

J|A|C|S

A R T I C L E S

Published on Web 00/00/0000

¹⁹F NMR Studies of the Native and Denatured States of Green Fluorescent Protein

Farid Khan,^{†‡} Ilya Kuprov,[§] Timothy D. Craggs,[†] P. J. Hore,^{§,*} and Sophie E. Jackson^{†,*}

Contribution from the Chemistry Department, Lensfield Road, Cambridge CB2 1EW, United Kingdom, and University of Oxford, Department of Chemistry, Physical and Theoretical Chemical Laboratory, South Parks Road, Oxford, OX1 3QZ, United Kingdom

Received January 26, 2006; E-mail: sej13@cam.ac.uk (S.E.J.); peter.hore@chem.ox.ac.uk (P.J.H.).

Abstract: Biosynthetic preparation and ¹⁹F NMR experiments on uniformly 3-fluorotyrosine-labeled green fluorescent protein (GFP) are described. The ¹⁹F NMR signals of all 10 fluorotyrosines are resolved in the protein spectrum with signals spread over 10 ppm. Each tyrosine in GFP was mutated in turn to phenylalanine. The spectra of the Tyr → Phe mutants, in conjunction with relaxation data and results from ¹⁹F photo-CIDNP (chemically induced dynamic nuclear polarization) experiments, yielded a full ¹⁹F NMR assignment. Two ¹⁹F-Tyr residues (Y92 and Y143) were found to yield pairs of signals originating from ring-flip conformers; these two residues must therefore be immobilized in the native structure and have ¹⁹F nuclei in two magnetically distinct positions depending on the orientation of the aromatic ring. Photo-CIDNP experiments were undertaken to probe further the structure of the native and denatured states. The observed NMR signal enhancements were found to be consistent with calculations of the HOMO (highest occupied molecular orbital) accessibilities of the tyrosine residues. The photo-CIDNP spectrum of native GFP shows four peaks corresponding to the four tyrosine residues that have solvent-exposed HOMOs. In contrast, the photo-CIDNP spectra of various denatured states of GFP show only two peaks corresponding to the ¹⁹F-labeled tyrosine side chains and the ¹⁹F-labeled Y66 of the chromophore. These data suggest that the pH-denatured and GdnDCI-denatured states are similar in terms of the chemical environments of the tyrosine residues. Further analysis of the sign and amplitude of the photo-CIDNP effect, however, provided strong evidence that the denatured state at pH 2.9 has significantly different properties and appears to be heterogeneous, containing subensembles with significantly different rotational correlation times.

Introduction

Green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is widely used as a fluorescent marker in many areas of biological research, owing to its unique photophysical properties.¹ The 238-residue protein undergoes an autocatalytic post-translational cyclization and oxidation of the polypeptide chain around residues Ser65, Tyr66, and Gly67, forming an extended and rigidly immobilized conjugated π -system, the chromophore, which emits green fluorescence.² No cofactors are necessary for either the formation or the function of the chromophore,³ which is embedded in the interior of the protein surrounded by an 11-stranded β -barrel, Figure 1.^{4,5} GFP has a high (~70%) fluorescence quantum yield due to the lack of vibrational relaxation, because in the native state, the chromophore is rigidly held and shielded from bulk solvent.^{4,5} On protein denaturation, the chromophore remains chemically intact but fluorescence is lost. The green fluorescence is, therefore, a sensitive probe of the state of the protein.

GFP has been extensively engineered to modify and improve its spectroscopic and physical properties to facilitate its use as a marker of gene expression and protein localization, an indicator of protein-protein interactions, and its use as a biosensor.¹ In all cases, GFP needs to fold efficiently to function in these different biological assays. In addition to its use as a reporter protein, GFP is also a model system for photophysical studies² and for folding studies on large single domain proteins.^{8–10} Despite the widespread use of GFP and level of interest in this protein, relatively little is known about its folding either in vitro or in vivo. A recent study by Kuwajima and co-workers has provided the most detailed information to date.⁸

workers has provided the most detailed information to date.⁸

[†] Chemistry Department.[‡] Current address: Protein Technologies Laboratory, Babraham Institute, Babraham, Cambridge CB2 4AT, UK.[§] University of Oxford.(1) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.(2) Zimmer, M. *Chem. Rev.* **2002**, *102*, 759–781.(3) Reid, B. G.; Flynn, G. C. *Biochemistry* **1997**, *36*, 6786–6791.(4) Ormo, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273*, 1392–1395.(5) Yang, F.; Moss, L.; Phillips, G. *Nat. Biotech.* **1996**, *14*, 1246–1251.(6) Battistutta, R.; Negro, A.; Zanotti, G. *Proteins: Struct. Func., Gen.* **2000**, *41*, 429–437.(7) DeLano, W. L. *World Wide Web* **2002**, <http://www.pymol.org>.(8) Enoki, S.; Saeki, K.; Maki, K.; Kuwajima, K. *Biochemistry* **2004**, *43*, 14238–48.(9) Fukuda, H.; Arai, M.; Kuwajima, K. *Biochemistry* **2000**, *39*, 12025–12032.(10) Stepanenko, O. V.; Verkhusha, V. V.; Kazakov, V. I.; Shavlovsky, M. M.; Kuznetsova, I. M.; Uversky, V. N.; Turoverov, K. K. *Biochemistry* **2004**, *43*, 14913–14923.

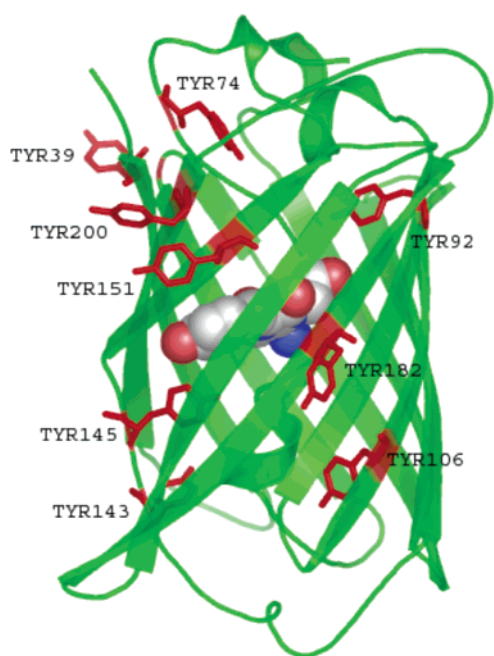


Figure 1. Ribbon diagram of the backbone of GFP showing the chromophore (space filled) and tyrosine residues (red). The structure is that of the GFPuv used in this study.⁶ The figure was created using PyMol.⁷

In their study, fluorescence and far-UV circular dichroism were used to probe folding and unfolding and a model proposed in which GFP folds through several intermediate states. Further studies using complementary techniques and probes are clearly necessary to obtain a complete understanding of the folding pathway of this large, complex, and important protein.

NMR spectroscopy has proven to be a powerful method of studying structured, partially structured, and denatured states of proteins, including molten globules and kinetic intermediates.¹¹ In the cases where NMR can be used, it provides valuable residue-specific information on structure and dynamics and therefore complements optical techniques such as fluorescence and circular dichroism that only report on global properties. ¹H, ¹³C, and ¹⁵N are the most commonly used nuclei. Quite recently, however, much progress has been made on the study of proteins using ¹⁹F NMR.^{12–17}

A distinctive feature of ¹⁹F NMR spectroscopy is the strong chemical shielding anisotropy (CSA) of the fluorine nucleus. The nuclear relaxation rate due to CSA is quadratic in magnetic induction, and for a protein the size of GFP at 10–20 T, the transverse ¹⁹F relaxation is dominated by the CSA term and is very fast. This leads to broad spectral lines (50–400 Hz) and reduced detection sensitivity. Fast transverse relaxation of ¹⁹F in proteins also makes magnetization transfer steps impossible in all but a few two-dimensional (2D) NMR experiments.¹⁸

Despite these seemingly formidable difficulties, ¹⁹F NMR of fluorine-labeled biological macromolecules is gaining popularity, the reason being that the ¹⁹F nucleus has a large chemical shift

dispersion and its chemical shielding is very sensitive to changes in molecular environment.¹⁹ In a protein containing 10 fluorinated tyrosines, one may generally expect that all 10 will be resolved in the native-state spectrum and that any structural transformation will be accompanied by significant changes in chemical shifts. Analysis of these changes can give direct information on the kinetics of the structural transformation in question.¹² In addition to the information obtained from chemical shifts, structural information can also be gathered from relaxation data. ¹⁹F relaxation behavior is well understood and, in principle, allows the extraction of local and global motion correlation times, order parameters and rotational diffusion tensors,¹⁸ which are useful as indicators of the local mobility and stability of the macromolecule.

