Absorption Intensities of Organic Molecules from Electronic Structure Calculations versus Experiments: The Effect of Solvation, Method, Basis Set, and Transition Moment Gauge

Jorge C. Garcia-Alvarez* and Samer Gozem*

Department of Chemistry, Georgia State University, Atlanta, Georgia 30302, United States

E-mail: jgarciaalvarez1@student.gsu.edu; sgozem@gsu.edu

Abstract

Recently, we derived experimental oscillator strengths (OSs) from well-defined UV-visible absorption spectral peaks of 100 molecules in solution. Here, we focus on a subset of transitions with the highest reliability to further benchmark the OSs from several wave function methods and density functionals. We consider multiple basis sets, transition moment gauges (length, velocity, and mixed), and solvent corrections. Most transitions in the comparison set come from conjugated molecules and have $\pi \rightarrow \pi^*$ character. We use an automated algorithm to assign computed transitions to experimental bands. OSs computed using the Tamm-Dancoff approximation (TDA), CIS, or EOM-CCSD exhibited a strong gauge dependence, which is diminished in linear response theories (TD-DFT, TD-HF, and to a smaller degree LR-CCSD). OSs calculated from TD-DFT with PCM solvent models are systematically larger than apparent

OSs derived from experimental spectra. For example, f_{comp} from hybrid functionals and PCM have mean absolute errors that are ∼10% of $n \cdot f_{exp}$, where n is a solvent refractive index factor that arises from the energy flux of the radiation field in a dielectric (solvent). Theoretical cavity field corrections considering spherical cavities do not improve the agreement between computed and experimental data. Corrections that account for the molecular shape and the direction of transition dipole moments should be more appropriate.

Introduction

Among quantum mechanics' earliest successes was its ability to explain and predict spectra, from black-body radiation to atomic emissions. The agreement between calculations and spectroscopic quantities remains an important and widely used metric for assessing the accuracy of quantum chemical theories and models. Many benchmark studies have provided data on the accuracy of computed electronic transition energies in molecules. For instance, vertical energies computed with time-dependent density functional theory (TD-DFT) can lie within a fraction of an electron volt from experimental λ_{max}.[1] Specifically for $\pi \rightarrow \pi^*$ transitions in organic dyes, TD-DFT computations have typical deviations in the 0.15-0.25 eV range.[2] Protocols aimed at reproducing experimental adiabatic excitation energies and that take into account vibrational zero-point energies have achieved chemical accuracy (errors under 1 kcal·mol^{-1} or 0.043 eV) on small molecules using systematically improvable but more time-consuming *ab initio* excited-state methods.[3]

Compared to electronic transition energies, far fewer studies have looked into the ability of computational methods to accurately reproduce absorption intensities. This is the focus of the present work. We start by briefly reviewing some of the benchmark studies for oscillator strengths (OS, or f values) in the literature.

Most studies focus on comparing f values computed using one method to another suitable computational reference. For instance, Silva-Junior et al.[4] compared TD-DFT (BP86,

B3LYP, and BHLYP) and DFT-based multireference configuration interaction (DFT/MRCI) f values to the best theoretical estimates from *ab initio* methods such as MS-CASPT2, CC2, and CCSD.[4,5] They focused on optically allowed transitions from 28 medium-sized organic molecules and found that TD-DFT generally underestimates the *ab initio* OSs, while DFT/MRCI f values were comparable, with a mean absolute deviation in the range 0.06-0.08.[4]

Similarly, Caricato et al.[6] assessed the performance of TD-DFT functionals, RPA, CIS, and CIS(D) relative to EOM-CCSD calculations in a set of 11 small organic molecules containing alkenes, carbonyls, and azobenzenes. They analyzed a total of 69 states: 30 valence and 39 Rydberg in nature. They found significant variations between functionals and a marked dependence of the error magnitude on the molecule.[6]

A few studies have directly compared computed and experimental f values. Chrayteh et al.[7] studied the excited state properties of 13 small molecules in the gas phase using the CC-expansion and extrapolating to the complete basis set limit. They found that their computations fall within experimental errors for transitions with experimentally reproducible f values.[7] Jacquemin et al.[8] compared f values computed with the Bethe–Salpeter equation (BSE) formalism (combined with the GW approximation) to experimentally-derived f values of 30 anthraquinones in dichloromethane.[9] The BSE/GW calculations reproduced the experimental trend. They report a R^2 value of 0.819 for the linear regression between the computed and experimental data, even though the calculation did not include solvation effects.[8]

