

Theoretical Investigation of Positional Substitution and Solvent Effects on n-Cyanoindole Fluorescent Probes

Salsabil Abou-Hatab and Spiridoula Matsika*

Department of Chemistry, Temple University, Philadelphia, PA 19122, USA

E-mail: smatsika@temple.edu

Abstract

The absorption and fluorescence of indole and n-cyanoindole derivatives are modeled in gas phase and aqueous solution using high level quantum mechanical methods and implicit solvation. These molecules have been experimentally examined as fluorescent probes for studying the structure, function, and hydration status of proteins, and it is found that substitution of the cyano group on different positions of indole has diverse effects on the absorption and fluorescence spectra in water solvent. Our calculations predict that in absorption the L_b excited state is lower in energy than the L_a state for all positional isomers in gas phase and in solution. In fluorescence, however, water solvent causes level inversion leading to emission from the L_a excited state for indole and n-cyanoindole derivatives with the cyano on the six-membered ring. On the other hand, when cyano substitution is on the five-membered ring L_a is not stabilized enough and emission occurs from the L_b excited state. In addition, we predict that the relatively high fluorescence intensity of 4-cyanoindole in aqueous solution results from minimization of radiationless decay pathways since both absorption and

fluorescence occur from the lowest excited state (unlike the other derivatives).

Introduction

Fluorescence microscopy and spectroscopy are non-invasive techniques commonly used for visualizing, monitoring and studying the structure, function, and dynamics of chemical and biological systems.¹⁻⁷ The sensitivity of the photophysical properties of the fluorophores used in these devices to the local environment is exploited to optimize signaling and detection.^{2,6} A few natural biological molecules in living organisms such as aromatic amino acids in proteins can fluoresce.⁷ As a result, they are employed as naturally occurring fluorophores for studying cell and enzymatic processes in vivo. Amongst the amino-acids, tryptophan (Trp) is the most widely used biosensor due to its relatively high fluorescence quantum yield and environmental sensitivity. However, its absorption spectrum overlaps with other aromatic amino acids such as phenylalanine and tyrosine making it difficult to selectively excite, and it has complex decay kinetics.⁷ For these reasons unnatural amino-acid derivatives such as functionalized tryptophans were explored as potential fluorescent probes.

The addition of functional groups is a commonly used technique to tune the fluorescence intensity and quantum yield of conjugated aromatic chromophores.⁸⁻¹⁰ Novel cyano-substituted tryptophans have been explored as potential probes for protein structure and functions. 6 and 7-cyanotryptophan (6-TrpCN and 7-TrpCN) were synthesized and shown to be much brighter fluorescent probes compared to tryptophan.¹ In particular, 6-TrpCN was used in FRET experiments as a fluorescence donor for the study of protein conformational events.¹ The exposure of 5-cyanotryptophan (5-TrpCN)¹¹ and 7-TrpCN¹² to water was found to significantly quench their fluorescence making them potential probes for the local hydration status of proteins. Likewise, the optically active chromophore of tryptophan, indole, has been explored in the search of optimizing fluorescent probes because it is the active chromophore in Trp. Indole is involved in various bacterial processes such as antibi-

otic resistance, biofilm formation, quorum formation and sensing.^{13,14} In addition, indole is found in phytohormones that regulate the growth of plants known as indole auxins, hallucinogenic drugs such as tryptamines, and is a precursor to neurotransmitters in the central nervous system such as serotonin.^{15,16} Functionalizing the different positions of indole rings (see Figure 1) with an electron withdrawing group has been found to tune its absorption and fluorescence spectra differently.¹⁷ Substitution of a cyano group on position 4 was found to have a significantly longer fluorescence lifetime and high quantum yield in water solvent relative to other positions of substitution.¹⁸ Moreover, 5, 6, and 7-cyanoindole probes were found to be highly sensitive to the polarity of the solvent.¹⁸

To understand the unique behavior of these cyanoindole and cyanotryptophan isomers, one must closely examine the changes in the electronic structure and photophysical properties of indole in different media.^{19,20} Indole has two low-lying excited states with $\pi\pi^*$ character denoted L_b and L_a according to the Platt nomenclature.²¹ The absorption and radiative decay of indole is mainly affected by the dependence of the nature of the L_b and L_a excited states on the polarity of the micro-environment.²²⁻²⁴ The L_b state of indole is found to have a small dipole moment like that of the ground state with electron density delocalized across the long-axis of indole (as seen in Figure 1). The L_a excited state is ionic-like in nature due to its large dipole moment depicted by an electron delocalization in the HOMO across the short-axis of the molecule.²⁰ The L_a excited state is found to be more sensitive to solvents of increasing polarity relative to the L_b excited state.²⁴

There have been many theoretical studies of indole and its photophysical properties.^{20,22,25-27} In the Franck Condon region the L_b is found to be lower in energy than the L_a excited state in all media. The energetic ordering of the L_a and L_b excited states then changes upon excited state relaxation to the S_1 minimum in different media. Indole is found to fluoresce from the $S_1(L_b)$ in vacuum whereas in aqueous solution, the polar solvent stabilizes the L_a excited state, resulting in its energy being lower than the L_b state. As seen in Figure 1, the Transition Dipole Moments (TDM) of $L_b(1\pi\pi^*)$ and $L_a(2\pi\pi^*)$ are nearly orthogonal in

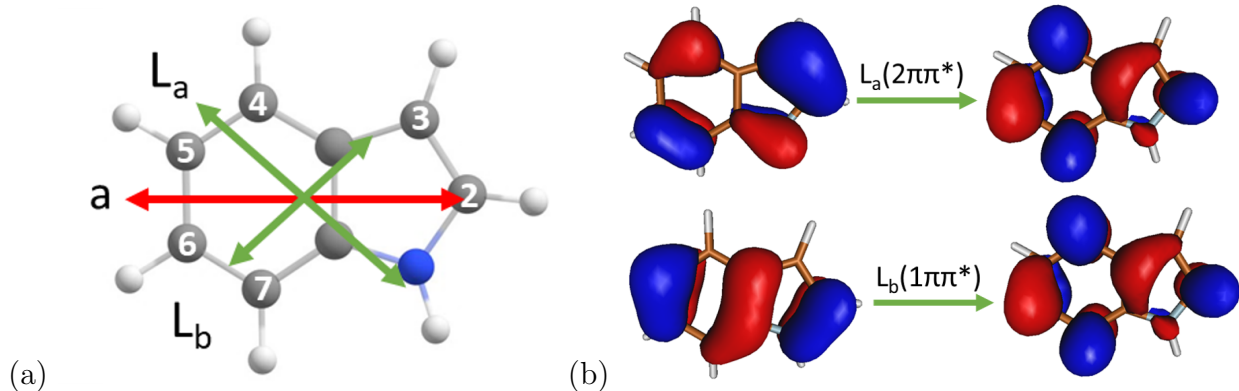


Figure 1: (a) Schematic diagram of numbered position of substitution on the ring of indole and the Transition Dipole Moment (TDM) vectors of the two low-lying $\pi\pi^*$ excited states of indole. (b) The molecular orbitals responsible for these $L_b(1\pi\pi^*)$ and $L_a(2\pi\pi^*)$ transitions computed at the CASSCF level.

vacuum and there is a higher probability of populating the $L_a(2\pi\pi^*)$ state than the $L_b(1\pi\pi^*)$. Addition of a weakly electron withdrawing group such as fluorine on different positions of indole has been found to dramatically influence the orientation of the TDM, the magnitude of the permanent dipole moment, and consequently the L_a/L_b nature of the S_1 minima.^{28–30} As a result, the orientation of the TDM and magnitude of the permanent electric dipole moment have been used to determine the character of the S_1 minima for substituted indoles.

In this present work we will investigate the effect of substitution of a strong electron withdrawing functional group, cyano, on different positions n of indole (denoted n -CNI) using quantum mechanics and implicit solvation. The work is motivated by the above described experimental studies, which have examined experimentally the absorption and fluorescence properties of the cyanosubstituted indoles.^{12,17,18} We aim to explain why the absorption and fluorescence spectra change drastically when varying the position of substitution of the cyano group on the ring of indole. More specifically, we want to understand why substitution on position 4 of indole has the brightest fluorescence in water solvent relative to substitution on the other carbons of indole. Likewise, studying the substitution of cyano on positions 5 and 7 of indole may help explain why the fluorescence of 5 and 7-TrpCN are significantly quenched as the ratio of water solvent increases.

Methods

We modeled indole and n-cyanoindole (n-CNI) fluorescent probes by varying the cyano electron-withdrawing group on the carbons of indole’s pyrrole and benzene ring where n represents the positions of substitution numbered 2 through 7 as seen in Figure 1. Geometry optimization and vibrational frequencies were computed using B3LYP/6-31G(d) in gas phase to localize the ground state minima.