Another unusual property of the fluorine nucleus, namely the large ¹⁹F hyperfine coupling constants of fluorinated aromatic radicals, makes it particularly useful in NMR experiments with photochemical pumping of magnetization through the photo-CIDNP (chemically induced dynamic nuclear polarization) effect.^{20–21} CIDNP is the non-Boltzmann population of nuclear spin states produced in chemical reactions that proceed through radical pair intermediates.²² It is manifested by strongly enhanced NMR absorption or emission and has been used to detect solvent-accessible tyrosine, tryptophan, and histidine residues in proteins.^{23–25} In high magnetic fields, the amplitude of the photo-CIDNP effect is directly proportional to the hyperfine coupling constant in the intermediate radical. Fluorine-containing molecules are quite unique in this respect because the ¹⁹F hyperfine constants in aromatic radicals are large, owing to both the high magnetogyric ratio of the ¹⁹F nucleus and the strong electronegativity of the fluorine atom.²⁶

In this paper, we present NMR studies on the native and denatured states of GFP, which provide important information on the initial and final states of the protein on the folding pathway. We describe the strategies employed to produce 3-fluorotyrosine-labeled GFP. Protein engineering is used systematically to replace each of the tyrosines in turn to facilitate the assignment of the ¹⁹F NMR spectrum. ¹⁹F photo-CIDNP techniques are then used to characterize the native and denatured states of GFP. In particular, the solvent accessibility of the 10 tyrosines in GFP is assessed in both the native state and the acid- and guanidinium chloride-induced denatured states, providing important information on the starting point of kinetic refolding studies.

Materials and Methods

Construction of the trGFPuv Expression Vector. The gene encoding GFPuv (Clontech) was cloned into a modified pRSET vector (Invitrogen). GFPuv has three mutations (F99S, M153T, V163A) that enhance solubility and fluorescence.²⁷ The GFPuv gene was subsequently truncated to its minimal domain for green fluorescence (Met1-

(11) Redfield, C. In *Methods in Molecular Biology*; Protein NMR Techniques, Vol. 278; Humana Press: Totowa, NJ, 2004; pp 233–254.278
 (12) Bann, J. G.; Frieden, C. *Biochemistry* **2004**, *43*, 13775–13786.
 (13) Hull, W. E.; Sykes, B. D. *Biochemistry* **1974**, *13*, 3431–3437.
 (14) Li, H.; Frieden, C. *Biochemistry* **2005**, *44*, 2369–2377.
 (15) Shu, Q.; Frieden, C. *J. Mol. Biol.* **2004**, *345*, 599–610.
 (16) Sykes, B. D.; Hull, W. E. *Methods Enzymol.* **1978**, *49*, 270–295.
 (17) Wang, X.; Mercier, P.; Letourneau, P.-J.; Sykes, B. D. *Prot. Sci.* **2005**, *14*, 2447–2460.
 (18) Peng, J. W. *J. Magn. Res.* **2001**, *153*, 32–47.

(19) Gakh, Y. G.; Gakh, A. A.; Gronenborn, A. M. *Magn. Res. Chem.* **2000**, *38*, 551–558.
 (20) Goetz, M. *Concepts Magn. Reson.* **1995**, *7*, 69–86.
 (21) Stob, S.; Kaptein, R. *Photochem. Photobiol.* **1989**, *49*, 565–577.
 (22) Goetz, M. *Adv. Photochem.* **1997**, *23*, 63–163.
 (23) Hore, P. J.; Broadhurst, R. W. *Prog. Nucl. Magn. Reson. Spectrosc.* **1993**, *25*, 345–402.
 (24) Kaptein, R. *Bio. Magn. Reson.* **1982**, *4*, 145–91.
 (25) Kaptein, R.; Dijkstra, K.; Nicolay, K. *Nature* **1978**, *274*, 293–4.
 (26) Brinkman, M. R.; Bethell, D.; Hayes, J. J. *Chem. Phys.* **1973**, *59*, 3431–4.
 (27) Cramer, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C. *Nat. Biotech.* **1996**, *14*, 315–319.

138 lle229)²⁸ by the introduction of a stop codon at position 230 using PCR
139 (polymerase chain reaction) techniques. The resulting plasmid (ptrG-
140 FPuv) was fully sequenced. The wild-type GFP referred to throughout
141 this paper is a pseudo wild-type GFP corresponding to the trGFPuv
142 construct described above.

143 **Site-Directed Mutagenesis of Tyrosine Residues.** All 10 tyrosine
144 residues were systematically mutated to phenylalanine in ptrGFPuv
145 using the QuikChange kit from Stratagene. All mutations were
146 confirmed by DNA sequencing of the truncated GFP gene.

147 **Expression and Purification of Fluorotyrosine-Labeled GFP**
148 **Mutants.** The incorporation of 3-fluorotyrosine was successfully
149 achieved using a modification of the method of Kim and co-workers.²⁹
150 In this approach, the herbicide glyphosate was used to inhibit the
151 shikimate pathway in cultures of *Escherichia coli*. Briefly, single
152 colonies of transformed *E. coli* cells (BL21 DE3) harboring ptrGFPuv
153 were picked from TYE ampicillin plates and were used to inoculate 5
154 mL of 2 × TY media (containing 0.1 mg mL⁻¹ ampicillin) and were
155 grown overnight in a shaker at 37 °C. This overnight culture was used
156 to inoculate 400 mL of 2 × TY media containing of 0.1 mg mL⁻¹
157 ampicillin. This preculture was left to grow to a cell density with an
158 A₆₀₀ of between 0.8 and 1.0 in a shaker at 37 °C. The 2 × TY preculture
159 was spun down (50 mL) at 25 °C for 10 min at 4000 rpm (SLC4000
160 rotor, Sorvall). Meanwhile, two 50-mL aliquots of minimal media were
161 taken from a flask containing 500 mL of M9 minimal media. One 50-
162 mL aliquot was used to resuspend the preculture cell pellet and was
163 added to the remaining 400 mL minimal media. This main culture was
164 allowed to grow to a cell density with an A₆₀₀ of 0.6–0.7 in a shaker
165 at 25 °C before GFP expression was induced by the addition of IPTG
166 to a final concentration of 0.5 mM (for the Y66F and Y74F, the
167 temperature was dropped to 18 °C). For selective fluorotyrosine
168 labeling, 35 mg of 3-fluoro-D,L-tyrosine (96%, Lancaster), 30 mg of
169 L-phenylalanine (Sigma), and 30 mg of L-tryptophan (Sigma) with 0.5
170 g of glyphosate (99.5% N-(phosphonomethyl)glycine, Sigma) were
171 dissolved in the remaining 50 mL of minimal medium. This labeling
172 mixture was added 30 min after induction to the cell culture, which
173 was then left to grow overnight. The cultures were harvested by
174 centrifugation (Sorvall, SLC4000 rotor, 15 min at 8000 rpm at 4 °C)
175 and resuspended into 50 mL of buffer A (50 mM Tris, pH 8.0).
176 The cells were lysed either by pulsed sonication on ice for 4 min or in a
177 cell homogenizer (Glen Creston, UK) by pressure disruption at 4 °C
178 and then sonicated for 1 min to make the lysate less viscous. The lysate
179 was centrifuged at 18 000 rpm (SS34, Sorvall) at 4 °C for 45 min. The
180 supernatant was then pooled and loaded onto a Q-Sepharose (Pharma-
181 cia) column at 2 mL min⁻¹ until the green protein was visibly bound
182 to the top of the column. After washing with 6 column volumes of
183 buffer A (50 mM Tris, pH 8.0), trGFPuv was eluted with a 0–60 min
184 gradient using buffer B (50 mM Tris, pH 8.0, 0.5 M NaCl). The green
185 fractions were pooled and concentrated (Vivaspins 20, Vivascience)
186 and washed through with PBS. The concentrated protein (~10 mL)
187 was then loaded onto a HiLoad 26/60 Superdex G75 column (Amer-
188 sham Biosciences), pre-equilibrated in buffer A, and run at 2 mL min⁻¹.
189 The visibly green fractions that eluted were pooled. Proteins were pure
190 as assessed by sodium–dodecyl-sulfate–polyacrylamide gel electro-
191 phoresis (SDS–PAGE) and LC-MS. The labeling efficiency was also
192 confirmed by LC-MS (for reference, the fully labeled, pseudo-wild-
193 type protein (truncated GFPuv with no further mutation) showed a single
194 HPLC peak of 25 806 Da corresponding to the incorporation of 10
195 3-fluorotyrosines). Pooled fractions were concentrated to 1 mM and
196 either stored at –80 °C or used immediately.