In a study that focused on N_2, CO, formaldehyde, ethylene, and benzene, Tawada et al.[10] found that LC-functionals (LC-BOP, LC-BLYP, and LC-PBEOP) were able to correctly reproduce the order of magnitude of the experimental f values in N_2 Rydberg transitions. On the other hand, pure functionals (BOP, BLYP, and PBEOP) underestimated f values by two orders of magnitude, and B3LYP underestimated them by one order of magnitude. With a less marked difference in CO's $\sigma \rightarrow \pi^*$ transition, LC-TDDFT still outperformed

B3LYP and pure functionals (the worst performing again). In formaldehyde, ethylene, and benzene, results became more mixed, with LC-TDDFT greatly overestimating C2H4's $\pi \rightarrow \pi^*$ transition.[10]

Miura, Aoki, and Champagne[11] focused on lowest-energy dipole-allowed (mainly $\pi \rightarrow \pi^*$) transitions in benzene, phenol, aniline, and fluorobenzene. They compared seven functionals (SVWN, BLYP, PBE, TPSS, B3LYP, PBE0, and BHandHLYP) to available gas-phase experimental f values, and to RPA, CIS, CCS, CC2, and CCSD calculations. They investigated the effect of basis set (Pople's 6-31G*, 6-311G*, 6-311G**, and 6-311++G**; and Dunning's cc-pVDZ to cc-pV5Z, aug-cc-pVDZ to aug-cc-pVQZ, and d-aug-cc-pVDZ to d-aug-cc-pVTZ) on the energies, f values, and character of the transition computed with B3LYP and PBE0. An important conclusion from their work is that diffuse basis functions are important for both energies and OSs. For example, the decrease in excitation energy (in agreement with experiment) is more pronounced going up in the cc-pVXZ than in the aug-cc-pVXZ series of basis sets. The corresponding f values increase with the former and decrease (in agreement with experiment) with the latter series. The bulk of their calculations were carried out with the 6-311++G** basis set. When expanding their benchmark to include chlorobenzene, anisole, and phenetole and comparing their calculations to experiments, they found that the computations correlated well with the experimental data but were not in quantitative agreement. For example, TD-B3LYP calculations yielded a slope of 1.48 and a y-intercept of -0.26 when plotted against experimental data, but the correlation coefficient (R) value is 0.94. Other functionals gave comparable results.[11]

For more information about oscillator strength benchmark studies in the literature published before 2013, we refer the reader to Ref. 1.

Recently, we generated a collection of f values from experimental UV-Vis absorption spectra of 100 organic molecules in solution.[12] A total of 164 OSs were obtained by integrating the attenuation coefficient $\epsilon(\tilde{\nu})$ over the limit of well-defined bands in the spectra. Transitions were categorized as either very high (VH), high (H), medium (M), low (L), or very low (VL)

confidence on the basis of the reproducibility and quality of the fitting. We refer the reader to Ref. 12 for more details on the fitting, integration, categorization, and a discussion of the sources of error of these f values. While errors in experimental OSs are difficult to quantify, we expect that errors in the condensed phase should be smaller than errors in the gas phase.[12,13]

Here, we employ this benchmark set to compare OSs computed using several TD-DFT functionals and wave function methods. Aspects that affect the comparison between theory and experiments, such as experimental deviations from the Beer-Lambert Law, solvent effects, and the f value dependence on the energy of the electric transition, are discussed.

The manuscript is structured in four sections. To provide a framework for how computed and experimental OSs can be compared, The first section presents a concise background explaining how f is obtained from absorption experiment observables (Experimental oscillator strength) and from quantum theory (Theoretical oscillator strength). A third subsection discusses cavity field corrections in the literature. The next section details the approach used to compute f in this work (Computed absorption transitions) and outlines how they are compared with the experimental data (Statistical analysis). The last two sections present the results of the benchmark and conclusions, respectively.

Modern quantum chemical methods have become valuable tools for early-stage screening of novel dyes.[14] Chromophores with strong absorption and emission have applications in areas spanning solar energy and light-emitting devices to bioimaging. For such applications, we emphasize the importance of selecting computational methods that accurately predict relative transition strengths as well as transition energies.

Theoretical background

Experimental oscillator strength

A detailed discussion of the experimental aspects and theory of UV-vis absorption spectroscopy can be found, for example, in Refs. 15 and 16. Here we briefly summarize some main points that connect the experimental observables to the oscillator strength and provide context for later discussion on the validity of the expressions used.

