Ab initio investigation of the □1La and □1Lb excited states of indole
by Lee Stuart Slater

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Chemistry
Montana State University

Abstract:
Indole, the chromophore of the amino acid tryptophan, has two low-lying kk* singlet excited states denoted 1La and 1Lb. The differential sensitivity of the 1La and 1Lb to molecular environment enables indole to be exploited as a probe of protein structure and dynamics. Optimized excited state geometries are calculated using configuration interaction - singles (CIS) and MP2-corrected CIS (CIS-MP2) theory with the 3-21G basis set. The CIS/3-21G excited state properties are compared with experiment. It is found that the 1Lb permanent dipoles agree well with the vacuum measurements but the 1La dipole is underestimated compared to previous INDO/S calculations and solvent measurements. CIS and CIS-MP2 vertical excitation energies and CIS transition properties are calculated using a progression of basis sets. The CIS transition energies are slightly overestimated with the CIS-MP2 energies grossly overestimated. The basis set effects on the state energies are dramatic. The 1Lb transition dipoles and oscillator strengths agree better with experiment than the calculated 1La properties. CIS and CIS-MP2 adiabatic potential energy surface sections are calculated. In the CIS/3-21 G manifold an avoided crossing exists but the potential well minima order is reversed. In the CIS-MP2/3-21G manifold the potential well mimima order is correct but a surface crossing is present. There is general agreement with previous INDO/S calculations.

AB INITIO INVESTIGATION OF THE ^1L_a AND ^1L_b EXCITED STATES OF INDOLE

by

Lee Stuart Slater

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Chemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

July, 1994

APPROVAL

of a thesis submitted by

Lee Stuart Slater

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

22-JUL-94
Date

Patrick R Calls
Chairperson, Graduate Committee

Approved for the Major Department

8/3/94
Date

David M Dooley
Head, Major Department

Approved for the College of Graduate Studies

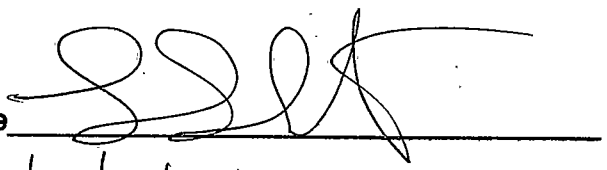
8/4/94
Date

R L Brown
Graduate Dean

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Signature 

Date 7/22/94

ACKNOWLEDGEMENTS

I thank Dr. Patrik Callis for introducing me to excited state calculations and providing me an opportunity to advance my career in theoretical chemistry. Dr. Callis helped me appreciate critical thinking and an aggressive approach to problem solving. Without his expertise in excited state analysis, this work would not have been possible. Dr. Lee Spangler has also provided valuable insight into the analysis of my calculations. He has helped me appreciate the importance of checking theoretical results with intuition and known experimental results.

I thank the rest of my graduate committee, Dr. Edward Dratz and Dr. Arnold Craig, for useful discussions concerning research and career options in chemistry.

Finally, I thank my Mother, Father, and the rest of my family and friends for support and good times through the years.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF TABLES

LIST OF TABLES - Continued

LIST OF FIGURES

LIST OF FIGURES - Continued

ABSTRACT

Indole, the chromophore of the amino acid tryptophan, has two low-lying $\pi\pi^*$ singlet excited states denoted ^1L_a and ^1L_b. The differential sensitivity of the ^1L_a and ^1L_b to molecular environment enables indole to be exploited as a probe of protein structure and dynamics. Optimized excited state geometries are calculated using configuration interaction - singles (CIS) and MP2-corrected CIS (CIS-MP2) theory with the 3-21G basis set. The CIS/3-21G excited state properties are compared with experiment. It is found that the ^1L_b permanent dipoles agree well with the vacuum measurements but the ^1L_a dipole is underestimated compared to previous INDO/S calculations and solvent measurements. CIS and CIS-MP2 vertical excitation energies and CIS transition properties are calculated using a progression of basis sets. The CIS transition energies are slightly overestimated with the CIS-MP2 energies grossly overestimated. The basis set effects on the state energies are dramatic. The ^1L_b transition dipoles and oscillator strengths agree better with experiment than the calculated ^1L_a properties. CIS and CIS-MP2 adiabatic potential energy surface sections are calculated. In the CIS/3-21G manifold an avoided crossing exists but the potential well minima order is reversed. In the CIS-MP2/3-21G manifold the potential well mimima order is correct but a surface crossing is present. There is general agreement with previous INDO/S calculations.

I. INTRODUCTION

Indole, the chromophore of the amino acid tryptophan, is responsible for a large fraction of the UV absorption and fluorescence in most proteins.[1]

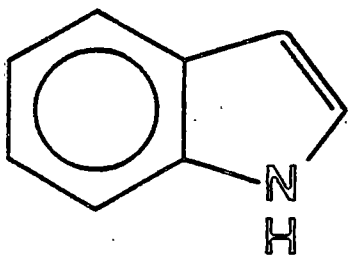


Figure 1. Indole

The strong 280 nm absorption band results from two nearly degenerate vibronic transitions involving two excited states denoted ^1L_a and ^1L_b. The ^1L_a and ^1L_b are singlet, $\pi\pi^*$ excited states having A' symmetry for indole with C_s molecular symmetry.

The ^1L_a and ^1L_b states have distinctive structure and properties. The ^1L_a transition is a strong transition having an oscillator strength of approximately 0.12 while the ^1L_b transition is comparatively weak with an oscillator strength of about 0.012.[2,3,4] The ^1L_a and ^1L_b transition dipoles are approximately perpendicular. The ^1L_a transition dipole is found to be -54° from the long axis in indole (negative angle toward the nitrogen) while the ^1L_b is 45° from the same axis.[3,5,6] The ^1L_a and ^1L_b states have differential solvent interactions resulting in a highly environment-dependent transition energy order. This effect is thought to be due

primarily to the large difference in the ^1L_a and ^1L_b excited state dipole moments. In polar solvent, the larger ^1L_a dipole is thought to quickly reorient the local solvent environment after excitation causing a greater redshift (stabilization) than the smaller ^1L_b dipole resulting in an ^1L_a fluorescence transition energy less than the ^1L_b transition energy.[7] In vapor phase, the larger ^1L_a dipole is unsolvated giving a larger transition energy than the ^1L_b transition. This environment dependence has been exploited in tryptophan as a sensitive probe of protein environment and solvation.[8] Methyl substitution in indole can also result in transition energy shifts and reordering.[9]

The near degeneracy of the ^1L_a and ^1L_b states is accidental but can be understood in terms of a crosslink perturbation and an aza substitution electrostatic perturbation.[10] Initially, the nine member aromatic $C_9H_9^-$ has the ^1L_a and ^1L_b states degenerate. Crosslinking $C_9H_9^-$ to form the fused five and six member ring system, indenide, splits the ^1L_a and ^1L_b considerably. However, the aza substitution in the five member ring to form indole brings the ^1L_a and ^1L_b into near degeneracy again.

Interpreting indole spectra is complicated by the near degeneracy of the ^1L_a and ^1L_b transitions. Pioneering work using polarization resolved two-photon fluoresence excitation spectroscopy on jet-cooled (~2K) indole has been accomplished by the Callis and Spangler research groups.[11,12] Using the fact that the ^1L_a and ^1L_b transitions have different two-photon absorptivity ratios for circularly:linearly polarized light, these studies have allowed the assignment of

transition character to the vibronic peaks in the overlapping spectra. With this important advance, new complexities have arisen. The spectral congestion of peaks having ^1L_a character a few hundred cm^{-1} above the ^1L_b origin coupled with the observation of more ^1L_a character in the ^1L_b origin of 3-methylindole suggest an avoided crossing in the ^1L_a-^1L_b manifold. An avoided crossing has been demonstrated theoretically using semi-empirical QCPE calculations.[13] The first two peaks having ^1L_a character in the indole spectra were thought initially to be the ^1L_a split origin. Recent evidence, such as the lack of significant shift in "^1L_a origin" due to complexation with polar molecules in the jet-cooled experiments[14] and disagreement in the intenstiy of the ^1L_a origin in simulated spectra[13,15] suggest a different ^1L_a origin. A new explanation involving Herzberg-Teller coupling and/or Born-Oppenheimer breakdown is evolving in our group.