197 **Mass Spectrometry.** Electrospray ionization mass spectrometry was
198 performed using a Micromass/Waters Quattro LC instrument connected
199 to a Prodigy C8 (250 mm × 10 mm) reverse phase (RP) column for

200 protein analysis. Papain digested peptide mass analysis was done
201 according to the method of Cody and co-workers³⁰ on a Jupiter RP–
202 C18 (250 mm × 10 mm) column using a gradient of 0–100%
203 acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA) buffer over 30 min.

204 **¹⁹F NMR and CIDNP.** Conventional ¹⁹F NMR spectra were
205 recorded with 5K – 10K scans, with a 12 kHz sweep width and free
206 induction decays of 4K points. The pulse width was 60°, and the pulse
207 spacing was 1.5 s (the ¹⁹F T₁ times in GFP range from 0.5 to 2 s; T₂
208 times are 10–100 ms). The relatively short relaxation delay, which
209 was essential for sensitivity reasons, has the consequence that the
210 relative intensities of the ¹⁹F signals are not proportional to numbers
211 of spins as they would be in a fully relaxed spectrum. However, the
212 distortion is not severe and at no point are the signal intensities
213 interpreted quantitatively.

214 377 MHz ¹⁹F NMR measurements were performed on 1 mM protein
215 samples in 10% D₂O with PBS (phosphate-buffered saline) buffer at
216 300 K using a Bruker DRX400 spectrometer equipped with a 5-mm
217 QNP probe. ¹⁹F chemical shifts were referenced to external trifluoro-
218 acetic acid.

219 564 MHz ¹⁹F NMR and photo-CIDNP spectra were recorded at 300
220 K on a Varian Inova 600 (14.1 T) NMR spectrometer equipped with
221 a 5 mm ¹⁹F{¹H} z-gradient probe. The light source was a Spectra
222 Physics BeamLok 2080 argon ion laser, operating in single-line mode
223 at 5 W output power at a wavelength of 514 nm. A mechanical shutter
224 (NM Laser Products LS200) controlled by the spectrometer was used
225 to produce 100 ms light pulses. The light was focused into a 6-m length
226 of optical fiber (Newport F-MBE) using a Newport M-5X objective
227 lens. The other end of the fiber was attached (via Newport SMA
228 connectors) to a 2-m section of the same fiber whose stepwise-tapered
229 tip³¹ was held inside a 5-mm NMR tube by a truncated Wilmad WGS
230 5BL coaxial insert.

231 Before each photo-CIDNP experiment, the sample was repeatedly
232 dialyzed against 50 mM phosphate buffer in D₂O at pH 7.2 (uncorrected
233 for deuterium isotope effect) for several days to complete the proton
234 exchange of both the solvent and the protein. Deuterated guanidinium
235 chloride (GdnDCI), NaOD, and DCl were used as necessary to reach
236 the prescribed conditions. During each scan of the photo-CIDNP
237 experiment, a 0.2 mM fluorotyrosinated GFP sample containing 1 mM
238 flavin mononucleotide as photosensitizer was irradiated for 100 ms and
239 subjected to a 90° radio frequency pulse on ¹⁹F followed by immediate
240 acquisition of the free induction decay. After application of a shifted
241 Gaussian window function, zero-filling, and Fourier transformation,
242 the spectra were analyzed using mixed Lorentzian–Gaussian line-fitting.
243 Photo-CIDNP spectra are shown as light minus dark difference spectra
244 recorded with 16 scans, with a 60 s delay between scans, 12 kHz sweep
245 width, and 4K points in the free induction decays. The large signal
246 enhancement afforded by the CIDNP effect means that high-quality
247 spectra can be obtained using relatively few scans together with a
248 relaxation delay that is much longer than the ¹⁹F T₁s. The photo-CIDNP
249 spectra are thus “fully relaxed”. In general, the CIDNP intensities in
250 spectra recorded with a continuous wave laser reflect the spin–lattice
251 relaxation rates of the individual nuclei in the protein; however, as the
252 laser irradiation time (100 ms) is much shorter than the shortest T₁
253 (~500 ms), the observed intensities are not distorted by relaxation in
254 this case.³² There are well-established differences between time-resolved
255 ¹H CIDNP spectra, obtained with a nanosecond laser and microsecond
256 time delays, and spectra recorded with a continuous wave laser and
257 much lower time-resolution.³³ However, these effects are much smaller
258 for ¹⁹F than for ¹H CIDNP because of the very fast ¹⁹F spin–lattice

(30) Cody, C. W.; Prasher, D. C.; Westler, W. M.; Prendergast, F. G.; Ward, W. W. *Biochemistry* **1993**, *32*, 1212–1218.

(31) Kuprov, I.; Hore, P. J. *J. Magn. Res.* **2004**, *171*, 171–175.

(32) Hore, P. J.; Egmond, M. R.; Edzes, H. T.; Kaptein, R. *J. Magn. Res.* **1982**, *49*, 122–150.

(33) Morozova, O. B.; Turkovskaya, A. V.; Sagdeev, R. Z.; Mok, K. H.; Hore, P. J. *J. Phys. Chem. B* **2004**, *108*, 15355–15363.

(28) Li, X.; Zhang, G.; Ngo, N.; Zhao, X.; Kain, S. R.; Huang, C. C., *J. Biol. Chem.* **1997**, *272*, 28545–28549.

(29) Kim, H. W.; Perez, J. A.; Ferguson, S. J.; Campbell, I. D. *FEBS Lett.* **1990**, *272*, 34–36.

259 relaxation in the fluoro-tyrosine radical resulting from the highly
260 anisotropic ^{19}F hyperfine interaction;³⁴ we believe them to be negligible
261 here.

