the chromophore vibrational modes using the ONIOM method has allowed us to model how spectroscopic transitions are modified in the Green Fluorescent Protein. ONIOM-(QM:MM) reduces the expense of normal mode calculations by computing most of the second derivatives at the MM level. New developments for the efficient solution of the CPHF equations, including contributions from electrostatic interactions with environment charges, have allowed large QM blocks to be embedded within the MM environment. This has allowed

computation of the 3N-6 vibrational frequencies of the entire GFP protein for three different QM model sizes, both with and without including contributions arising from electronic embedding of the QM region. The frequencies and dipole derivative vectors obtained from this computational analysis have been used to interpret experimental results that derive from vibrational spectroscopy of GFP.

Using computationally constructed difference spectra, we have shown that the experimental spectrum can be reasonably reproduced even within the harmonic approximation. EE is important to describe some modes, such as the frequency of the C=O mode. Through the comparison of computed I_2 -A difference spectra with the experimental spectra, assignments have been made which support those determined using isotopic labeling and vacuum HBDI normal mode calculations.

The change in chromophore mode frequency between mechanically and electronically embedded calculations allows us to compute the vibrational Stark shift explicitly. This can be used as a new tool for analysis of experimental frequency shifts of local dielectric probes. Mode delocalization and coupling complicates the correlation of electrostatic field and frequency shift.

Analyzing mode angles (VTMAs) showed that chromophore vibrational modes respond to both structural and electronic perturbations produced by the protein environment. The coupling between the chromophore and its environment often generates several modes for the protein matching one for the isolated chromophore. For matching computed mode angles with experiment, this proves to be a particular problem, because each of the full system modes matching a particular chromophore vibration can have a very different mode angle. This is because the direction of the vibrational transition dipole vector (μ_2) is highly sensitive to the extent to which the environment couples. In other words, a single computed vibrational mode and its mode angle cannot be compared directly with experiment. This is due to inherent mode mixing, rather than inaccuracy of the calculations. In other words, the sensitivity of these mode angles to geometry indicates that this property will have a high conformational dependence (in contrast to the difference spectrum) and that future attempts to match with experiment may need to take this into account by averaging over snapshots obtained from MD simulations.

■ ASSOCIATED CONTENT

S Supporting Information

Methodology for constructing electrostatic potential maps, additional figures referenced here and Cartesian coordinates and energies of vacuum HBDI in neutral and anionic forms and GFP A and I_2 structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: m.bearpark@imperial.ac.uk.

*E-mail: j.vanhor@imperial.ac.uk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.V.T. and M.J.B. acknowledge support from EPSRC via grant agreement EP/I003304/1. J.V.T. also acknowledges support from the European Research Council via Grant Agreement No.

208650 and BBSRC via grant agreement BB/H004238/1. J.T.S. and P.M.C. acknowledge financial support from the National Science Foundation (CHE-1026369). We thank Dr. Peng Tao for helpful discussions concerning ONIOM and Mr. Craig Young and Mr. Andy Forester for their early contributions to this work. We also thank Dr. David Mendive Tapia, Mr. Clyde Fare, and Dr. Magdalena Bak-Maier for helpful discussions. All calculations were carried out using the Imperial College High Performance Computing Service.

NOMENCLATURE - ACRONYMS

HBDI, 4-hydroxybenzylidene-2,3-dimethylimidazolinone; GFP, Green Fluorescent Protein; ONIOM, our own n-layered integrated molecular orbital and molecular mechanics; QM, quantum mechanics; MM, molecular mechanics; ME, mechanical embedding; EE, electronic embedding; μ_1 , S_1/S_0 electronic transition dipole moment; μ_2 , vibrational transition dipole moment; VTMA, vibrational transition mode angle between μ_1 and μ_2 ; E, electric field vector; $\Delta\mu$, Stark tuning vector; M1, Model 1 (Figure 1); M2, Model 2 (Figure 1); M3, Model 3 (Figure 1); GFP_A, neutral ground state of GFP; GFP₁₂, anionic ground state of GFP; PES, Potential Energy Surface; CPHF, Coupled Perturbed Hartree-Fock; SCF, Self Consistent Field; $\langle M \rangle$, trace of matrix M; M^x , derivative of M with respect to perturbation x; G(M), contraction of matrix M with two electron integrals; P, density matrix; F, Fock matrix; S, overlap matrix; h, one-electron core Hamiltonian matrix; e , eigenvalue of Fock matrix; A, unperturbed terms; B^x, perturbed terms; F^(x), integral derivative terms of the derivative Fock matrix; W^x, (PFP)^x; U^x, orbital rotation matrix; S_{oo}^x, occupied-occupied block of the derivative overlap matrix (obtained from use of density matrix as projection operator S_{oo}^x = PS^xP)