A typical UV-Vis absorption experimental setup in the condensed phase (Fig. 1) involves nearly monochromatic, collimated electromagnetic radiation traveling through a cuvette containing the molecule of interest dissolved in a solvent. A sensor measures the intensity of light after it traverses the cuvette. This intensity is compared to a reference, typically light that has traveled through an identical cuvette filled only with the solvent to account for the intensity reduction resulting from reflection, scattering, or absorption by molecules other than the solute.[16] [1]

The intensity of light of wavenumber $\tilde{\nu}$ reaching the sample, $I_0(\tilde{\nu})$, and the intensity leaving it, $I(\tilde{\nu})$, will determine the spectral absorbance $A(\tilde{\nu})$, a magnitude commonly used to describe the reduction in the light intensity:

$$A(\tilde{\nu}) = \log\left[\frac{I_0(\tilde{\nu})}{I(\tilde{\nu})}\right]. \tag{1}$$

The absorbance of a solute of interest is then found from the difference between the absorbance obtained for the solution (sample) and the pure solvent (reference):[16]

$$A(\tilde{\nu}) = A_{Sample}(\tilde{\nu}) - A_{Reference}(\tilde{\nu}) \tag{2}$$

[1] Depending on the spectrophotometer used and the wavelength of interest, the light reaching the cuvette will exhibit a certain degree of polarization arising from reflection and refraction in the optical elements of the monochromator. For example, one of the spectrophotometers used for some of the molecules in the benchmark, the Cary model 14, was found to have varying degrees of polarization, going from fairly constant in the UV, to more varying in the visible, to sharp maxima in the IR.[17] This effect should not be important for an isotropic distribution of molecules (such as molecules in solution) but becomes relevant for an anisotropic or partially oriented sample.[15]

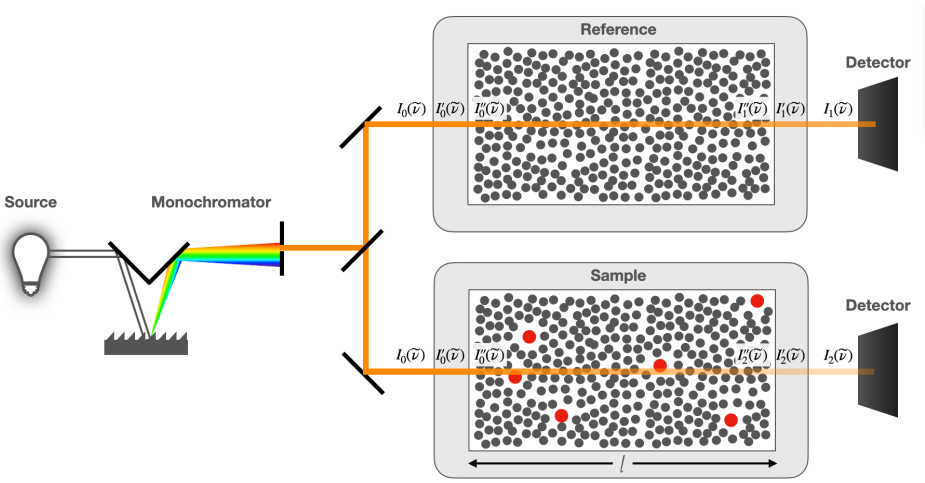

Figure 1: A scheme of the typical absorption experiment setup. Here, $I_0(\tilde{\nu})$ is the intensity of the incident light in air entering the cuvettes, $I_1(\tilde{\nu})$ is the intensity of light leaving the reference cuvette and reaching the detector, and $I_2(\tilde{\nu})$ is the intensity of light leaving the sample cuvette and reaching the detector. The terms with primes indicate changes in the intensity of light as it enters and exits the cuvette wall.

Assuming the solution and solvent reference are measured under identical conditions, effects such as scattering and reflection by the solvent and cuvette walls that affect light intensity cancel out, and the expression for the absorbance of the solute simplifies to:

$$A(\tilde{\nu}) = \log\left[\frac{I_0(\tilde{\nu})}{I_2(\tilde{\nu})}\right] - \log\left[\frac{I_0(\tilde{\nu})}{I_1(\tilde{\nu})}\right] = \log\left[\frac{I_1(\tilde{\nu})}{I_2(\tilde{\nu})}\right], \tag{3}$$

where $I_0(\tilde{\nu})$ is the intensity of light reaching the cuvettes, $I_2(\tilde{\nu})$ is the intensity leaving the cuvette with the sample, and $I_1(\tilde{\nu})$ is the intensity of light leaving the cuvette with the reference.[2] The absorbance measured in this way will correspond to a reduction in intensity