To understand these spectral complexities and other interesting questions involving indole-solvent complex formation, the electronic structure and potential energy surfaces of indole need to be assessed. The initial electronic structure calculations of the ^1L_a and ^1L_b excited states have been performed using semi-empirical INDO/S methods.[16] With the advent of new theoretical methods and advances in algorithm and hardware development, systematic ab initio excited state molecular orbital procedures are emerging with applications becoming feasible for "large" molecules.[17,18] An initial ab initio study on indole has been accomplished.[19] Lacking a complete excited state investigation, the scope of this study was to determine the effect of a dipole reaction field on the transition

energies and properties of the ^1L_a and ^1L_b excited states of indole. Further ab initio studies are warranted to understand the electronic structure of the ^1L_a and ^1L_b excited states in detail and determine existence of an avoided crossing in the ^1L_a-^1L_b manifold. Consistency with the initial INDO/S picture should help establish a unified theoretical view of the ^1L_a and ^1L_b excited states of indole.

Problem Statement

This thesis will address the calculation of the optimized ground state wavefunction and electronic properties of indole using a series of basis sets. The effect of MP2 correlation is evaluated using the 6-31G* basis. The ^1L_a and ^1L_b optimized excited states and properties are calculated at the CIS/3-21G level of theory using different density matrices. The CIS-MP2/3-21G method is used to assess the effect of higher correlation on the optimized geometry. The study also addresses the effect of basis set on CIS and CIS-MP2 vertical transition energies and CIS transition properties. The existence of an avoided crossing in the ^1L_a-^1L_b manifold at the CIS/3-21G and CIS-MP2/3-21G level of theory is determined.

II. THEORETICAL BACKGROUND

A. Approximate Solutions to the Schrödinger Equation

The theoretical investigation of the electronic structure of molecules[20,21] involves determining approximate solutions to the Schrödinger equation

$$\hat{H}|\Phi_\alpha\rangle = \mathscr{E}_\alpha|\Phi_\alpha\rangle \quad \alpha = 0,1,2,...$$

where the electronic Hamiltonian operator in atomic units for N electrons and M nuclei

$$\hat{H} = -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}$$

results from the application of the Born-Oppenheimer approximation. Due to the larger rest mass of the nuclei compared to the electron mass, the nuclear kinetic energy operators can be neglected and the quantal electronic distributions are determined in the electrostatic field of fixed nuclei.

The exact eigenvalues of the Schrödinger equation correspond to the electronic energies of the ground state ($\alpha = 0$) and the electronic excited states ($\alpha = 1,2,...$) of the molecule. The electronic energies have parametric dependence on the nuclear coordinates

$$\mathscr{E}_\alpha = \mathscr{E}_\alpha(\{R_A\}).$$

The corresponding exact eigenfunctions are the electronic wavefunctions which describe the motion of all the electrons in the molecule for each electronic state.

The electronic wavefunctions have explicit dependence on the electron coordinates but parametric dependence on the nuclear coordinates

$$|\Phi_\alpha\rangle = |\Phi_\alpha(\{r_i\};\{R_A\})\rangle.$$

Under the Born-Oppenheimer approximation the potential for nuclear motion is provided by the total state energy (for fixed nuclei)

$$\mathscr{E}_{tot,\alpha} = \mathscr{E}_\alpha + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}$$

The double sum represents the nuclear repulsion which was omitted from the electronic Hamiltonian. This function of the nuclear coordinates, obtained by solving the electronic problem, represents the potential surface on which the nuclei move.

For computationally viable approximate solutions, the Schrödinger equation can be cast into the equivalent matrix representation. This can be accomplished by adopting an approximate or trial wavefunction which is written as a finite linear expansion of the orthonormal basis functions $\{|\Psi_i\rangle\ i=1,...,n\}$

$$|\Phi'_\alpha\rangle = \sum_{j=1}^{N} a_j^\alpha |\Psi_j\rangle \quad \alpha = 0,1,...,N-1 \tag{4}$$

where the coefficients are variational parameters required to obey $\sum(a_i^\alpha)^* a_i^\beta = \delta_{\alpha\beta}$. With this constraint on the variational coefficients and the orthonormality of the basis functions, the trial wavefunctions are orthonormal. The expectation value

$$E_\alpha = \langle \Phi'_\alpha | \hat{H} | \Phi'_\alpha \rangle$$

is the approximate energy (eigenvalue) of the trial wavefunction and is a function of the variational parameters. The variation principle assures that for orthonormalized trial wavefunctions, the approximate energies are bounded below by the exact energies[20]

$$E_\alpha \geq \mathscr{E}_\alpha \quad \alpha = 0,1,...,N-1$$

The linear variational method determines the variational parameters which minimize E_α. Defining the matrix representation of the electronic Hamiltonian in the $|\Psi_j\rangle$ basis

$$H_{ij} = \langle \Psi_i | \hat{H} | \Psi_j \rangle$$

and using the orthonormality of the basis functions, application of the linear variational method results in

$$\sum_i H_{ij} a_j^\alpha = E_\alpha a_j^\alpha$$

or

$$\mathbf{H}\mathbf{a}^\alpha = E_\alpha \mathbf{a}^\alpha \quad \alpha = 0,1,...,n-1$$

where $\mathbf{a}^\alpha$ is the coefficient eigenvector $(a_1^\alpha\ a_2^\alpha ... a_n^\alpha)^t$ containing the coefficients which minimize the expectation value for the energy. Introducing a diagonal matrix $E_{\alpha\beta} = E_\alpha \delta_{\alpha\beta}$ and the matrix of eigenvectors $A_{i\alpha} = a_i^\alpha$, the n relations can be written as

$$\mathbf{HA} = \mathbf{AE}.$$

For a complete basis $\{|\Psi_i\rangle\}$, the eigenvalues of H are exactly equal to the eigenvalues of $\hat{H}$. Since a finite basis (noncomplete) is used, the linear variation method is equivalent to solving the eigenproblem, $\hat{H}|\Phi_\alpha\rangle = \mathcal{E}_\alpha|\Phi_\alpha\rangle$ in a finite subspace spanned by $\{|\Psi_i\rangle\ i=1,...,n\}$. A large body of computational mathematics has been developed to efficiently diagonalize H and thus determine the eigenvalues and corresponding eigenvectors.

B. Hartree-Fock Theory

The basis functions $\{|\Psi_i\rangle\ i=1,...,n\}$, are most commonly defined in quantum chemistry as Slater determinants. The Slater determinant

$$|\chi_i\chi_j\cdots\chi_k\rangle$$

is a many-electron function , defined in terms of spin orbitals, which satisfies the Pauli Exclusion Principle and the antisymmetry principle. The spin orbitals are one-electron functions which are constructed as products of a spatial function ψ, termed a molecular orbital, and an α or β spin function.

$$\chi_i(x) = \begin{matrix} \\ \text{or} \\ \\ \end{matrix} \begin{matrix} \psi_i(r)\alpha(\omega) \\ \\ \psi_i(r)\beta(\omega) \end{matrix}$$

The spin functions are the eigenfunctions of the S^2 and the S_z operators. The molecular orbital (MO) describes the spatial distribution of an electron such that $|\psi_i(r)|^2\ dr$ is the probability of finding an electron in the volume element dr centered at r.

The spin orbitals are determined from the Hartree-Fock (HF) eigenvalue equation

$$f(i)\chi(x_i) = \varepsilon\chi(x_i)$$

where ε is the spin orbital energy for the i^{th} electron and

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + v^{HF}(i)$$

is the effective one-electron Fock operator. Here $v^{HF}(i)$ is the Hartree-Fock potential which is the average potential experienced by the i^{th} electron due to the presence of the other electrons in their spin orbitals. The derivation of the HF equations results from the functional variation of the spin orbitals to minimize the expectation value of the electronic energy. Since f(i) depends on the spin orbitals of the other electrons (i.e. f(i) depends on its own eigenfunctions through v^{HF}), it is nonlinear and requires an iterative solution. The solution to the eigenvalue problem for all electrons is termed the self-consistent-field (SCF) method and yields a set of orbital energies $\{\varepsilon_k\}$ and a set $\{\chi_k\}$ of orthonormal spin orbitals.

