Probing the Ground State Structure of the Green Fluorescent Protein Chromophore Using Raman Spectroscopy†

Alasdair F. Bell,‡ Xiang He,‡ Rebekka M. Wachter,§ and Peter J. Tonge*,‡,||

Department of Chemistry, Graduate Program in Biophysics, and Graduate Program in Molecular and Cellular Biochemistry, SUNY at Stony Brook, Stony Brook, New York 11794-3400, and Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403

Received November 19, 1999; Revised Manuscript Received February 3, 2000

ABSTRACT: We present Raman spectra, obtained using 752 nm excitation, on wild-type GFP and the S65T mutant of this intrinsically fluorescent protein together with data on a model chromophore, ethyl 4-(4-hydroxyphenyl)methylidene-2-methyl-5-oxoimidazolacetate. In the pH range 1−14, the model compound has two macroscopic pK_as of 1.8 and 8.2 attributed to ionization of the imidazolinone ring nitrogen and the phenolic hydroxyl group, respectively. Comparison of the model chromophore with the chromophore in wild-type GFP and the S65T mutant reveals that the cationic form, with both the imidazolinone ring nitrogen and the phenolic oxygen protonated, is not present in these particular GFP proteins. Our results do not provide any evidence for the zwitterionic form of the chromophore, with the phenolic group deprotonated and the imidazolinone ring nitrogen protonated, being present in the GFP proteins. In addition, since the position of the Raman bands is a property exclusively of the ground state structure, the data enable us to investigate how protein−chromophore interactions affect the ground state structure of the chromophore without contributions from excited state effects. It is found that the ground state structure of the anionic form of the chromophore, which is most relevant to the fluorescent properties, is strongly dependent on the chromophore environment whereas the neutral form seems to be insensitive. A linear correlation between the absorption properties and the ground state structure is demonstrated by plotting the absorption maxima versus the wavenumber of a Raman band found in the range 1610−1655 cm^{-1}.

The green fluorescent protein (GFP)[1] from the jellyfish *Aequorea victoria* has found many applications in molecular biology because the chromophore is an integral part of the protein sequence. Consequently, in situ expression of GFP results in an intrinsically fluorescent protein that can be fused to another protein and used to report on protein expression and transport within cells (see refs *1* and *2* for recent reviews). Wild-type GFP autocatalytically generates the 4-hydroxybenzylidene-imidazolinone chromophore, depicted in Figure 1, via a posttranslational internal cyclization of the Ser^{65}-Tyr^{66}-Gly^{67} tripeptide followed by 1,2-dehydrogenation of the tyrosine (*3*, *4*). Clearly, given the range of biotechnological applications (*1*, *2*), there is a great deal of interest in understanding the absorption and emission spectra of GFPs with a view to the rational design of new mutants with specific spectroscopic properties. A number of such mutants have been found using random mutagenesis, but with the availability of high-resolution crystallographic data (*5*−*11*) combined with spectroscopic techniques, it should be possible to design new mutants with desirable properties.

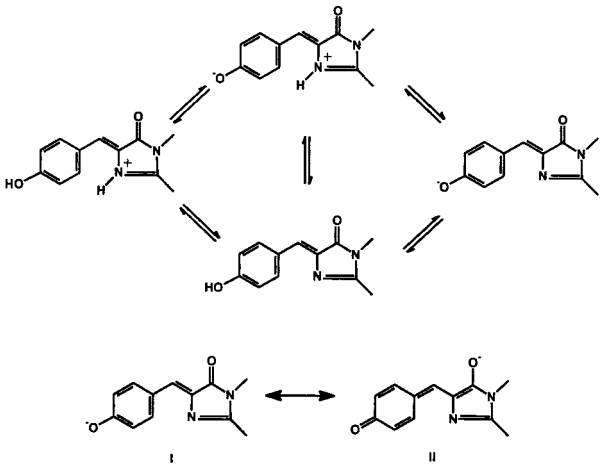

FIGURE 1: Structural formula of cationic, neutral, zwitterionic, and anionic forms of wild-type GFP chromophore and two possible resonance structures for the anionic form.

In principle, four protonation states are available to the GFP chromophores as shown in Figure 1. The absorption spectrum of wild-type GFP contains two main bands at about

† This work was supported by NSF Grant MCB9604254, by U.S. Army Research Office Grant DAAG559710083 to P.J.T., and by NIH Grant AI44639. The NMR facility at SUNY Stony Brook is supported by grants from the NSF (CHE9413510) and from the NIH (1S10RR554701).

* To whom correspondence should be addressed. Telephone: (631) 632 7907. Fax: (631) 632 7960. Email: Peter.Tonge@sunysb.edu.

‡ Department of Chemistry, SUNY at Stony Brook.

§ Institute of Molecular Biology and Department of Physics, University of Oregon.

|| Graduate Program in Biophysics and Graduate Program in Molecular and Cellular Biochemistry, SUNY at Stony Brook.

[1] Abbreviations: GFP, green fluorescent protein; BFP, blue fluorescent protein; YFP, yellow fluorescent protein; 4-OH-Cin-CoA, 4-hydroxycinnamoyl-CoA; MCAD, medium chain acyl-CoA dehydrogenase; DMSO, dimethyl sulfoxide.

10.1021/bi992675o CCC: $19.00
Published on Web 03/24/2000

395 and 475 nm, and these are usually attributed to the neutral and anionic forms of the chromophore, respectively (*12*, *13*). Recently, another form of the chromophore, named the I-form, has been shown by hole-burning spectroscopy to contribute to the absorption spectrum of wild-type GFP as a broad wing to the red of the 475 nm peak (*14*). The zwitterionic form of the chromophore, with the ring nitrogen protonated and the phenolic oxygen deprotonated, has also been proposed to be important on the basis of quantum chemical calculations (*15*−*17*), but no experimental evidence has yet been found for its existence (*11*). The cationic form with both the ring nitrogen and the phenolic oxygen protonated is expected to exist only at pH values well below the physiological range. Exposure to radiation at either 395 or 475 nm causes wild-type GFP to fluoresce at around 500 nm, explaining its characteristic green color. Since the pK_a for the phenolic oxygen of the chromophore is lowered in the excited state, deprotonation of the neutral form occurs prior to emission. Thus, the neutral and anionic forms have rather similar fluorescent spectra. A number of mutants of GFP have already been discovered which have a range of different absorbance and emission properties. For example, the S65T (*18*) and T203I (*19*) mutants have chromophores that exist predominantly in the anionic and neutral forms, respectively. In addition, blue fluorescent protein (BFP) (*8*, *19*) and yellow fluorescent protein (YFP) (*9*) mutants have been identified that exhibit blue- and red-shifted absorption and emission spectra relative to wild-type GFP, respectively. Recently, Matz et al. have cloned six new fluorescent proteins that are homologous to GFP, two of which exhibit even more red-shifted absorption and emission spectra than YFP (*20*).