[2]Upon normal incidence on a surface separating two media, the transmitted electric field intensity vector is $\mathbf{E_t} = 2\eta_t/(\eta_t + \eta_i)\mathbf{E_i}$, where $\mathbf{E_i}$ is the incident electric field intensity vector, $\eta_i = \sqrt{\mu_i/\epsilon_i}$ is the intrinsic impedance of the medium of the incident waves, $\eta_t = \sqrt{\mu_t/\epsilon_t}$ is the intrinsic impedance of the medium of the reflected waves, μ_i and μ_t are the relative permeabilities of the medium of the incident wave and the medium of the reflected wave, respectively, and ϵ_i and ϵ_t are the dielectric constant of the incident medium and the reflected medium, respectively. Therefore, the intensities in air are related to those in solution by the same multiplicative constant. That is, $I_1''(\tilde{\nu}) = cI_1'(\tilde{\nu})$ and $I_2''(\tilde{\nu}) = cI_2'(\tilde{\nu})$, where c is a common constant.

given by the Beer-Lambert law,

$$A(\tilde{\nu}) = \log\left[\frac{I_1(\tilde{\nu})}{I_2(\tilde{\nu})}\right] = \varepsilon(\tilde{\nu}) \cdot c_M \cdot l, \tag{4}$$

where $\varepsilon(\tilde{\nu})$ is the decadic molar extinction coefficient (also called absorption or attenuation coefficient), c_M is the molar concentration of the solute in the solution, and l is the path length of the light through the solution. The linear correlation between absorbance $A(\tilde{\nu})$ and concentration c_M in the Beer-Lambert law requires that:[15,18] (i) the solute molecules do not aggregate, (ii) light scattering by the solute molecules is negligible, (iii) $I_0(\tilde{\nu})$ is small and multiphoton processes, excited states populations, and photochemical reactions are negligible, (iv) the spectral bandwidth of the light used is narrow compared to the transition bandwidth, i.e., the light leaving the monochromator spans only a narrow range of frequencies.[3]

In molecules, the spectral band for a given electronic transition is spread over a wide range of frequencies by each electronic level's vibrational and rotational substructures. Solvent effects further broaden the line shape of transitions. The transition probability is obtained by measuring the full range of $\varepsilon(\tilde{\nu})$ as a function of $\tilde{\nu}$ for that specific excitation.

Since A is dimensionless, the units for ε are determined by the units of c_M and l, typically moles per liter (mol/L) and centimeters (cm), respectively. Therefore, ε is usually reported in units of $L/(\text{mol} \cdot \text{cm})$ or equivalently in $M^{-1}cm^{-1}$.

The reduction in intensity can be equivalently expressed in terms of the absorption cross-section of the solute in the specific solvent $\sigma(\tilde{\nu})$ as:

$$I(\tilde{\nu}) = I_0(\tilde{\nu})10^{-\varepsilon(\tilde{\nu})c_M l} = I_0(\tilde{\nu})e^{-\sigma(\tilde{\nu})n'l}, \tag{5}$$

[3]A narrow spectral bandwidth (i.e., being as close as possible to monochromatic) can be critical when obtaining atomic or highly resolved vibronic spectra. The monochromator spectral bandwidth should be smaller than the absorption bandwidth of the sample to resolve it properly. This is less of a concern for broad-band absorptions such as the ones in the experimental benchmarks used in this work.

where n' is the number density (number of molecules per unit volume). The following expression can be used to convert from cross-sections expressed in cm^{-2} and attenuation coefficients in $cm^{-1}M^{-1}$:

$$\varepsilon = \frac{10^{-3}N_A}{\ln(10)}\sigma_{\text{abs}}, \tag{6}$$

where N_A is Avogadro's number.

Delving into the derivation of the Beer-Lambert law in terms of the absorption cross-section offers valuable molecular-level insights into the approximations inherent in the law. A detailed discussion can be found in Ref. 19, where the law is derived by equating the probability of a photon traversing the sample length l (without being absorbed) to the probability of encountering no molecules within a cylinder of base equal to the molecular absorption cross-section $\sigma(\tilde{\nu})$ and of height l. Although we will focus on OSs as our primary metric, we will revisit the cross-section perspective later as we describe solvent effects theoretically.

The connection between oscillator strength f and the experimental metrics of attenuation is given by f's historical origin as a link between classical electromagnetic dispersion theory and quantum theory. In classic electrodynamics, the propagation of a plane, monochromatic, linearly polarized wave in an isotropic, nonmagnetic medium, with the constitutive relations $\mathbf{D} = \tilde{\epsilon}\mathbf{E}$ and $\mathbf{B} = \mathbf{H}$, [4] is given by

$$\mathbf{E} = \text{Re}\left\{\mathbf{E_0}e^{i\left[2\pi\tilde{\nu}c\left(t-\frac{\tilde{n}}{c}x\right)\right]}\right\}, \tag{7}$$

where c is the speed of light in vacuum, and $\tilde{n}$ is a complex index of refraction:[5]

$$\tilde{n}(\tilde{\nu}) = n(\tilde{\nu}) - i\kappa(\tilde{\nu}). \tag{8}$$

The real part of (8), $n(\tilde{\nu})$, quantifies the phase velocity (the usual refractive index) while

[4] The magnetic permeability of typical solvents μ is practically equal to vacuum's μ_0.[15] By using Gaussian units[21] then $\mu \approx \mu_0 = 1$. Also, we have that $|H_0| = |E_0|$.