The single determinant formed from the N lowest energy spin orbitals is termed the Hartree-Fock ground state wavefunction $|\Psi_0\rangle = |\chi_1\chi_2\cdots\chi_N\rangle$ and has energy $E_0 = \langle\Psi_0|\hat{H}|\Psi_0\rangle$. This single determinant wavefunction incorporates exchange correlation between pairs of electrons having parallel spins but lacks any correlation between pairs of electrons with opposite spins. For this reason, single determinant wavefunctions are termed uncorrelated wavefunctions. The other spin orbitals produced in the SCF method are virtual orbitals and are used to form

excited configurations also called excited or substituted Slater determinants. Singly excited configurations are formed by substituting an occupied spin orbital χ_i from the HF wavefunction with a virtual orbital χ_a and are denoted $|\Psi_{ia}\rangle$. Similarly, doubly excited configurations are formed by substituting two occupied spin orbitals χ_i, χ_j with two virtual spin orbitals χ_a, χ_b and are represented as $|\Psi_{ijab}\rangle$. Higher excited configurations are formed and represented analogously. Excited configurations are used in electron correlation methods.

Since the spin orbital is a product of an MO and a spin function, it is possible to integrate out the spin component leaving expressions involving only the MOs. A common approach for representing the MOs is to decompose them into finite linear expansions of predetermined atomic basis functions $\{\phi_\mu(r)\ \mu = 1,2,...,k\}$.

$$\psi_i(r) = \sum_{\mu=1}^{k} C_{\mu i}\phi_\mu(r) \qquad i=1,2,...k$$

Such a set of basis functions is termed a basis set. The common basis sets used in this study are STO-3G, 3-21G, 6-31G, 6-31+G, 6-31G*, and 6-31+G*. They are listed in order of increasing number of basis functions and are described more fully below. A basis set of k functions produces 2k spin orbitals, N which are occupied and 2k-N which are virtual. The larger the basis set, the greater the flexibility of the MO expansion. As a result, the electronic energy of the molecule usually decreases upon increase in basis set size. In the limit of an infinite, complete basis set, the electronic energy approaches the HF limit. Finite basis sets will give an energy above the HF limit. Certain calculated properties of the wavefunction

have strong dependence on the inclusion of particular types of basis functions. The details of this dependency will be discussed in the Results and Discussion chapter.

The STO-3G is a minimal basis set having one basis function per electron. It uses a linear combination of three Gaussian type functions (primitive) to model Slater type orbitals. The other basis sets are split-valence using a linear combination of two atomic basis functions for every valence electron but only one basis function for the core electrons. The 3-21G basis uses three Gaussian primitives to model the one basis function representing the core electrons with two and one Gaussian primitives to represent the valence basis functions. The basic 6-31G series has the same number of valence and core electron basis functions but has an increased number of gaussian primitives to better model the basis functions - 6 Gaussian primitives for the core electron basis functions with 3 and 1 Gaussian primitives for the valence basis functions. The basic 6-31G basis set can be modified by the inclusion of polarization basis functions and diffuse basis functions. The polarization functions denoted by "*", allow electron distributions to be polarized from the presence of nonuniform electric fields in nonspherical molecular environments. Diffuse functions denoted by "+", represent highly diffuse electron distributions present in anions, diffuse lone pairs, and excited states having significant anti-bonding character. The polarization functions used in the basis sets listed are six 3d-functions of symmetry $3d_{xx}$, $3d_{yy}$, $3d_{zz}$, $3d_{xy}$, $3d_{yz}$, and $3d_{zx}$ centered on each heavy atom. The diffuse functions for the listed basis sets

are a single set of diffuse s-type and p-type functions centered on each heavy atom.

With the introduction of a basis, the HF differential equations are transformed into the algebraic Roothaan-Hall equations which are solved by standard matrix techniques. Defining the overlap matrix

$$S_{\mu\nu} = \int dr \phi_\mu^*(1)\, \phi_\nu(1)$$

and the Fock matrix

$$F_{\mu\nu} = \int dr \phi_\mu^*(1)\, f(1)\, \phi_\nu(1)$$

where f(1) represents the spatial Fock operator with spin integrated out and ϕ_μ are the basis functions, the Roothaan-Hall equations are

$$\sum_\nu F_{\mu\nu} C_{\nu i} = \varepsilon_i \sum_\nu S_{\mu\nu} C_{\nu i} \qquad i = 1,2,\ldots,k$$

or

$$FC = SC\varepsilon$$

where C is the k x k matrix of the MO expansion coefficients $C_{\mu i}$ and ε is the diagonal matrix of orbital energies ε_i. Solving this matrix equation produces the HF MOs $\{\psi_i\}$ and orbital energies $\{\varepsilon_i\}$ which minimize the HF energy E_0. To do this an explicit expression for the Fock matrix is needed.

Defining the SCF density matrix as

$$P_{\mu\nu} = 2\sum_{a}^{N/2} C_{\mu a} C_{\nu a}^{*}$$

and the one-electron Hamiltonian operator

$$h(1) = -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{|r_1 - R_A|}$$

then, the Fock matrix can be decomposed as

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + G_{\mu\nu}$$

where

$$H_{\mu\nu}^{core} = \int dr_1 \phi_\mu^*(1) h(1) \phi_\nu(1)$$

is fixed for a given basis and

$$G_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma}[(\mu\nu|\sigma\lambda) - \frac{1}{2}(\mu\lambda|\sigma\nu)]$$

is the two-electron part of the Fock matrix depending on the density matrix and the two electron integrals

$$(\mu\nu|\lambda\sigma) = \int dr_1 dr_2 \phi_\mu^*(1) \phi_\nu(1) r_{12}^{-1} \phi_\lambda^*(2) \phi_\sigma(2).$$

The calculation and manipulation of the two-electron integrals form the bottleneck in SCF calculations. For k basis functions, the number of distinct two-electron integrals is $O(k^4/8)$. The main difference between ab initio and semi-empirical calculations is that all the two-electron integrals are evaluated in ab initio calculations while semi-empirical methods neglect many integrals but formulate other schemes to model the electron repulsion neglected by the omission of these

integrals.

C. Electron Correlation Methods

For most closed shell molecules in their ground state the HF method, using a single determinant wavefunction, is very successful in describing the energy and electronic structure of the molecule. This is attributed to two factors. First, the antisymmetrization of the Hartree product wavefunction to produce the Slater determinant introduces exchange correlation between electrons of the same spin creating a Fermi hole around each electron. Second, the potential $v^{HF}(i)$ of the HF method accounts for the average coulombic interactions of an electron with the "sea" of another electrons in the molecule. Both of these factors are highly dominant in producing the proper electronic energy and charge distributions. But the HF method fails to account for the interactions of electrons of the opposite spin and the pairwise instantaneous coulombic interactions of any two electrons. These omissions can have a large effect on the calculation of accurate energies and wavefunctions for the ground state of molecules having several degenerate or nearly degenerate electron configurations (excited configurations). Also, treatments beyond the HF method are essential for the calculation of potential energy surfaces and excited states for any molecule.

The difference in the exact, nonrelativistic electronic energy and the energy of the HF wavefunction, as the basis set approaches completeness, is termed the

correlation energy. Two types of electron correlation that contributes to the correlation energy and electronic structure have been identified[22,23]. Nondynamical electron correlation arises when two or more excited configurations are degenerate or nearly degenerate in energy and thus have relatively large coefficients in the CI expansion of the wavefunction. This is the main failure of a single determinantal wavefunction approach, such as the HF method, and seems to be more important for the high energy valence electrons. Dynamical correlation is represented by the pairwise instantaneous electron-electron coulombic interactions. It is the dominant correlation for core electron interactions having close proximity due to large nuclear attraction and also for core-valence electron interactions. Two techniques are commonly used to account for these types of electron correlation in molecules.