262 **Relaxation Analysis.** The relaxation of ^{19}F nuclei in GFP-sized
263 diamagnetic molecules in solution occurs primarily as a result of
264 rotational modulation of two anisotropic interactions: the dipole–dipole
265 interaction with nearby protons (DD mechanism) and the Zeeman
266 interaction with the applied magnetic field (chemical shift anisotropy,
267 or CSA mechanism). A Bloch–Redfield–Wangsness relaxation theory
268 treatment³⁵ results in the following expressions for the transverse
269 relaxation rate of the fluorine nucleus:

$$\left(\frac{1}{T_2}\right)_{\text{DD}} = \frac{\delta_{\text{DD}}^2 + 3\eta_{\text{DD}}^2}{72} [4J(0) + 6J(\omega_{\text{H}}) + 3J(\omega_{\text{F}}) + 6J(\omega_{\text{H}} + \omega_{\text{F}}) + J(\omega_{\text{H}} - \omega_{\text{F}})]$$

$$\left(\frac{1}{T_2}\right)_{\text{CSA}} = \frac{\delta_{\text{CSA}}^2 + 3\eta_{\text{CSA}}^2}{72} [4J(0) + 3J(\omega_{\text{F}})] \quad (1)$$

270 where δ_{A} and η_{A} denote the axially and rhombicity of the point dipolar
271 interaction ($\text{A} = \text{DD}$) and the Zeeman interaction ($\text{A} = \text{CSA}$):

$$\delta_{\text{A}} = 2A_{zz} - (A_{yy} + A_{xx}) \quad \eta_{\text{A}} = A_{yy} - A_{xx}$$

$$\delta_{\text{DD}} = 3 \frac{\gamma_{\text{H}} \gamma_{\text{F}} \hbar \mu_0}{r_{\text{HF}}^3 4\pi} \quad \eta_{\text{DD}} = 0$$

$$\delta_{\text{CSA}} = [2\sigma_{zz} - (\sigma_{yy} + \sigma_{xx})] B_0 \gamma_{\text{F}} \quad \eta_{\text{CSA}} = (\sigma_{yy} - \sigma_{xx}) B_0 \gamma_{\text{F}} \quad (2)$$

272 r_{HF} is the proton–fluorine distance, ω_{H} and ω_{F} are the ^1H and ^{19}F
273 Larmor frequencies, σ_{ii} are the eigenvalues of the shielding tensor, γ_{H}
274 and γ_{F} are magnetogyric ratios, and B_0 is the magnetic field strength.
275 We use the Lipari–Szabo restricted motion model for the spectral
276 density $J(\omega)$ ^{36,37}

$$J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S^2) \tau_e}{1 + \omega^2 \tau_e^2} \right) \quad \tau_e = \left(\frac{1}{\tau_c} + \frac{1}{\tau_i} \right)^{-1} \quad (3)$$

277 in which S^2 is an order parameter indicating the extent to which motion
278 is restricted, τ_c is the global molecular rotational correlation time
279 (assumed to be isotropic), and τ_i is the internal motional correlation
280 time of the amino acid residue in question. Note that τ_i does not include
281 contributions from fluoro–tyrosine ring flips (see below), which are
282 too slow to make a significant contribution to the relaxation.

283 At magnetic field strengths of 10–15 T, the typical Zeeman
284 interaction anisotropy of an aromatic ^{19}F nucleus is $\delta_{\text{CSA}} = (5-8) \times$
285 10^5 rad s^{-1} ,³⁸ compared to $\delta_{\text{DD}} = (1-2) \times 10^5 \text{ rad s}^{-1}$, meaning that
286 $\delta_{\text{DD}}^2 + 3\eta_{\text{DD}}^2 \ll \delta_{\text{CSA}}^2 + 3\eta_{\text{CSA}}^2$. The DD contribution to the relaxation
287 rate may, therefore, be neglected.

288 It has relatively recently become possible to obtain the chemical
289 shielding anisotropy parameters from ab initio calculations.³⁸ Values
290 of the anisotropy of the ^{19}F chemical shielding tensor were obtained
291 using the Gaussian03 program³⁹ at three different levels of theory (Table
292 1). The CSGT DFT B3LYP/cc-pVDZ method is recommended by
293 Tormena et al.⁴⁰ for accurate calculation of ^{13}C chemical shifts. The
294 intermediate GIAO DFT B3LYP/6-311++G(2d,2p) method is a
295 general-purpose technique for chemical shift calculation that has been
296 shown to provide reasonably accurate results for most Period II

Table 1. Computed Parameters of the ^{19}F Chemical Shift Tensor in 3-Fluorotyrosine

| method (geometry/NMR) | $ 2\sigma_{zz} - (\sigma_{xx} + \sigma_{yy}) $ /ppm | $ \sigma_{xx} - \sigma_{yy} $ /ppm |
|--|--|---------------------------------------|
| DFT B3LYP 6-311++G(2d,2p)/ CSGT DFT B3LYP cc-pVDZ | 238 | 44 |
| DFT B3LYP 6-311++G(2d,2p)/ GIAO DFT B3LYP 6-311++G(2d,2p) | 271 | 54 |
| DFT B3LYP 6-311++G(2d,2p)/ GIAO HF 6-311++G(2d,2p) | 210 | 44 |

297 elements. Although, in general, one would expect a better estimate from
298 a higher level method, ^{19}F chemical shift calculations are known to
299 disobey this rule, and the least computationally expensive in vacuo
300 GIAO HF/6-311++G(2d,2p) method yields values of the ^{19}F chemical
301 shifts that are superior to those obtained with DFT and MP2.⁴¹ We
302 therefore chose to use the GIAO HF value for the relaxation analysis.
303 DFT methods are known to overestimate ^{19}F shielding in fluorinated
304 aromatics,³⁸ as is also seen here.

305 Once the chemical shift tensor is known, measuring relaxation times
306 at two different magnetic fields allows extraction of the order parameters
307 and the correlation times. It is particularly convenient to use T_2 values
308 for fluorinated proteins, both because of signal-to-noise and degassing
309 difficulties associated with T_1 measurements and because, for signals
310 broader than approximately 50 Hz, accurate estimates of T_2 s may be
311 obtained simply from line widths (the broadening arising from field
312 inhomogeneity being negligible).

Results

313
314 **Mass Spectrometric Analysis of Labeled and Nonlabeled**
315 **GFPs.** The incorporation of the fluorinated tyrosine analogue
316 into wild-type and mutant GFP was analyzed using liquid
317 chromatography electro-spray ionization mass spectrometry (LC-
318 ESI-MS). In all cases, the major mass product obtained
319 corresponded to the protein with a fully cyclized chromophore
320 that had a loss of 20 Da consistent with dehydration and aerial
321 oxidation. The high-resolution HPLC step allowed the separation
322 of the small quantities of unlabeled protein from fluorinated
323 protein. In all cases, using peak-area integration, the expected
324 fluorinated protein was the major product ($> 95\%$ labeled) (data
325 not shown). In addition, the labeled and nonlabeled proteins
326 were enzymatically digested using papain and separated, and
327 the resulting peptides were identified using LC-ESI-MS and,
328 in each case, compared. No unlabeled peak was detected in the
329 fluorinated protein digest HPLC chromatogram, suggesting that
330 the ^{19}F –Tyr labeling efficiency was very high ($> 95\%$
331 fluorinated chromo-peptide).

332 **^{19}F NMR of Tyr \rightarrow Phe Mutants.** The ^{19}F assignments were
333 obtained using a combination of site-directed mutagenesis,
334 calculated solvent accessibilities and photo-CIDNP spectroscopy,
335 corroborated by NMR line-fitting analysis and line width
336 measurements. Wild-type GFP has 10 tyrosine residues distrib-
337 uted throughout the structure of the protein (Figure 1). The ^{19}F
338 NMR spectrum of wild-type GFP shows a distinctive spread of
339 ^{19}F chemical shifts from -50 to -60 ppm (Figure 2), as
340 expected for a folded protein. To assign the ^{19}F NMR spectrum,
341 each of the 10 tyrosine residues was mutated, one at a time, to
342 phenylalanine. If the mutation does not significantly perturb the
343 structure, the spectrum of each of the tyrosine mutants should
344 lack one or two signals compared to the spectrum of fully ^{19}F –

(34) Kuprov, I.; Goez, M.; Abbott, P. A.; Hore, P. *J. Rev. Sci. Instrum.* **2005**, *76*, 84–103.