[5] Rigorously, all magnitudes in (8) depend also on temperature T and the (number) concentration of molecules ($n = n(\tilde{\nu}, T, n')$, $\kappa = \kappa\tilde{\nu}, T, n')$).

the imaginary part, $\kappa(\tilde{\nu})$, quantifies absorption in the medium. The average light intensity at a depth x is given by:

$$I = \frac{1}{2} E_0^2 e^{-4\pi\tilde{\nu}\kappa x}, \tag{9}$$

or in terms of the intensity at $x = 0$, I_0:

$$I = I_0 e^{-4\pi\tilde{\nu}\kappa x}. \tag{10}$$

Comparing (10) to (5) then:

$$4\pi\tilde{\nu}\kappa = \sigma(\tilde{\nu})n' = \ln 10\varepsilon(\tilde{\nu})c_M \tag{11}$$

In classical dispersion theory, the interaction of electromagnetic radiation with the medium is described by a model that has electrons harmonically bound to positive charges (e.g., nuclei). These oscillating electrons have characteristic frequencies and damping constants, resulting in distinctive oscillations when the frequency of the incident field is close to the characteristic frequency (in analogy to the resonant character of atomic electronic transitions). The polarization induced in the medium by the external field, without accounting for any ordering of the permanent molecular dipoles, is described in terms of the displacements induced in the harmonically bound electrons.[22–24 25 26] In this way, the model accounts for the dielectric constant $\tilde{\epsilon}$ and determines both the real and imaginary parts of the refractive index $\tilde{n} = \sqrt{\tilde{\epsilon}}$.

In 1921, in the context of the development of quantum mechanics, Ladenburg introduced the oscillator strength[27] as a quantity that represents the fraction of the total number of atoms/molecules that have a "dispersion electron," i.e., those electrons oscillating with the characteristic frequency of a given electronic transition.

From the relations above, f-values can be expressed in terms of the integrated absorption intensity of a band as:[15,18]

$$f_{exp} = 10^3 \ln 10 \frac{m_e c^2}{\pi e^2 N_A} \int_{band} \varepsilon(\tilde{\nu}) d\tilde{\nu}, \tag{12}$$

where $\varepsilon(\tilde{\nu})$ is expressed in $M^{-1}cm^{-1}$ and $d\tilde{\nu}$ in cm^{-1}.

Two comments regarding equation (12) must be made. First, it is common to include the (real) refractive index of the medium in the denominator of (12) to account for the effect of the solvent on the electric field "felt" by the solute.[18,15] This correction is expected to make f values obtained in different solvents directly comparable. The factor $1/n$ is recognized to be a rough approximation[15] for a complicated problem with several authors proposing different corrections. This is discussed further at the end of the section.

Second, while the OS is a well-defined magnitude for electronic transitions in atoms or even for a line corresponding to a vibronic transition in a molecule, it does not have the same clear meaning for a molecular spectral band.[28,29,24] The main obstacles are the temperature-dependent population of energy sublevels, and the spread of frequencies over which molecular transitions are possible.

Nonetheless, f values obtained according to equation (12), as a function of the integrated intensity if nothing else, quantify the probability of an electronic transition in a way suitable to compare to theoretical probability computations.

Theoretical oscillator strength

In the linear regime (where the Beer-Lambert law is valid), the reduction in light's intensity as it goes through a sample is attributed exclusively to one-photon processes. At these intensities, light can be treated classically while the light–molecule interaction can be described using time-dependent perturbation theory. Such a semiclassical treatment of the interaction is presented at length in several quantum mechanics textbooks such as Refs. 30–33 as well as in books focused on spectroscopy such as Refs. 15 and 29. To frame our comparison with experimental strengths, a minimal discussion of the key points in the derivation is provided

below.