Although the absorbance and fluorescence properties of wild-type GFP are relatively insensitive to changes in pH (*12*), the S65T mutant responds rapidly and reversibly to changes in solvent pH (*21*), with a chromophore pK_a of 6.0 (*11*). Crystal structures of the S65T protein at both high and low pH have been solved (*5*, *11*), and the changes in the arrangement of side chains close to the phenolic hydroxyl have been interpreted in terms of titration of this group. Thus, this particular mutant has been included in the present study to provide Raman data on the anionic form of the chromophore in a fully folded protein. In contrast, titration does not occur in the wild-type GFP without raising the pH to a value where protein denaturation occurs (*12*). Crystal structures on the wild-type protein seem to indicate that a complex hydrogen bonding network around the chromophore may be responsible for the lack of pH dependence of the chromophore in the physiological range (*6*, *7*).

Vibrational spectroscopy can provide new information on the ground state structure of this interesting chromophore (*22*). In the present study, we use Raman spectroscopy with 752 nm excitation to examine a model chromophore, wild-type GFP, and the S65T mutant. This choice of excitation wavelength allows selective intensity enhancement of vibrational bands originating in the chromophore over the rest of the protein modes without the problems sometimes associated with strictly on-resonance Raman experiments such as fluorescence, photoisomerization, or sample degradation. From a comparison of the Raman spectra of a model chromophore with wild-type GFP and the S65T mutant at various pH values, we assign Raman bands characteristic for the cationic, neutral, and anionic forms of the chromophores and investigate the role of the protein environment in modulating the spectral properties.

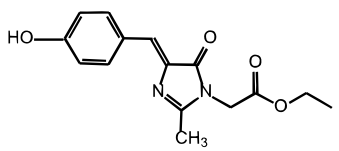

FIGURE 2: Structural formula of the model chromophore ethyl 4-(4-hydroxyphenyl)methylidene-2-methyl-5-oxoimidazolacetate.

MATERIALS AND METHODS

Protein Preparation. All GFP variants were expressed and purified as described (*5*). Samples for Raman spectroscopy were exchanged into the appropriate buffer by three dilution/reconcentration cycles. The absorption spectrum was checked in each case to determine whether neutral or anionic forms of the chromophore were present.

Synthesis of Model Chromophore. 4-*tert*-Butyldimethylsiloxybenzaldehyde (4-TBDMSOC_6H_4CHO) was synthesized from 4-hydroxybenzaldehyde by O-silyation with *tert*-butyldimethylchlorosilane (*23*). All other materials used were obtained from Aldrich and used without further purification. The model chromophore, ethyl 4-(4-hydroxyphenyl)methylidene-2-methyl-5-oxoimidazolacetate (Figure 2), was synthesized essentially following the method of Niwa et al. (*13*). ^{1}H NMR (300 MHz, CD_3OD) δ 8.02 (d, J = 8.4 Hz, 2H), 7.04 (s, 1H), 6.84 (d, 8.7 Hz, 2H), 4.49 (s, 2H), 4.24 (q, 7.2 Hz, 2H), 3.66 (s, 1H, OH), 2.33 (s, 3H), 1.29 (t, 7.2 Hz, 3H); MS (EI, 70 eV) *m*/*z* (relative intensity) 288 (M^+, 100), 146 (25), 100 (34), 55 (39); M^+ calculated for $C_{15}H_{16}N_2O_4$ 288.11.

Raman Spectroscopy. The Raman spectra presented here were acquired using an instrument based in part on that described recently by Dong et al. (*24*). The key feature of this instrument is that all the components have been optimized for operation in the near-IR to reduce problems associated with large fluorescence backgrounds that can plague Raman studies on biological systems.

A model 890 Ti:sapphire laser (Coherent, Santa Clara, CA), pumped by an Innova 308C argon ion laser (Coherent), provided 700 mW of near-IR excitation. For our system, 752 nm excitation provided the best balance between Raman intensity throughout the spectral region of interest and maximum obtainable laser power. It was necessary to place a holographic band-pass filter (Kaiser Optical Systems, Inc., Ann Arbor, MI) at the exit port of the Ti:sapphire laser to remove spurious lines that were found to interfere with the Raman spectra of protein samples which have strong Rayleigh scattering relative to the Raman scattering. The laser beam was focused with a 25 mm f/8.0 lens into the base of a rectangular quartz cell, and the Raman scattering was collected in a 90° geometry.

The Raman scattered light was collimated by a near-IR coated 85 mm f/1.4 camera lens (Kaiser Optical Systems, Inc.) before passing through a super notch plus holographic filter (Kaiser Optical Systems, Inc.), which effectively suppressed the Rayleigh scattered light. The collimated light was then focused onto the entrance slit of the spectrograph by a second near-IR coated 85 mm f/1.4 camera lens (Kaiser Optical Systems, Inc.). To maximize throughput, we used a

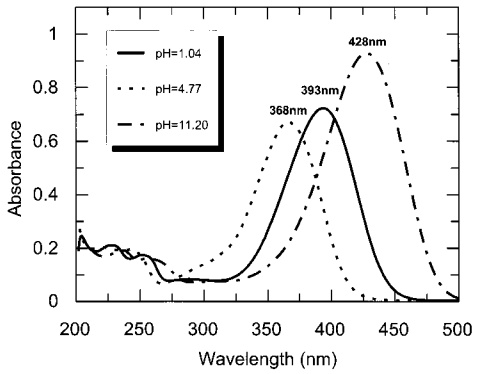

FIGURE 3: Characteristic absorption spectra for the cationic (1 M HCl), neutral (acetate buffer, pH 4.5, 20 mM), and anionic (5% 1 M NaOH) forms of the model chromophore.

single grating f/1.4 Kaiser Holospec to disperse the Raman scattered light. Finally, a red-sensitive, back-thinned CCD camera (Princeton Instruments, model no. EHRB1024) was used to detect the Raman scattered intensity. This detector has a peak quantum efficiency of about 80% and allowed us to measure the spectral region between about 400 and 2400 cm^{-1} in a single run.

The Raman measurements were made by adding approximately 60 μL of protein to a 2 mm by 2 mm cell and collecting data for 8 min. Subsequently, the same volume of buffer was placed in the same cell without making any changes to the optical alignment or to the cell position. Raman spectra were then calculated by performing a computer subtraction of the spectrum of the protein in buffer minus the spectrum of the buffer. The protein concentration was around 60 μM in each case. The difference spectra presented here were wavenumber-calibrated against cyclohexanone and are accurate to ± 2 cm^{-1}. All spectral manipulations were carried out using Win-IR software, and data acquisition was performed using WinSpec (Princeton Instruments, Trenton, NJ). Under the conditions used for acquiring good quality difference spectra on protein systems, the resolution of our system was approximately 8 cm^{-1}.