1. Møller-Plesset Perturbation Theory

In this study, perturbation theory is used to calculate the ground state correlation energy in indole. Another common approach is configuration interaction but this method is much more computationally intensive than perturbation theory. Configuration interaction, to be discussed in the next section, is used for the excited state electron correlation only. The use of Rayleigh-Schrödinger perturbation theory (also called many body perturbation theory) with the HF Hamiltonian, $\hat{H}_0$, as the zeroth-order Hamiltonian is termed Møller-Plesset perturbation theory (MPPT). In this theory, the generalized electronic Hamiltonian,

$\hat{H}_\lambda$ is the sum

$$\hat{H}_\lambda = \hat{H}_0 + \lambda \hat{V}$$

Here, the perturbation is defined by

$$\lambda \hat{V} = \lambda(\hat{H}-\hat{H}_0)$$

where $\hat{H}$ is the exact electronic Hamiltonian and λ is a dimensionless parameter. Now, $|\Psi_\lambda\rangle$ and E_λ, the exact ground state wavefunction (within a given basis for λ=1) corresponding to the Hamiltonian $\hat{H}_\lambda$, can be expanded in a power series according to many body perturbation theory

$$|\Psi_\lambda\rangle = |\Psi^{(0)}\rangle + \lambda|\Psi^{(1)}\rangle + \lambda^2|\Psi^{(2)}\rangle + \ldots$$

and

$$E_\lambda = E^{(0)} + E^{(1)} + E^{(2)} + \ldots$$

Implementations of the theory set λ=1 (to give $\hat{H}_\lambda = \hat{H}$, the exact Hamiltonian) and truncate both series to various orders. For this study, truncation after the second order, known as MP2, is used. Applying the generalized Hamiltonian to the truncated wavefunction and energy series and solving results in

$$|\Psi_\lambda^{MP2}\rangle = |\Psi^{(0)}\rangle + |\Psi^{(1)}\rangle$$

where $|\Psi^{(0)}\rangle = |\Psi_0\rangle$ is the HF wavefunction and

$$|\Psi^{(1)}\rangle = \sum_{S>0} (E_0 - E_S)^{-1} V_{S0} |\Psi_S\rangle$$

where V_{S0} are the matrix elements involving the perturbation operator and E_S is the eigenvalue corresponding to a substituted determinant $|\Psi_S\rangle$. By Brillouin's Theorem ($\langle\Psi_S|\hat{H}|\Psi_0\rangle = 0$, where S is a single substitution) and other properties of matrix elements over higher substituted determinants[20], V_{S0} vanishes unless S corresponds to a double substitution. For the energy expansion,

$$E_\lambda^{MP2} = E^{(0)} + E^{(1)} + E^{(2)}$$

where $E^{(0)} + E^{(1)}$ is the HF energy and

$$E^{(2)} = -\sum_S^D (E_0 - E_S)^{-1} |V_{S0}|^2$$

Since $|\Psi_S\rangle$ is the double substitution ij $\rightarrow$ ab, V_{S0} reduces to

$$V_{S0} = \langle ij||ab\rangle$$

where

$$\langle ij||ab\rangle = \int \chi_i^*(1)\chi_j^*(2)\, r_{12}^{-1} [\chi_a(1)\chi_b(2) - \chi_b(1)\chi_a(2)]\, d\tau$$

is an antisymmetrized two-electron integral over spin orbitals. The final form for the second order contribution to the electronic energy is

$$E^{(2)} = -\frac{1}{4}\sum_{ijab} (\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)^{-1} |\langle ij||ab\rangle|^2$$

This has been noted to be probably the simplest approximate expression for the correlation energy. As a theoretical model, MP2 theory is size consistent but not variational.

2. Configuration Interaction

As discussed in section II.A. of this chapter, the wavefunctions for the ground and excited states can be written as a linear expansion of the basis functions $\{|\Psi_i\rangle\ i=1,...,n\}$. It can be shown that for the N-electron problem, excited configurations (substituted Slater determinants) can be used as the basis functions[20]. The procedure of representing the wavefunction with excited configurations and using the linear variational method in determining the energies and coefficients of this linear expansion is known as configuration interaction (CI). The general form of the ground state full CI wavefunction is

$$|\Phi_0\rangle = a_0|\Psi_0\rangle + \sum_{ia} a_{ia}|\Psi_{ia}\rangle + \sum_{\substack{i<j \\ a<b}} a_{ijab}|\Psi_{ijab}\rangle + \sum_{\substack{i<j<k \\ a<b<c}} a_{ijkabc}|\Psi_{ijkabc}\rangle + \cdots$$

where $|\Psi_{ia}\rangle$, $|\Psi_{ijab}\rangle$, $|\Psi_{ijkabc}\rangle$, etc. are the singly, doubly, triply, etc. excited configurations and the coefficients a_{ia}, a_{ijab}, a_{ijkabc} are the CI coefficients determined with linear variational theory. The excited state CI wavefunction omits the initial HF wavefunction term. This is a simplified form of the wavefunction. In implementations of CI theory, spin adapted configurations (linear combinations of the excited configurations) are always used to keep the wavefunction an eigenfunction of the total spin angular momentum squared operator, $\mathcal{S}^2$. Also, excited configurations of different spin symmetry are omitted since no mixing occurs between wavefunctions of different spin symmetry. The CI wavefunction can be connected to MPPT[21]. The first order contribution to the CI coefficient a_S, where S represents a substitution, is

$$a_S^{(1)} = (E_0 - E_S)^{-1} V_{S0}$$

Using the full CI wavefunction and a finite basis set, the calculation is termed exact within the subspace spanned by the one-electron basis. The difference between the energy of this exact calculation and the HF energy using the same basis is termed the basis set correlation energy. As the basis set approaches completeness, the wavefunctions and energies approach the exact non-relativistic solutions to the Schrödinger equation.

a. Configuration Interaction - Singles

Limiting the use of full CI is that C(2k,N) distinct N-electron excited configurations comprise the wavefunction. For scale, using the 3-21G basis on indole, the full CI wavefunction will have $O(10^{25})$ excited configurations. Clearly, the number of excited configurations needs to be reduced for practical implementations of CI. To do this in a non-biased, unprejudiced manner, the wavefunction is restricted to certain types of excited configurations. In this study, the CI wavefunction is composed of singly excited configurations only. This is termed full configuration interaction in the space of the single substitutions or CI-Singles (CIS)[17]. The wavefunction is then written as

$$|\Psi_{CIS}\rangle = \sum_{ia} a_{ia} |\Psi_{ia}\rangle$$

The CI coefficients are the normalized eigenvectors of the Hamiltonian matrix

$$\langle \Psi_{ia} | \hat{H} | \Psi_{jb} \rangle = [E_{HF} + \epsilon_a - \epsilon_i] \delta_{ij} \delta_{ab} - (ja||ib)$$

where (jalib) is the usual antisymmetrized two-electron integral in terms of spatial basis functions and E_{HF} is the sum of the HF energy (E_0) and the nuclear potential energy. The eigenvalues E_{CIS} are the CI-Singles total energies for various excited states and are given by

$$E_{CIS} = E_{HF} + \sum_{ia} a_{ia}^2 (\epsilon_a - \epsilon_i) - \sum_{ijab} a_{ia} a_{jb} (ja||ib)$$

In the Gaussian92 implementation of CIS theory, the eigenvectors and corresponding eigenvalues are calculated with the iterative Davidson diagonalization method.

There are some important points to note about the CIS method. The CIS method produces excited state wavefunctions orthogonal to the HF ground state wavefunction by Brillouin's Theorem, $\langle \Psi_{ia} | \hat{H} | \Psi_0 \rangle = 0$. Next, the method is variational and size consistent. Finally, the CIS method leads to a well-defined and differentiable energy for which analytical gradient techniques can be applied to determine properties and optimized excited state geometries very efficiently.

b. Higher electron correlation

The CIS method, involving only singly excited configurations, is an adequate zeroth-order approximation to some excited states. However, the omission of higher excited configurations (doubly and triply excited configurations, etc.) can lead to poor representation of the energies and properties derived from

wavefunctions of certain excited states. An expression which attempts to access the influence of higher excited configurations on the energy has been suggested[17]. Recall the second order correction to the HF energy of the ground state (see section II.C.1.). The expression includes summation over singly excited configurations but those matrix elements vanish by Brillouin's Theorem. This expression for the ground state correlation energy suggests that if the zeroth order wavefunction is a CIS wavefunction, then a plausible second order correction will involve single and double substitutions from the reference state. This implies double and triple substitutions from the CIS wavefunction. The energy correction is given by

$$\Delta E_{CIS\text{-}MP2} = -\frac{1}{4}\sum_{ijab} \frac{\langle \Psi_{CIS}|\hat{H}|\Psi_{ijab}\rangle^2}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j - \Delta_{CIS}}$$

$$-\frac{1}{36}\sum_{ijkabc} \frac{\langle \Psi_{CIS}|\hat{H}|\Psi_{ijkabc}\rangle^2}{\epsilon_a + \epsilon_b + \epsilon_c - \epsilon_i - \epsilon_j - \epsilon_k - \Delta_{CIS}}$$

where $\Delta_{CIS} = E_{HF} - E_{CIS}$ is the CIS vertical excitation energy. This energy correction, when added to the CIS energy, defines a second order total energy for an excited state, $E_{CIS\text{-}MP2}$. The difference in $E_{CIS\text{-}MP2}$ and the ground state E_{MP2} is the CIS-MP2 vertical excitation energy. The procedure is referred to as CIS-MP2 theory. For full details see reference 17. It is important to note that this is only a suggested energy correction with no wavefunction correction derived. This, as well as the $O(k^6)$ running time, severly limits the utility of the method. It has

been successfully used in this study to demonstrate the dependence of the ^1L_a and ^1L_b excited states of indole on higher electron correlation but fails in several important aspects.