A linearly polarized plane wave, such as (7), can be expressed in terms of the vector potential $\mathbf{A}$. Under the Coulomb gauge, and in the absence of charges, $\mathbf{A}$ relates to $\mathbf{E}$ and $\mathbf{H}$ via $\mathbf{E} = -(1/c)\partial\mathbf{A}/\partial t$ and $\mathbf{H} = \nabla \times \mathbf{A}$. Expressed as a function of time (t) and position ($\mathbf{r}$), $\mathbf{A}$ is given by:

$$\mathbf{A}(\mathbf{r}, t) = 2A_0\mathbf{u}\cos[(\mathbf{k}\cdot\mathbf{r}) - \omega t] \tag{13}$$

$$\mathbf{A}(\mathbf{r}, t) = A_0\mathbf{u}e^{-i(\mathbf{k}\cdot\mathbf{r})}e^{i\omega t} + A_0\mathbf{u}e^{i(\mathbf{k}\cdot\mathbf{r})}e^{-i\omega t}, \tag{14}$$

where $\mathbf{u}$ is the direction of polarization of the wave, $\mathbf{k}$ is the wave vector that points in the direction of propagation of the wave, and ω is the angular frequency. The magnitude of the wave vector, $|\mathbf{k}|$, is related to the wavelength of light λ by $|\mathbf{k}| = 2\pi/\lambda$ and to ω and the speed of light in vacuum c as $|\mathbf{k}| = \omega/c$, while the angular frequency $\omega = 2\pi\nu$ is related to the wave number in vacuum as $\omega = 2\pi\tilde{\nu}c$. The constant A_0 represents the intensity of light and can be related to the average number of monochromatic photons of energy $\hbar\omega$ per unit volume, $N_{photons}$:[31]

$$A_0 = \sqrt{\frac{2\pi\hbar c^2 N_{photons}}{\omega}}. \tag{15}$$

The probability per unit time P_{nm} of a molecule absorbing a photon of the incident field, described by (14), and undergoing a transition from an initial state m to a final state n is given by:[6]

$$P_{nm} = \frac{2\pi}{\hbar}\left|\frac{A_0 e}{m_e c}\langle n|\mathbf{u}\cdot\hat{\mathbf{p}}\cdot e^{i(\mathbf{k}\cdot\mathbf{r})}|m\rangle\right|^2 \delta(E_n - E_m - \hbar\omega) \tag{16}$$

where e and m_e are the charge and mass of an electron, respectively, $\hat{\mathbf{p}}$ is the linear momentum operator, E_m and E_n are the energies of the initial and final states, respectively, $\delta(x)$ is Dirac's delta function, and $\langle n|\mathbf{u}\cdot\hat{\mathbf{p}}\cdot e^{i(\mathbf{k}\cdot\mathbf{r})}|m\rangle$ is the transition moment integral.

[6]From equations (16) to (23), n represents the final state of a system, while in the rest of the manuscript n refers to a solvent refractive index.

Expression (16) is obtained as a first-order approximate solution in a time-dependent perturbation-theory treatment of the interaction. The perturbation used, $-\frac{e}{mc}\mathbf{A}\cdot\hat{\mathbf{p}}$, is the most relevant term from the classical Hamiltonian for the interaction of an electron with an electromagnetic field characterized by the vector potential $\mathbf{A}$ (and the scalar potential ϕ.) The Dirac's $\delta(x)$ in (16) represents the resonant character of the transition and conservation of energy.

The position $\mathbf{r}$ in (14), and consequently in (16), is measured from an arbitrary origin. Choosing it to be located on the molecule is convenient to carry out the integration in (16), since the integrand will be non-vanishing only in the proximity of the molecule, where the wavefunctions are different from zero. Given the relatively small dimensions of a molecule compared to the wavelength of the electromagnetic radiation, $\mathbf{k}\cdot\mathbf{r} \ll 1$ in the relevant volume. It is convenient then, to expand the exponential term $e^{i(\mathbf{k}\cdot\mathbf{r})}$ in (16) using the definition of an exponential function in the complex plane:

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \tag{17}$$

where $z \in \mathbb{C}$. This gives the infinite series:

$$e^{i(\mathbf{k}\cdot\mathbf{r})} = 1 + i(\mathbf{k}\cdot\mathbf{r}) - \frac{(\mathbf{k}\cdot\mathbf{r})^2}{2!} - i\frac{(\mathbf{k}\cdot\mathbf{r})^3}{3!} + \cdots \tag{18}$$

Taking $e^{i(\mathbf{k}\cdot\mathbf{r})} \approx 1$ results in what is known as the dipole approximation:

$$P_{nm} = \frac{2\pi}{\hbar}\left|\frac{A_0 e}{m_e c}\mathbf{u}\cdot\langle n|\hat{\mathbf{p}}|m\rangle\right|^2 \delta(E_n - E_m - \hbar\omega) \tag{19}$$