RESULTS AND DISCUSSION

Characterization of Model Chromophore. To provide a basis for the interpretation of the Raman spectra of GFP proteins, a model chromophore, ethyl 4-(4-hydroxyphenol)-methylidene-2-methyl-5-oxo-1-imidazolacetate (Figure 2), has been synthesized. There are four possible protonation states available to this model chromophore, and to chromophores within GFP proteins (Figure 1). In the pH range 1−14, the absorption spectra identify three distinct forms of the model chromophore corresponding to different protonation states as shown in Figure 3. These three forms have characteristic absorption maxima at 393, 368, and 428 nm and correspond to the cationic, neutral, and anionic forms of the model chromophore, respectively, as shown in Table 1. Although it is possible that the absorption spectrum of the neutral form, with a maximum at 368 nm, actually corresponds to the zwitterionic form (Figure 1), this is unlikely since to date no experimental evidence for protonation of the ring nitrogen has been presented (*11*). In addition, although quantum chemical calculations have predicted that the zwitterionic form of the chromophore may be important, they also predict that this form would exhibit an absorption spectrum that was significantly red-shifted to a value close to or greater than the anionic form (*15*−*17*, *25*). Clearly, this does not agree with our experimental results. Furthermore, the absorption and Raman spectra of the 4-methoxy derivative of the model chromophore (not shown here) are almost identical to those of the 4-hydroxy derivative, and a pK_a of 1.8 ± 0.1 was determined for the imidazolinone ring nitrogen. The fact that the zwitterionic form is not available to the 4-methoxy derivative and that the pK_a values do not change significantly when the phenolic group is substituted confirms that the zwitterionic form is not present in significant quantities in the model chromophore. Therefore, in the following discussion on the model chromophore, the zwitterion is not considered further. Of course, it is possible that the zwitterionic form is important in the GFP proteins since the Glu222 side chain is positioned close to the imidazolinone ring nitrogen and may donate a proton to the chromophore.

Table 1: Absorption and Emission Data for the Model Chromophore, Folded and Denatured Wild-Type GFP, and the S65T Mutant

	absorption (nm)	emission (nm)	pK_a
model chromophore	393, 368, and 428[a]	*b*	1.8 and 8.2
wild-type GFP	395 and 475	504	*c*
denatured wild-type GFP	384 and 448[d]	−	8.1[d]
S65T	394 and 489	511	6.0[e]

[a] Absorption maxima for cationic, neutral, and anionic forms of the chromophore, respectively. [b] The model chromophore is not fluorescent in aqueous solution at room temperature (*13*). [c] The chromophore in wild-type GFP is not titratable unless the protein is denatured. [d] Taken from ref *12*. [e] Taken from ref *11*.

The macroscopic pK_as describing ionization of the model chromophore are determined by absorption spectroscopy to be 1.8 ± 0.1 and 8.2 ± 0.1, respectively, as shown in Figure 4. For both pK_a determinations, the data were found to fit well to a single ionization. Based on the expectation that there is no zwitterion present, the pK_a at 1.8 is assigned to the imidazolinone ring nitrogen and the pK_a at 8.2 to the phenolic oxygen. The pK_a for the phenolic oxygen compares to a pK_a of 10.9 for the phenolic hydroxyl group of tyrosine, from which the chromophore in GFP is derived, which means that the extended delocalized π-electron skeleton in the model chromophore reduces the pK_a of this group by 2.7 units. The pK_a for the phenolic oxygen in the model chromophore is also very close to the value of 8.1 reported for the denatured wild-type protein (*12*). Interestingly, the absorption maximum of the cationic form of the model chromophore (393 nm) coincides with that of the neutral form of wild-type GFP (395 nm) which may indicate that this form is relevant to the GFP proteins.

An earlier study by Niwa et al. characterized the absorption spectra of the model chromophore as a function of solvent (*13*). This study revealed almost no solvent dependence for the absorption maxima of the neutral form of the chromophore; however, for the anionic form, absorption maxima were reported at 424, 435, 443, and 462 nm for water, ethanol, 2-propanol, and DMSO, respectively. Clearly, the dependence on the solvent does not correlate simply with the dielectric constant. However, these absorption data can be understood by considering the ability of the solvents to stabilize the negative charge on the phenolic oxygen in the

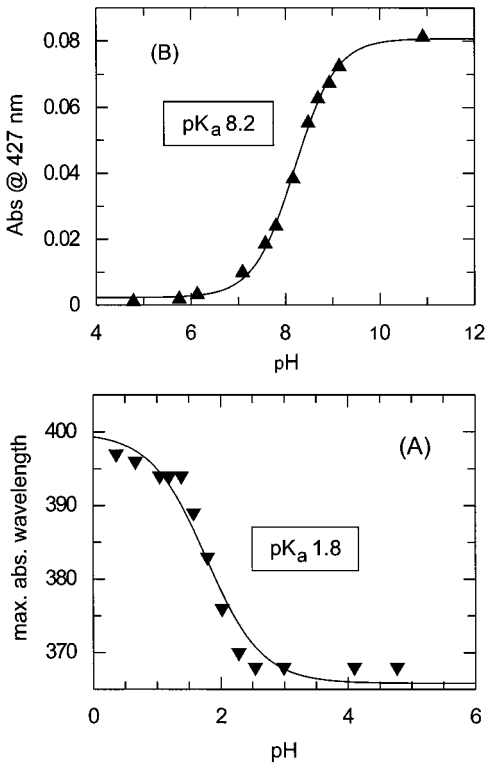

FIGURE 4: Determination of the macroscopic pK_a values for (A) the imidazolinone ring nitrogen and (B) the phenolic oxygen of the model chromophore.

anionic form of the chromophore. DMSO is known to have poor anion solvating properties (*26*), and this will promote structures where the negative charge on the phenolic oxygen is delocalized over the chromophoric skeleton with a concomitant red-shift in the absorption maximum. On the other hand, water and other protic solvents have good anion solvating properties (*26*) which will tend to pin the negative charge on the phenolic oxygen, resulting in reduced delocalization and a concomitant blue-shift in the absorption maximum.