Absorption

We begin our investigation by performing a benchmark study of the absorption and fluorescence of indole with various quantum mechanical methods in order to identify a reliable method to study the behavior of these n-CNI probes. **Four vertical excitation energies were computed using both single reference and multireference methods.** The single reference methods Equation of Motion Coupled Cluster Singles and Doubles (EOM-CCSD), and hybrid exchange correlation functional CAM-B3LYP of Density Functional Theory (DFT) were used with basis set 6-311+G(d) in gas phase. **The aug-cc-pVTZ basis set was also used with EOM-CCSD in calculating the excited states of indole in order to check the effect of basis set on the excitation energies.** Excited state energies require larger and often more diffuse basis sets compared to ground state geometry optimizations in order capture the more diffuse **wavefunctions.** The excited states computed at the EOM-CCSD level were characterized as L_a or L_b by performing a Natural Orbital Transition (NTO) calculation using Qchem5.0 software package.³¹

The excited states were also computed at the Complete Active Space Self-Consistent Field (CASSCF) with the addition of Complete Active Space Perturbation Theory (CASPT2) with basis set aug-cc-pVDZ using a (12,11) active space (12 electrons in 11 orbitals) for indole in gas phase. **The Dunning basis set was used in this case since these basis sets are more compatible with CASSCF/CASPT2.** The orbitals involved in the active space are shown in Supporting Information (Figure S2). For indole all the 9 π orbitals were included in the

active space. In addition, the σ and σ^* orbitals along NH were also included, providing an active space of 11 orbitals of which 6 are occupied. The excited L_a and L_b states were computed with CASPT2 only in gas phase because the molpro2015 package³² does not include an implicit solvation model that is compatible with this method. Oscillator strengths were computed using the scalar of the square of the transition dipole moments at the CASSCF level and the energies calculated at the CASPT2 level of theory.

Similarly, vertical excitations for all n-CNI systems from the ground state minima were computed in gas phase with EOM-CCSD/6-311+G(d) and CASPT2/aug-cc-pVDZ. An active space of (14,13) was used to compute excited states for n-CNI probes. In addition to previously described orbitals for indole we included two more orbitals (one occupied and one virtual) to describe the CN substituent. Photophysical properties were computed and analyzed the same way as described in the benchmark of indole.

Fluorescence

Excited state optimizations were performed using CAM-B3LYP and EOM-CCSD with basis set 6-311+G(d) in gas and aqueous phase. We would like to model indole in gas phase and solution, and it is known that it emits from two different excited state minima in the two phases. Since CAM-B3LYP did not give the observed L_b minimum for indole in gas phase it was necessary to perform optimization with EOM-CCSD. Optimization of the S_1 and S_2 excited states with CAM-B3LYP/6-311+G(d) in gas phase and with implicit solvation converges to the L_a excited state minimum. Optimization of the S_1 state with EOM-CCSD in gas phase converged to the L_b minimum for all n-CNI except for 4-CNI which converges to L_a minimum. Frequency calculations were computationally prohibitive for EOM-CCSD, so we assume that the converged stationary points are true minima rather than saddle points. Comparison of the obtained geometries with previously obtained geometries for indole confirm this assumption. Single point calculations at the gas phase minima (L_a minima from CAM-B3LYP and L_b minima from EOM-CCSD) were used to calculate the

vertical emission energies. The EOM-CCSD/6-311+G(d) in gas and aqueous phase and CASPT2/aug-cc-pVDZ in gas phase methods were used for these calculations.

Polarizable Continuum Model

Solvation effects for bulk aqueous solution were included implicitly with a polarizable continuum model (PCM). The Integral Equation Formalism variant of the Polarizable Continuum Model (IEFPCM) which is the default Self-Consistent Reaction Field (SCRF) method in the Gaussian package³³ was used.

Based on the Franck-Condon principle the nuclei move much slower than the electrons because they are more massive, and as a result, upon absorption, the nuclei of the solvent do not have enough time to respond to the electron density of the excited state while the electrons of the solvent become polarized based on the excited state density. In continuum models this is represented by defining the solvent to have a fast and slow component that correspond to the motion of the electrons and nuclei respectively. The vertical excited state energies in solution are thus determined by non-equilibrium solvation where only the response of the electrons of the solvent is allowed to change with excitation (fast component defined by the dielectric of the solvent).

Fluorescence occurs at a slower timescale than absorption which allows for the solvent to reorient its dipoles to the dipole of the solute at its excited state equilibrium geometry. The energy at the excited state minimum is then computed by using equilibrium solvation for the excited state. The slow/inertial charges of the solvent reaction field in equilibrium with the excited state density are saved and then are kept frozen in a subsequent non-equilibrium solvation energy calculation of the ground state energy.

For EOM-CCSD/PCM we used the perturbation theory energy (PTES) approach where only part of the CCSD reduced one particle density matrix is used for the correlation reaction field.³⁴⁻³⁶ The state-specific solvation approach was used. At the TDDFT level only non-equilibrium solvation was used within the linear response approach.³⁷

The Gaussian package (gaussian 09 and gaussian 16)^{33,38} was used for the DFT and EOM-CCSD calculations. QChem was used to obtain the dipole moments and NTOs at the EOM-CCSD/PCM level.³¹ MOLPRO was used for the CASSCF and CASPT2 calculations.³²

Results

Benchmark

Since indole has been studied extensively, it is used to benchmark the quantum mechanical methods which will later be used to predict the absorption and fluorescence of n-CNI fluorescent probes. The excitation energies of indole to the two low-lying excited $\pi\pi^*$ states were computed with single reference methods, EOM-CCSD, Density Functional Theory with CAM-B3LYP, and multi-reference wavefunction-based methods CASSCF, and CASPT2 in gas phase. The vertical excitation energies were also studied in aqueous solution implicitly using PCM. The vertical excitation energies to the L_b and L_a excited states are reported in Table 1 at all levels of theory along with the corresponding oscillator strengths, f , and the computed electric dipole moments μ of indole at the ground state (GS) and excited states.

To capture the spectroscopic profile of indole the methods must satisfy four general observables: (1) The order of the $1\pi\pi^*$ and $2\pi\pi^*$ excited states at the absorption in vacuum and water solvent should be L_b and L_a respectively; (2) the character of the brightest absorbing state in both mediums is L_a ; (3) the energy gap between the two states is observed to be 0.5 eV in gas phase and decreases to 0.3 eV upon solvation;^{20,23} and (4) indole fluoresces from the L_b excited state minimum in gas phase and undergoes energy level inversion and emits from the L_a state in aqueous solution.

All of the methods are found to agree with experiment in predicting the order of the excited states accurately. CAM-B3LYP underestimates the energy gap between the two excited states by an order of magnitude at the absorption. EOM-CCSD on the other hand, predicts the gap in vacuum to be 0.49 eV (very close to the experimental value) and in

Table 1: Vertical excitation energies, E , and corresponding oscillator strengths, f , of the $L_b(1\pi\pi^*)$ and $L_a(2\pi\pi^*)$ excited states of indole benchmarked in gas and aqueous phases with various methods using the 6-311+G(d) basis set unless otherwise specified. The magnitudes of the permanent electric dipole moment, μ , for the ground state (GS) and two low-lying $\pi\pi^*$ states are also reported in Debye. Previous results at the algebraic diagrammatic construction scheme of the polarization propagation (ADC(2) and ADC(3)) with cc-pVDZ basis set are also reported, as well as experimental values when available.

	GS	$L_b(1\pi\pi^*)$			$L_a(2\pi\pi^*)$		
	μ (D)	E (eV)	f	μ (D)	E (eV)	f	μ (D)
Gas Phase							
CAM-B3LYP	2.151	4.992	0.038	2.527	5.039	0.110	3.795
EOM-CCSD	2.134	4.858	0.027	2.230	5.344	0.108	5.260
EOM-CCSD/aug-cc-pVTZ		4.816	0.027		5.233	0.102	
CASPT2/aug-cc-pVDZ		4.352	0.028		4.809	0.086	
ADC(2) ²²		4.92	0.04	2.4	5.25	0.11	5.2
ADC(3) ²²		4.70	0.03	2.2	5.25	0.12	5.1
Exp	2.09 ³⁹	4.37 ²³	0.045 ⁴⁰	2.3 ⁴¹	4.88 ²³	0.123 ⁴⁰	5.4 ⁴²
Aqueous Solution							
CAM-B3LYP	2.961	4.941	0.117	5.490	5.011	0.072	3.191
EOM-CCSD	2.889	4.879	0.024		5.219	0.088	
EOM-CCSD/aug-cc-pVTZ		4.840	0.024		5.183	0.094	
ADC(2)/COSMO ²²		4.90		3.6	5.10		5.9
Exp ²⁰		4.31			4.59		

aqueous solution to be 0.34 eV (again very close to experiment). As expected, CASSCF overestimate the excited state energies, however, the addition of dynamical second order perturbation theory corrects this and gives the most accurate prediction of these energetics in gas phase with a small error of 0.13 eV. The comparison between ADC(2) and ADC(3) from previous work²² highlights that correlation energy is very important for the energetics of these states. The excitation energies are best predicted by CASPT2, which predicts both excitation energies within 0.07 eV of the experimental gas phase energies. The EOM-CCSD method on the other hand has an error of about 0.5 eV when the 6-311+G(d) basis set is used. The discrepancy with experiment decreases to about 0.4 eV when the larger aug-cc-pVTZ basis set is used. It has been shown before in the literature that EOM-CCSD for electronic excitation energies overestimates their energy by 0.3-0.5 eV.⁴³⁻⁴⁵ In addition to basis set effects, other sources of error include the exclusion of triple excitations and lack of vibronic effects. It has also been shown that inclusion of implicit solvation model PCM to EOM-CCSD increases the error by an average of 0.2 eV for solvated systems.⁴⁴ PCM also includes approximations for the polarization of the solvent and neglects contributions from special effects such as H-bonding. The error however is consistent for all systems (systematic) and that is why we are able to obtain correct trends, as will be discussed later. The difference between the 6-311+G(d) and aug-cc-pVTZ basis sets is small while the computational cost increases significantly when going to aug-cc-pVTZ, so it is seen that it is not worth using this basis set for all substituents. Since EOM-CCSD/6-311+G(d) predicts the gap and, as we will see in the next section, all the trends correctly, it will still be used in this work to lead us to important qualitative conclusions, especially since we cannot add solvation in the CASPT2 results.