Expression (19) provides the probability per unit time of an induced transition on a specific molecule. Such a probability will depend on the orientation of the molecule relative to the polarization of the EM wave, as indicated by the product $\mathbf{u}\cdot\langle n|\hat{\mathbf{p}}|m\rangle$. Therefore P_{nm} will be maximal when $\mathbf{u}$ and $\langle n|\hat{\mathbf{p}}|m\rangle$ are aligned in the same direction. On the other hand,

for a perpendicular orientation of these vectors, P_{nm} will be 0. For any given molecule, $0 \leq \mathbf{u} \cdot \langle n|\hat{\mathbf{p}}|m\rangle \leq |\langle n|\hat{\mathbf{p}}|m\rangle|$ (since $|\mathbf{u}| = 1$). The different orientation of the molecules relative to the polarization of the field results in an average value of $|\langle n|\hat{\mathbf{p}}|m\rangle|^2/3$ for the product $|\mathbf{u} \cdot \langle n|\hat{\mathbf{p}}|m\rangle|^2$ when considering all the molecules in an isotropic sample.[34]

The commutation relation between the position $\hat{r}$ and momentum $\hat{p}$ operators can be used to relate the transition moments associated with each:

$$\langle n|\hat{\mathbf{p}}|m\rangle = im_e\omega_{nm}\langle n|\hat{\mathbf{r}}|m\rangle \tag{20}$$

The equality in (20) only holds for exact wavefunctions. The approximate nature of $\langle n|$ and $|m\rangle$ makes the transition probabilities computed from $\langle n|\hat{p}|m\rangle$ (referred to as dipole velocity formulation) or $\langle n|\hat{r}|m\rangle$ (dipole length formulation) differ. These two formulations are perhaps the most widely used, though in general the transition dipole moments can be expressed in terms of other operators.[35,36]

The oscillator strength f_{nm} for a transition from an initial state m to a final state n is obtained from P_{nm} and can be expressed in either the length formulation (superscript lg), the velocity formulation (vg), or a mixed formulation (mx), as follows:[36,37]

$$f_{nm}^{lg} = \frac{2m_e\omega_{nm}}{3\hbar}\,|\langle n|\hat{\mathbf{r}}|m\rangle|^2 \tag{21}$$

$$f_{nm}^{vg} = \frac{2}{3\hbar m_e\omega_{nm}}\,|\langle n|\hat{\mathbf{p}}|m\rangle|^2 \tag{22}$$

$$f_{nm}^{mx} = \frac{2i}{3\hbar}\,|\langle n|\hat{\mathbf{r}}|m\rangle\langle m|\hat{\mathbf{p}}|n\rangle| \tag{23}$$

The expressions above do not explicitly account for solvent effects. The presence of a solvent influences transition probabilities in three ways: (i) the solvent may chemically alter the solute (e.g., tautomerization, acid-base reactions, complexation, etc.), (ii) the solvent's

electrostatic potential acts on the solute, and (iii) the solvent affects the incident electromagnetic field that drives the electronic transition on the solute.[38,39]

Effects (i) and (ii) would affect the probability (per unit time) of an induced electronic transition, as given by equation 19, by altering the wavefunctions of the absorbing molecule in the ground and excited state, therefore affecting the transition dipole moment $\langle n|\hat{\mathbf{p}}|m\rangle$. On the other hand, effect (iii) would modify the intensity of the perturbation along the direction of the transition dipole moment which is represented in equation 19 by the projection of the term $A_0\mathbf{u}$ along the direction of $\langle n|\hat{\mathbf{p}}|m\rangle$.

If considered in the context of the Beer-Lambert law derivation, where the decrease in light intensity dI as it travels a distance dx is given by $dI = -I\sigma n' dx$, effects (i) and (ii) would affect the absorption cross-section σ, while effect (iii) would result in a replacement of the light intensity I, proportional to the square of the electric field (E^2), with a more complicated function of the incident field: $dI = -f(E)\sigma_{\text{solution}} n' dx$ (where n' is the number of molecules per unit volume).

The first two interactions (i) and (ii) exist in the absence of the incident field. For absorption measurements, solvents and experimental conditions are chosen to prevent as much as possible chemical alteration of the solute, and effect (i) will not be discussed here. The description of the second effect, dating back to Onsager's "reaction field,"[40] is historically linked to the description of effect (iii). The reaction field that acts on the solute molecule originates from the polarization of the dielectric medium caused by the solute molecule itself. The reaction field has been the subject of intense research resulting in several *continuum* solvation models that treat the solute-solvent interaction representing the latter with a continuum dielectric material. These models are widely used and implemented in many electronic structure software packages.[41,42] Effect (ii) can also be treated with calculations including *explicit* solvent molecules around the solute of interest. In this manuscript we assume that effect (ii) is adequately described by the *polarizable continuum model (PCM)* used in our computations (see the methods section for further details). By default, those

methods do not account for effect (iii) in most widely used implementations. We describe that effect in some more detail next to understand how it affects experimentally measured and computed absorption intensities.