(35) Goldman, M. *J. Magn. Res.* **2001**, *149*, 160–87.

(36) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.

(37) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.

(38) de Dios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1994**, *116*, 7453–4.

(39) Frisch, M. J.; G. W. et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(40) Tormena, C. F.; da Silva, G. V. *J. Chem. Phys. Lett.* **2004**, *398*, 466–470.

(41) Sanders, L. K.; Oldfield, E. *J. Phys. Chem. A* **2001**, *105*, 8098–8104.

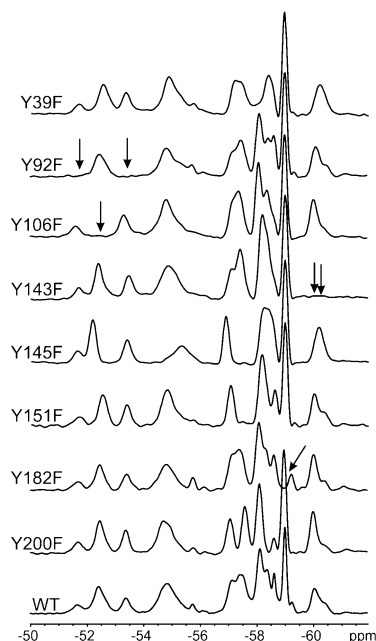


Figure 2. ¹⁹F spectra (564 MHz) of wild-type and Tyr → Phe mutant GFPs used for signal assignment. For Y92F, Y106F, Y143F, and Y182F, the arrows indicate the loss of a signal corresponding to that residue. For Y92F and Y143F, two signals are lost as these residues populate two distinct rotameric states (see section on ring-flip conformers).

Tyr labeled wild-type protein. As will be seen, the number of missing signals is determined by the existence of ring-flip conformers.

For Y92F, Y106F, Y143F, and Y182F, the spectra clearly lack one or two signals allowing these fluorotyrosines to be assigned immediately, Figure 2. In the other cases, however, the mutation resulted in noticeable chemical shift changes (e.g., the Y145F mutant, Figure 2), most likely a result of small structural perturbations caused by loss of the polar hydroxyl group when the 3-fluorotyrosine residue is mutated to phenylalanine. Note that because of the short relaxation delay between signal acquisitions, the NMR intensities in Figure 2 are weighted by the ¹⁹F relaxation rates and, therefore, vary from residue to residue.

For Y66, which forms part of the chromophore, the mutant protein (Y66F) could not be prepared in sufficient yield to acquire an NMR spectrum, most likely because the Tyr → Phe mutation in the chromophore disrupts the extensive hydrogen bond network and destabilizes the protein. However, Y66 may be assigned on the basis of its relaxation behavior. After the posttranslational cyclization, Y66 is incorporated into the chromophore, which is known to be a very rigidly immobilized structure.^{4,5} It should therefore be expected that Y66 would have the slowest motion of all 10 tyrosines and, therefore, the broadest NMR line. By far, the broadest signal in the spectrum is located around -55 ppm (Figure 2); we assign this peak to Y66. However, it is clear from the majority of the spectra in Figure 2 that this resonance is asymmetric, suggesting a contribution from another tyrosine residue. As this signal partially vanishes in the spectrum of the Y74F mutant (Figure 3), we attribute it to a superposition of peaks from Y66 and Y74.

Table 2 gives the approximate chemical shifts of the 11 resolvable ¹⁹F NMR lines in the wild-type spectrum (two lines for each of Y92 and Y143, one line for Y66/Y74, and one line for each of the other six tyrosine residues).

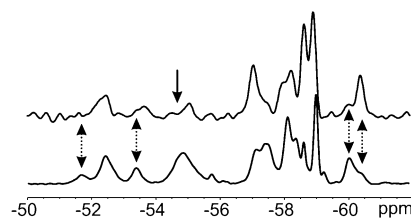


Figure 3. ¹⁹F NMR spectra of the Y74F mutant (above) and the wild-type protein (below). The Y74F spectrum was recorded at a different temperature (5 °C) and, consequently, shows a different ring-flip conformer distribution for Y92 and Y143 compared to that in Figure 2 (double-headed dashed arrows). Note the attenuated signal at around -55 ppm in the Y74F spectrum (single-headed solid arrow).

The as yet unassigned NMR signals, which fall in the range -57.1 to -58.4 ppm, are labeled A–D (Table 2); the unassigned residues at this point are Y39, Y145, Y151, and Y200.

Solvent Accessibility Data. The generation of CIDNP in tyrosine residues in proteins involves an electron transfer from an amino acid residue to the excited photosensitizer,⁴² a process that only occurs with the necessary kinetics if the residue is exposed to the bulk solvent.^{43,44} Solvent accessibility data were therefore calculated using the web-based program GetArea⁴⁵ and the crystal structure of GFP⁶ (PDB code 1b9c), using either 0.14 or 0.30 nm as the radius of the solvent probe (to model accessibility by water and the flavin, respectively). Both whole-residue accessibilities and atomic accessibilities were calculated. Although there have been attempts in the past to correlate CIDNP intensities with whole-residue accessibilities,^{46,23} it is more likely to be the accessibility of the orbital from which the electron is removed that determines the magnitude of the nuclear polarization. Accordingly, the solvent accessibilities of the highest occupied molecular orbital (HOMO) of the 10 tyrosine residues were computed as the weighted sum of atomic accessibilities, the weights corresponding to the HOMO density on each atom.

Only four of the 10 tyrosines (Y39, Y151, Y182, Y200) have significant HOMO accessibilities (21.7, 12.8, 11.8, and 9.9%, respectively); the others are essentially buried (accessibility ≤ 1%). This result suggests that only Y39, Y151, Y182, and Y200 should show significant polarization in a photo-CIDNP experiment.

Photo-CIDNP of the Native State. ¹⁹F photo-CIDNP spectra were recorded for wild-type and mutant GFP proteins. GFP absorbs strongly at the usual photo-CIDNP excitation wavelength of 488 nm, so the illumination system was modified to use fiber optics specifically designed to deal with optically dense samples.³¹ It was found that the UV lines of the argon ion laser degrade GFP, so the experiment was further adjusted to operate in a single-line, rather than multi-line, laser mode at a wavelength of 514 nm.

Most of the mutants exhibit four emissively polarized overlapping peaks in the ¹⁹F photo-CIDNP spectrum (Figure 4). In addition to signals attributable, by their chemical shifts and line widths, to the native states of the proteins, there appears to be a small admixture of unfolded, misfolded, or degraded

(42) Tsentalovich, Y. P.; Lopez, J. J.; Hore, P. J.; Sagdeev, R. Z. *Spectrochim. Acta, Part A* **2002**, *58*, 2043–2050.

(43) Berliner, L. J.; Kaptein, R. *Biochemistry* **1981**, *20*, 799–807.

(44) Kaptein, R. J. *Chem. Soc. D: Chem. Commun.* **1971**, *14*, 732–733.

(45) Fraczkiewicz, R.; Braun, W. J. *Comput. Chem.* **1998**, *19*, 319–333.

(46) Feeney, J.; Roberts, G. C. K.; Kaptein, R.; Birdsall, B.; Gronenborn, A. M.; Burgen, A. S. V. *Biochemistry* **1980**, *19*, 2466–2472.