The electrostatic dipole moments, μ , of indole are also investigated in order to study the effect that water solvent has on the fluorescence of n-CNI. The dipole moment of the L_b excited state was found to be similar to that of the ground state of indole of approximately 2 D whereas the L_a is often referred to as a disguised charge-transfer state due to its large

charge separation reflected by a high dipole moment of 5.4 D.^{46,47} The difference in the dipole moments for the two excited states is not captured by CAM-B3LYP. EOM-CCSD is found to predict this property with most accuracy and thus can be a reliable method in studying the solvation effect on n-cyanoindole probes.

Table 2: Vertical emission energies of indole and corresponding oscillator strengths, f , computed as single point calculation from the L_b and L_a minima. Previous ADC(3) results are also reported,²² as well as experimental values when available.

	L_a minimum				L_b minimum			
	S_1 (eV)	f	S_2 (eV)	f	S_1 (eV)	f	S_2 (eV)	f
Gas Phase								
CAM-B3LYP	4.464	0.159	4.907	0.003				
EOM-CCSD	4.757	0.072	4.794	0.103	4.541	0.04	5.201	0.091
CASPT2	4.261	0.016	4.293	0.120	4.042	0.035	4.667	0.070
ADC(3) ²²	4.54	0.04	4.60	0.12	4.42	0.03	5.03	0.13
Exp ⁴⁸					4.20			
Aqueous Phase								
CAM-B3LYP	4.172	0.365	4.905	0.020				
EOM-CCSD	4.2899	0.095	4.648	0.002				
ADC(3) ²²	4.07				4.67			
Exp ¹⁷	3.59							

To evaluate the performance of the methods in predicting the fluorescence of indole in gas and aqueous phase, we optimized the $L_b(1\pi\pi^*)$ and $L_a(2\pi\pi^*)$ excited states with CAM-B3LYP/6-311+G(d). At this level of theory both states optimized to the same S_1 minimum which was characterized to be an L_a equilibrium geometry based on comparison with previously published geometries and molecular orbitals. The geometry is given in Supporting Information (SI). The inability for CAM-B3LYP to optimize to the L_b excited state in gas phase and predict the dipole moment of the excited states accurately makes it an inadequate method to study the fluorescence of indole with. Optimization of the S_1 state using EOM-CCSD leads to a minimum with L_b character in accordance with experiment. The vertical emission of indole from the L_b minimum and oscillator strength are reported in Table 2 at various levels of theory. In the gas phase EOM-CCSD overestimates the energy of the emission (at the L_b minimum) by about 0.34 eV whereas CASPT2 underestimates

it by 0.16 eV. The adiabatic energy of the L_b state in indole is found at the EOM-CCSD level to be 4.69 eV, while at the CASPT2 level it is 4.12 eV. Experimentally this energy is found to be 4.37 eV, between the two theoretical values. In aqueous solution we only have EOM-CCSD results which overestimate the emission energy (from the L_a minimum) by 0.7 eV. For the remaining of this work the absorption and fluorescence of n-CNI probes will be studied with EOM-CCSD and CASPT2.

Absorption

Bright Excited State

In order to examine the absorption spectra of n-CNI derivatives we calculated their excited states using EOM-CCSD and CASPT2 which were determined to be the best approaches for indole. We focus the following discussion on the two low-lying $\pi\pi^*$ states. It should be mentioned that in indole the Rydberg states have also been found to play an important role in gas phase. In order to determine their role in the other n-CNI systems we also computed them theoretically, and they are reported in SI (Figure S1, Tables S1-S4). It is clear from these results that the Rydberg states are destabilized by CN substitution, and are not expected to play a role here. So, we continue by only focusing on the $\pi\pi^*$ states. The energies and corresponding oscillator strengths of the lowest two $\pi\pi^*$ states for all n-CNI derivatives in the gas phase are reported in Table 3 while results in solution are given in Table 4. Both in gas phase and in aqueous solution the first excited state has L_b character while the second $\pi\pi^*$ state has L_a character. NTOs describing the two states are shown in SI (Figures S7,S8). Almost in all cases the oscillator strength is higher for the L_a state (with 4-CNI and 7-CNI in aqueous solution notable exceptions), although the ratio of the L_a over L_b oscillator strengths changes dramatically between substituted systems. In order to compare with experiment we use λ_{max} of the experimental spectra reported by Hilaire et al.¹⁷ and compare with the theoretical state that has the highest oscillator strength. We term this the bright excited state in the following discussion. However, as mentioned already,

in some cases the oscillator strengths of the $1\pi\pi^*$ and $2\pi\pi^*$ states are close in magnitude and using the term "bright" is not always the best description since both states have similar "brightness".

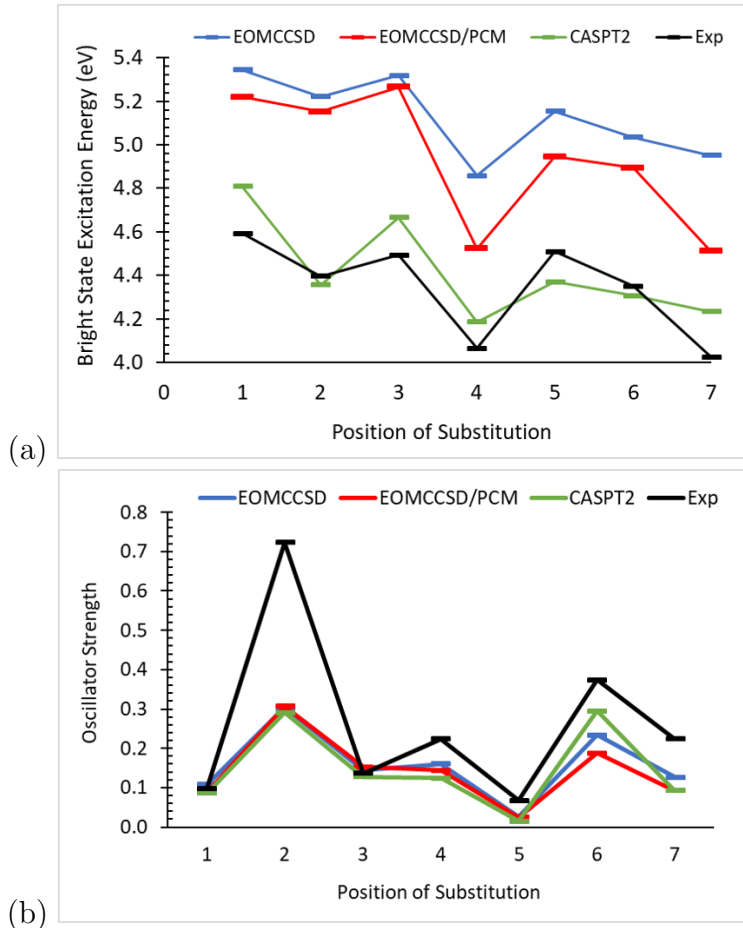


Figure 2: (a) Experimental absorption maxima (black), taken from Ref.,¹⁷ and theoretical bright excitation energies for all positional isomers of cyanoindole in gas and aqueous phase. (b) Scaled experimental molar extinction coefficients, ϵ , (by setting ϵ value of indole equal to the theoretical EOM-CCSD/PCM value) and theoretical oscillator strengths of the bright state.

The bright excited state of n-CNI is plotted in Figure 2(a) as a function of position of substitution of the cyano group in vacuum and aqueous solution. Both methods predict the bright state to be $L_a(2\pi\pi^*)$ in gas and aqueous phase except for 4-CNI which is predicted at the EOM-CCSD/PCM level to absorb more to the $L_b(1\pi\pi^*)$ state in water solvent. 7-CNI has almost identical oscillator strengths for the two states in solution, although the oscillator

strength for L_b is slightly higher, so we use this state in Figure 2. The qualitative comparison with experiment depicted in Figure 2 shows that in general theory predicts the experimental trend. The comparison is particularly good between experiment and EOM-CCSD/PCM. A more quantitative comparison is attempted with correlation plots between theory and experiment, which are shown in SI (Figure S6). The energetics predicted by coupled cluster correlate more with the experimental results (correlation plots have an $R^2 = 0.82$ for gas phase and 0.85 in PCM) compared to CASPT2’s performance ($R^2 = 0.64$).

The energy is red-shifted upon CN substitution for all positions of indole, as observed experimentally. Strongest shifts are observed for addition of CN on positions 4 and 7 along the short axis of benzene ring. There are two reasons for the strong red shift in these two systems. Both L_a and L_b states are red shifted by substitution of the electron donating group on positions 4 and 7. As has been discussed by Callis^{46,47} this is because of electron density moving from the five membered ring to the six membered ring (and especially at positions 4 and 7) during the transition. The relative intensities of the L_a and L_b states are also affected by substitution however, and for 4-CNI and 7-CNI in solution L_b becomes more intense than L_a . As a result in these two systems the bright state is even further redshifted because it now corresponds to S_1 rather than S_2 .

The oscillator strengths of the bright state computed with various methods are plotted in Figure 2(b) as a function of position of substitution. The experimental molar extinction coefficient is scaled by setting its value for indole equal to the EOM-CCSD/PCM theoretical oscillator strength and scaling all other values accordingly. The methods reproduce the experimental trend at least qualitatively. EOM-CCSD reproduces this trend with the most precision with R^2 value around 0.86 in the correlation plots for EOM-CCSD/PCM, whereas, the oscillator strength computed at the CASPT2 level has a weaker correlation with the experimental molar extinction coefficient with R^2 value of 0.73. 2-CNI and 6-CNI are found to have the highest oscillator strength for the L_a transition. This indicates that the probability of populating the $L_a(2\pi\pi^*)$ excited state should be highest for 2-CNI than all the other

isomers as observed experimentally in aqueous solution by the molar extinction coefficient. Even though theory predicts 2-CNI to have the highest oscillator strength it does not predict it at the same degree as is observed experimentally.