D. Molecular Electronic Properties

1. Computational Methods

The calculation of molecular properties is accomplished by two methods. In computing the quantum mechanical expectation value of an observable, a suitable operator representing the molecular property is formulated. For instance, the dipole moment operator is

$$\hat{\mu} = -\sum_{i=1}^{N} r_i + \sum_{A} Z_A R_A.$$

Then the expectation value

$$\langle M \rangle = \langle \Phi_\alpha | \hat{M} | \Phi_\alpha \rangle$$

is evaluated directly with the use of matrix element expressions involving the two-electron integrals. Here, $|\Phi_\alpha\rangle$ is the exact electronic wavefunction for a particular ground or excited state. An alternate approach to the calculation of molecular properties, involving the derivatives of the electronic energy, follows from the Hellmann-Feynman theorem. For an exact perturbed wavefunction $|\Phi_\alpha(\mu)\rangle$ which is an eigenfunction of the perturbed Hamiltonian $\hat{H}(\mu)$, both dependent on a perturbation parameter proportional to the strength of the perturbation,

$$\frac{dE(\mu)}{d\mu} = \langle \Phi_\alpha(\mu) | \frac{\partial \hat{H}}{\partial \mu} | \Phi_\alpha(\mu) \rangle$$

Then, adopting a Hamiltonian of the form

$$\hat{H}(\mu) = \hat{H} + \mu \hat{M}$$

where $\hat{H}$ is the electronic Hamiltonian of the unperturbed molecule having the wavefunction $|\Phi_\alpha\rangle$,

$$\left(\frac{dE(\mu)}{d\mu}\right)_{\mu=0} = \langle \Phi_a | \hat{M} | \Phi_a \rangle = \langle M \rangle.$$

The equivalence of these two procedures given by this expression also holds for certain approximate theoretical methods. For example, the equality is valid for the HF method even with a finite basis[24,25]. In general, the equality is not valid for incomplete electron correlation methods such as MP2 or CIS[26]. It should be noted that derivative methods have been deemed superior since they represent the wavefunction response to a perturbation and are thus more closely related to experimental observables[26,27]. Also, analytic derivatives exist for the MP2 and CIS implementations in Gaussian92 making this method of calculating molecular properties attractive.

2. The Density Matrix

Many molecular electronic properties can be expressed in terms of the

density matrix; the key dependency of the expressions for the energy derivatives and expectation values is the density matrix. The one particle (one electron) reduced density matrix (1PDM) is defined as

$$\gamma_\alpha(x_1,x_1') = N\int dx_2...dx_N\ \Phi_\alpha(x_1,x_2,...,x_N)\Phi_\alpha^*(x_1',x_2,...,x_N)$$

where Φ_α is a normalized ground or excited state wavefunction. The density matrix $\gamma(x_1,x_1')$ results from integrating the probability distribution (charge density) function $\Phi_\alpha^*\Phi_\alpha$ over the space and spin coordinates of the N-1 electrons having positions $x_2,...,x_N$ and thus depends only on two continuous indices. For the case that $\Phi_\alpha = \Psi_0$, the HF ground state wavefunction, it can be shown that

$$\gamma^{HF}(x_1,x_1') = \sum_a^N \chi_a(x_1)\chi_a^*(x_1')$$

With introduction of a basis and integrating over spin, the diagonal element $\gamma^{HF}(x_1,x_1)$ reduces to

$$\rho(r) = \sum_{\mu\nu} P_{\mu\nu}\phi_\mu(r)\phi_\nu(r)$$

where

$$P_{\mu\nu}^{HF} = 2\sum_a^{N/2} C_{\mu a}C_{\nu a}^*$$

is the discrete representation of the HF or SCF density matrix. With this last expression, given a predetermined basis set, the HF ground state charge density is completely specified.

The one particle density matrix involves integration (averaging) over all electrons except one and thus represents the density of one electron independent of the explicit location of the other electrons in the system. Corrections to the 1PDM to produce a "correlated" density have been made. The generalized density matrix (GDM) is an effective correlated density matrix which is defined as the sum of the 1PDM and the Z-vector[28]. The Z-vector is obtained in solving a single set of coupled perturbed Hartree-Fock (CPHF) equations for the orbital rotation parameters necessary in CI gradient calculations[29,30] and acts as a relaxation correction to the 1PDM. This relaxation is captured through nonzero occupied-virtual (OV) density matrix elements which are interpreted, in the case of the CIS GDM, as orbital relaxation following the initial gross charge rearrangement due to electronic excitation[17]. These OV elements, which are zero in the 1PDM, are found from solving a single set of CPHF equations. The Z-vector method is not limited to CI calculations, but is applicable to all correlation methods including MP2 gradient calculations[27,29,30]. It has been demonstrated that it is the GDM, and not the 1PDM, which allows realistic computation of charge distributions, orbital populations, and molecular electronic properties[17,28] and is a better representation for correlated densities. In Gaussian 92, derivative methods are used in conjunction with the CIS and ground state MP2 methods while expectation value methods are used with the HF method in the calculation of molecular properties for the ground and excited states[31].

3. Electronic Transition Dipole Moment and Oscillator Strength

An important quantity in characterizing molecular electronic transitions is the electronic dipole transition moment

$$\mu_{\alpha'\alpha} = \langle \Psi_{\alpha'} | \hat{\mu}_e | \Psi_\alpha \rangle$$

The electronic transition moment (or transition dipole moment) can be interpreted as a measure of the dipole associated with the electronic charge density migration during the electronic transition. It is the transition dipole moment which couples with the electric field vector of the light to stimulate photon absorbtion and emission processes. The origin of the electronic transition moment lies in the ability to partition the vibronic transition moment

$$\langle \Psi_{\alpha'v'} | \hat{\mu} | \Psi_{\alpha v} \rangle$$

characteristic of the intensity of the vibronic transition, into separate factors, one dependent on the electronic wavefunctions and the other on the vibrational wavefunctions. The Born-Oppenheimer approximation allows the separability of the vibronic wavefunction into a product of an electronic wavefunction and a vibrational wavefunction

$$\Psi_{\alpha v} = \Psi_\alpha \Psi_v$$

Also, the dipole moment operator, defined above, can be partitioned into an electronic term and a nuclear term

$$\hat{\mu} = \hat{\mu}_e + \hat{\mu}_n$$

With the orthogonality of the electronic wavefunctions and the assumption that the electronic transition moment does not vary significantly with nuclear configuration, the vibronic transition moment can be factored

$$\langle \Psi_{\alpha' v'} | \mu | \Psi_{\alpha v} \rangle \approx \mu_{\alpha' \alpha} S_{v' v}$$

where $S_{v'v}$ is the overlap integral between two vibrational states. The insignificant invariance of the electronic transition moment is valid about the equilibrium ground state geometry from which the vertical transition originates according to the Franck-Condon Principle. The electronic transition moment can be written conveniently as

$$\mu_{\alpha' \alpha} = -\int \rho_{\alpha' \alpha} r d\tau$$

where

$$\rho_{\alpha' \alpha} = 2^{-1/2} \sum_{ia} a_{ia} \chi_i^* \chi_a$$

is the transition density between the HF ground state and any single reference CI wavefunction. This definition includes single reference CI wavefunctions beyond the CIS level of theory since the matrix elements between the HF wavefunction and higher excited configurations vanish. In terms of this form of the electronic transition moment, for an electronic transition to be allowed, the transition density must have a dipole which is the electronic transition dipole moment[32].

The intensity of a transition is more commonly quantitated by the oscillator

strength. For a nondegenerate transition, the oscillator strength is

$$f_{\alpha'\alpha} = (4\pi m_e / 3e^2\hbar)\, \nu\, \mu^2_{\alpha'\alpha}$$

where ν is the vertical excitation energy in wavenumbers (cm^{-1}) and $\mu_{\alpha'\alpha}$ is in Debyes (D). Intense, allowed electronic transitions have oscillator strengths of O(1) while weak, forbidden electronic transitions have oscillator strengths of $O(10^{-5})$ to $O(10^{-4})$.

E. Population Analysis

Mulliken population analysis allows the quantitative, fractional allocation of electron density among the atoms and bonds of a molecule in terms of a density matrix (1PDM or GDM), $P_{\mu\nu}$ for the ground and excited states[21].

