# Density Functional and MP2 Calculations of Spin Densities of Oxidized 3-Methylindole: Models for Tryptophan Radicals

**G. M. Jensen,[†] D. B. Goodin,[†] and S. W. Bunte*,[‡]**

*Department of Molecular Biology, MB8, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037, and U.S. Army Research Laboratory, Attn: AMSRL-WT-PC, Aberdeen Proving Ground, Maryland, 21005-5066*

*Received: June 28, 1995; In Final Form: October 11, 1995*[⊗]

*Ab initio* calculations have been carried out on 3-methylindole, and the cation and neutral radicals of 3-methylindole, using density functional theory (DFT), the Becke3−Lee−Yang−Parr functional, and the 6-31G*, 6-31+G*, 3-21G*, and TZ2P basis sets. Optimized geometries, vibrational frequencies, and for the radicals, atomic spin densites are calculated. The latter are compared to experimental spin densities recently determined for the tryptophan-191 radical of compound ES of the enzyme cytochrome-*c*-peroxidase. The DFT spin densities for the cation radical of 3-methylindole are in excellent agreement with the data for the tryptophan-191 radical, which supports the conclusion that the tryptophan radical is a cation radical. The results are compared to calculations using second-order Møller−Plesset theory (MP2) and the 6-31G** basis set. The MP2 spin densities are in significantly worse agreement with the experimental spin densities.

## Introduction

The peroxidases comprise a class of enzymes that catalyze the $H_2O_2$-mediated oxidation of an enormous range of biological substrates.[1] The heme-containing enzyme cytochrome-*c*-peroxidase (CCP) from yeast reacts with $H_2O_2$ to form a two-electron-oxidized intermediate: compound ES. This intermediate in turn accepts electrons sequentially from two ferrous cytochromes *c*. One oxidizing equivalent in compound ES is stored as an oxyferryl (Fe(IV)═O) heme, while the other is stored as a radical species on an amino acid side chain. This is distinct from many other peroxidases, where the second oxidizing equivalent is stored on the porphyrin macrocycle as a $\pi$-cation radical. Identification of this amino acid and the nature of the radical proved to be problematical, because the presence of a multiplicity of spin-coupled states between the $S = 1/2$ radical with the $S = 1$ oxyferryl heme[2] results in an unusual EPR line shape and EPR properties. Intensive studies of compound ES of CCP by site-directed mutagenesis, EPR and ENDOR spectroscopies, X-ray crystallography, and theoretical calculations have concluded unambiguously that the radical is housed on the indole side chain of tryptophan-191.[2,3] It has also been difficult to determine whether the tryptophan radical is a neutral radical or a cation radical, resulting from oxidation with or without loss of a proton, respectively. Indole radicals are formed at highly oxidizing potentials (∼1 V vs SHE) and exist in solution as an equilibrium of neutral and cationic forms.[4] The p$K_a$ of this equilibrium is ∼4, and therefore the presence of a cation radical in CCP at physiological pH must involve some special stabilization of the cationic form of the radical by the protein. Indirect evidence of a cation-stabilizing propensity at the site of the indole side chain of tryptophan-191 has been obtained in the observation that cationic heterocyclic molecules and buffer cations bind to a cavity created by mutating the indole side chain to hydrogen (tryptophan-191 → glycine).[5] Complete characterization of the properties of the radical of CCP is required for a detailed understanding of the mechanism of action of this enzyme. Further, protein engineering, especially at the site of this radical, may lead to modified enzymes capable of alternative oxidation pathways for organic synthesis or bioremediation.

In the most recent ENDOR study of CCP,[3f] the utilization of various isotopically labeled (at tryptophan) CCPs has led to the assignment of spin densities for some atoms in the indole ring of the tryptophan-191 radical. To aid in the interpretation of these ENDOR results in terms of a cation or neutral radical, one must adduce reliable theoretical models for both forms. Calculations of spin densities that have been used for this purpose have so far been limited to Hückel−McLachlan molecular orbital (HMMO)[6] and restricted to open-shell Hartree−Fock (ROSHF)[3d] calculations, which are now known to be unlikely to provide reliable spin densities for $\pi$-radicals.[7]

Accurate *ab initio* methods for the determination of ground state molecular properties generally, and spin densities in particular, will require inclusion of electron correlation, sufficient basis set size and sophistication, and for medium sized organic molecules, practicability in terms of disk space, computer time, and code availability. The most popular correlated *ab initio* method has utilized the second-order Møller−Plesset (MP2) perturbation theory. MP2 methods are, however, very expensive computationally, and most medium sized organic molecules of interest are currently beyond the bounds of practicability. Very recently, methods using density functional theory (DFT) have become available in the program Gaussian92/DFT[8] and provide correlated methods of high accuracy. These include standard local functionals[9] as well as two new hybrid density functionals (B3LYP and B3P86)[10] of the type introduced by Becke.[10a] Stephens and co-workers[11] have calculated the geometries, force fields, and vibrational absorption and vibrational circular dichroism (VCD) spectra of a variety of closed-shell molecules using the aforementioned functionals in addition to using SCF and MP2 methods. Taking advantage of the sensitivity of predicted VCD spectra to the quality of the force field used, they were able to demonstrate that the hybrid functional B3LYP was competitive with MP2 methods in terms of accuracy and that the LSDA and BLYP functionals were significantly less reliable. They also demonstrated that for B3LYP calculations

* To whom correspondence should be addressed.
[†] The Scripps Research Institute.
[‡] U.S. Army Research Laboratory.
[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1995.