Table 3: Vertical excitation energies of the $L_b(1\pi\pi^*)$ and $L_a(2\pi\pi^*)$ states for indole and n-CNI isomers and their corresponding oscillator strengths in parenthesis computed with CASPT2/aug-cc-pVDZ and EOM-CCSD/6-311+G(d) in gas phase.

	EOM-CCSD/6-311+G(d)		CASPT2/aug-cc-pVDZ	
	$L_b(1\pi\pi^*)$	$L_a(2\pi\pi^*)$	$L_b(1\pi\pi^*)$	$L_a(2\pi\pi^*)$
In	4.858 (0.027)	5.344 (0.108)	4.352 (0.028)	4.809 (0.086)
2-CNI	4.684 (0.085)	5.221 (0.304)	3.942 (0.094)	4.357 (0.290)
3-CNI	4.910 (0.033)	5.318 (0.146)	4.382 (0.029)	4.665 (0.127)
4-CNI	4.648 (0.067)	4.858 (0.160)	3.970 (0.047)	4.186 (0.124)
5-CNI	4.716 (0.012)	5.153 (0.026)	3.988 (0.015)	4.370 (0.014)
6-CNI	4.648 (0.024)	5.034 (0.234)	3.998 (0.020)	4.307 (0.295)
7-CNI	4.577 (0.083)	4.952 (0.126)	4.028 (0.033)	4.233 (0.092)

Table 4: Vertical excitation energies of the $L_b(1\pi\pi^*)$ and $L_a(2\pi\pi^*)$ states for indole and n-CNI isomers and their corresponding oscillator strengths in parenthesis at the EOM-CCSD/6-311+G(d) level in aqueous phase using PCM.

	$L_b(1\pi\pi^*)$	$L_a(2\pi\pi^*)$
In	4.879 (0.024)	5.219 (0.088)
2-CNI	4.653 (0.092)	5.151 (0.307)
3-CNI	4.949 (0.028)	5.265 (0.153)
4-CNI	4.523 (0.145)	4.717 (0.072)
5-CNI	4.700 (0.008)	4.946 (0.025)
6-CNI	4.560 (0.057)	4.893 (0.188)
7-CNI	4.510 (0.098)	4.756 (0.092)

L_a and L_b states and resulting spectra

A more complete picture of the spectra needs consideration of both excited states. The two peaks associated with the $\pi\pi^*$ transitions to the L_a and L_b excited states are expected to be present in the absorption spectra of indole and indole derivatives, although they are not always observed. When the energy gap between the two excited states is small, the second peak is most likely hidden under the envelop of the broadness of the brighter absorbing state.

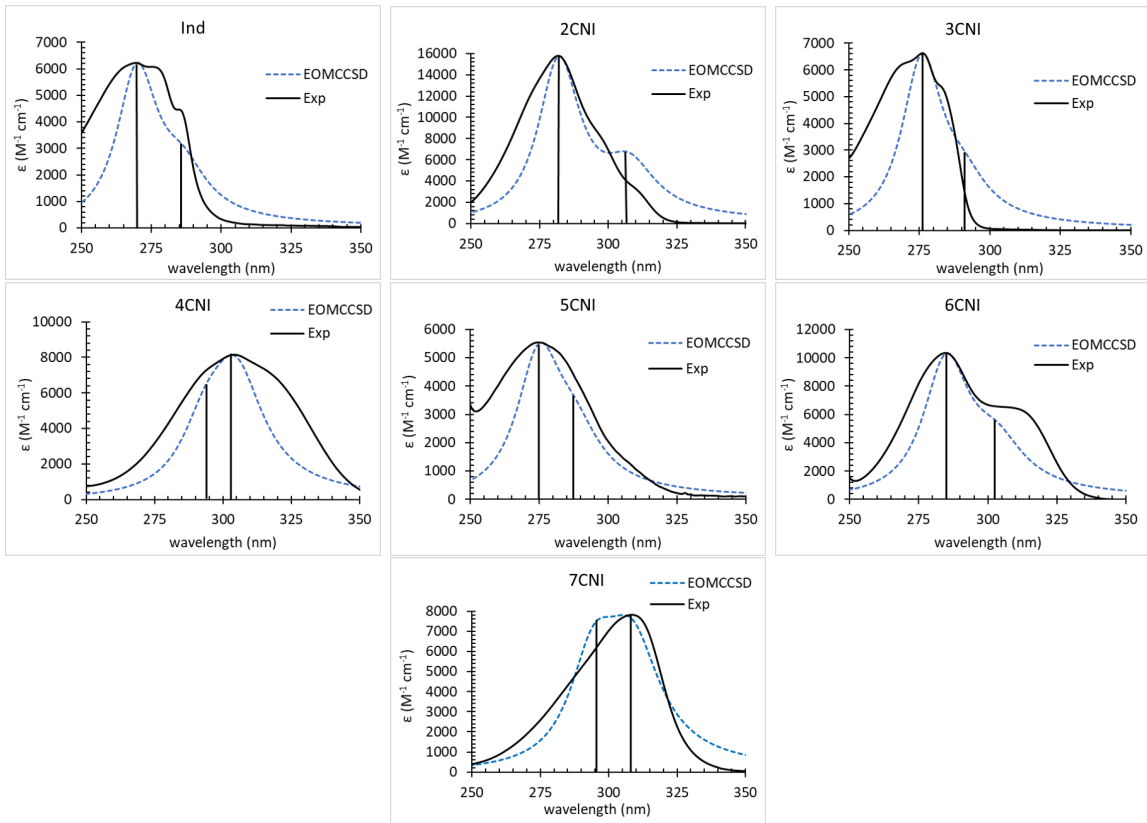


Figure 3: Absorption spectra of indole and n-CNI in aqueous solution computed at the EOM-CCSD/PCM/6-311+G(d) level simulated by mapping the excitation of the L_b and L_a excited states to a Lorentzian distribution with linewidth of 0.4 eV (dotted line). The oscillator strength of the bright state is scaled to the molar extinction coefficient of the molecule and the energies are converted to wavelengths, and shifted to the experimentally measured absorption maximum in water solvent (black solid line). The position of each of the two peaks for the absorption to the L_b and L_a excited states is represented by solid black vertical line to the peak maximum. Experimental spectra are taken from Ref.¹⁷

Figure 3 shows the simulated spectra calculated using EOM-CCSD/PCM/6-311+G(d) and superimposed on the experimental ones. The spectra are calculated using the oscillator

strengths and a Lorentzian distribution with linewidth of 0.4 eV. It is clear that depending on the gap between the two states and their relative oscillator strengths the shape of the spectra changes. In 2-CNI, 6-CNI and Indole the separation is large enough that a shoulder is clear in the theoretical and experimental spectra. Experimentally, the spectrum of 6-CNI shows the clearest separation of the peaks. Theoretically, 2-CNI has the largest separation. The theoretical spectrum of 7-CNI shows two peaks with almost the same oscillator strengths, and this leads to the asymmetry observed experimentally in the absorption peak. So the experimental spectrum in this case confirms our prediction that the S_1 state absorbs more in solution. On the other hand, the calculated 4-CNI spectrum does not show the second peak since the two states are closer in energy. Overall the calculated spectra are in good, although not perfect, agreement with experiment. The gas phase simulated spectra superimposed on the experimental ones are shown in SI (Figure S3), and their shape does not agree as well with the experimental spectra. This highlights that the solvent is important for the spectra.

Indole and 6-CNI have two distinguishable peaks at the absorption spectrum in aqueous solution associated with the large separation between the L_a and L_b electronic excited states, and they can be used to benchmark our theory in a more quantitative way. We use these two as a ruler for the performance of the methods in accurately predicting the L_a/L_b gap. The energy gap between the $L_b(1\pi\pi^*)$ and $L_a(2\pi\pi^*)$ states computed at the EOM-CCSD/6-311+G(d), EOM-CCSD/PCM/6-311+G(d) and CASPT2/aug-cc-pVDZ level of theory are plotted in SI (Figure S4) as a function of position of substitution of the cyano group. Theory is found to predict the energy gap of indole and 6-CNI reasonably well. The gap is found to be larger for substitution on positions 2 (about 0.5 eV) and smallest (0.2 eV) when the cyano group is substituted on position 4 of indole.