Table 2. ^{19}F Chemical Shifts and Partial Assignments for the 10 ^{19}F -Tyr Residues in Wild-Type Fluorotyrosinated GFP

| chemical shift/ppm assignment | -51.7 Y92(I) | -52.4 Y106 | -53.4 Y92(II) | -54.9 Y66/74 | -57.1 A | -57.4 B | -58.1 C | -58.4 D | -59.0 Y182 | -60.1 Y143(I) | -60.4 Y143(II) |
|----------------------------------|-----------------|---------------|------------------|-----------------|------------|------------|------------|------------|---------------|------------------|-------------------|
|----------------------------------|-----------------|---------------|------------------|-----------------|------------|------------|------------|------------|---------------|------------------|-------------------|

Table 3. Assignment of the ^{19}F Chemical Shifts, the Values of the Generalized Order Parameters, the Residue Accessibilities, and HOMO Accessibility Data for the 10 ^{19}F -Tyr Residues in Wild-Type Fluorotyrosinated GFP^a

| residue | chemical shift ^b ppm | S^2 | SASA ^c % | SASA ^d % | HOMO accessibility ^e |
|---------------|---------------------------------|--------------------|---------------------|---------------------|---------------------------------|
| Tyr92(I) | -51.65 ± 0.04 | 0.58 ± 0.03 | 0.6 | 0.0 | 0.0 |
| Tyr106 | -52.40 ± 0.08 | 0.61 ± 0.04 | 0.0 | 0.0 | 0.0 |
| Tyr92(II) | -53.36 ± 0.04 | 0.51 ± 0.03 | 0.6 | 0.0 | 0.0 |
| Tyr66/74 | -54.86 ± 0.10 | 1.00 ^f | 0.5/1.9 | 0.0 | 0.0 |
| Tyr145 | -57.06 ± 0.04 | 0.49 ± 0.04 | 6.4 | 2.6 | 0.0 |
| Tyr151 | -57.41 ± 0.06 | 0.40 ± 0.03 | 35.3 | 21.8 | 12.8 |
| Tyr39 | -58.11 ± 0.04 | 0.38 ± 0.03 | 45.0 | 35.1 | 21.7 |
| Tyr200 | -58.36 ± 0.02 | 0.41 ± 0.06 | 21.2 | 12.8 | 9.9 |
| Tyr182 | -58.98 ± 0.02 | 0.21 ± 0.02 | 31.4 | 17.3 | 11.8 |
| Tyr143(I) | -60.08 ± 0.06 | 0.44 ± 0.04 | 22.8 | 10.6 | 1.0 |
| Tyr143(II) | -60.37 ± 0.06 | 0.44 ± 0.04 | 22.8 | 10.6 | 1.0 |

^a The residues whose signals appear in the photo-CIDNP spectra of the native protein are shown in bold. ^b The signal of residual free 3-fluorotyrosine was used as a chemical shift reference, at -58.60 ppm. ^c Solvent accessible surface area (SASA) calculated using a 0.14 nm probe. ^d SASA calculated using a 0.30 nm probe. ^e Atomic accessibility weighted by the Mulliken atomic populations for the HOMO (see Supporting Information), the latter obtained from a DFT B3LYP 6-311G(d,p)** ab initio calculation using the GAMESS program.⁴⁷ The assignments of Y39 and Y200 are tentative, as described in the text. ^f The Y66 residue is assumed to be rigidly immobilized in the protein core.

422 GFP (the sharp signal marked by a cross in the spectra of Y200F
423 and Y151F in Figure 4). The exact nature of the admixture is
424 unknown; it could not be separated by HPLC or dialysis, appears
425 to have higher side chain mobility than GFP (hence the sharper
426 NMR signal) and to have much greater solvent accessibility of
427 its fluorotyrosine(s) (accounting for the strong signal in the
428 CIDNP spectra).

429 The polarized signal at -58.98 ppm (wild-type chemical shift)
430 was previously assigned to Y182 (Figure 2) and is one of the
431 four residues with significant HOMO accessibility (11.8%).
432 Comparison of the various CIDNP spectra in Figure 4 indicates
433 that B–D are polarizable but that A is not. Peak A can therefore

be assigned to Y145, the only unassigned residue with negligible
HOMO accessibility.

Peak B is absent from the NMR and CIDNP spectra of the
Y151F mutant and can thus be assigned to Y151. The two
remaining peaks, C and D, cannot unambiguously be assigned
to the remaining two residues (Y39 and Y200). Judging by the
line shapes of the NMR peaks close to -58.2 ppm in the Y200F
and Y151F spectra, it appears that D is absent from the former,
suggesting that D should be assigned to Y200 and, by elimina-
tion, C to Y39. However, C and D are so strongly overlapping
in the Y151F spectrum that this conclusion cannot be definitive.

Table 3 gives the final assignments, the chemical shifts
obtained from line-fitting, side chain accessibilities (which will
be discussed further below), and the values of the generalized
order parameters obtained from a relaxation analysis (see below).
The standard deviations reported in Table 3 reflect the range of
chemical shifts observed for each residue in the different
mutants. Figure 5 shows the results of the spectral line-fitting
procedure, together with the assignments. Note that because of
the short relaxation delay between signal acquisitions, the NMR
intensities in this spectrum, as in Figures 2–4, are weighted by
the ^{19}F relaxation rates and, therefore, vary from residue to
residue.

^{19}F Relaxation Data. The chromophore of GFP is a rigidly
immobilized structure^{4,5} which must therefore tumble with the
overall molecular correlation time of the protein. It would
therefore be reasonable to assume that it has a unit order
parameter, $S^2 = 1$. Using this assumption and the ab initio ^{19}F
shielding tensor parameters given in Table 1, it is possible to
extract the overall molecular rotation correlation time τ_c and
the order parameters for each residue from the experimental T_2
data at 377 and 564 MHz. The resulting value of τ_c is 14.8 ns,
in excellent agreement with the value of 13.5 ns obtained from
a HYDRONMR⁴⁸ calculation using the X-ray diffraction

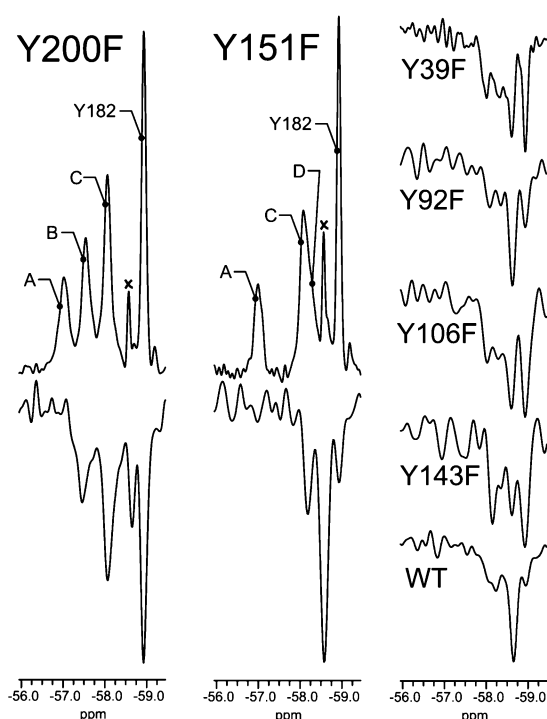


Figure 4. ^{19}F photo-CIDNP spectra of GFP mutants and the wild-type protein. In all cases, the ^{19}F polarizations are emissive. The NMR spectra of Y200F and Y151F are shown above the corresponding photo-CIDNP spectra (on the left of the Figure).

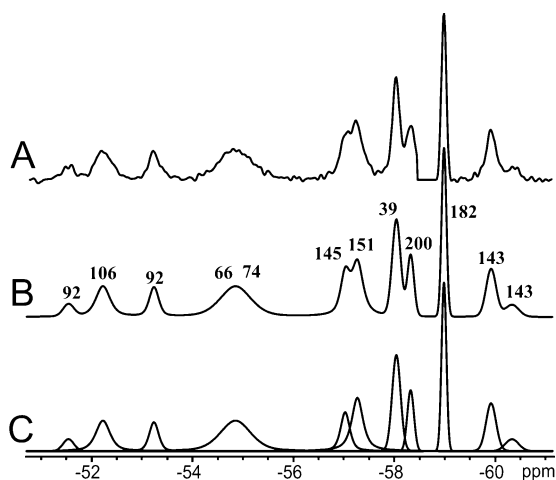


Figure 5. (A) ¹⁹F spectrum (564 MHz) of GFP with Lorentzian-to-Gaussian resolution enhancement. (B) Line-fitting. (C) Individual lines in the fitting. The narrow resonance of free 3-fluorotyrosine at -58.7 ppm has been zeroed to facilitate least squares line-fitting. The assignments of Y39 and Y200 are tentative, as described in the text.