The Raman spectra of the cationic, neutral, and anionic forms of the model chromophore in water are shown in Figure 5. In addition, we have examined the solvent dependence of the anionic form of the model chromophore by taking Raman spectra in water, methanol, 2-propanol, and DMSO as displayed in Figure 6. [The Raman spectra of the neutral form of the model chromophore were not found to display any solvent dependence (data not shown).] In each case, the solvent contained 5% v/v 1 M NaOH to ensure full ionization of the model chromophore. This set of data indicates that there are ground state structural changes as a function of solvent in the anionic form of the model chromophore but not in the neutral form.

The assignment of the Raman bands in the model chromophore that are sensitive to the protonation state of the phenolic hydroxyl and the imidazolinone ring nitrogen relies on attributing the absorption bands at 393, 368, and 428 nm to the cationic, neutral, and anionic forms, respectively, as discussed above. By this criteria, the Raman spectra of the model chromophore in 1 M HCl, in pH 4.5 acetate buffer, and in aqueous 1 M NaOH (5% v/v) (Figure 5) represent the cationic, neutral, and anionic forms of the chromophore, respectively. Clearly, although there are some

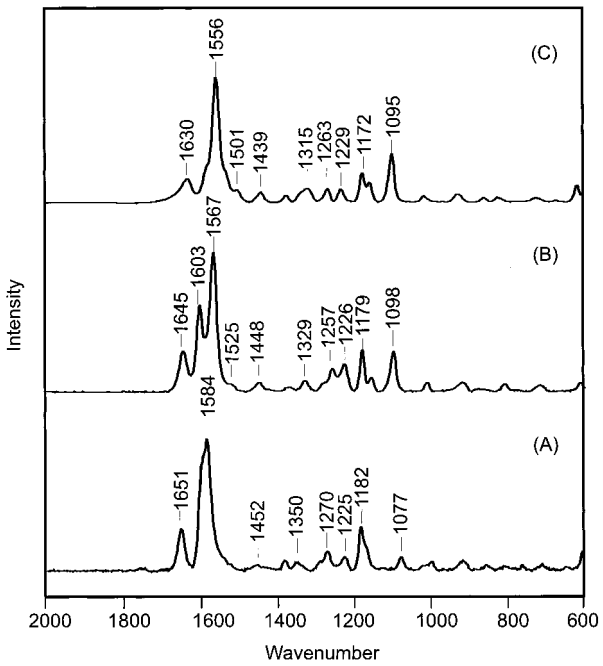

FIGURE 5: Raman spectra of the model chromophore in (A) 1 M HCl, (B) sodium acetate buffer (pH 4.5, 20 mM), and (C) 5% v/v 1 M NaOH aqueous solution. Solvent background has been subtracted. The concentrations were 350 μM, and the spectra were accumulated for 8 min.

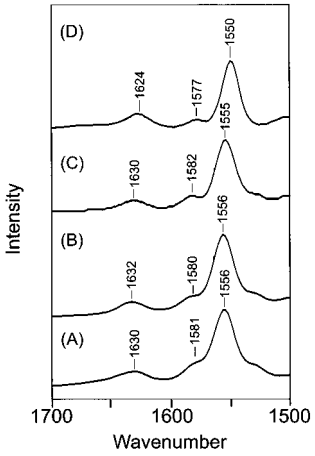

FIGURE 6: Raman spectra of the model chromophore between 1500 and 1700 cm^{-1} in (A) water, (B) methanol, (C) 2-propanol, and (D) DMSO. In each case, the solvent background has been subtracted. The concentrations were 350 μM, and the spectra were accumulated for 8 min.

general similarities between the Raman spectra of the three forms, there are also key differences.

The most striking differences between the cationic, neutral, and anionic forms of the model chromophore lie in the spectral region above 1500 cm^{-1} (Figure 5). Each of the different protonated forms gives a clear signature in the double bond stretching region between 1500 and 1700 cm^{-1} which can be used to distinguish between them. In particular, the cationic form has Raman bands at 1651, 1595 (shoulder), 1584, and 1539 cm^{-1}, the neutral form has bands at 1645, 1603, 1567, and 1525 cm^{-1}, and the anionic form has bands at 1630, 1577 (shoulder), 1556, 1535 (shoulder), and 1501 cm^{-1} (Table 2). We have recently examined ionized and unionized forms of 4-hydroxycinnamoyl-CoA (4-OH-Cin-CoA) bound to human medium chain acyl-CoA dehydrogenase

Table 2: Position in Wavenumbers of Characteristic Raman Marker Bands for the Cationic, Neutral, and Anionic Forms of the Model Chromophore

cationic form	neutral form	anionic form
1651	1645	1630
1595 (sh)[a]	1603	1577 (sh)
1584	1567	1556
1539 (weak)	1525 (weak)	1535 (sh)
		1501 (weak)

[a] sh: shoulder.

(MCAD) (*27*). Since the 4-OH-Cin-CoA is structurally similar to the wild-type GFP chromophores, but without the heterocyclic imidazolinone ring, this system provides a useful starting point for interpreting the Raman spectra. Two bands at 1597 and 1558 cm^{-1} characterize the protonated form of the 4-OH-Cin-CoA bound to MCAD. Upon deprotonation, these bands are replaced by bands at 1581 and 1546 cm^{-1}. Comparing the MCAD and model chromophore data (Figure 5), we can see that there is some similarity between the enzyme-bound 4-OH-Cin-CoA and the model chromophore. Thus, the neutral form marker bands for the model chromophore at 1603 and 1567 cm^{-1} and the anionic form marker bands at 1577 and 1556 cm^{-1} may correspond to bands with similar normal mode compositions in the un-ionized and ionized forms of 4-OH-Cin-CoA bound to MCAD, respectively. Similarly, the additional band at 1501 cm^{-1} in the anionic form of the model chromophore is also observed at 1497 cm^{-1} in the ionized form of 4-OH-Cin-CoA bound to MCAD. These observations support the conclusion that deprotonation of the phenolic hydroxyl is responsible for the observed pK_a at 8.2. In contrast, the cationic form has more similarity to 4-OH-Cin-CoA free in aqueous solution.

These data outlined above suggest that 4-OH-Cin-CoA bound to MCAD provides a reasonable model for the GFP chromophores in this spectral region, indicating that the normal modes are mainly localized on the 4-hydroxybenzylidene ring of the model chromophore. Consequently, it seems probable that the increased conjugation in the model chromophore results in a greater degree of π-electron delocalization than observed for 4-OH-Cin-CoA free in solution since the Raman spectra of the model chromophore in aqueous solution (except for the cationic form) more closely resemble those observed for 4-OH-Cin-CoA when bound within a polarizing enzyme active site. This effect can probably be ascribed to the presence of the imidazolinone ring in the model chromophore.