One perspective of the electron population involves partitioning the electron density into atoms within the molecule. The gross atomic electron population is defined as

$$q_A = \sum_{\mu}^{A} q_\mu$$

where the sum is over all basis functions centered on atom A and

$$q_\mu = P_{\mu\mu} + \sum_{\nu \neq \mu} P_{\mu\nu} S_{\mu\nu}$$

is a gross population for the basis function ϕ_μ. It follows that this sum over all basis functions is the total electron population, N. The diagonal density matrix

element $P_{\mu\mu}$ represents the number of electrons directly associated with the normalized basis function ϕ_μ. The off diagonal term $P_{\mu\nu}S_{\mu\nu}$, $\mu \neq \nu$, arbitrarily represents 1/2 of the overlap electron population between the basis functions ϕ_μ and ϕ_ν; the total overlap electron population between ϕ_μ and ϕ_ν being

$$Q_{\mu\nu} = P_{\mu\nu}S_{\mu\nu} + P_{\nu\mu}S_{\nu\mu} = 2P_{\mu\nu}S_{\mu\nu}.$$

When appropriate, the gross atomic electron population can be further partitioned into π and σ electron populations by summing over the basis functions having π and σ symmetry

$$\sum_{\mu}^{A} q_\mu = \sum_{\mu}^{A_\pi} q_\mu + \sum_{\mu}^{A_\sigma} q_\mu.$$

Another perspective of the electron population which emphasizes the bonding within the molecule is the internuclear overlap population

$$q_{AB} = \sum_{\mu}^{A}\sum_{\nu}^{B} Q_{\mu\nu}$$

where one sum is over the basis functions centered on atom A and the other sum is over the basis function on atom B. Large positive q_{AB} represents strong bonding while large negative magnitudes represent strong antibonding. Like the gross atomic populations, q_{AB} can be partitioned into π and σ components.

To assess what portion of the gross atomic population is due to strictly to atomic electron density and not overlap population, a modified atomic population is calculated. This is defined as the gross atomic population with the dominant

adjacent atom overlap population subtracted. This is only an approximation since overlap population with nonadjacent atoms, while small in magnitude, is still present.

F. Potential Energy Surfaces and Avoided Crossings

Under the Born-Oppenheimer approximation the nuclei of a molecule move on the potential energy surface (PES) of a specified ground or excited state. The total energy (see section II.A.), as viewed as a function of the nuclear coordinates, constitutes the PES for a specified eigenvalue of the Hamiltonian matrix. The full PES is a high dimensional surface having the same dimension as the number of nuclear degrees of freedom in the molecule or complex plus another for the energy coordinate. Often, only a section of this multi-dimensional surface is necessary to investigate the important behavior of the system. In this study, the high dimension of the full PES is reduced to a planar section to observe the behavior of the surfaces for a possible avoided crossing. Since a potential avoided crossing may occur between the ^1L_a and ^1L_b states, the PES sections in this study represent a plot of the CIS or CIS-MP2 energy in the plane containing the ^1L_b-^1L_a difference vector connecting the ^1L_a and ^1L_b optimized geometries.

In many singlet and triplet manifolds avoided crossings between potential energy surfaces are common (no mixing between surfaces of differing spin symmetry can result in an avoided crossing, however). The avoided crossing

results from degeneracy in two or more unperturbed energy surfaces causing a surface repulsion in the perturbed system characterized by time-independent perturbation theory[33]. To rigorously establish the existence of an avoided crossing, the difference of Hamiltonian matrix elements for the ^1L_a and ^1L_b states in the diabatic representation

$$\Delta H = H_{11} - H_{22} = \langle \phi_1 | H | \phi_1 \rangle - \langle \phi_2 | H | \phi_2 \rangle$$

must be examined for a change of sign as the geometry passes through the avoided crossing region (ACR). Along the seam (point for 2-D sections) for which $\Delta H = 0$, the additional condition

$$H_{12} = \langle \psi_1 | H | \psi_2 \rangle \neq 0$$

must be met where H_{12} is the perturbation and ψ is the adiabatic representation of the wavefunction for the ^1L_a and ^1L_b states[34,35].

In this study, the CIS wavefunctions, generated under the B.O. approximation, represent adiabtic wavefunctions. Diabatic wavefunction representations are not explicitly constructed, although a straightforward method exists[36]. With the adiabatic wavefunctions produced in the CIS method, only perturbed energy surfaces are observed in the event of an avoided crossing. Without explicit diabatic states, the degeneracy of the unperturbed energy surfaces in the diabatic repesentation must be deduced by examining the change in the wavefunction character as the nuclear geometry is incremented through the ACR.

III. PROCEDURES AND METHODS

The computational quantum chemical calculations in this study use the Gaussian 92[37] (G92) suite of codes resident on the Cray Y-MP C90 supercomputer at the Pittsburgh Supercomputing Center (PSC) in Pittsburgh, Pennsylvania. The application of all HF and MP2 ground state methods[21] and CIS and CIS-MP2 excited state methods[17] used in this study are discussed in the G92 User's Guide and G92 Programmer's Reference. The MOs and density differences are plotted using InsightII v. 2.2.0 (Biosym Technologies, San Diego). The potential energy surfaces are plotted using Quattro Pro v. 4.00 (Borland International).

All calculations in this study are constrained to a planar indole geometry implying C_s symmetry in the ground and the ^1L_a and ^1L_b excited states. For indole under a planar nuclear configuration, 2N-3 = 29 degrees of freedom result where N = 16 is the number of atoms. An internal coordinate representation is used within G92. The standard numbering system of indole is given in figure 2.

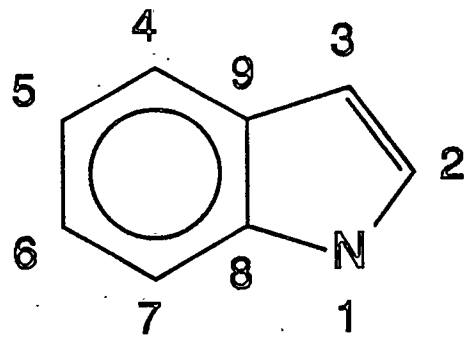


Figure 2. Numbering of Indole

The hydrogen number corresponds to the heavy atom number.

The internal coordinates used are

Bond Lengths:

N1-C2, C2-C3, C3-C9, C4-C9, C4-C5, C5-C6, C6-C7, C8-C9, N1-H1, C2-H2, C3-H3, C4-H4, C5-H5, C6-H6, C7-H7

Bond Angles:

C8-C9-C3, C9-C3-H3, C9-C3-C2, C3-C2-H2, C3-C2-N1, C8-N1-H1, C8-C9-C4, C9-C4-H4, C9-C4-C5, C4-C5-H5, C4-C5-C6, C5-C6-H6, C5-C6-C7, C8-C7-H7

The above listed internal coordinates were determined to be linearly independent using the "FOpt" option and link 0 command "%KJOB L103" within G92.

The CIS and CIS-MP2 calculations included all electrons. The ground state MP2 optimizations used the frozen-core approximation. Standard basis sets were used without modification. The HF (SCF), CIS GDM, MP2 GDM, and 1PDM were kept in terms of canonical HF MOs to calculate the total electron density planes. The MO and density xy planes parallel to the molecular plane were calculated at 0.2 Å intervals extending for 5 Å in the ±x and ±y directions from the center of charge.

A. Ground State Optimizations

1. Starting Geometry

The starting geometry for all HF and MP2 ground state optimizations was adapted from a tryptophan crystal structure[38]. The ring was made planar and all bond lengths to hydrogens were taken as 1.08 Å. The complete starting structure is provided in table 16. of the Appendix.

2. Methods/Options Used

The HF wavefunctions constructed at the starting geometry using the STO-3G, 3-21G, 6-31G, and 6-31G* basis sets and the MP2/6-31G* wavefunction also using this starting geometry, were optimized with the standard "OPT" option. This performs a Berny optimization[39] to achieve the minimum energy ground state structure. The SCF procedure was not explicitly specified in the input deck. This meant that G92 decided whether to run the SCF direct or not. For ground state indole and these basis sets, the conventional algorithm was used which stores the two-electron integrals in the AO representation externally instead of using the direct method which recomputes the needed integrals.

3. Resource Allocations

For all optimizations, 4 megawords of memory were allocated. Based on single point calculations, the following CPU times were allocated: 10 min for

HF/STO-3G, 35 min for HF/3-21G, 1.5 hrs for HF/6-31G and HF/6-31G*, and 2 hrs for MP2/6-31G*. The MP2/6-31G* optimization was restarted from the existing checkpoint file since more CPU time than allocated was required.