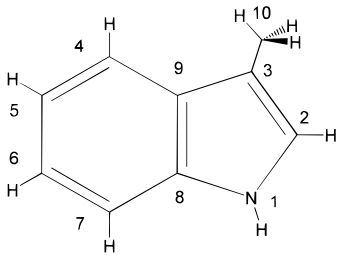


**Figure 1.** Atom numbering for 3-methylindole (skatole).

the 6-31G* basis set[12] is sufficient for accurate results. DFT calculations of open-shell systems have yielded good results for a variety of properties.[7,13] Where the new hybrid functionals have been used on open-shell systems, Adamo et al.[7a] conclude that for ground state geometries and for thermochemical, infrared, and hyperfine parameters of a variety of molecules, the B3LYP method is the current functional of choice and that the local functionals are significantly less reliable. On the other hand, Qin and Wheeler noted that both hybrid and local functionals provided reasonable spin densities for phenoxy radical[7b] at the 6-31G* basis set level.

With the aforementioned studies in mind, we have chosen to carry out DFT calculations using the B3LYP functional and the 6-31G* basis set on the tryptophan analog 3-methylindole (skatole) (**1**) (Figure 1), 3-methylindole neutral radical (**2**), and 3-methylindole cation radical (**3**). Geometries, vibrational frequencies, and spin densities are calculated. To examine the basis set dependence of the B3LYP results, we have also calculated the geometries and atomic spin densities for **1**, **2**, and **3** utilizing the 6-31+G*, 3-21G*, and TZ2P basis sets. For comparison, the geometries and spin densities are also calculated at the MP2 level and the 6-31G** basis set. These results are compared to experimental data and to earlier calculations.

## Methods

Calculations presented here were carried out using the GAUSSIAN92/DFT or GAUSSIAN94[8] suite of programs running on a Cray2 or a Silicon Graphics Inc. Power Challenge Array located at the U.S. Army Research Laboratory. The starting geometry for 3-methylindole was a 3-21G* optimized geometry that was calculated at the SCF level of approximation using CADPAC 5.0[14] (in UNICHEM) on the Scripps Cray YMP. The geometries, Mulliken spin densities, and harmonic force fields were calculated using the spin-restricted B3LYP functional for the closed-shell 3-methylindole (skatole) (**1**) and the spin-unrestricted B3LYP functional for the open-shell neutral radical of 3-methylindole (**2**) and the open-shell radical cation of 3-methylindole (**3**). B3LYP geometries for **1**, **2**, and **3** and atomic spin densities for **2** and **3** were determined using the TZ2P, 6-31G*, 6-31+G*, and 3-21G* basis sets. In addition, the vibrational frequencies of **1**, **2**, and **3** were calculated using the 6-31G* basis set. The "fine" integration grid was used in all of the DFT calculations. *Ab initio* MO geometries and Mulliken spin densities were also calculated at the spin-restricted MP2 level of theory for **1** and at the spin-unrestricted MP2 level of theory for **2** and **3**. Unfortunately, we were unable to obtain MP2 vibrational frequencies due to a lack of disk resources. All of the MP2 calculations were obtained using the frozen core approximation and the 6-31G** basis set. Because **2** and **3** each have a doublet multiplicity, the expectation value of the $S^2$ operator is 0.75. However, as the UHF wave function is not an eigenfunction of the $S^2$ operator, it is usually contaminated with higher multiplicities (quartets, etc.), resulting in $\langle S^2 \rangle$ being greater than 0.75. In the case of the B3LYP calculations, $\langle S^2 \rangle$ was determined to range from 0.7596 to 0.7694 for **2** and **3** (see Table 1) for all basis sets employed, indicating a good representation of the doublet state. We note that for the DFT technique, the interpretation of $\langle S^2 \rangle$ is not straightforward[7b] because the calculations yield electron densities, not electronic wave functions. For the MP2 calculations on **2** and **3**, $\langle S^2 \rangle$ was calculated to be 0.9476 and 0.8478, respectively, indicating a somewhat worse representation of the doublet state relative to the B3LYP calculations.

## Results and Discussion

B3LYP geometries and energies of **1**, **2**, and **3**, are shown for each of the basis sets employed in this study in Table 1. The B3LYP/TZ2P geometries for the indole portion of **1** are compared to the geometry reported by Chadwick,[15] which was obtained by averaging the bond angles and distances of the structures of 34 3-substituted indoles from the Cambridge Crystallographic Database. Overall, the geometry of **1** compares extremely well to experiment, with an average difference between the calculated and experimental geometries of 0.005 Å. Generally, the predicted geometry exhibits longer bonds than the experimental geometry. The largest deviation between theory and experiment is for the C6−C7 bond, which is 0.013 Å longer than experiment. The results obtained using the 3-21G*, 6-31G*, and 6-31+G* basis sets lead to geometries differing from experiment, on average, by 0.013, 0.010, and 0.011 Å, respectively. The MP2/6-31G** geometries and energies for **1**, **2**, and **3**, are also shown in Table 1. The MP2/6-31G** geometry for **1** is very similar to the B3LYP/TZ2P and B3LYP/6-31G* geometries, differing by an average of only 0.005 and 0.003 Å, respectively. The MP2/6-31G** geometry for **1** differs from that of Chadwick[15] by 0.009 Å, with the largest deviation being 0.017 Å for the C6−C7 bond. The previously reported ROSHF[3d] geometry for indole differs from experiment by an average of 0.020 Å. To our knowledge, there are no experimental geometries available for **2** or **3**. Both B3LYP/TZ2P and MP2/6-31G** calculations predict significant changes upon oxidation from **1** to either **2** or **3**. These predicted differences are qualitatively similar but quantitatively different between the two methods. Going from **1** to **2** the B3LYP/TZ2P prediction shows a lengthening of the C2−C3 bond by 0.069 Å and a shortening of the N1−C2 bond by 0.068 Å. Going from **1** to **2** the MP2 prediction shows lengthening of the C2−C3 and C3−C9 bonds by 0.028 and 0.030 Å, respectively, and a shortening of the N1−C2 bond by 0.062 Å. Going from **1** to **3** the B3LYP/TZ2P prediction shows a lengthening of the C2−C3, C8−N1, and C6−C7 bonds by 0.061, 0.032, and 0.027 Å, respectively, and a shrinking of the N1−C2, C3−C9, and C7−C8 bonds by 0.053, 0.021, and 0.024 Å, respectively. Going from **1** to **3** the MP2/6-31G** prediction shows a lengthening of the C2−C3, C3−C9, and C4−C5 bonds by 0.028, 0.024, and 0.027 Å, respectively, and a shrinking of the N1−C2, and C9−C4 bonds of 0.029 and 0.028 Å, respectively.