468 structure (PDB code 1B9C). Resulting values of S^2 for the other
 469 nine fluorotyrosine residues of GFP are listed in Table 3. The
 470 four residues that appear in the photo-CIDNP spectra, namely
 471 Y39, Y151, Y182, and Y200, have the narrowest NMR signals
 472 and have order parameters in the range 0.21–0.41; the other
 473 six residues, which have near-zero HOMO accessibilities, all
 474 have larger S^2 (0.44–0.61). Moreover, the same four fluoroty-
 475 rosines correspond to the ¹⁹F resonances with chemical shifts
 476 closest to that of 3-fluorotyrosine (-58.7 ppm). Taken together,
 477 these results corroborate the assignments presented above.
 478 Residues with high HOMO accessibilities are close to the
 479 surface of the GFP molecule, are likely to be less constrained
 480 by neighboring side chains, and should therefore exhibit greater
 481 internal mobility.

482 **Ring-Flip Conformers.** An interesting feature of 3-fluoro-
 483 tyrosinated proteins is the existence of fluorotyrosine ring-flip
 484 conformers. Introduction of fluorine into the phenyl moiety of
 485 tyrosine breaks its symmetry, and therefore, two ¹⁹F signals per
 486 fluorotyrosine residue can be expected, because the chemical
 487 environment of the two possible fluorine positions will, in
 488 general, be different.

489 Two such signal pairs (Y92 and Y143) are observed in the
 490 wild-type GFP spectrum (Figure 2); the rightmost pair, belong-
 491 ing to Y143, displays characteristic behavior: the relative
 492 intensity of the two signals varies from mutant to mutant, the
 493 signals sometimes coalesce into one apparent signal, and both
 494 signals vanish simultaneously when the residue is mutated (as
 495 in the Y143F spectrum, Figure 2). The same behavior is also
 496 demonstrated by Y92; in this case, however, the signals of the
 497 two conformers are substantially further apart. All spectra also
 498 contain from two to five low-intensity peaks of variable
 499 intensity, which are likely to be weakly populated rotamers of
 500 the other fluorotyrosine residues. On the basis that the ring-
 501 flipping must be slow compared to the difference in ¹⁹F NMR
 502 frequencies to observe separate lines from the two conformers,
 503 we can put upper limits on the flipping rate of ~ 160 s⁻¹ for
 504 Y143 and ~ 960 s⁻¹ for Y92. In the relaxation analysis described
 505 above, we have assumed that the exchange is slow enough that
 506 its effect on the line widths can be neglected. Any significant

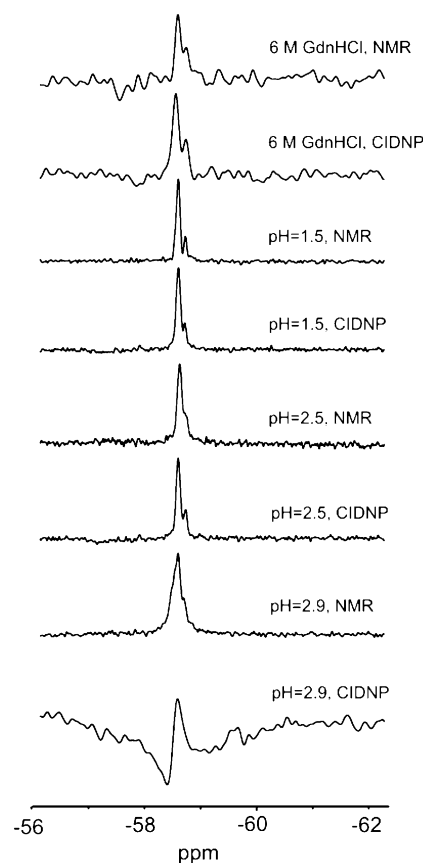


Figure 6. ¹⁹F NMR and photo-CIDNP spectra of denatured states of GFP. The NMR spectra were recorded with 256 scans, the CIDNP spectra with 16 scans.

contribution from chemical exchange would have made the lines 507
 asymmetric and non-Lorentzian. 508

Photo-CIDNP of Denatured GFP. ¹⁹F photo-CIDNP spectra 509
 were recorded for denatured samples of GFP, in acid solution 510
 (pH 1.5, 2.5, and 2.9) and in 6 M GdnDCI at pH 7.2. These 511
 conditions were chosen to replicate the initial conditions in 512
 ongoing fluorescence kinetic studies. 513

The ¹⁹F NMR spectra of GdnDCI and acid-denatured states 514
 are shown in Figure 6 along with the corresponding photo- 515
 CIDNP spectra. In all cases, two peaks are clearly resolved; 516
 the major peak corresponds to the nine fluorotyrosines that are 517
 evidently in chemically equivalent average environments in the 518
 denatured state. The minor peak results from the single 519
 fluorotyrosine (Y66) that forms part of the chromophore and 520
 has a different chemical environment due to the chemical 521
 modification of the side chain during chromophore formation. 522
 These results suggest that the denatured states (at least in terms 523
 of the chemical environments of the tyrosines) are similar in 524
 the four different denaturing conditions. 525

However, the photo-CIDNP spectra (Figure 6) provide more 526
 information on the nature of these denatured states and show 527
 that the pH 2.9 denatured state is qualitatively different from 528
 the other three denatured states. 529

Extensive experiments on 3-fluorotyrosine (not shown) reveal 530
 that for small motional correlation times (less than about 1 ns), 531
 the sign of the ¹⁹F CIDNP enhancement conforms to Kaptein's 532
 rules⁴⁴ and the ¹⁹F is absorptively polarized. For correlation 533
 times greater than about 1 ns at 14.1 T, the polarization changes 534
 sign to emission. Briefly, the origin of this effect has been traced 535

536 to a cross-correlation relaxation pathway involving the ^{19}F
 537 hyperfine tensor and the g -tensor of the intermediate fluoroty-
 538 rosyl radical, which becomes fast enough when $\tau_e < 1$ ns to
 539 interfere with the production of CIDNP in the fluorotyrosyl-
 540 flavin radical pair. The sign of the ^{19}F CIDNP polarization
 541 therefore provides a probe of the local motional correlation time.
 542 Full details of this phase change will be published separately
 543 in due course.

544 This relaxation effect explains the phases of the ^{19}F enhance-
 545 ments shown in Figures 4 and 6. The surface tyrosine residues
 546 of the folded GFP molecule have negative CIDNP enhance-
 547 ments, indicating that the effective rotational correlation time
 548 is larger than approximately 1 ns. In the unfolded states,
 549 however, the enhancement is positive, reflecting much less
 550 constrained motion of the polypeptide chain. The partially folded
 551 pH 2.9 state features a sharp positive signal (an unfolded, or
 552 less restricted ensemble of structures) on top of a broad negative
 553 signal (a folded, or more restricted ensemble).