Transition Dipole Moments

The orientation of the transition dipole moment vectors is important in identifying the character of excited states experimentally. So, it is important to examine how it is affected

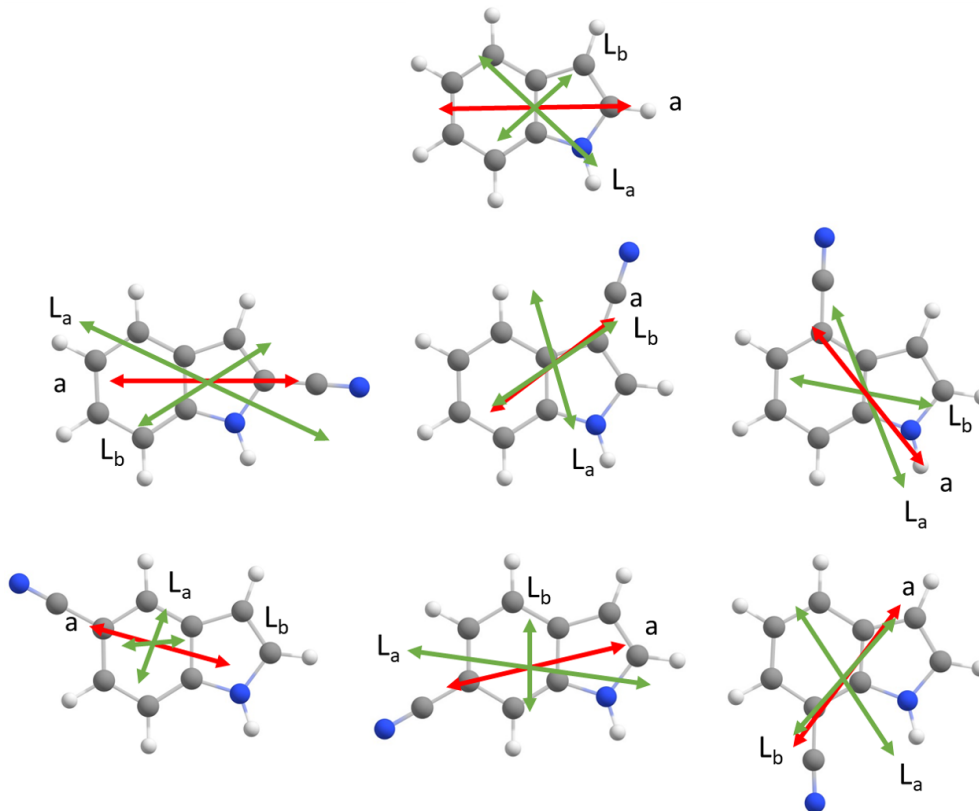


Figure 4: TDM vectors of the L_a and L_b excited states with respect to the inertia a -axis vector, a , computed at EOM-CCSD/6-311+G(d) level in gas phase

by CN substitution. In addition we can compare our results to the experimental studies when possible. The transition dipole moment (TDM) vectors for the L_a/L_b excited states computed at the EOM-CCSD/6-311+G(d) level are shown in Figure 4 while the vectors computed at the CASSCF level are shown in SI (Figure S13). The results are similar between the two methods. The magnitude of the TDM vectors for the L_a/L_b excited states and TDM angle θ measured from their respective inertia a -axis, are tabulated in the Supporting Information.

The angle θ for the excitation to the L_b and L_a electronic state of indole is computed to be $+39^\circ$ and -43° at the EOM-CCSD level, respectively, in good agreement with the experimentally reported dipole directions of $+42^\circ$ and -46° .⁴⁹ The direction of the TDM has been measured experimentally in the gas phase for 3-CNI and 5-CNI.^{50,51} The angle of the transition dipole moment vector with the main inertial a -axis in 3-CNI is found to be 15.3° .⁵⁰ In our calculations we find a value of 0.18° for the L_b state and 71° for the L_a state

at the EOM-CCSD level. These values change to 3° and 74° at the respective minima of these states (which is more relevant to what is measured experimentally). The computed value of L_b agrees better with experiment, enabling identification of the observed state as L_b . For 5-CNI the experimental value is 3° ⁵¹ while our calculations show 18° for L_b and 85° for L_a . The experimentally observed minimum corresponds to what we assign as the L_b state. It should be noted that our assignment is based on the NTOs and as a result the TDMs have rotated significantly when compared to the ones in indole. Ref.⁵¹ assigns this state to L_a , although we are describing the same state. Both the NTOs and the TDMs can change significantly by substitution, so we think it may not be as appropriate to use the strict notation of L_a and L_b states for the substituted systems. We will discuss this in the next section as well. Addition of the cyano group rotates the TDM of the states. The changes are more significant for substitution of CN on the six-membered ring.

Fluorescence

We now focus on fluorescence properties. The geometries of the L_a and L_b minima are shown in SI (Figures S19, S20 and Tables S14, S15). The distortions are similar for all n-CNI molecules. The L_b geometry is less distorted from the ground state geometry. The five membered ring does not change much while the six membered ring expands in general. The L_a minimum geometry shows larger distortions, and the single-double bonds in general alternate from the ones in the ground state, i.e. single bonds become double and vice versa.

The energy of the S_1 state is lower at the L_b minimum for all molecules in the gas phase (except 4-CNI which does not have an L_b minimum). This is shown in the adiabatic energies calculated for both geometries and shown in SI (Table S10). So we assign L_b as the global minimum on S_1 and calculate emission from there at the EOM-CCSD and CASPT2 levels of theory. In the case of 4-CNI the L_a minimum is used. The vertical emission energies of n-CNI are plotted in Figure 5(a) as a function of position of substitution of the CN group.

There is limited experimental information on gas phase fluorescence of the cyano sub-

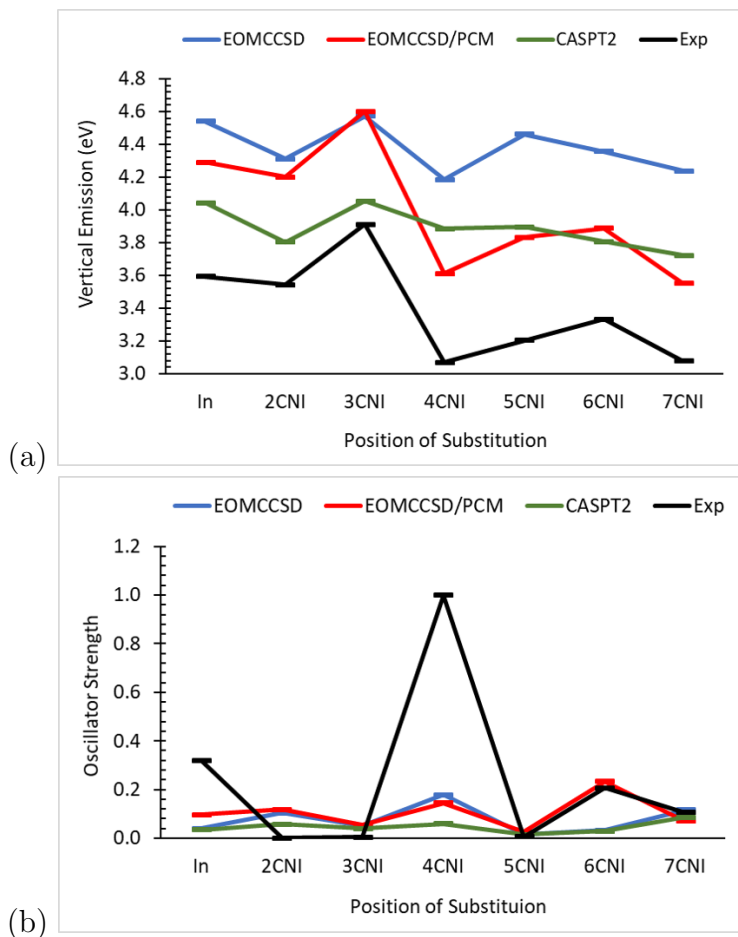


Figure 5: (a) Vertical emission energies computed at the EOM-CCSD/6-311+G(d) level in gas and aqueous solution (using PCM) and CASPT2/aug-cc-pVDZ in gas phase, and (b) corresponding oscillator strengths. All gas phase results are computed at the L_b minimum, while for EOM-CCSD/PCM results the L_b minimum is used for 2-CNI and 3-CNI and the L_a minimum for all other systems. Experimental values are taken from Ref.¹⁷

stituted indoles. It is known that indole and 3-CNI fluoresce from the L_b state⁵⁰ while a rotationally resolved spectroscopy study found that 5-CNI emits from L_a .⁵¹ In our calculations all optimized minima at EOM-CCSD except 4-CNI have L_b character, while for 4-CNI the character is L_a . This comparison indicates that we disagree with experiment on the minimum of the 5-CNI chromophore. However, as discussed in Section on TDMs, upon inspection of the properties reported experimentally we believe that we actually agree on the minimum. It appears that the difference is in how we label L_a and L_b states. We make our assignment based on the NTOs and how well they match the corresponding orbitals in indole. In Ref.⁵¹

molecular orbitals were used instead of NTOs, but there was very strong mixing between the configurations, so the molecular orbitals cannot clearly identify the states. The origins reported experimentally for 3-CNI and 5-CNI are 4.38 eV and 4.29 eV respectively.^{50,51} The values that we obtain theoretically are 4.73 eV and 4.59 eV, respectively at the EOM-CCSD level, which are both blueshifted by 0.38 eV compared to experiment. The CASPT2 values are 4.14 eV and 4.49 eV for 3-CNI and 5-CNI, closer to experiment but with the trend being reversed (i.e. the adiabatic energy for 3-CNI is predicted to be lower than that of 5-CNI).

For the solvated calculations in order to determine whether the minimum is L_a or L_b we did single point calculations using EOM-CCSD/PCM at the L_b geometries obtained from gas phase EOM-CCSD and at the L_a geometries obtained with TDDFT. The lowest S_1 energy at these geometries again was assigned as the S_1 global minimum. The energies in Hartree for S_1 at the two minima are shown in SI (Table S9). These results indicate that indole, 2-CNI and 3-CNI have an L_b minimum while all the other molecules have L_a minimum in solution. The energy difference between L_a and L_b for indole however is very small (less than 0.05 eV), while it is known experimentally that indole emits from L_a . In order to address this discrepancy we performed an additional optimization for indole at the EOM-CCSD/PCM level and proved that L_a is indeed the lowest energy minimum. In summary, all fluorescence properties in solution discussed from now on use the L_b minimum for 2-CNI and 3-CNI and the L_a minimum for all other chromophores. Fluorescence properties in gas phase use the L_b minimum in all cases.