B. Excited State Optimizations

1. Starting Geometries

Since it was suspected that an avoided crossing may be present in the potential energy surfaces, it was critical for the state being optimized to have the lowest excited state energy. If not, the lack of convergence (CIS method) or convergence to the wrong state (CIS-MP2 method) may result. For the CIS-MP2 optimizations, the optimized HF/3-21G ground state geometry had the ^1L_a state lowest in energy so this was used for the starting geometry in the CIS-MP2 ^1L_a optimization. Since the ^1L_b energy was higher at this geometry, a modification to this geometry was needed to get the ^1L_b energy lower. It seemed reasonable that a starting geometry similar to the ^1L_b optimized geometry may have the ^1L_b state energy lower than the ^1L_a state energy. Using the INDO/S density differences as a guide to predict the optimized ^1L_b geometry, the following bond length elongations (in angstroms) were made to the HF/3-21G otimized ground state geometry: (N1-C2, 1.40), (C4-C9, 1.42), (C4-C5, 1.45), (C5-C6, 1.45), (C6-C7, 1.40), (C8-C9, 1.50). This modified geometry, which produced the ^1L_b state energy lower than the ^1L_a state energy, was used as the starting geometry for the

CIS-MP2 ^1L_b optimization. It should be noted that in determining the identity of a state, the transition dipole must be examined. The CIS coefficients were not reliable indicators of state identity since ^1L_a-^1L_b state mixing was present at these starting geometries. For the CIS ^1L_a and ^1L_b optimizations the corresponding CIS-MP2 optimized geometries were used as the starting geometries.

2. Methods/Options Used

At the above stated initial geometries, the CIS and CIS-MP2 geometry optimizations were performed using the 3-21G basis set. The standard "OPT" Berny optimization was specified in both methods. In contrast to the ground state optimizations, the excited state optimizations had SCF=Direct explicitly specified. For the CIS-MP2 method, the CIS must be run in MO basis which is the default and fastest CIS option. This excludes running the CIS direct which is nice for restarts, however. The CIS run in MO basis with SCF=Direct is stated to be a very efficient combination and was chosen for both optimization methods. By default, three CIS excited states were calculated at each optimization step. This allowed monitoring of state mixing during the optimization. Since the state being optimized is lowest in energy, the "root=1" CIS option is always chosen. The CIS options "densities" and "AllTransitionDensities" were specified to write the 1PDM and the transition densities between all 3 states to the checkpoint file. Since the CIS-MP2 is an energy correction only, these CIS options are only applicable to the CIS optimizations. The "geom=check" command was used to read the z-matrix of the

starting geometry from the optimized HF/3-21G checkpoint file. The "geom=modify" command is used to modify the HF/3-21G starting geometry in the CIS-MP2 ^1L_b optimization. The "guess=read" command is used to specify that the initial guess to the HF wavefunction is read from the optimized HF/3-21G checkpoint file. In overlay 8, IOp(6)=4 needs to be set to make sure all the required MO integrals are produced and in overlay 9, IOp(50)=1 is set to specify the CIS-MP2 method.

The population analysis for the CIS/3-21G ^1L_a and ^1L_b states is done at their corresponding optimized geometries using the CIS GDM specified with the "density=current" option. The "pop=bonding" option is specified to get both the standard Mulliken population analysis and a bonding Mulliken population analysis (total overlap populations) for which only density matrix terms involving pairs of basis functions on different centers are retained. A ground state population analysis of the HF/3-21G optimized wavefunction is used as the reference in the density difference calculations.

3. Resource Allocations

The CIS and CIS-MP2 optimizations had 8 megawords of memory allocated. The CIS-MP2 optimizations, which were performed first, required many restarts until "optimized" structures (as defined in the results chapter) were determined. The CPU time allocated for the ^1L_a CIS-MP2 optimization was 1.5 hrs initially and a 5 hr restart was run. The convergence problems of the CIS-MP2 method had

not been discovered at this point and the majority of the 5 hr restart was unnecessary. The optimization was expected to converge fully and terminate preventing the unnecessary use of CPU time. The ^1L_b CIS-MP2 optimization was monitored more closely with a 0.25 hr initial run and three 0.5 hr restarts performed. However, since the CIS is run in the MO basis, it is necessary to recompute and tranform the integrals at each restart making restarts costly. The CIS optimizations were allocated an overestimated 2 hr to avoid a restart since full convergence was expected.

C. Vertical Excitations

1. Transition Geometry

The MP2/6-31G* optimized ground state geometry was used for the vertical excitation calculations in accord with the Franck-Condon Principle. This geometry was chosen since optimized structures of molecules at this level of theory are usually found to be in close agreement with experiment[28].

2. Methods/Options Used

The CIS and CIS-MP2 vertical transition energies were calculated using the following progression of basis sets: STO-3G, 3-21G, 6-31G, 6-31+G, 6-31G*, and 6-31+G*. Both the CIS and CIS-MP2 energies for an excited state were calculated in a single job step since the CIS-MP2 energy is a correction to the CIS energy.

A separate job step for the ^1L_a and ^1L_b must be run because the CIS-MP2 correction is only applied to a single state, even though three CIS states (by default) are calculated within the job step. This is accomplished by specifying "root=1" in the first job step and "root=2" in the second step. Once the first job step has diagonalized the CIS matrix for the energies and wavefunctions, the second step need only read the specified CIS wavefunction off the checkpoint file before performing the CIS-MP2 correction. With the exception of the OPT command, all the options pertinent to the CIS-MP2 excited state geometry optimizations are used for the vertical excitation energy calculations. In the "guess=read" command, the use of the 6-31G* HF wavefunction in the MP2/6-31G* checkpoint file as an initial guess for the HF wavefunction of a different basis poses no problems. An initial guess HF wavefunction using a different basis set can be projected onto the new basis. The same is not true for the CIS wavefunction.

The vertical excited state density differences (CIS GDM - SCF DM, CIS GDM - 1PDM) for the ^1L_a and ^1L_b excited states and the MOs were calculated with the CUBE keyword at the MP2/6-31G* optimized geometry. For the density difference and MO calculations, the geometry was read from the checkpoint file using the "geom=check" command and the wavefunction was read using the "guess=(read,only)" command. The density differences were calculated from the total electron densities using the 6-31+G* basis (the excited states are better modeled with diffuse functions) while the MOs were calculated using the 6-31G*

basis. The total electron densities were produced with the "Cube=(Density,Total,Full)" command. To specify the CIS GDM in calculating the total electron density, the "Density=(Chk,Current)" is used and to specify the 1PDM, the "Density=(Chk,CIS=n)" command is used where n is the excited state of interest. Accessing the CIS 1PDM from the checkpoint file assumes that the CIS command "Densities" was used in the excited state calculation. The MOs were produced using the "Cube=Orbitals" command from the SCF density by default. The total electron densities were examined at a parallel plane 0.6 Å above the molecular plane. The MOs were calculated in a parallel plane 0.7 Å above the molecular plane. The density difference and MO plane grids were 101 by 101 points across with an interval size of 0.1 Å and centered at the molecular center of charge given in the G92 standard orientation. The plots of the MOs used a contour range and separation of (-0.1 to 0.1, 0.01). The (CIS GDM - SCF DM) density differences used a contour range and separation of (-0.01 to 0.01, 0.001) while the (CIS GDM - 1PDM) density differences used (-0.001 to 0.001, 0.0001).

3. Resource Allocations

For the vertical transition calculations eight megawords of memory were allocated for all basis sets. The vertical excitation calculations had the following CPU time allocations: 10 min for STO-3G and 3-21G; 1 hr for 6-31G, 6-31+G, and 6-31G*; and 2hr for 6-31+G*. The density and MO plots had 2 megawords of memory allocated and 90 s of CPU time.

D. Potential Energy Surfaces

1. Geometry Incrementation

Once the optimized geometries for the ^1L_a and ^1L_b states have been calculated, the optimized geometry difference vector (difference vector) connecting the minima can be constructed. This is done by subtracting the 29-dimension vector representing the ^1L_b optimized geometry (optimized geometry vector) from the ^1L_a optimized geometry vector to form the ^1L_b-^1L_a difference vector. Each component of the optimized geometry vector is a generalized internal coordinate of the optimized excited state.