REFERENCES

- (1) Remorino, A.; Korendovych, I. V.; Wu, Y.; DeGrado, W. F.; Hochstrasser, R. M. *Science* **2011**, 332, 1206–9.
- (2) Hamm, P. *J. Biol. Phys.* **2009**, 35, 17–30.
- (3) Bredenbeck, J.; Helbing, J.; Kolano, C.; Hamm, P. *ChemPhysChem* **2007**, 8, 1747–56.
- (4) Cervetto, V.; Hamm, P.; Helbing, J. *J. Phys. Chem. B* **2008**, 112, 8398–405.
- (5) Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 14190–6.
- (6) Garrett-Roe, S.; Hamm, P. *Acc. Chem. Res.* **2009**, 42, 1412–22.
- (7) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, 96, 2036–41.
- (8) Karunakaran, V.; Benabbas, A.; Youn, H.; Champion, P. M. *J. Am. Chem. Soc.* **2011**, 133, 18816–27.
- (9) Karunakaran, V.; Denisov, I.; Sligar, S. G.; Champion, P. M. *J. Phys. Chem. B* **2011**, 115, 5665–77.
- (10) Gruia, F.; Ionascu, D.; Kubo, M.; Ye, X.; Dawson, J.; Osborne, R. L.; Sligar, S. G.; Denisov, I.; Das, A.; Poulos, T. L.; Terner, J.; Champion, P. M. *Biochemistry* **2008**, 47, 5156–67.
- (11) Li, H.; Jensen, J. H. *Theor. Chem. Acc.* **2002**, 107, 211–219.
- (12) Nonella, M.; Mathias, G.; Eichinger, M.; Tavan, P. *J. Phys. Chem. B* **2003**, 107, 316–322.
- (13) Freindorf, M.; Shao, Y.; Brown, S. T.; Kong, J.; Furlani, T. R. *Chem. Phys. Lett.* **2006**, 419, 563–566.
- (14) Reiher, M.; Neugebauer, J. *J. Chem. Phys.* **2003**, 118, 1634.
- (15) Herrmann, C.; Neugebauer, J.; Reiher, M. *J. Comput. Chem.* **2008**, 29, 2460–70.
- (16) Bowman, J. M.; Zhang, X.; Brown, A. *J. Chem. Phys.* **2003**, 119, 646.
- (17) Thomas, M.; Brehm, M.; Fligg, R.; Vohringer, P.; Kirchner, B.; Vöhringer, P. *Phys. Chem. Chem. Phys.* **2013**, 15, 6608.
- (18) Van Thor, J. J.; Ronayne, K. L.; Towrie, M.; Sage, J. T. *Biophys. J.* **2008**, 95, 1902–12.
- (19) Luin, S.; Voliani, V.; Lanza, G.; Bizzarri, R.; Amat, P.; Tozzini, V.; Serresi, M.; Beltram, F. *J. Am. Chem. Soc.* **2009**, 131, 96–103.
- (20) Chattopadhyay, A.; Boxer, S. G. *J. Am. Chem. Soc.* **1995**, 117, 1449–1450.
- (21) Suydam, I. T.; Boxer, S. G. *Biochemistry* **2003**, 42, 12050–5.
- (22) Hush, N. S.; Reimers, J. R. *J. Phys. Chem.* **1995**, 99, 15798–15805.
- (23) Reimers, J. R.; Hush, N. S. *J. Phys. Chem. A* **1999**, 103, 10580–10587.
- (24) Rosell, F. I.; Boxer, S. G. *Biochemistry* **2003**, 42, 177–83.
- (25) Olsen, S. *J. Chem. Theory Comput.* **2010**, 6, 1089–1103.
- (26) Olsen, S.; McKenzie, R. H. *J. Chem. Phys.* **2011**, 134, 114520.
- (27) Drobizhev, M.; Hughes, T. E.; Stepanenko, Y.; Wnuk, P.; O'Donnell, K.; Scott, J. N.; Callis, P. R.; Mikhaylov, A.; Dokken, L.; Rebane, A. *Sci. Rep.* **2012**, 2, 688.
- (28) Bublitz, G. Ü.; Ortiz, R.; Marder, S. R.; Boxer, S. G. *J. Am. Chem. Soc.* **1997**, 119, 3365–3376.
- (29) Tresewan, W.; Hirao, H.; Morokuma, K.; Hannongbua, S. *J. Mol. Model.* **2012**, 18, 2227–40.
- (30) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, 105, 1959.
- (31) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, 100, 19357–19363.
- (32) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *J. Mol. Struct.: THEOCHEM* **1999**, 461–462, 1–21.
- (33) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, 21, 1419–1432.
- (34) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, 24, 760–9.
- (35) Montgomery, J. A., Jr.; Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Morokuma, K.; Frisch, M. J.; Komaromi, I. *J. Chem. Theory Comput.* **2006**, 2, 815–826.
- (36) Cui, Q.; Karplus, M. *J. Chem. Phys.* **2000**, 112, 1133.
- (37) Frisch, M.; Head-Gordon, M.; Pople, J. *Chem. Phys.* **1990**, 141, 189–196.
- (38) Izmaylov, A. F.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2006**, 125, 104103.
- (39) Stratmann, R. E.; Burant, J. C.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1997**, 106, 10175.
- (40) Burant, J. C.; Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Chem. Phys. Lett.* **1996**, 258, 45–52.
- (41) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *Chem. Phys. Lett.* **1996**, 257, 213–223.
- (42) Strain, M.; Scuseria, G.; Frisch, M. *Science (80-)* **1996**, 271, 51–53.
- (43) Osamura, Y. *J. Chem. Phys.* **1982**, 77, 383.
- (44) Van Thor, J. J.; Georgiev, G. Y.; Towrie, M.; Sage, J. T. *J. Biol. Chem.* **2005**, 280, 33652–9.
- (45) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, 35, W522–5.
- (46) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *J. Chem. Theory Comput.* **2011**, 7, 525–537.
- (47) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, 117, 5179–5197.
- (48) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; 2010, University of California, San Francisco.
- (49) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, 25, 1157–1174.