To our knowledge, there is no complete, assigned, experimental vibrational spectrum of **1**, and for **2** or **3** there exists no information at all. However, several modes for tryptophan have been assigned and experimental frequencies measured.[16] These are shown in Table 2. B3LYP/6-31G* frequencies corresponding to these modes for **1**, **2**, and **3** are also shown in Table 2. The "zero" frequencies were all less than 10 $cm^{-1}$ in these calculations. Comparison of the differences between the predictions for **1** and the experimental frequencies reveals very good agreement, with the predicted frequencies differing by an

**TABLE 1: Geometries and Energies of 3-Methylindoles[a]**

| | | B3LYP/TZ2P | | | B3LYP/6-31G* | | | B3LYP/6-31+G* | | |
|---|---|---|---|---|---|---|---|---|---|---|
| bond | exptl[b] | **1** | **2** | **3** | **1** | **2** | **3** | **1** | **2** | **3** |
| N1−C2 | 1.380 | 1.383 | 1.315 | 1.330 | 1.385 | 1.320 | 1.336 | 1.387 | 1.321 | 1.336 |
| C2−C3 | 1.358 | 1.366 | 1.435 | 1.427 | 1.371 | 1.439 | 1.431 | 1.373 | 1.440 | 1.431 |
| C3−C9 | 1.436 | 1.439 | 1.434 | 1.418 | 1.443 | 1.438 | 1.423 | 1.444 | 1.439 | 1.425 |
| C9−C4 | 1.398 | 1.401 | 1.391 | 1.406 | 1.406 | 1.396 | 1.412 | 1.407 | 1.398 | 1.412 |
| C4−C5 | 1.381 | 1.384 | 1.399 | 1.392 | 1.389 | 1.404 | 1.398 | 1.391 | 1.406 | 1.400 |
| C5−C6 | 1.396 | 1.405 | 1.391 | 1.392 | 1.410 | 1.396 | 1.398 | 1.412 | 1.398 | 1.398 |
| C6−C7 | 1.372 | 1.385 | 1.406 | 1.412 | 1.390 | 1.411 | 1.418 | 1.392 | 1.413 | 1.419 |
| C7−C8 | 1.395 | 1.394 | 1.376 | 1.370 | 1.399 | 1.382 | 1.377 | 1.400 | 1.383 | 1.378 |
| C8−C9 | 1.410 | 1.418 | 1.420 | 1.416 | 1.423 | 1.426 | 1.421 | 1.423 | 1.426 | 1.421 |
| C8−N1 | 1.374 | 1.375 | 1.411 | 1.407 | 1.379 | 1.414 | 1.409 | 1.380 | 1.414 | 1.410 |
| C3−C10 | | 1.496 | 1.484 | 1.482 | 1.499 | 1.490 | 1.487 | 1.500 | 1.490 | 1.488 |
| energy[c] | | −403.283 717 | −402.632 474 | −403.015 998 | −403.135 526 | −402.489 663 | −402.876 955 | −403.150 952 | −402.504 794 | −402.884 264 |
| $\langle S^2\rangle$[d] | | 0.0000 | 0.7665 | 0.7613 | 0.0000 | 0.7676 | 0.7614 | 0.0000 | 0.7671 | 0.7612 |

| | | B3LYP/3-21G* | | | MP2/6-31G** | | |
|---|---|---|---|---|---|---|---|
| bond | exptl[b] | **1** | **2** | **3** | **1** | **2** | **3** |
| N1−C2 | 1.380 | 1.401 | 1.342 | 1.347 | 1.381 | 1.319 | 1.352 |
| C2−C3 | 1.358 | 1.373 | 1.435 | 1.435 | 1.377 | 1.405 | 1.405 |
| C3−C9 | 1.436 | 1.449 | 1.453 | 1.431 | 1.434 | 1.464 | 1.458 |
| C9−C4 | 1.398 | 1.403 | 1.392 | 1.407 | 1.407 | 1.385 | 1.379 |
| C4−C5 | 1.381 | 1.390 | 1.408 | 1.403 | 1.388 | 1.406 | 1.415 |
| C5−C6 | 1.396 | 1.411 | 1.395 | 1.395 | 1.412 | 1.402 | 1.403 |
| C6−C7 | 1.372 | 1.391 | 1.412 | 1.420 | 1.389 | 1.397 | 1.397 |
| C7−C8 | 1.395 | 1.398 | 1.378 | 1.375 | 1.400 | 1.399 | 1.410 |
| C8−C9 | 1.410 | 1.429 | 1.428 | 1.425 | 1.422 | 1.415 | 1.409 |
| C8−N1 | 1.374 | 1.385 | 1.439 | 1.420 | 1.376 | 1.392 | 1.374 |
| C3−C10 | | 1.504 | 1.492 | 1.489 | 1.495 | 1.483 | 1.478 |
| energy[c] | | −400.922 051 | −400.263 755 | −400.659 804 | −401.887 029 | −401.236 389 | −401.623 918 |
| $\langle S^2\rangle$[d] | | 0.0000 | 0.7694 | 0.7596 | 0.0000 | 0.9476 | 0.8478 |