Figure 5a, which compares the theoretical emission energies to experimental λ_{max} for fluorescence, makes it obvious that the solvation effect is much more important in fluorescence compared to absorption, since the gas phase results show poor correlation with experiment while the solvated ones show very good correlation. This is primarily because of the different character of the minima. A correlation plot with experiment gives a value of $R^2 = 0.98$ for EOM-CCSD/PCM, indicating an impressive correlation. This highlights the importance of the solvent in reproducing the experimental trends. The emission is blue shifted for 3-

CNI and red-shifted for all other n-cyanoindoles in agreement with the experimental emission spectra. The largest red shift is observed for 4-CNI and 7-CNI, in agreement with experiment. Experimentally it has also been speculated that 3-CNI emits from the L_b state since its emission shows very limited dependence on solvent.¹⁷ Our results confirm that indeed 3-CNI emits from L_b , although we find that 2-CNI also emits from L_b .

The corresponding oscillator strengths are compared to the measured fluorescence intensity of the probes in water solvent in Figure 5(b). As we would have expected using only the radiative oscillator strengths from our calculations is not sufficient to reproduce the experimental trends quantitatively. This is because there are many other competitive radiationless processes that are likely occurring as well, so the radiative intensities do not provide the whole picture. The calculations predict that 6-CNI would have the most intense fluorescence while experimentally 4-CNI is much more intense than all other chromophores. In the discussion we will discuss how we can explain the 4-CNI strong emission qualitatively.

The TDMs and NTOs of the excited states at the minima are reported in SI. It is interesting to observe that often the TDM vectors have rotated significant compared to their orientation at absorption. This is especially true at the L_a minima which differ more from the ground state minima. Also, the angle between the S_1 and S_2 TDM vectors changes indicating some mixing between what is initially identified as a pure L_a or L_b state. For example, for indole at absorption the angle between the TDM of L_a and L_b is 83° , making them almost perpendicular, while at the L_a minimum this angle is only 27° . The NTOs reflect similar changes in the character of the wavefunction when going from absorption to emission.

Solvation Effects

Some of the cyano derivatives we are studying here are strongly influenced by solvation effects. A question that we would like to address is how much of this effect is reproduced by the simple PCM solvation model employed in the current calculations. The solvatochromic

shift predicted by the PCM model depends specifically on the electrostatic interactions between the solvent and the difference of the dipole moments between the excited and ground state. This correlation is shown in Figure 6. The dipole moment difference is larger for the L_a state compared to the L_b state, so all the solvatochromic shifts for L_a are larger. For the L_a state, furthermore, the dipole moment difference is higher for 5-CNI, 6-CNI, and 7-CNI while it is the smallest for 2-CNI and 3-CNI. As expected the solvatochromic shifts follow the same pattern, and the larger shifts seen in 5-CNI and 7-CNI (0.2 eV).

Solvatochromic shifts for emission are larger since there is enough time for both the electronic and nuclear degrees of freedom of the solvent to respond to the electron density of the excited state. These are shown in Figure 6c. The largest solvatochromic shift is for 7-CNI, which in this case is about 0.8 eV.

Stokes Shift

There are no detailed experimental data to compare the calculated solvatochromic shifts, so it is hard to access the accuracy of the PCM model. On the other hand, there are Stokes shifts reported for all derivatives, and we can use that quantity to examine how well our model reproduces it. The Stokes shift includes contributions from both the vibrational relaxation and the solvatochromic shift.

Figure 7 shows the calculated Stokes shift for each chromophore in comparison to the experimental Stokes shift from Ref.¹⁷ The gas phase, as well as, PCM results are shown. Similarly to the fluorescence results, the agreement is much better between the EOM-CCSD/PCM results and experiment compared to the gas phase calculations, and this is expected since the solvatochromic shift is missing in the gas phase results. The different minima for fluorescence also exaggerate the solvent effect. The Stokes shift is larger for the 4,5,6,7-CNI systems which fluoresce from the L_a state, and it is smaller for 2,3-CNI which fluoresce from L_b in solution.

It should also be pointed out that the simple PCM model is sufficient to reproduce

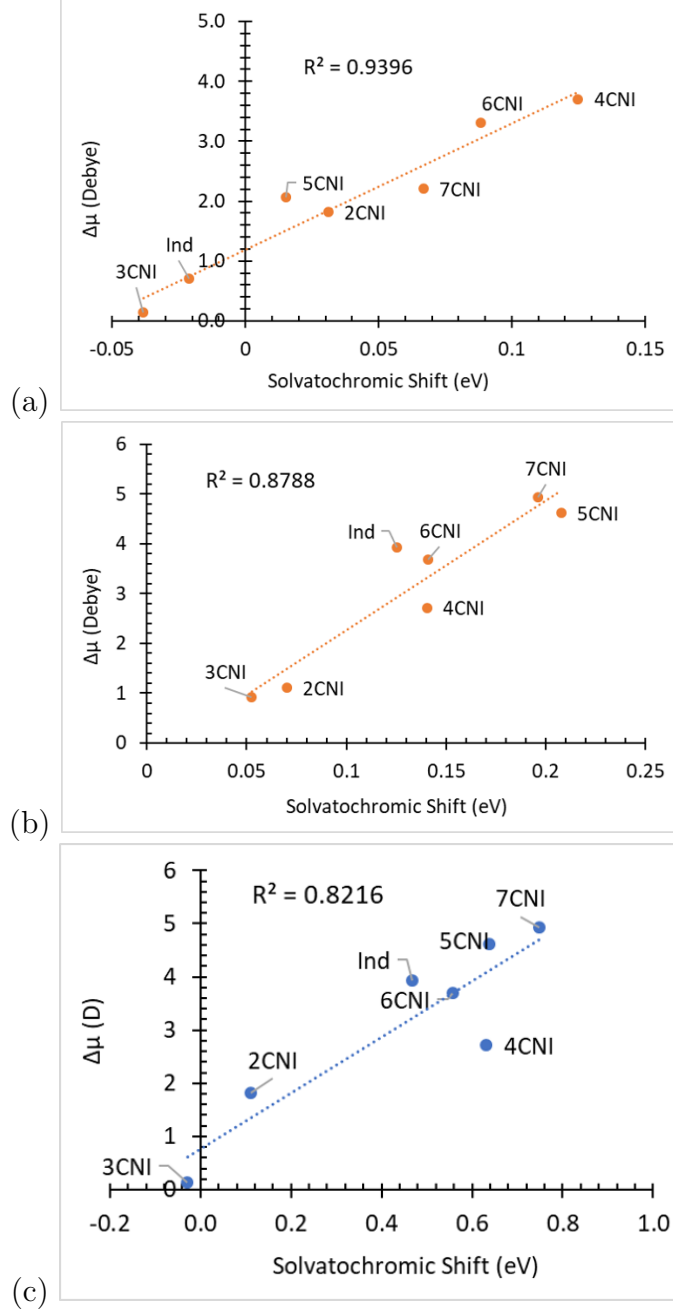


Figure 6: Solvatochromic shift in the absorption to the (a) $L_b(1\pi\pi^*)$ and (b) $L_a(2\pi\pi^*)$ excited states correlated with the difference in dipole moment $\Delta\mu$ between the excited state μ_e and ground state μ_g computed in aqueous solution at the EOM-CCSD/6-311+G(d) level. (c) Solvatochromic shift in the emission of Ind and 4-CNI to 7-CNI from L_a excited state and of 2-CNI and 3-CNI from the L_b excited state at the EOM-CCSD/6-311+G(d) level plotted with the corresponding dipole moments difference $\Delta\mu$ computed at the absorption in aqueous solution.

the basic trends observed experimentally in the Stokes shifts almost quantitatively. The largest deviation from the experiment values is 0.2 eV (much smaller than the errors seen in absorption and emission) while in some cases the theoretical results match exactly the experiment. This indicates that an explicit solvent is not necessary for reproducing these shifts. It is also possible that the effects are the same in absorption and emission, so they cancel out when taking the difference in the Stokes shift. Below we will discuss which properties seem to require explicit solvation.

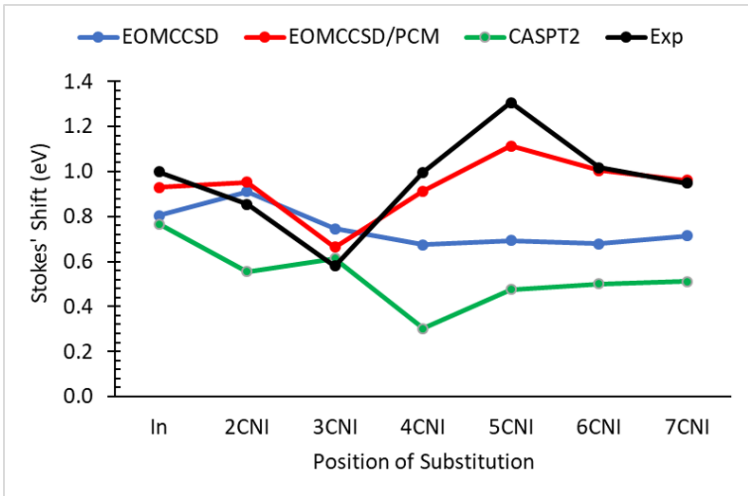


Figure 7: Stokes shift trend for cyanoindole fluorescent probes predicted at different levels of theory against the experimental Stokes shift (from Ref.¹⁷)

Discussion

Figure 8 summarizes our findings for the n-CNI systems in solution, and provides a rational explanation for the strong emission observed in 4-CNI. For all systems at absorption the first excited state is L_b while the second is L_a . The L_a state is usually much brighter than the L_b and in these cases the systems absorb mostly to $S_2(L_a)$. The main exception is 4-CNI where S_1 is more intense than S_2 . The result is that for all chromophores but 4-CNI there has to be internal conversion from S_2 to S_1 and fluorescence will occur from S_1 according to Kasha's rule. During relaxation from S_2 to S_1 additional radiationless decay pathways