Below 1500 cm^{-1}, there are a number of other changes that occur in the Raman spectra of the different forms of the model chromophore (Figure 5). Most notably, the bands appearing at 1452, 1350, 1270, and 1182 cm^{-1} in the cationic form are shifted to 1448, 1329, 1257, and 1179 cm^{-1} in the neutral form and to 1439, 1315, 1263, and 1172 cm^{-1} in the anionic form. The band around 1260 cm^{-1} has previously been assigned to a C−O stretching mode of the 4-hydroxy group in aromatic model compounds, and the shift to higher wavenumber going from the neutral to the anionic form would be consistent with an increase in double bond character of this group in the anionic form (*28*−*30*). The band appearing at 1448 cm^{-1} probably contains a large contribution from a CH_2 scissors motion, and the band near 1175 cm^{-1} is likely a C−H phenyl bending mode.

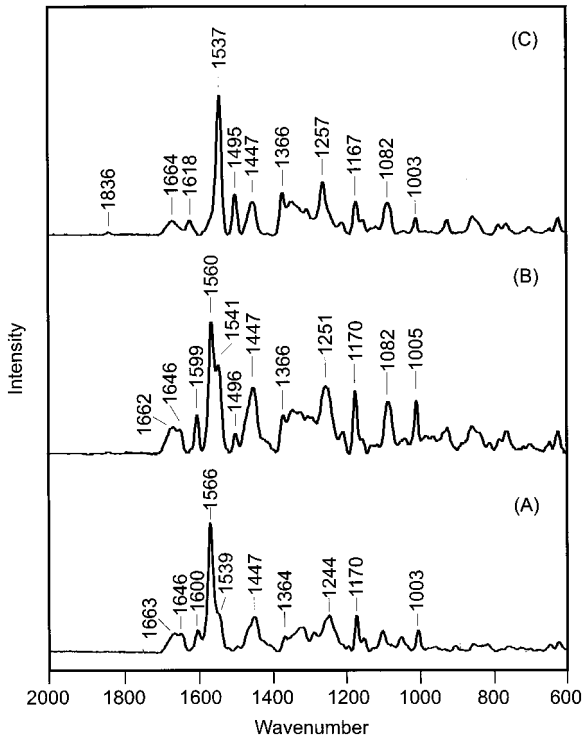

FIGURE 7: Raman spectra of (A) wild-type GFP in Tris buffer (pH 8.0, 20 mM) plus 140 mM NaCl, (B) S65T in sodium acetate buffer (pH 5.0, 20 mM) plus 140 mM NaCl, and (C) S65T in Tris buffer (pH 8.0, 20 mM) plus 140 mM NaCl. In each case, the solvent background has been subtracted. The protein concentrations were 60 μM, and the spectra were accumulated for 8 min.

More detailed assignment of the bands in the Raman spectra will have to await the results of isotope labeling studies and ab initio calculations. Undoubtedly, these studies will reveal complex coupling of the vibrational coordinates within this conjugated system due to the highly delocalized nature of the GFP chromophores.

Influence of Protein Environment on GFP Chromophores from Raman Spectroscopy. Since we are using 752 nm excitation, it is expected that the Raman spectra of wild-type GFP and the S65T mutant, shown in Figure 7, will be dominated by the pre-resonance-enhanced contributions from the 4-hydroxybenzylidene-imidazolinone chromophore (Figure 1) with weaker bands perhaps attributable to the remainder of the protein. Comparison of the Raman data for the model chromophore (Figure 5) with wild-type GFP and the S65T mutant (Figure 7) reveals that indeed the vast majority of the bands can be assigned with confidence to the chromophore unit. However, there are exceptions; specifically, bands at around 1660 and 1005 cm^{-1} in the protein spectra are assigned to the amide I mode and an aromatic side-chain mode of the protein, respectively. In addition, a weak broad background observed between about 1220 and 1350 cm^{-1} is attributable to contributions from the amide III modes of the protein, and the increase in Raman intensity of the band near 1448 cm^{-1} in the protein spectra is probably due to side chain CH_2 groups.

In the proteins, the chromophore has been removed from bulk solvent, and the absorption and emission properties of the chromophore are controlled by the specific interactions with the protein environment. For instance, crystal structures point to the importance of a number of hydrogen bonding interactions between the chromophore and the protein

environment (*5*−*11*). The predominant form of the chromophore in wild-type GFP is neutral, and the main interaction of the chromophore π-system consists of a hydrogen bond between the imidazolinone carbonyl oxygen and Arg96 (*7*). The phenolic hydroxyl, on the other hand, does not appear to be in direct contact with a protein group in the major form of the chromophore, and instead is hydrogen bonded to a water molecule (*7*). His148, with somewhat poorly defined electron density, is close by and could share a proton with the phenolic hydroxyl in the minor form of the chromophore, which constitutes roughly 15% of the protein and has been thought to be the anion (*31*).

The S65T mutant is known to respond rapidly and reversibly to pH changes (*21*), with a chromophore pK_a of 6.0 (*11*). Crystal structures at both high and low pH (*5*, *11*) have been solved, and the results are consistent with titration of the chromophore's phenolic hydroxyl. At pH 8.0, both the Thr203 as well as the His148 side chains are directly hydrogen bonded to the chromophore phenolate. When the pH is lowered to 4.6, both the side chain of Thr203 and the imidazole ring of His148 move away from the chromophore, and the phenolic hydroxyl is now hydrogen bonded to the backbone carbonyl of Thr203 and a water molecule, consistent with chromophore protonation at this site. As with all other GFP structures solved to date, the S65T mutant also contains a hydrogen bond from the imidazolinone oxygen to Arg96. In the anionic form of the chromophore, this bond appears to be somewhat stronger than in the neutral form, as judged by a decrease of hydrogen bonding distance from 3.0 Å (pH 4.6) to 2.7 Å (pH 8.0) (*11*).

The Raman spectra of wild-type GFP at pH 8.0 and of the S65T mutant at pH 5.0 and 8.0 are shown in Figure 7. For wild-type GFP, it is thought that over the physiological pH range the chromophore exists as an approximately 6:1 mixture of neutral to anionic forms and that this ratio is essentially pH independent (*12*). Consequently, the Raman spectrum is expected to be dominated by bands attributable to the neutral form of the chromophore. This is important because it is the anionic form of the chromophore that is responsible for the observed fluorescence since when the neutral form of GFP proteins absorb light there is an excited state deprotonation that occurs prior to fluorescence emission (*31*). The Raman spectra of the neutral and anionic forms of the model chromophore (Figure 5) correspond very well with the spectra of S65T at pH 5.0 (and wild type at pH 8.0) and S65T at pH 8.0, respectively. This indicates that neither the cationic nor the zwitterionic forms of the chromophore are present in the GFP proteins. Thus, although the absorption data indicate that the cationic form of the chromophore may be present in the GFP proteins, the Raman spectra do not support this possibility. Similarly, the Raman spectra do not provide any evidence that the zwitterionic form of the chromophore is present in the GFP proteins. We cannot exclude the possibility that the zwitterion is present since we do not have a model Raman spectrum of this particular form. However, our data are only fully consistent with the absorption bands at 395 and 475 nm representing neutral and anionic forms of the model chromophore in GFP proteins in the ground state.