[a] Bond lengths in angstroms. **1** = 3-methylindole; **2** = 3-methylindole neutral radical; **3** = 3-methylindole cation radical. Numbering as in Figure 1. [b] Experimental geometry of Chadwick[15] resulting from averaging 34 3-substituted indoles from the Cambridge Crystallographic Database. [c] Total energies in Hartrees. [d] Eigenvalue of the $S^2$ operator; see the Methods section.

**TABLE 2: Selected Vibrational Frequencies of 3-Methylindoles[a]**

| | | experimental frequencies | | | B3LYP/6-31G* | | |
|---|---|---|---|---|---|---|---|
| mode[b] | description[c] | A[d] | B[e] | C[f] | 1 | 2 | 3 |
| W1 | benz $\nu_{8a}$ + N1−C8 str | 1622 | 1614 | | 1680 | 1609 | 1623 |
| W2 | benz $\nu_{8b}$ | 1575 | 1578 | | 1637 | 1543 | 1647 |
| W3 | C2−C3 pyrrole str | 1555 | 1550 | | 1615 | 1455 | 1295[g] |
| W4 | benz $\nu_{19b}$ | 1496 | | | 1502 | 1487 | 1501 |
| W5 | benz $\nu_{19a}$ | 1462 | | | 1462 | 1431 | 1467 |
| W6 | N1−C2−C3 str + N1−H bend | 1361 | | | 1389 | 1393 | 1388 |
| W7 | pyrrole ring breath. | 1342 | 1345 | | 1336 | 1387 | 1379 |
| W8 | C3−C9 str + N1−H bend | 1305 | | | 1282 | 1331 | 1323 |
| W10 | C−H + C3−C10 str | 1238 | | | 1258 | 1264 | 1213 |
| W13 | sim to benz $\nu_{9b}$ C−H bend | 1127 | | | 1160 | 1171 | 1123 |
| W16 | benz C−C str | 1016 | 1009 | | 1045 | 1040 | 1035 |
| W17 | sim to benz $\nu_{12}$ + N1−H | 880 | 877 | | 889 | 872 | 874 |
| W18 | indole ring breath. | 762 | 760 | | 778 | 769 | 769 |
| | NH str | | | 3420 | 3668 | | 3606 |
| | benz $\nu_{20a}$ | | | 2950 | 3197 | 3206 | 3229 |
| | $CH_3$ vibration | | | 2860 | 3033 | 3027 | 3036 |
| | N1−H + C2−H bend | | | 1090 | 1116 | 1203 | |
| | C2−H out-of-plane bend | | | 805 | 785 | 892 | 987 |
| | benz $\nu_{11}$ | | | 740 | 755 | 765 | 778 |

[a] Frequencies in $cm^{-1}$. Atom numbering as in Figure 1. **1** = 3-methylindole, **2** = 3-methylindole neutral radical, **3** = 3-methylindole cation radical. [b] Tryptophan mode.[16a] [c] Approximate description of the tryptophan mode.[16a] Other mode assignments from this work. [d] Frequencies from Takeuchi and Harada[16b] and Su, Wang, and Spiro.[23] [e] Frequencies from Sweeney and Asher.[24] [f] Frequencies from Aldrich FTIR spectra.[16c] [g] This represents a combination of modes, but does possess significant C2−C3 motion.

average of 3.1% from the experimental frequencies, the former being generally larger. The level of agreement is comparable to that seen by Stephens and co-workers.[11] The most significant change in the vibrational frequencies in going from **1** to **2** or **3** can be seen in mode W3, which is principally the C2−C3 stretch. In **1**, this mode is calculated to be at 1615 $cm^{-1}$, in **2** it shifts to 1455 $cm^{-1}$, and in **3,** it decreases even further to 1295 $cm^{-1}$. In **3**, this mode is a mixture of several modes; however, it does contain significant C2−C3 movement. The reason for this change is not obvious, but as noted below, the largest difference in the calculated spin densities between **2** and **3** occurs on C2.