The difference vector is scaled by the inverse of the desired number of intervals between the minima. This produces a linear geometry increment between the excited state minima. For the surfaces in this study, ten intervals between the minima were chosen. This produces 11 geometry points between (and including) the minima. An additional four increments in the negative direction of the ^1L_b-^1L_a difference vector starting at the ^1L_b minima and 4 increments along the direction of the ^1L_b-^1L_a difference vector starting at the ^1L_a minima produce an additional 8 geometry points for a total of 19 geometry points. Two sets of 19 geometry points over which to calculate the ^1L_a and ^1L_b electronic state energy result since the CIS and CIS-MP2 ^1L_b-^1L_a difference vectors are distinct.

2. Surface Construction

For the calculation of the CIS ^1L_a-^1L_b singlet manifold, individual input decks at each geometry point were constructed. Since the CIS method as implemented in G92, by default, calculates three excited states and some of their properties, the information for both the ^1L_a and ^1L_b states at a selected geometry point is acquired with one input deck. In the CIS-MP2 calculation, individual input decks for each excited state ("root=1" and "root=2") at each geometry point must be constructed since the CIS-MP2 method is applied to only one excited state at a time. Each of these input decks performs a single point CIS or CIS-MP2 excited state calculation at a geometry modified from the HF/3-21G optimized ground state geometry. A program was written to perform the geometry increments discussed in the preceeding section and write each geometry point (generalized coordinates) to a file in the format acceptable for the "geom=modify" command. A VMS DCL command file was written to append each geometry point file to a G92 input deck containing the necessary keyword command line to perform an excited state calculation at each modified geometry.

Each input job, which runs one input deck, was submitted in batch mode to the Cray Y-MP C90. To speed the calculation, four input jobs to be processed were kept on the Cray queue. Since the resulting output files were rather large (approximately 80 blocks with "pop=minimal"), the relevant excited state information was extracted from each output file as it was produced and then the output file was discarded. The important information to keep is: 1) excited state

and ground state energies, 2) transition dipoles and oscillator strengths, 3) excited configuration coefficients.

The excited state information was then merged into one file creating an excited state summary file from which the potential energy surface section was plotted. The excited state summary file was read into Quattro Pro as a text file and a Quattro Pro macro was written to convert the text data into numerical columns. The CIS or CIS-MP2 state energy and transition dipole angle at each geometry point was plotted as a function of the fraction of the difference vector corresponding to the geometry point.

3. Resource Allocations

For each CIS and CIS-MP2 single point calculation, 8 megawords of memory and 10 minutes of CPU time was allocated. At least approximately 320 blocks (4 X 80 blocks) of disk storage was neccessary on the VMS front end to prevent jobs being held on the Cray in the event the output file processing was interrupted.

IV. RESULTS AND DISCUSSION

A. Ground State Optimizations

1. Convergence and Timings

The optimized ground state geometries of indole (C_s symmetry) were obtained relatively easily using a series of basis sets: STO-3G, 3-21G, 6-31G, and 6-31G*. The convergence criteria of maximum force, RMS force, maximum displacement, and RMS displacement within Gaussian 92 were met indicating a converged geometric structure for the HF and MP2 methods and all basis sets considered.

The optimization timings are given in table 1. Increasing the number of basis functions results in increased CPU time required. The MP2 method requires substantially larger CPU time due to a dominant partial integral transformation which scales as $O(Nk^4)$[21], where N is the number of electrons and k is the number of basis functions. Within the HF method the number of optimization steps is constant indicating an independence of basis set (for the basis sets used) in the rate of structure convergence. The MP2 method actually requires fewer optimization steps (though nearly the same) than the HF method. This suggests that the method used for optimization and not the basis determines the speed of structure convergence. The choice of basis will affect the SCF convergence rate within each optimization step, however. Optimizations using different basis sets

having the same number of basis functions require the same time indicating the independency on the number of gaussian primitives.

Table 1. Ground State Geometry Optimization Timings

method/basis	basis functions	optimization steps[a]	CPU[b] (min)
HF/STO-3G	52	6	2
HF/3-21G	95	6	15
HF/6-31G	95	6	15
HF/6-31G*	149	6	82
MP2/6-31G*	149	5	144

[a]All optimizations start from crystal coordinates (see text).
[b]Cray Y-MP C90 running Gaussian92

2. Optimized Ground State Geometric Structure

The optimized internuclear distances are listed in table 2 and compared with one experimental crystal structure[38]. The theoretical geometries are relatively consistent but do not agree well with the crystal structure for most bond lengths. Bond length alternation in the calculated structures is observed in the benzene ring bonds not common to the pyrrole ring. The HF/STO-3G geometry is in most discord with the HF/split valence basis geometries in the bonds involving nitrogen (N1-C2 and N1-C8). This is probably due to the enhanced ability of the split valence basis to produce carbon orbitals in the anisotropic environment which overlap the contracted nitrogen orbitals better producing shorter bond lengths. The

Table 2. Ground State Optimized Internuclear Distances (in Å)

	HF/STO-3G	HF/3-21G	HF/6-31G	HF/6-31G*	MP2/6-31G*	expt[a]
N1-C2	1.400	1.388	1.383	1.373	1.380	1.377
C2-C3	1.342	1.349	1.353	1.348	1.376	1.344
C3-C9	1.450	1.446	1.444	1.442	1.430	1.451
C8-C9	1.401	1.404	1.405	1.401	1.423	1.380
C4-C9	1.406	1.394	1.398	1.399	1.408	1.412
C4-C5	1.371	1.375	1.379	1.379	1.388	1.397
C5-C6	1.410	1.401	1.404	1.404	1.413	1.386
C6-C7	1.370	1.374	1.379	1.375	1.389	1.391
C7-C8	1.403	1.390	1.393	1.394	1.401	1.400
N1-C8	1.396	1.377	1.379	1.371	1.377	1.391

[a]Crystal structure (see text)

effect of polarization functions in going from the 6-31G basis to 6-31G* basis is also most pronounced for these bonds enabling a shift of electron density into the carbon-nitrogen internuclear region resulting in bond lengths about 0.01 Å shorter. The polarization functions have little effect on the other bond lengths. Varying the number of gaussian primitives in going from the 3-21G basis to the 6-31G basis the geometry varies insignificantly by less than 0.006 Å.

Compared to the HF/6-31G* structure, the MP2 correlation tends to lengthen all bonds by about 0.01 Å except for the C3-C9 and N1-C8 bonds which are involved in the linkage of the enamine moiety to the benzene ring. The C3-C9 bond length is contracted by about 0.01 Å and the N1-C8 bond is invariant. The C2-C3 and C8-C9 MP2/6-31G* bond lengths elongate significantly from both the HF and crystal values. This is an interesting result in light of the observation that MP2/6-31G* structures of many other molecules are found to closely agree with their experimental structure. This is thought to be due to a cancellation of errors induced by anharmonicity and MP2 bond length extension.[28] Interestingly, it is the C2-C3 and C8-C9 bonds which have near perfect agreement with another recently reported SCF ground state structure calculated using a double-zeta level basis with K-shell orbitals replaced by compact effective core potentials.[19] The rest of the MP2/6-31G* bond lengths are contracted by over 0.01 Å compared to this structure. The MP2/6-31G* geometry is used as the ground state structure in excited state calculations involving the Franck-Condon approximation. The complete optimized structure is provided in table 17 in the Appendix.

3. Ground State Dipole Moments

The calculated permanent dipole moments at the optimized geometries are provided in table 3. The ground state of the indole molecule has C_s symmetry and the dipole moment has nonzero x,y components which can equivalently be expressed in terms of a magnitude and angle from an arbitrary axis chosen to be the long axis. The magnitudes are in Debyes (D) and the direction is in degrees away from the long axis with a negative angle and positive dipole end toward the nitrogen. The approximate long axis is calculated from each optimized geometry as the vector connecting the midpoint of the C5-C6 bond and the C2 atom.

The dipole magnitudes and directions are qualitatively independent of the various geometries, density matrices, and methods used in the dipole calculation. The dipole magnitude tends to increase with basis and level of theory. The HF/STO-3G//SCF calculated dipole has a smaller magnitude than the calculated dipoles using a split valence basis. Also, it has a direction about 3 degrees farther away from the long axis. The HF//SCF dipoles using a split valence basis vary slightly in magnitude but have nearly equal directions. The effect of MP2 correlation using the SCF density tends to increase the magnitude to very good agreement with experiment[40] while preserving the direction. This agreement may be coincidental since the experimental magnitude is a benzene solution measurement. The effect of using the MP2 density increases the magnitude while slightly decreasing the angle.

Justifying the MP2 correlation effects on the dipole moment is not easily