554 Discussion

555 **^{19}F -fluorotyrosine Labeling Strategy.** Many strategies now
 556 exist for the incorporation of nonnatural amino acids into
 557 proteins.^{49–51} Here, we have modified one of these methods²⁹
 558 to incorporate 3-fluorotyrosine efficiently into GFP for ^{19}F NMR
 559 studies. The method uses glyphosate to inhibit the biosynthesis
 560 of aromatic amino acids in *E. coli*, the culture media being
 561 supplemented with fluorinated tyrosine and nonfluorinated
 562 phenylalanine and tryptophan. The same method can be used
 563 to introduce fluorinated phenylalanine and tryptophan (F.K. and
 564 S.E.J., unpublished results). The efficiency of labeling and the
 565 extent of incorporation of labeled tyrosine into GFP was assessed
 566 using LC-MS and shown to be greater than 95%. Interestingly,
 567 Budisa's group also successfully replaced the tyrosines in the
 568 enhanced variant of GFP (eGFP) with 2- and 3-fluorotyrosine
 569 derivatives using a Tyr-auxotrophic strain of *E. coli*, work which
 570 was published recently.⁵² We also tried this method but were
 571 less successful. In their study, Budisa and co-workers crystal-
 572 lized the 3-fluorotyrosine-labeled eGFP and found the structure
 573 to be indistinguishable from the nonlabeled GFP.⁵³

574 **Characterization of the Native State of GFP by ^{19}F NMR.**
 575 The ^{19}F NMR spectrum of ^{19}F -Tyr-labeled wild-type GFP
 576 shows multiple peaks corresponding to the 10 tyrosine residues
 577 of GFP, and the spectrum shows a distinctive spread of chemical
 578 shifts as expected for a natively folded protein. A full assignment
 579 was made using a combination of Tyr \rightarrow Phe mutants, photo-
 580 CIDNP spectroscopy, and accessibility calculations, corroborated
 581 by relaxation data. Two of the tyrosine residues (Y92 and
 582 Y143) were each found to give rise to two ^{19}F peaks corre-
 583 sponding to the two ring-flip conformers, suggesting that the
 584 side chains of these residues are in a rigid environment. This is
 585 not unexpected, as these side chains are buried in the native
 586 structure. Surprisingly, the signals from Y106 and Y145, which
 587 are also buried (Table 3), do not appear to be split to such a
 588 degree (there is, however, evidence of minor populations of other
 589 ring-flip conformers). This may be because the two conformers

590 have fortuitously identical chemical shifts (although this seems
 591 unlikely), or because one conformer is highly favored over the
 592 other due to specific packing interactions in the native state.
 593 Alternatively, although the crystal structure indicates that these
 594 side chains are buried and rigidly held, local breathing motions
 595 might conceivably allow the conformers to interconvert rapidly.
 596 Nevertheless, the splitting of the ^{19}F signal for Y92 and Y143
 597 gives a sensitive probe of the local conformational flexibility
 598 of the protein, which can be utilized in future unfolding/folding
 599 experiments.

600 Further experiments were undertaken to characterize the
 601 native and denatured states of the protein under a variety of
 602 experimental conditions. Four fluorotyrosine residues were
 603 found to exhibit nuclear polarization and were assigned using
 604 the calculated solvent accessibilities of the molecular orbitals
 605 responsible for donating electrons to the photoexcited triplet
 606 flavin. Only four of the tyrosine residues have significant
 607 HOMO accessibilities ($\geq 10\%$), the others being $\leq 1\%$.
 608 Previous attempts to correlate whole-residue solvent accessibili-
 609 ties (SASA) with observed photo-CIDNP effects have not
 610 always been totally successful.²³ In the case of GFP, we also
 611 find that the SASAs, whether calculated with a 0.14 nm or a
 612 0.30 nm probe, are qualitatively in disagreement with the
 613 observed CIDNP effects. The SASA data in Table 3 suggest
 614 that five, rather than four, fluorotyrosine residues should be
 615 polarizable in GFP, i.e., Y39, Y151, Y182, Y200, and Y143.
 616 Although Y143 has very similar overall solvent accessibility
 617 to Y200, the HOMO accessibilities differ by an order of
 618 magnitude. Inspection of the crystal structure⁶ shows that the
 619 part of Y143 which protrudes into the solvent is an unreactive
 620 $\text{C}_\beta\text{H}_2\text{-C}_\alpha\text{H-NH}$ fragment that bears little HOMO electron
 621 density, whereas it is the reactive aromatic side chain that is
 622 exposed for Y200.

623 **Characterization of the Denatured States of GFP.** In
 624 general, it is considerably more difficult to obtain detailed
 625 structural information on the denatured states of proteins than
 626 on native states. Here, we have used ^{19}F photo-CIDNP to probe
 627 the environment of the tyrosine side chains of GFP under
 628 unfolding conditions (low pH and high concentrations of
 629 chemical denaturant). The photo-CIDNP spectra of denatured
 630 GFP at pH 1.5, 2.5 and 6 M GdnDCI are all very similar (Figure
 631 6) and exhibit a positive enhancement. In contrast, the pH 2.9
 632 denatured state exhibits both a positive and a negative polariza-
 633 tion. As the CIDNP amplitude is related to the correlation time,
 634 as summarized above, it would appear that there is some
 635 heterogeneity in the size/mobility of the molecule at this pH.
 636 This could be due to two structures in equilibrium, one more
 637 compact than the other. This is consistent with far-UV circular
 638 dichroism studies of the pH 2.9 denatured state which has
 639 considerably more secondary structure than the denatured states
 640 at pH 1.5 or in 6 M GdnDCI, see Supporting Information.
 641 Residual structure, such as this, may be very important in the
 642 folding mechanism of GFP, effectively restricting conforma-
 643 tional space at a very early stage during the folding process. It
 644 may also explain the results of folding measurements on GFP
 645 that suggest that folding is considerably faster from the pH 2.9
 646 denatured state than from the pH 1.5 denatured state (data not
 647 shown).

648 In conclusion, we have fully assigned the ^{19}F NMR spectrum
 649 of GFP including the identification of two fluorotyrosine ring-

(49) Bayley, H.; Jayasinghe, L. *Mol. Membr. Biol.* **2004**, *21*, 209–220.

(50) Hahn, M. E.; Muir, T. W. *Trends Biochem. Sci.* **2005**, *30*, 26–34.

(51) Petersson, E. J.; Brandt, G. S.; Zacharias, N. M.; Dougherty, D. A.; Lester, H. A. *Methods Enzymol.* **2003**, *360*, 258–273.

(52) Pal, P. P.; Bae, J. H.; Azim, M. K.; Hess, P.; Friedrich, R.; Huber, R.; Moroder, L.; Budisa, N. *Biochemistry* **2005**, *44*, 3663–3672.

650 flip conformers which provide important probes of the dynamics
651 of the protein. Further, we have shown that photo-CIDNP
652 techniques can be successfully applied to ¹⁹F-labeled proteins
653 and provide important structural information on denatured states.
654 A strong correlation has been established between the observed
655 CIDNP effect and the solvent accessibility of the HOMO of
656 the tyrosines. The sign and amplitude of the CIDNP effects have
657 proved particularly useful in identifying differences in the four
658 denatured states of GFP, providing important structural informa-
659 tion complementary to other spectroscopic methods.

660 **Acknowledgment.** This work was funded in part by the
661 Welton Foundation. F.K. and T.D.C. were funded by MRC and
662 BBSRC Ph.D. studentships, respectively. F.K., T.D.C., and
663 S.E.J. thank Prof. Sir Alan Fersht and Prof. Chris Dobson for
664 access to biophysical equipment and helpful discussions. I.K.

thanks the Scatcherd European Foundation and Hill Foundation 665
for a Ph.D. studentship. I.K. and P.J.H. thank the Oxford 666
Supercomputing Centre for a generous allocation of CPU time. 667
P.J.H. acknowledges the financial support of INTAS (Project 668
No. 02-2126), the Royal Society (International Joint Project 669
Grant Program), and the BBSRC. 670

Supporting Information Available: Expression and purifica- 671
tion of fluorotyrosine labeled GFP mutants; far-UV circular 672
dichroism spectra; complete list of authors for ref 39; atomic 673
accessibilities and Mulliken atomic populations for the HOMO 674
of tyrosine. This material is available free of charge via the 675
Internet at <http://pubs.acs.org>. 676

JA060618U 677