Calculated Mulliken spin densities of **2** and **3** using the B3LYP functional and the TZ2P, 6-31G*, 6-31+G*, and 3-21G* basis sets are reported in Table 3. The B3LYP/TZ2P spin densities for **2** and **3** are significantly different from each other, providing the possibility of experimental discrimination of the two species. Both have significant spin density on the

**TABLE 3: Spin Densities of Neutral and Cation Radicals of 3-Methylindole**

| | | B3LYP/TZ2P | | B3LYP/6-31G* | | B3LYP/6-31+G* | | B3LYP/3-21G* | |
|---|---|---|---|---|---|---|---|---|---|
| atom | exptl[a] | **2**[b] | **3** | **2** | **3** | **2** | **3** | **2** | **3** |
| N1 | 0.14 | 0.307 | 0.138 | 0.299 | 0.136 | 0.296 | 0.140 | 0.369 | 0.171 |
| C2 | 0.35 | −0.125 | 0.194 | −0.113 | 0.189 | −0.122 | 0.172 | −0.177 | 0.149 |
| C3 | 0.41 | 0.603 | 0.366 | 0.588 | 0.360 | 0.617 | 0.418 | 0.614 | 0.388 |
| C4 | n.d. | 0.190 | 0.254 | 0.194 | 0.259 | 0.228 | 0.282 | 0.175 | 0.268 |
| C5 | −0.07 | −0.052 | −0.091 | −0.052 | −0.089 | −0.057 | −0.093 | −0.033 | −0.086 |
| C6 | n.d. | 0.150 | 0.185 | 0.149 | 0.179 | 0.160 | 0.186 | 0.122 | 0.162 |
| C7 | n.d. | 0.005 | 0.059 | 0.010 | 0.070 | 0.014 | 0.075 | 0.031 | 0.086 |
| C8 | n.d. | 0.056 | 0.032 | 0.053 | 0.022 | −0.002 | −0.017 | 0.030 | 0.010 |
| C9 | n.d. | −0.123 | −0.106 | −0.128 | −0.113 | −0.127 | −0.155 | −0.129 | −0.132 |
| C10 | | −0.058 | −0.028 | −0.044 | −0.020 | −0.053 | −0.017 | −0.053 | −0.029 |

| | | MP2/6-31G** | | ROSHF[c] | | HMMO[d] | |
|---|---|---|---|---|---|---|---|
| atom | exptl[a] | **2** | **3** | **2** | **3** | **2** | **3** |
| N1 | 0.14 | 0.734 | 0.438 | 0.77 | 0.08 | 0.30 | 0.04 |
| C2 | 0.35 | −0.682 | −0.204 | −0.004 | 0.24 | 0.04 | 0.42 |
| C3 | 0.41 | 0.942 | 0.726 | 0.13 | 0.40 | 0.39 | 0.39 |
| C4 | n.d. | 0.109 | 0.260 | 0.013 | 0.09 | 0.09 | 0.10 |
| C5 | −0.07 | 0.051 | 0.047 | 0.02 | 0.001 | 0.03 | −0.04 |
| C6 | n.d. | 0.020 | −0.043 | 0.005 | 0.08 | 0.06 | 0.09 |
| C7 | n.d. | 0.130 | 0.280 | 0.04 | 0.04 | 0.08 | −0.01 |
| C8 | n.d. | −0.135 | −0.189 | −0.001 | 0.04 | | |
| C9 | n.d. | −0.144 | −0.241 | 0.03 | 0.017 | | |
| C10 | | −0.113 | −0.082 | | | | |

[a] Experimental spin densities for tryptophan-191 of cytochrome-*c*-peroxidase determined from ENDOR experiments by Huyett et al.[3f] "n.d." = not determined. Atom numbering as in Figure 1. [b] **2** = 3-methylindole neutral radical; **3** = 3-methylindole cation radical. [c] ROSHF calculations of Krauss and Garmer.[3b] [d] HMMO calculations of Margoliash et al.[6]

benzene ring, where the change in spin density distribution is relatively small between **2** and **3**. The pyrrole ring, however, is significantly affected by the retention or loss of the N1 hydrogen. Both species have significant spin density on N1 and C3, though these are quantitatively different, while C2 goes from negative to significantly positive spin density from **2** to **3**.

The calculated spin densities using the B3LYP functional and the 6-31G*, 6-31+G*, and 3-21G* basis sets change an average (per atom) of 0.010, 0.015, and 0.025 for **2** and 0.006, 0.023, and 0.019 for **3**, respectively, when compared to the results obtained with B3LYP/TZ2P. Clearly, there is little apparent gain in accuracy in using the TZ2P basis set versus the 6-31G* basis. However, the inclusion of diffuse functions (6-31+G* basis set) does appear to have a somewhat larger effect.

B3LYP/TZ2P Mulliken spin densities for **2** and **3** are compared to experiment[3f] in Figure 2. Table 3 also shows the ROSHF[3d] and HMMO[6] calculations cited above for comparison. For the experimental spin densities of atoms in tryptophan-191 of CCP that have been determined, there is excellent agreement with the calculated spin densities for **3** and much poorer agreement with those calculated for **2**. These results are in agreement with the conclusion of Huyett et al.[3f] that the tryptophan radical in CCP is indeed a cation radical. In addition, these calculations also indicate that the spin density distribution for the radical is probably not appreciably perturbed by the protein environment. The average deviation between theory and experiment for these atoms is 0.056 for **3** and 0.213 for **2**.