A comparison of the Raman spectra of the neutral form of the model chromophore (Figure 5B) and wild-type GFP (Figure 7A) reveals that the protein environment causes almost no change in the position of the Raman bands above 1500 cm^{-1}. However, a weak shoulder at 1539 cm^{-1} in the wild-type GFP Raman spectrum is attributable to a small population of the anionic form. This particular band is shifted by 19 cm^{-1} to lower wavenumber relative to the same band in the anionic form of the model chromophore.

The most direct comparison of the Raman data can be made between the neutral and anionic forms of the model chromophore (Figure 5B,C) and the S65T mutant (Figure 7B,C) since it has a titratable chromophore, unlike the wild-type GFP which is nearly pH-insensitive over a broad range (*12*). The Raman spectrum of the neutral form of S65T resembles that of the neutral form of the model chromophore and of wild-type GFP. The extra band in the neutral form of the S65T mutant at 1541 cm^{-1} is due to the anionic form since we do not have total conversion at pH 5.0 (Table 1). Unfortunately, at lower pH values, the S65T mutant starts to become unstable, and it is important to maintain a fully folded protein to probe the interactions between the chromophore and its environment. We estimate that the Raman scattering cross section for the anionic form marker band near 1540 cm^{-1} is about twice as large as the corresponding neutral form marker band around 1560 cm^{-1}. Thus, the small populations of the anionic form are overrepresented by a factor of two in our spectra. In contrast, we can see that the main anionic form marker bands in the model chromophore at 1630, 1556, and 1501 cm^{-1} have been shifted to 1618, 1537, and 1495 cm^{-1}, respectively, in the S65T mutant. In addition, we observe a weak band near 1836 cm^{-1} only for the anionic form in the S65T mutant. This frequency is probably too high for this band to be assigned to a carbonyl stretching mode. However, in benzene derivatives, the C−H bending coordinates are known to produce characteristic overtone modes in this region (*32*, *33*), which may provide an explanation for the presence of this band.

The absorption data listed in Table 1 reveal a similar trend in the optical properties as observed for the Raman spectra. For the neutral forms of the chromophore, the absorption maxima fall in a fairly narrow range between 368 and 395 nm. In addition, the solvent dependence of the neutral form of the model chromophore was examined by Niwa et al. and found to undergo only slight variations between 368 and 373 nm as a function of solvent (*13*). In contrast, for the anionic form, there are a wide range of absorption maxima, between 428 and 515 nm (515 nm for YFP; see ref *9*), and a strong solvent dependence was found for the model chromophore (428 nm in water to 462 nm in DMSO). In addition, the protein environment shifts the absorption maximum of the anionic form from 428 nm in H_2O for the model chromophore to 475 nm in wild-type GFP and even higher (489 nm) in the S65T mutant (Table 1).

Combining the information outlined above, we can make the following conclusions. The specific interactions with the protein environment significantly lower the wavenumber of the Raman bands, reflecting a change in the ground state structure of the chromophore. This modulation of the structure seems only to be important for the anionic form of the chromophore. The ground state structural shift is primarily mediated by hydrogen bonding to the imidazolinone carbonyl group from Arg96, providing a pull of electron density from the chromophore, and interactions at the 4-hydroxy group, at least in the anionic form, providing a

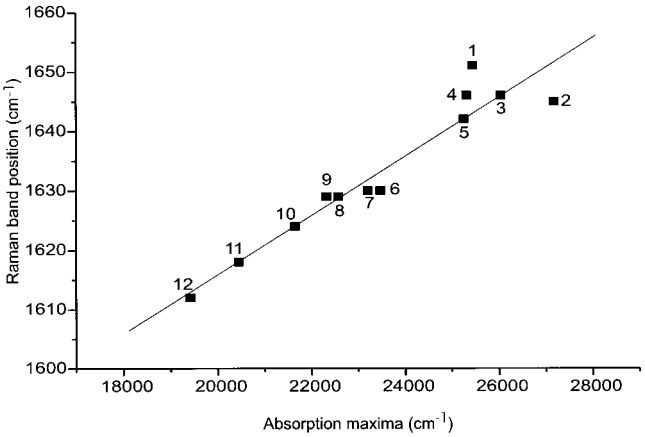

FIGURE 8: Plot of absorption maxima in wavenumber versus Raman band position in wavenumber for cationic, neutral, and anionic forms of GFP chromophores. (1) Cationic form of the model chromophore; (2) neutral form of the model chromophore; (3) neutral form of wild-type GFP; (4) neutral form of S65T; (5) neutral form of YFP H148Q (data not shown here); (6) anionic form of the model chromophore in water; (7) anionic form of the model chromophore in methanol; (8) anionic form of the model chromophore in 2-propanol; (9) anionic form of the model chromophore in DMSO; (10) alkali-denatured wild-type GFP (data not shown here); (11) anionic form of S65T; and (12) anionic form of YFP H148Q (data not shown here).

push of electron density into the chromophore. The nature of these particular interactions would seem to suggest that the changes observed in the Raman spectrum are reflecting a shift from a benzenoid-like resonance structure (Figure 1, form I) toward a quinonoid-like resonance structure for the anion (Figure 1, form II). For the neutral form of the chromophore, the protein environment seems to have a smaller effect on the ground state structure, as evidenced by the similarity in the Raman spectra, even though the chromophore has been removed from solvent and buried within the protein. This observation can be readily understood since, in contrast to the anionic form, resonance structures are not easily accessible to the neutral form of the chromophores. However, the absorption maximum does red-shift from 368 nm for the model chromophore in water to 395 nm in wild-type GFP. This indicates that excited state effects are dominating since there is no evidence from the Raman data for involvement of the cationic form of the chromophore that absorbs at 393 nm.

Linear Correlation between Ground State Structure and Absorption Maxima. The absorption and emission properties of GFP chromophores are governed by the difference between ground and excited state energies and thus the molecular structures in these two states. In contrast, the positions of the vibrational Raman bands are a property exclusively of the ground state molecular structure. Therefore, by combining the two techniques, we can start to disentangle the relative contributions from interactions with the protein environment on the ground and excited state structures, and to correlate the observed changes with the absorption and emission properties of the chromophore.