- (50) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (51) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821–39.
- (52) 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.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, D. J.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Parandekar, P. V.; Mayhall, N. J.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian Development Version*, Revision H.13; Gaussian, Inc.: Wallingford, CT, 2010.
- (53) Altoè, P.; Cembran, A.; Olivucci, M.; Garavelli, M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20172–7.
- (54) Andruniów, T.; Olivucci, M. *J. Chem. Theory Comput.* **2009**, *5*, 3096–3104.
- (55) Helms, V.; Straatsma, T. P.; McCammon, J. A. *J. Phys. Chem. B* **1999**, *103*, 3263–3269.
- (56) Esposito, A. P.; Schellenberg, P.; Parson, W. W.; Reid, P. J. *J. Mol. Struct.* **2001**, *569*, 25–41.
- (57) Clemente, F. R.; Vreven, T.; Frisch, M. J. In *Quantum Biochemistry*; Matta, C. F., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 61–83.
- (58) Kennis, J. T. M.; Larsen, D. S.; van Stokkum, I. H. M.; Vengris, M.; van Thor, J. J.; van Grondelle, R. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17988–93.
- (59) Barabanschikov, A.; Demidov, A.; Kubo, M.; Champion, P. M.; Sage, J. T.; Zhao, J.; Sturhahn, W.; Alp, E. E. *J. Chem. Phys.* **2011**, *135*, 015101.
- (60) Ansbacher, T.; Srivastava, H. K.; Stein, T.; Baer, R.; Merkx, M.; Shurki, A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 4109–17.
- (61) Vreven, T.; Morokuma, K. *J. Chem. Phys.* **2000**, *113*, 2969–2975.
- (62) Vreven, T.; Morokuma, K. *Theor. Chem. Acc.* **2003**, *109*, 125–132.
- (63) Bearpark, M. J.; Larkin, S. M.; Vreven, T. *J. Phys. Chem. A* **2008**, *112*, 7286–95.
- (64) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502–16513.
- (65) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (66) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–41.
- (67) Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8362–7.
- (68) Van Thor, J. J. *Springer Ser. Fluoresc.* **2012**, *11*, 183–216.
- (69) Van Thor, J. J. *Chem. Soc. Rev.* **2009**, *38*, 2935–50.
- (70) Van Thor, J. J.; Zanetti, G.; Ronayne, K. L.; Towrie, M. *J. Phys. Chem. B* **2005**, *109*, 16099–108.
- (71) He, X.; Bell, A. F.; Tonge, P. J. *J. Phys. Chem. B* **2002**, *106*, 6056–6066.
- (72) Boxer, S. G. *J. Phys. Chem. B* **2009**, *113*, 2972–2983.
- (73) Brewer, S. H.; Franzen, S. *J. Chem. Phys.* **2003**, *119*, 851.
- (74) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science* **2006**, *313*, 200–4.
- (75) Sandberg, D. J.; Rudnitskaya, A. N.; Gascón, J. A. *J. Chem. Theory Comput.* **2012**, *8*, 2817–2823.
- (76) Webb, L. J.; Boxer, S. G. *Biochemistry* **2008**, *47*, 1588–98.
- (77) Bazewicz, C. G.; Lipkin, J. S.; Smith, E. E.; Liskov, M. T.; Brewer, S. H. *J. Phys. Chem. B* **2012**, *116*, 10824–10831.
- (78) Wang, X.; He, X.; Zhang, J. Z. *H. J. Phys. Chem. A* **2013**, *117*, 6015–6023.
- (79) Gascón, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 175–186.
- (80) Gready, J. E.; Hush, N. S. *Chem. Phys.* **1978**, *31*, 467–483.
- (81) Bauschlicher, C. W. *Chem. Phys. Lett.* **1985**, *118*, 307–310.
- (82) Usman, A.; Mohammed, O. F.; Nibbering, E. T. J.; Dong, J.; Solntsev, K. M.; Tolbert, L. M. *J. Am. Chem. Soc.* **2005**, *127*, 11214–5.
- (83) Zhang, W.; Lan, Z.; Sun, Z.; Gaffney, K. *J. Phys. Chem. B* **2012**, *116*, 11527–36.
- (84) Luber, S.; Adamczyk, K.; Nibbering, E. T. J.; Batista, V. S. *J. Phys. Chem. A* **2013**, *117*, 5269–79.