The B3LYP/TZ2P results presented here represent accurate gas phase results and should therefore be applicable to any tryptophan or indole radical. EPR studies of the tryptophan radical in DNA photolyase[17] which utilized isotopically labeled tryptophan were interpreted as originating from a cation radical. These conclusions were based on comparison to the HMMO[6] calculations, which yielded very small spin density on C2 and large spin density on N1 for the neutral radical, with the opposite being the case for the cation radical (Table 3). The B3LYP/

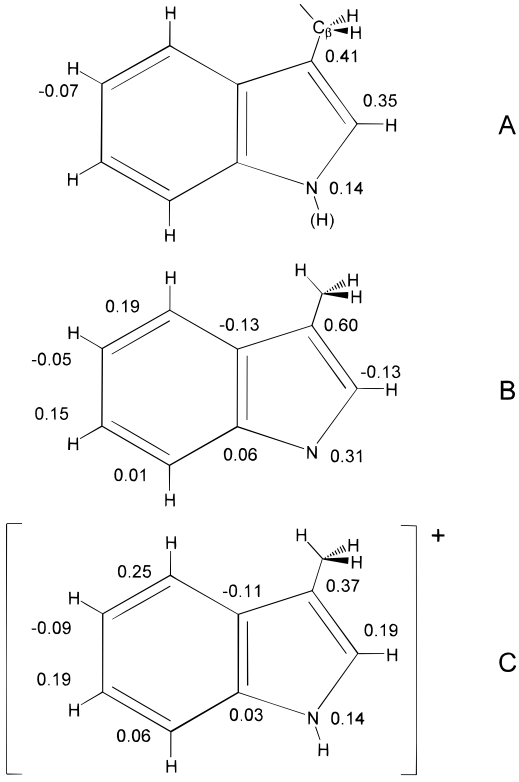


**Figure 2.** (A) Experimental spin densities[3f] for tryptophan-191 of cytochrome-*c*-peroxidase; (B) calculated Mulliken spin densities for the neutral radical of 3-methylindole **2**; and (C) calculated Mulliken spin densities for the cation radical of 3-methylindole **3**. B and C utilize the Becke3LYP functional and the TZ2P basis set.

TZ2P calculations show significant spin density on N1 for both species (albeit twice as large for **2**) and significant negative spin density on C2 for the neutral species. While the tryptophan radical in DNA photolyase may indeed be a cation radical, the interpretation of the isotope/EPR study may not be so straightforward.

Photochemically induced dynamic nuclear polarization (photo-CIDNP) experiments[18] on the tryptophan cation radical have revealed that significant spin density exists on C2, C3, C4, C6, and N1 and that no significant spin density exists on C5 and C7. For example, McCord et al.[18a] deduced proton isotropic hyperfine coupling constants from photo-CIDNP experiments on the tryptophan cation radical in the following relative order of absolute magnitude: C3 ≫ C2 ∼ C4 ∼ C6 > N1 ≫ C5 > C7, and with N1, C2, C3, C4, and C6 being positive, C5 negative, and C7 essentially zero. This is in excellent accord with the B3LYP/TZ2P spin densities calculated for **3**.

Comparing the HMMO[6] spin densities for indole to the experimental spin densities for tryptophan-191 of CCP, the average deviation from experiment is 0.055 for the cation radical and 0.148 for the neutral radical, making the choice of cation or neutral radical less clear, though still correct. The average deviation for the ROSHF[3d] theory and the CCP experiment is 0.063 for the cation radical and 0.316 for the neutral radical, which is in slightly worse agreement for both **2** and **3** with experiment than the B3LYP/TZ2P numbers. The similar agreement with experiment appears largely fortuitous. On the pyrrole ring, the ROSHF[3d] calculations differ most (when compared to the B3LYP/TZ2P calculations) for the neutral radical, where they predict zero spin density on C2, too much spin density on N1, and a change in spin density in the wrong direction for C3 going from neutral to cation. The HMMO[6] calculations show no change in spin density on C3, small positive spin density on C2, and not enough spin density on N1 for the cation. More dramatically, the spin density distribution for both of the ROSHF[3d] calculations severely underestimate the significant spin density in the benzene ring predicted by the B3LYP/TZ2P calculations on both **2** and **3** and observed in the photo-CIDNP experiments[18] for the tryptophan cation radical. The HMMO[6] calculations do show significant spin density on C4 and C6 of the benzene ring. For the relevant labeling experiments, the EPR spectra of the tryptophan radicals are predicted to exhibit a greater spectral extent by the B3LYP/TZ2P calculations than either of the older calculations.

On the basis of results in the recent literature[7] and the agreement with the CCP and photo-CIDNP experimental data for the cation radical, we conclude that the errors in the B3LYP/TZ2P calculated spin densities presented above are relatively small. Errors can come from the use of Mulliken population analysis for determining atomic spin densities, from spin contamination (see the Methods section), and from deficiencies in the B3LYP functional.

MP2/6-31G** spin densities for **2** and **3** are also reported in Table 3. Comparing the predicted spin densities with the experimental spin densities for tryptophan-191 of CCP shows differences of 0.570 and 0.321 for **2** and **3**, respectively. The spin densities for both **2** and **3** are in substantially worse agreement with experiment than the B3LYP/TZ2P calculations. Qualitatively, the results are slightly better, as they predict a decrease, increase, and decrease in spin density on N1, C2, and C3, respectively, on going from **2** to **3**. The quantitative deviation from experiment results from the individual absolute spin densities being much too large. Additionally, while the MP2/6-31G** calculations predict spin density on the benzene ring, no significant spin density is predicted for C6, in disagreement with the photo-CIDNP experiments.[18] The failure of MP2 in this case may be due to problems identified in the prediction of other ground state properties for conjugated molecules at this basis set level[19] or, more likely, to the significantly higher level of spin contamination (see the Methods section) evidenced in the MP2 calculations.