In Figure 8 we provide evidence for a direct correlation between the ground state structure, as probed by Raman spectroscopy, and the absorption spectra of GFP chromophores. This is demonstrated by plotting the absorption maxima (cm^{-1}) against the position of a Raman band found between 1610 and 1655 cm^{-1} in the cationic, neutral, and anionic forms of the model chromophore, the anionic form of the chromophore in various solvents, wild-type GFP, and the S65T mutant (Figures 5−7) together with the YFP H148Q mutant (data not shown). The YFP mutants will be discussed in a separate manuscript, and the data for one of these mutants are included in Figure 8 to extend the correlation. This particular band is the most sensitive in the Raman spectrum to the chromophore environment and presumably is a highly delocalized mode encompassing both rings and the exocyclic methylene and carbonyl stretches. The linear (positive) correlation we observe indicates that the ground state structure is an important determinant of the optical properties for this class of chromophore and that excited state effects are not solely responsible for the observed absorption changes. This is an important observation since in many systems excited state effects are found to dominate the changes in optical properties because the excited state structure often has a greater charge separation making it more suseptible to environmental factors. We note that only the cationic form of the model chromophore does not follow this correlation, reinforcing the suggestion that this form is not present in the proteins studied here.

A number of other bands between 1500 and 1600 cm^{-1} also exhibit substantial wavenumber shifts on changes in protonation state (Figure 5). The bands in this region can probably be assigned to complex aromatic modes and do not exhibit a direct correlation between Raman band position and absorption maxima. For instance, for the cationic form of the model chromophore, the most intense band in this region appears at 1584 cm^{-1}; however, this band is shifted to lower wavenumber (1566 cm^{-1}) in the neutral form despite the fact that the absorption maximum for the cationic form is red-shifted relative to the neutral form. Without a deeper understanding of the origin of the normal modes in this region, perhaps with the help of ab initio calculations, it is difficult to explain this apparent discrepancy. It is likely that the band that appears between 1610 and 1650 cm^{-1} has a larger contribution from the exocyclic C═C and C═O stretches, i.e., is more delocalized over the chromophoric framework, which results in the direct correlation of the Raman band positions and absorption maxima.

The decrease in wavenumber of the Raman band is associated with a red-shift in the absorption maxima (Figure 8) which reflects a reduction in the energy gap between the ground and excited states. As stated in the previous section, the shift in the Raman band position may reflect a change in the ground state structure of the chromophores between benzenoid-like and quinonoid-like resonance structures (Figure 1). However, the direction of this change is dependent on the relative ground and excited state energies of the benzenoid and quinonoid-like structures and the exact nature of the interactions with the surrounding protein, which is at present unknown. Thus, the Raman band shifts may reflect either a stabilization or a destabilization of the ground state. If the Raman band shifts correspond to a ground state destabilization, then the excited state can be either stabilized or destabilized as long as any destabilization is smaller than that experienced by the ground state. However, if the Raman band shifts correspond to a ground state stabilization, then the excited state energies must experience an even greater

stabilization to explain the observed red-shift in the absorption maxima.

It is interesting to speculate on changes in the protein environment that may alter the absorption and fluorescence properties of the chromophores on the basis of the Raman studies. Clearly, altering the ground state structure will result in a change in the absorption maximum, and this may be achieved in a number of different ways. One important factor is the hydrogen bond between the imidazolinone carbonyl oxygen and the guanidinium side chain of Arg96. In addition, this side chain carries a positive charge which will help to stabilize resonance structures that have a negative charge on the imidazolinone carbonyl. For the S65T mutant, direct replacement of Arg96 with a cysteine resulted in an absorption blue-shift from 489 to 472 nm (*1*). If positioned correctly, the cysteine side chain could still be hydrogen bonded to the imidazolinone carbonyl; however, this side chain does not carry a positive charge with which to stabilize negative charge on the imidazolinone carbonyl. The cysteine mutant is still substantially red-shifted from the model chromophore in water (428 nm) (*13*), so, although important, the hydrogen bond and charge stabilization interactions with Arg96 are certainly not the only contributing factor to the red-shift of the anionic form of the chromophore due to the protein environment. Increasing the strength of this hydrogen bond would increase the pull of electron density from the carbonyl oxygen and stabilize formation of the partial negative charge at this position, i.e., resonance structure II in Figure 1. A decrease in distance from 3.0 to 2.7 Å is observed for this hydrogen bond upon ionization as deduced from the crystal structures of the neutral and anionic forms of the S65T mutant (*5*, *11*). This would seem to indicate that the anionic form of the model chromophore has increased quinonoid-like character relative to the neutral form. It is unlikely that a direct replacement of Arg96 would lead to a red-shift in absorption since the positively charged side chain presumably stabilizes the formation of negative charge on the imidazolinone carbonyl oxygen. However, the strength of the interaction may be modulated by looking further afield and identifying side chains that interact with Arg96 and are important in positioning this group relative to the chromophore.

Another set of key interactions are between the phenolic group and the protein environment. Recently, Rudik et al. demonstrated that ionization of the phenolic hydrogen in 4-OH-Cin-CoA bound to wild-type MCAD was suppressed (*27*). This effect was ascribed to the presence of an ionized glutamate (Glu99) positioned in the active site close to the phenolic end of this product analogue. However, the push of electron density provided by this glutamate was found to be partly responsible for the red-shift in the absorption maximum observed on binding for the neutral form of 4-OH-Cin-CoA. A similar situation is expected for the anionic forms of the GFP chromophores, where interactions that provide a push of electron density will force the negative charge to delocalize over the chromophoric skeleton rather than localizing on the phenolic oxygen, causing a red-shift in the absorption maxima. (Note that a red-shift would also be expected for the neutral forms of GFP chromophores.) This would be in agreement with the solvent-dependent absorption and Raman data discussed earlier (Figures 3 and 5). However, placing a negative charge in a position to interact with the phenolic oxygen will also raise the pK_a of the phenolic hydrogen. The side chain of His148 has already been identified from crystal structures (*5*−*11*) to be in a position to interact with the phenolic group. In particular, a S65T/H148D mutant was found to have the absorption maximum of the neutral form raised to 415 nm (+20 nm) while the anionic form was slightly blue-shifted to 487 nm (−2 nm). In addition, the pK_a of the phenolic hydrogen was raised by 1.8 units to a value of 7.8 for the S65T/H148D mutant relative to S65T (pK_a 6.0) (*11*). An alternative method for forcing the negative charge on the phenolic oxygen to delocalize over the chromophoric skeleton would be to remove some of the hydrogen bonding interactions between this group and the remainder of the protein. A number of such interactions have been identified from crystal structures and will have the effect of pulling electron density from the phenolic oxygen, thus stabilizing a charge localized on this atom. This could be achieved by replacing His148 and Thr203, which are both in position to interact with the phenolic oxygen (*5*−*11*), with side chains that lack any groups which could act as hydrogen bond donors. This would generate an aprotic region around the phenolic oxygen which may still raise the pK_a of the phenolic hydrogen but could also introduce a significant shift in the absorption maximum for the anion. It is possible that such an effect provides part of the observed red-shift for the YFP mutants where Thr203 has been substituted with an aromatic side chain. Recently, it has been shown that substitution of Thr203 with nonaromatic side chains will also red-shift the absorption maximum although not to quite the same extent (Boris Steipe, personal communication).