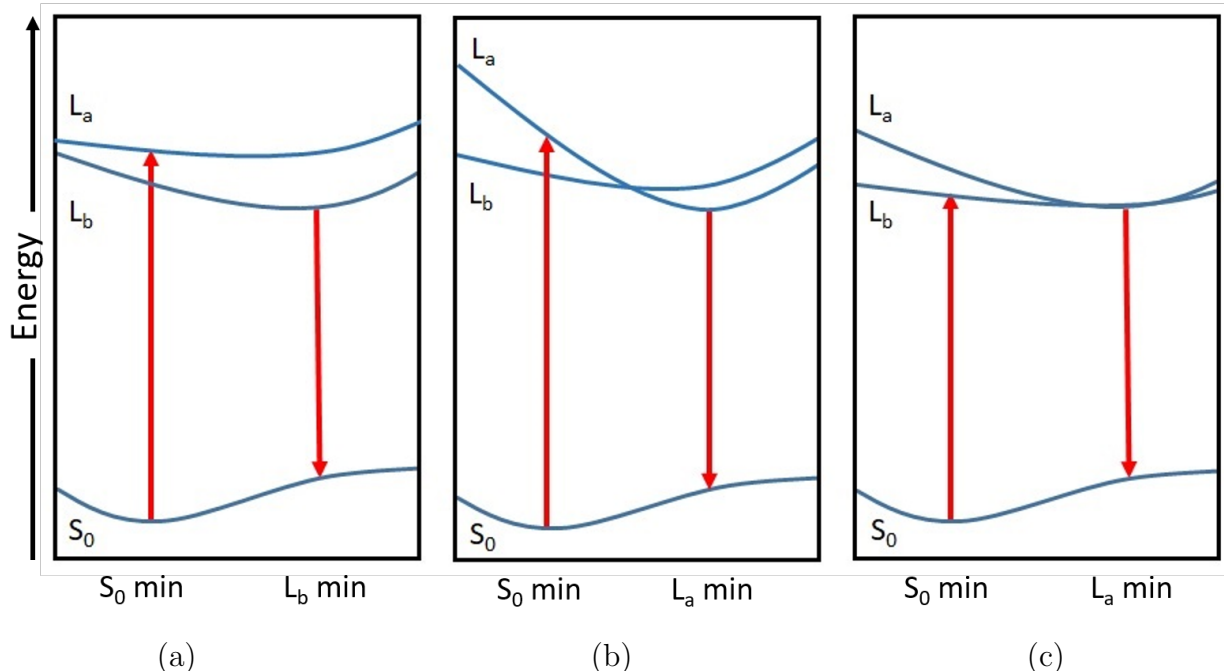


Figure 8: Radiative emission paradigms of n-CNI in aqueous solution. Cartoons represent the following systems: (a) 2-CNI and 3-CNI (b) Indole, 5-CNI, 6-CNI, (c) 4-CNI. 7-CNI is represented by both (b) and (c).

may be accessed leading to fluorescence quenching. In 4-CNI however both absorption and fluorescence occur from S_1 (see Figure 8c) making this a much more efficient process and minimizing radiationless decay. This we believe is a main reason 4-CNI has a more intense fluorescence compared to the other chromophores.

7-CNI needs special consideration as well. This molecule absorbs to both S_1 and S_2 almost equally in aqueous solution according to our calculations. According to the previous discussion on 4-CNI we would expect the emission intensity of 7-CNI to also be enhanced. This however is not observed experimentally. A reason may be that since it absorbs equally to S_1 and S_2 there are some radiationless pathways accessible here which are not accessible in 4-CNI. Furthermore, 7-CNI has a strong solvent dependence, and there may be specific interactions with the solvent that quench some of the fluorescence.

The current discussion on the 4,7-CNI fluorescence properties is based on our studies which include information on PES and radiative rates, but not radiationless rates. Further

studies, and particular conical intersection information and dynamical studies, are needed to prove our assumptions.

In 2-CNI and 3-CNI in aqueous solution the minimum on S_1 has L_b character (cartoon shown in Figure 8a) while for all the other chromophores it has L_a character (cartoon shown in Figure 8b). Since L_b is not as bright as L_a 2-CNI and 3-CNI are expected to have a weaker radiative emission. The experimental spectra show that these two chromophores have the weakest emission, in agreement with our results.

In the experimental fluorescence spectra the probe with the third weakest emission is 5-CNI. According to our results the oscillator strength of the S_1 state in this molecule is much smaller than all the other probes, and this indicates that the radiative component in this molecule is also responsible for the weak emission.

The PCM calculations reproduce the energetics of absorption and emission quite well. The trends seen experimentally for absorption maxima, emission maxima, and Stokes shifts are all reproduced in these calculations. On the other hand, the experimental intensities for fluorescence are not reproduced by the calculations. Our results show that the radiative intensity for fluorescence for 4-CNI and 6,7-CNI are very similar, but experimentally 4-CNI fluoresces much stronger than the others. Furthermore, the intensity of 5-CNI and 7-CNI is very sensitive to the solvent. This behavior indicates that there are specific, e.g. H-bonding, interactions between the solute and solvent that are responsible for this sensitivity, and our calculations cannot reproduce them. The effect of hydrogen bonding on the fluorescence intensity of 5-CNI, 6-CNI, and 7-CNI has been seen experimentally in studies where protic and aprotic solvents were used.¹⁷ Modeling using explicit solvent is needed to reproduce these effects, and we plan to carry out such calculations in the future. The current work however highlights the cases where specific interactions are responsible and the ones that do not seem to be affected by these interactions.

The situation in the gas phase is best depicted by cartoon (a) in Figure 8 which shows emission from L_b . Exception once again is 4-CNI which emits from L_a and is best described

by cartoon (b). What is worth noting is that the photophysics in the gas phase in all of these systems is quite different from the photophysics in aqueous solution, and studies in the gas phase are insufficient to produce, even qualitatively, the behavior in solution. The character of the emitting state inverts upon solvation, leading to the observed solvent dependence. Describing theoretically the relative stability between L_a and L_b minima is very challenging.

Conclusion

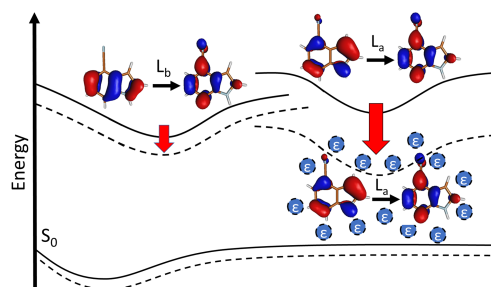
The absorption and fluorescence for indole and n-CNI were modeled in gas phase and aqueous solution using high level quantum mechanical methods. The trends in the experimental spectra are reproduced very well. We propose a rational paradigm explaining the effect of positional substitution of electron withdrawing cyano group on the absorption and fluorescence of these fluorescent probes. We provide new information about both L_a and L_b states and the gap between them. Deviations in computational results using the implicit solvation model from experimental fluorescence intensities suggest the involvement of intermolecular interactions between the probes and polar solvent (e.g. H-bonding) and/or the involvement of radiationless decay pathways. Explicit solvation and non-radiative decay pathways are needed to fully capture the fluorescence of probes that are found to have high solvent dependence.

Acknowledgements

The authors gratefully acknowledge the National Science Foundation (NSF, grant number: CHE-1800171) for funding. We also wish to thank Prof. Feng Gai and his group for sharing their experimental spectra with us and for fruitful discussions. This work used the Extreme Science and Engineering Discovery Environment (XSEDE),⁵² which is supported by National Science Foundation grant number ACI-1548562. The Temple University Owl’s Nest cluster was also used.

Supporting Information: Excitation energies at all levels of theory; spectra at the CASPT2 level; orbitals; correlation plots; additional information on transition dipole moments and dipole moments, and cartesian coordinates.

Table of Content Graphic



References

- (1) Talukder, P.; Chen, S.; Roy, B.; Yakovchuk, P.; Spiering, M. M.; Alam, M. P.; Madathil, M. M.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. Cyanotryptophans as Novel Fluorescent Probes for Studying Protein Conformational Changes and DNA-Protein Interaction. *Biochemistry* **2015**, *54*, 7457–7469.
- (2) Li, G. Y.; Han, K. L. The sensing mechanism studies of the fluorescent probes with electronically excited state calculations. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, 1–17.
- (3) Cohen, B. E.; Pralle, A.; Yao, X.; Swaminath, G.; Gandhi, C. S.; Jan, Y. N.; Kobilka, B. K.; Isacoff, E. Y.; Jan, L. Y. A fluorescent probe designed for studying protein conformational change. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 965–970.
- (4) Pazos, I. M. Spectroscopic Probes of Protein Structure, Dynamics, Hydration and Electrostatics. *Dissertation* **2015**, 1–198.
- (5) Giepmans, B. N.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y. The Fluorescent Toolbox for Assessing protein location and function. *Science* **2006**, *312*, 217–224.
- (6) Kobayashi, H.; Ogawa, M.; Alford, R.; Choyke, P. L.; Urano, Y. New Strategies for Fluorescent Probe Design in Medical Diagnostic Imaging. *Chem. Rev.* **2010**, *110*, 2620–2640.
- (7) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer, 2006.
- (8) Krygowski, T. M.; Stepień, B. T. Sigma- and Pi-Electron Delocalization: Focus on Substituent Effects. *Chem. Rev.* **2005**, *105*, 3482–3512.
- (9) Kistler, K.; Matsika, S. Cytosine in Context: A Theoretical Study of Substituent Effects on the Excitation Energies of 2-Pyrimidinone Derivatives. *J. Phys. Chem. A* **2007**, *111*, 8708–8716.