As this study was being completed, we became aware of a study of calculated spin densities on indole cation and neutral radical using the SVWN[9] functional and the 6-31G* basis set.[20] The results obtained for both species are qualitatively similar to those reported here, for both the pyrrole and benzene rings. The spin densities obtained in this study[20] for the atoms relevant to the CCP tryptophan-191 ENDOR data are, for the cation radical/neutral radical: 0.12/0.23 for N1, 0.14/−0.03 for C2, 0.30/0.48 for C3, and −0.04/−0.02 for C7. This results in average deviations between theory and experiment for the cation and neutral indole of 0.093 and 0.148, respectively, and are somewhat worse than those for **2** and **3** and B3LYP/TZ2P reported here. The most notable differences are in the quantitative values for C3 and the lack of significant negative spin density on C2 in the neutral radical. These differences may result from the absence of a methyl group on C3 or from the different choice of functional, or both.

A thorough evaluation of the accuracy of the B3LYP/TZ2P spin densities must await the determination of additional experimental spin densities. For CCP, this will require additional isotopic labeling schemes beyond those previously employed in the ENDOR studies[3f] to test the prediction of large spin densities on C4 and C6 (as well as the benzene ring in general). Data on an authentic neutral radical and spin density determinations for other known biological tryptophan radicals[17,22] are also desirable.

This study has provided a sound basis for the identification of the radical species in compound ES of CCP as the cation radical of tryptophan-191. The stability of this species near neutral pH remains to be explained and modeled quantitatively. Accurate *ab initio* parameters (electrostatic potential, partial charges, energies) coupled to a practicable and reasonably accurate solvent and protein model[21] should lead to a quantitative explanation of the factors responsible for the stability of the cation radical in CCP.

**Acknowledgment.** The authors thank C. S. Ashvar, Dr. C. F. Chabalowski, Dr. M. M. Fitzgerald, Professor B. M. Hoffman, Dr. K. J. Jalkanen, Dr. D. E. McRee, Dr. R. A. Musah, Professor P. J. Stephens, and S. K. Wilcox for helpful discussions and Professor P. J. Stephens, Professor R. A. Wheeler, and S. Walden for making results available prior to publication. Research on CCP at Scripps is supported by Grant GM41049 (to D.B.G.) from the National Institutes of Health. This work was supported in part by a grant of HPC time from the DoD HPC Center, U.S. Army Research Laboratory, on the SGI Power Challenge Array.

## References and Notes

(1) (a) Dawson, J. H. *Science* **1988**, *246*, 433−439. (b) Poulos, T. L. *Adv. Inorg. Biochem.* **1988**, *7*, 1−33.

(2) (a) Houseman, A. L. P.; Doan P. E.; Goodin, D. B.; Hoffman, B. M. *Biochemistry* **1993**, *32*, 4430−4443. For studies of the perturbation of this EPR signal by mutagenesis of neighboring amino acid residues see: (b) Fishel, L. A.; Farnum, M. F.; Mauro, J. M.; Miller, M. A.; Kraut, J.; Liu, Y.; Tan, X.; Scholes, C. P. *Biochemistry* **1991**, *30*, 1986−1996. (c) Goodin, D. B.; McRee, D. E. *Biochemistry* **1993**, *32*, 3313−3324. (d) McRee, D. E.; Jensen, G. M.; Fitzgerald, M. M.; Siegel, H. A.; Goodin, D. B. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 12847−12851.

(3) (a) Sivaraja, M.; Goodin D. B.; Smith M.; Hoffman, B. M. *Science* **1989**, *245*, 738−740. (b) Erman, J. E.; Vitello, L. B.; Mauro, J. M.; Kraut, J. *Biochemistry* **1989**, *28*, 7992−7995. (c) Scholes, C. P.; Liu, Y.; Fishel, L. A.; Farnum, M. F.; Mauro, J. M.; Kraut, J. *Isr. J. Chem.* **1989**, *29*, 85−92. (d) Krauss, M.; Garmer, D. R. *J. Phys. Chem.* **1993**, *97*, 831−836. (e) Fülöp, V.; Phizackerley, R. P.; Soltis, S. M.; Clifton, I. J.; Wakatsuki, S.; Erman, J.; Hajdu, J.; Edwards, S. L. *Structure* **1994**, *2*, 201−208. (f) Huyett, J. E.; Doan, P. E.; Gurbriel, R.; Houseman, A. L. P.; Sivaraja, M.; Goodin, D. B.; Hoffman, B. M. *J. Am. Chem. Soc.* **1995**, *117*, 9033−9041.

(4) Jovanovic, S. V.; Steenken, S.; Simic, M. G. *J. Phys. Chem.* **1991**, *95*, 684−687, and references therein.

(5) (a) Fitzgerald, M. M.; Churchill, M. J.; McRee, D. E.; Goodin, D. B. *Biochemistry* **1994**, *33*, 3807−3818. (b) Miller, M. A.; Han, G. W.; Kraut, J. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11118−11122. (c) Fitzgerald, M. M.; Trester M. L.; Jensen, G. M.; McRee, D. E.; Goodin D. B. *Protein Sci.* **1995**, *4*, 1844−1850.

(6) Hoffman, B. M.; Roberts, J. E.; Kang, C. H.; Margoliash, E. *J. Biol. Chem.* **1981**, *256*, 6556−6564.

(7) (a) Adamo, C.; Barone, V.; Fortunelli, A. *J. Chem. Phys.* **1995**, *102*, 384−393. (b) Qin, Y.; Wheeler, R. A. *J. Chem. Phys.* **1995**, *102*, 1689−1698. See also discussion and references in these papers.

(8) Frisch, M. W.; et al. *GAUSSIAN92/DFT*; *GAUSSIAN94* (Revision B.1); Gaussian, Inc.: Pittsburgh, PA, 1995.