CONCLUSIONS

The Raman studies on GFP outlined here have shown that high-quality Raman data on an intrinsically fluorescent protein can be obtained. An important step in this work was the synthesis of a model chromophore (Figure 2) which assists in dissecting the contributions that govern the absorption properties of GFP chromophores. We have determined the macroscopic pK_as for the imidazolinone ring nitrogen and phenolic hydroxyl of the model chromophore to be 1.8 and 8.2, respectively. In addition, Raman marker bands have been found for the cationic, neutral, and anionic forms of the model chromophore. These Raman band assignments have been transferred to the proteins to monitor the effect of interactions between the chromophore and the protein on the ground state structure of the chromophore. Since none of the cationic form Raman marker bands are present in the spectra of either wild-type GFP or the S65T mutant, we conclude that the protein spectra can be explained in terms of the neutral and anionic forms of the chromophore. In the GFP proteins, the neutral forms of the chromophore are found to have Raman spectra very similar to the neutral form of the model chromophore whereas shifts to lower wavenumber are observed for the anionic forms of the chromophore within the proteins relative to the model chromophore. The different behavior of the neutral and anionic forms of the chromophore can be explained in terms of the accessibility of resonance structures.

A linear correlation between the Raman wavenumber and absorption maxima has been established, and since the position of Raman bands is exclusively dependent upon the

ground state structure, this demonstrates that the ground state is an important determinant of the optical properties. Consequently, the ground state structure must be considered together with the excited state in order to gain a full appreciation of the spectral properties of the chromophore.

REFERENCES

1. Tsien, R. Y. (1998) *Annu. Rev. Biochem. 67*, 509−544.
2. Palm, G. T., and Wlodawer, A. (1999) *Methods Enzymol. 302*, 378−394.
3. Shimomura, O. (1979) *FEBS Letts. 104*, 220−222.
4. Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G., and Ward, W. W. (1993) *Biochemistry 32*, 1212−1218.
5. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. (1996) *Science 273*, 1392−1395.
6. Yang, F., Moss, L. G., and Phillips, G. N., Jr. (1996) *Nat. Biotechnol. 14*, 1246−1252.
7. Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormo, M., and Remington, S. J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2306−2311.
8. Wachter, R. M., King, B. A., Heim, R., Kallio, K., Tsien, R. Y., Ormo, M., and Remington, S. J. (1997) *Biochemistry 36*, 9759−9765.
9. Wachter, R. M., Elsliger, M. A., Kallio, K., Hanson, G. T., and Remington, S. J. (1998) *Structure 6*, 1267−1277.
10. Palm, G. J., Zdanov, A., Gaitanaris, G. A., Stauber, R., Pavlakis, G. N., and Wlodawer, A. (1997) *Nat. Struct. Biol. 4*, 361−365.
11. Elsliger, M., Wachter, R. M., Hanson, G. T., Kallio, K., and Remington, S. J. (1999) *Biochemistry 38*, 5296−5301.
12. Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., and Reeves, S. C. (1982) *Photochem. Photobiol. 35*, 803−808.
13. Niwa, H., Inouye, S., Hirano, T., Matsuno, T., Kojima, S., Ohashi, M., and Tsuji, F. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13617−13622.
14. Creemers, T. M. H., Lock, A. J., Subramaniam, V., Jovin, T. M., and Volker, S. (1999) *Nat. Struct. Biol. 6*, 557−560.
15. Voityuk, A. A., Michel-Beyerle, M., and Rosch, N. (1998) *Chem. Phys. Lett. 296*, 269−276.
16. Voityuk, A. A., Michel-Beyerle, M., and Rosch, N. (1998) *Chem. Phys. 231*, 13−25.
17. Weber, W., Helms, V., McCammon, J. A., and Langhoff, P. W. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 6177−6182.
18. Heim, R., Cubbitt, A. B., and Tsien, R. Y. (1995) *Nature 373*, 663−664.
19. Heim, R., Prasher, D. C., and Tsien, R. Y. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 12501−12504.
20. Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., and Lukyanov, S. A. (1999) *Nat. Biotechnol. 17*, 969−973.
21. Kneen, M., Farinas, J., Li, Y., and Verkman, A. S. (1998) *Biophys. J. 74*, 1591−1599.
22. Van Thor, J. J., Pierik, A. J., Nugteren-Roodzant, I., Xie, A., and Hellingwerf, K. J. (1998) *Biochemistry 37*, 16915−16921.
23. Corey, E. J., and Venkateswarlu, A. (1976) *J. Am. Chem. Soc. 94*, 6190−6191.
24. Dong, J., Dinakarpandian, D., and Carey, P. R. (1998) *Appl. Spectrosc. 52*, 1117−1122.
25. Weber, W., Helms, V., McCammon, J. A., and Langhoff, P. W. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 6177−6182.
26. Gutman, V. (1976) *Coord. Chem. Rev. 18*, 225−255.
27. Rudik, I., Bell, A. F., Tonge, P. J., and Thorpe, C. (2000) *Biochemistry 39*, 92−101.
28. Jagodzinski, P. W., Funk, G. F., and Peticolas, W. L. (1982) *Biochemistry 21*, 2193−2202.
29. Dong, J., Xiang, H., Luo, L., Dunaway-Mariano, D., and Carey, P. R. (1999) *Biochemistry 38*, 4198−4206.
30. Callender, R., Chen, D., Lugtenburg, J., Martin, C., Rhee, K. W., Sloan, D., Vandersteen, R., and Yue, K. T. (1998) *Biochemistry 37*, 3672−3681.
31. Chattoraj, M., King, B. A., Bublitz, G. U., and Boxer, S. G. (1996) *Proc Natl. Acad. Sci. U.S.A. 93*, 8362−8367.
32. Lin-Vien, D., and Colthup, N. B. (1991) in *The Handbook of infrared and Raman characteristic frequencies of organic molecules*, Academic Press, Boston.
33. Varsanyi G. (1969) *Vibrational spectra of benzene derivatives*, Academic Press, New York.

BI992675O