- (10) Abou-Hatab, S.; Spata, V. A.; Matsika, S. Substituent Effects on the Absorption and Fluorescence Properties of Anthracene. *J. Phys. Chem. A* **2017**, *121*, 1213–1222.
- (11) Markiewicz, B. N.; Mukherjee, D.; Troxler, T.; Gai, F. Utility of 5-Cyanotryptophan Fluorescence as a Sensitive Probe of Protein Hydration. *J. Phys. Chem B* **2016**, *120*, 936–944.
- (12) Mukherjee, D.; Ortiz Rodriguez, L. I.; Hilaire, M. R.; Troxler, T.; Gai, F. 7-Cyanoindole fluorescence as a local hydration reporter: Application to probe the microheterogeneity of nine water-organic binary mixtures. *Phys. Chem. Chem. Phys.* **2018**, *20*, 2527–2535.
- (13) Lee, J.; Lee, J. Indole as an intercellular signal in microbial communities. *FEMS Microbiol. Rev.* **2010**, *34*, 426–444.
- (14) Melander, R. J.; Minvielle, M. J.; Melander, C. Controlling bacterial behavior with indole-containing natural products and derivatives. *Tetrahedron* **2014**, *70*, 6363–6372.
- (15) Benamins, R.; Scheses, B. Auxin: The Looping Star in Plant Development. *Annu. Rev. Plant Biol.* **2008**, *59*, 443–465.
- (16) Catalán, J. The first UV absorption band of l-tryptophan is not due to two simultaneous orthogonal electronic transitions differing in the dipole moment. *Phys. Chem. Chem. Phys.* **2016**, *18*, 15170–15176.
- (17) Hilaire, M. R.; Mukherjee, D.; Troxler, T.; Gai, F. Solvent dependence of cyanoindole fluorescence lifetime. *Chem. Phys. Lett.* **2017**, *685*, 133–138.
- (18) Hilaire, M. R.; Ahmed, I. A.; Lin, C.-W.; Jo, H.; DeGrado, W. F.; Gai, F. Blue fluorescent amino acid for biological spectroscopy and microscopy. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6005–6009.
- (19) Arulmozhiraja, S.; Coote, M. L. 1L_a and 1L_b states of indole and azaindole: Is density functional theory inadequate? *J. Chem. Theory Comput.* **2012**, *8*, 575–584.

- (20) Serrano-Andrés, L.; Roos, B. O. Theoretical study of the absorption and emission spectra of indole in the gas phase and in a solvent. *J. Am. Chem. Soc.* **1996**, *118*, 185–195.
- (21) Platt, J. R. Classification of spectra of cata-condensed hydrocarbons. *J. Chem. Phys.* **1949**, *17*, 484–495.
- (22) Brisker-Klaiman, D.; Dreuw, A. Explaining level inversion of the L_a and L_b states of indole and indole derivatives in polar solvents. *ChemPhysChem* **2015**, *16*, 1695–1702.
- (23) Lami, H. Presence of a low-lying Rydberg band in the vapour phase absorption spectra of indole and 1-methyl indole. *Chem. Phys. Lett.* **1977**, *48*, 447–450.
- (24) Strickl, E. H.; Horwitz, J.; Billups, C. Near-ultraviolet absorption bands of tryptopan. Studies using indole and 3-methylindole as models. *Biochemistry* **1970**, *9*, 4914–4921.
- (25) Slater, L. S.; Callis, P. R. Molecular Orbital Theory of the 1La and 1Lb States of Indole. 2. An ab Initio Study. *J. Phys. Chem.* **1995**, *99*, 8572–8581.
- (26) Giussani, A.; Merchán, M.; Roca-Sanjuán, D.; Lindh, R. Essential on the photophysics and photochemistry of the indole chromophore by using a totally unconstrained theoretical approach. *J. Chem. Theory Comput.* **2011**, *7*, 4088–4096.
- (27) Sobolewski, A. L.; Domcke, W. Computational Studies of the Photophysics of Hydrogen-Bonded Molecular Systems. *J. Phys. Chem. A* **111**, 11725.
- (28) Brand, C.; Oeltermann, O.; Wilke, M.; Tatchen, J.; Schmitt, M. Ground and electronically excited singlet-state structures of 5-fluoroindole deduced from rotationally resolved electronic spectroscopy and ab initio theory. *ChemPhysChem* **2012**, *13*, 3134–3138.
- (29) Wilke, J.; Wilke, M.; Brand, C.; Meerts, W. L.; Schmitt, M. On the Additivity of Molecular Fragment Dipole Moments of 5-Substituted Indole Derivatives. *ChemPhysChem* **2016**, *5*, 2736–2743.

- (30) Wilke, J.; Wilke, M.; Brand, C.; Spiegel, J. D.; Marian, C. M.; Schmitt, M. Modulation of the La/Lb Mixing in an Indole Derivative: A Position-Dependent Study Using 4-, 5-, and 6-Fluoroindole. *J. Phys. Chem A* **2017**, *121*, 1597–1606.
- (31) Shao, Y. et al. Advances in quantum chemical methods and algorithms in the Q-Chem 3.0 program package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.
- (32) Werner, H.-J. et al. MOLPRO, version 2015.1, a package of ab initio programs. 2015; see.
- (33) Frisch, M. J. et al. Gaussian16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.
- (34) Caricato, M. Absorption and Emission Spectra of Solvated Molecules with the EOM-CCSDPCM Method. *J. Chem. Theory Comput.* **2012**, *8*, 4494–4502.
- (35) Caricato, M. Exploring Potential Energy Surfaces of Electronic Excited States in Solution with the EOM-CCSD-PCM Method. *J. Chem. Theory Comput.* **2012**, *8*, 5081–5091.
- (36) Caricato, M. CCSD-PCM: Improving upon the reference reaction field approximation at no cost. *J. Chem. Phys.* **2011**, *135*, 074113.
- (37) Cammi, R.; Mennucci, B.; Tomasi, J. Fast Evaluation of Geometries and Properties of Excited Molecules in Solution: A Tamm-Dancoff Model with Application to 4-Dimethylaminobenzonitrile. *J. Phys. Chem. A* **2000**, *104*, 5631–5637.
- (38) Frisch, M. J. et al. Gaussian 09, Revision B.01. 2009.
- (39) Caminati, W.; Di Bernardo, S. Microwave spectrum and amino hydrogen location in indole. *J. Mol. Struct.* **1990**, *240*, 253.
- (40) Britten, A. Z.; Lockwood, G. The indole chromophore: Separation of the 1La and 1Lb components by a methyl substitution perturbation technique. *Spectrochim. Acta* **1976**, *32A*, 1335.

- (41) Chang, C.; Wu, C.; Muirhead, A. R.; Lombardi, J. R. The dipole moment in the lowest singlet $\pi \rightarrow \pi^*$ state of indole determined by the optical Stark effect. *Photochem. Photobiol.* **1974**, *19*, 347–351.
- (42) Lami, H.; Glasser, N. Indoles solvatochromism revisited. *J. Chem. Phys.* **1986**, *84*, 597.
- (43) Kannar, D.; Szalay, P. Benchmarking Coupled Cluster Methods on Valence Singlet Excited States. *J. Chem. Theory Comput.* **2014**, *10*, 37573765.
- (44) Ren, S.; Harms, J.; Caricato, M. An EOM-CCSD-PCM Benchmark for Electronic Excitation Energies of Solvated Molecules. *J. Chem. Theory Comput.* **2017**, *13*, 117124.
- (45) Epifanovsky, E.; Kowalski, K.; Fang, P.-D.; Valiev, M.; Matsika, S.; Krylov, A. I. On the Electronically Excited States of Uracil. *J. Phys. Chem. A* **2008**, *112*, 9983 – 9992.
- (46) Callis, P. R. Molecular orbital theory of the 1L_b and 1L_a states of indole. *J. Chem. Phys.* **1991**, *95*, 4230–4240.
- (47) Callis, P. R. $1L_a$ and $1L_b$ Transitions of Tryptophan: Applications of Theory and Experimental Observations to Fluorescence of Proteins. *Methods Enzymol.* **1997**, *278*, 113–150.
- (48) Meng, X.; Harricharran, T.; Juszczak, L. J. A Spectroscopic Survey of Substituted Indoles Reveals Consequences of a Stabilized L_b Transition. *Photochem. Photobiol.* **2013**, *89*, 40–50.
- (49) Albinsson, B.; Nordén, B. Excited-state properties of the indole chromophore. Electronic transition moment directions from linear dichroism measurements: Effect of methyl and methoxy substituents. *J. Phys. Chem* **1992**, *96*, 6204–6212.
- (50) Schneider, M.; Hebestreit, M.-L.; Lindic, M. M.; Parsian, H.; Torres-Boy, A.; Álvarez Valtierra, L.; Meerts, W. L.; Kühnemuth, R.; Schmitt, M. Rotationally Resolved Elec-

- tronic Spectroscopy of 3-Cyanoindole and the 3-Cyanoindole-water Complex. *Phys. Chem. Chem. Phys.* **2018**, *20*, 23441–23452.
- (51) Oeltermann, O.; Brand, C.; Engels, B.; Tatchenc, J.; Schmitt, M. The structure of 5-cyanoindole in the ground and the lowest electronically excited singlet states, deduced from rotationally resolved electronic spectroscopy and ab initio theory. *Phys.Chem.Chem.Phys.* **2012**, *14*, 10266–10270.
- (52) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazelwood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **2014**, *16*, 62–74.