(9) The local functionals employed by Gaussian92/DFT include the standard LSDA (also called SVWN) functional, the local correlation functional VWN, and the BLYP functional combining the gradient correction of Becke and the LYP correlation functional: (a) Becke, A. D. *The Challenge of d and f Electrons*; Salahub, D. R., Zerner, M. C., Eds.; American Chemical Society: Washington, DC, 1989; Chapter 12, pp 165−179. (b) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211. (c) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785−789.

(10) The B3LYP (Becke3LYP) functional[11a] is a Gaussian92/DFT[8] implementation of the three-parameter hybrid functional introduced by Becke.[10a] It is a hybrid of Hartree−Fock exchange with local and gradient-corrected exchange and correlation terms. An alternative functional of this type available in Gaussian92/DFT is the B3P86 (Becke3P86) functional, which consists of Becke's three-parameter hybrid functional with the nonlocal correlation provided by the "Perdew 86"[10b] expression. (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (b) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(11) (a) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623−11627. (b) Stephens, P. J.; Devlin, F. J.; Ashvar, C. S.; Chabalowski, C. F.; Frisch, M. J. *Faraday Discuss.* in press. (c) Finley, J. W.; Stephens, P. J. *J. Mol. Struct. (THEOCHEM)*, in press.

(12) Hehre, W. J.; Radom, L.; Schleyer, P. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986. Note that we use 6d functions in TZ2P, 6-31G*, 6-31+G*, 3-21G*, and 6-31G**.

(13) (a) Laming, G. J.; Handy, N. C.; Amos, R. D. *Mol. Phys.* **1993**, *80*, 1121−1134. (b) Eriksson, L. A.; Malkin, V. G.; Malkina, O. L.; Salahub, D. R. *J. Chem. Phys.* **1993**, *99*, 9756−9763. (c) Eriksson, L. A.; Malkina, O. L.; Malkin, V. G.; Salahub, D. R. *J. Chem. Phys.* **1994**, *100*, 5066−5075. (d) Barone, V.; Adamo, C. *Chem. Phys. Lett.* **1994**, *224*, 432−438. (e) Austen, M. A.; Eriksson, L. A.; Boyd, R. J. *Can. J. Chem.* **1994**, *72*, 695−704.

(14) Amos, R. D. *Cambridge Analytic Derivatives Package*, Cambridge, Version 5.0, Cambridge: U.K., 1992.

(15) Chadwick, D. J. *Comprehensive Heterocyclic Chemistry: The Structure, Reactions, Synthesis, and Uses of Heterocyclic Compounds;* Katritzky, A. R., Rees, C. W., Eds.; Pergammon: Oxford, 1984; Chapter 3.04, pp 155−200.

(16) (a) Austin, J. C.; Jordan, T.; Spiro, T. G. *Biomolecular Spectroscopy A;* Clark, R. J. H., Hester, R. E., Eds.; Wiley: New York, 1993; Chapter 2, pp 55−127. (b) Takeuchi, H.; Harada, I. *Spectrochim. Acta* **1986**, *42A*, 1069−1078. (c) *The Aldrich Library FT-IR Spectra;* Aldrich: Milwaukee, 1989; Vol. 1(2), p 653D.

(17) Kim, S.-T.; Sancar, A.; Essenmacher, C.; Babcock, G. T. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8023−8027.

(18) (a) McCord, E. F.; Bucks, R. R.; Boxer, S. G. *Biochemistry* **1981**, *20*, 2880−2888. (b) Stob, S.; Kaptein, R. *Photochem. Photobiol.* **1989**, *49*, 565−577. (c) Hore, P. J.; Broadhurst, R. W. *Prog. NMR Spectrosc.* **1993**, *25*, 345−402.

(19) Simandiras, E. D.; Rice, J. E.; Lee, T. J.; Amos, R. D.; Handy, N. C. *J. Chem. Phys.* **1988**, *88*, 3187−3195.

(20) (a) Walden, S. E.; Wheeler, R. A. *Abstracts of Papers*, 50th Southwest Regional Meeting of the American Chemical Society, Ft. Worth, TX; American Chemical Society: Washington, DC, 1994; ORG 181. (b) Walden, S. E.; Wheeler, R. A. *Abstracts of Papers*, 209th National Meeting of the American Chemical Society, Anaheim, CA; American Chemical Society: Washington, DC, 1995; PHYS 191. (c) Walden, S. E.; Wheeler, R. A. *J. Phys. Chem.*, in press.

(21) (a) Churg, A.; Warshel, A. *Biochemistry* **1986**, *25* 1675−1681. (b) Langen, R.; Jensen, G. M.; Jacob, U.; Stephens, P. J.; Warshel, A. *J. Biol. Chem.* **1992**, *267*, 25625−25627. (c) Jensen, G. M.; Warshel, A.; Stephens, P. J. *Biochemistry* **1994**, *33*, 10911−10924.

(22) (a) Sahlin, M.; Lassmann, G.; Pötsch, S.; Slaby, A.; Sjöberg, B.-M.; Gräslund, A. *J. Biol. Chem.* **1994**, *269*, 11699−11702. (b) Sahlin, M.; Lassmann, G.; Pötsch, S.; Sjöberg, B.-M.; Gräslund, A. *J. Biol. Chem.* **1995**, *270*, 12361−12372.

(23) Su, C.; Wang, Y.; Spiro, T. G. *J. Raman Spectrosc.* **1990**, *21*, 435−440.

(24) Sweeney, J. A.; Asher, S. A. *J. Phys. Chem.* **1990**, *94*, 4784−4791.

JP951795C