Cavity field corrections

The effect of the solvent on the incident electromagnetic field "felt" by a solute molecule is usually described by considering a cavity that contains the solute inside a macroscopic dielectric medium that represents the solvent. Considering a dielectric medium upon which an external static (constant-in-time) electric field acts, and a hypothetical spherical cavity large enough that its inner region still can be described by the macroscopic constants of the surrounding medium, Lorentz obtained, for the local field F_L acting upon a charge inside the cavity:[43]

$$F_L = \frac{\epsilon_s + 2}{3} E \tag{24}$$

where E is the macroscopic electric field in the dielectric, and ϵ_s is the static dielectric constant which is assumed to be independent of E when the latter is small enough to prevent saturation effects.[44]

Onsager proposed the division of the local field into two components, a "cavity field" proportional to the external field as well as to the polarization induced by this field, and the aforementioned "reaction field" proportional to the dipole moment of the solute molecule. Considering a spherical cavity containing only a dipolar molecule, he obtained for the cavity field $F_{C,O}$:[40] [7]

$$F_{C,O} = \frac{3\epsilon_s}{2\epsilon_s + 1} E \tag{25}$$

Onsager's theory of polarization has further generalizations. For example, Kirkwood[45] considered explicitly the electric moments of the first shell of solvent molecules around a solute. His corrections are particularly relevant for polar solvents, since in the vicinity

[7] If Onsager's cavity is considered as a homogeneously polarized continuum, then the resulting field (reaction + cavity) acting on the solute molecule equals the Lorentz field (see for example Ref. 44)

of a given molecule, the surrounding molecules tend to maintain definite (either parallel or antiparallel) orientations of their dipoles.[46] Detailed derivations and discussion of these effects are provided in Ref. 44.

Other authors have generalized the local field corrections by considering non-spherical cavities. For example, Scholte[47] and Shibuya[48] considered ellipsoidal cavities and the fields along the principal axes of the ellipsoid.

Using the Maxwell relation connecting the refractive index of a medium to its dielectric constant ($n^2 = \epsilon$), the different cavity field corrections have been adapted to time-varying fields, for which the permanent dipoles of solvent molecules have no time to reorient. Using these local field expressions, several authors have proposed corrections that relate the "apparent" OS of a molecule when experiments are carried out in solvents with different refractive indexes. We will employ a common notation to summarize some of those corrections: we call f'' the OS measured (according to equation 12) for a molecule in a solvent of refractive index n, and f the OS of the same molecule when the refractive index of the medium is 1.

The first correction dates back to 1934. Chako, using a Lorentz field obtained:[49]

$$\frac{f''}{f} = \frac{(n^2+2)^2}{9n} \tag{26}$$

A limitation of this expression is that it predicts that f'' will always increase, at a fixed rate, when absorption experiments are carried out in solvents of higher refractive index. This is known not to be the case, with notable exceptions such as the $\pi \rightarrow \pi^*$ transition of β-carotene.[50] To correct this issue, Böttcher[51] and Schuyer[52] included the polarizability and radii of the solute molecules in their correction. Myers and Birge[50] considered a cylindrical cavity and the orientation of the transition moment relative to the cavity. Shibuya,[48] considering the local field along the principal axis of the ellipsoidal cavity he used, obtained for

a transition oriented along the principal axis k the correction:

$$\frac{f''_k}{f_k} = \frac{[s_k(n^2-1)+1]^2}{n}, \tag{27}$$

where s_k is a shape parameter equal to the depolarization factor along the corresponding axis. The value of s_k ranges from $0-1$, and $\sum_k^3 s_k = 1$.[53] Therefore, Shibuya's correction f''/f ranges from $1/n$ to n^3 depending on the molecular shape and relative orientation of the transition moment, and reduces to Chako's in the limit where the ellipsoid becomes a sphere. Ref. 48 contains a more extensive survey of the work in this area prior to 1983.

Other corrections depending only on the refractive index are those by Abe.[54] Using a Lorentz field Abe obtained the expression:[54]

$$\frac{f''}{f} = \frac{(n^2+2)^2}{9n^2}, \tag{28}$$

while when using an Onsager cavity field he obtained:

$$\frac{f''}{f} = \frac{9n^2}{(2n+1)^2}. \tag{29}$$

In the same paper Abe approximated Schuyer's expression as

$$\frac{f''}{f} = \frac{6n}{(2n^2+1)(n^2+1)}. \tag{30}$$

For this last expression, it is worth mentioning that instead of equation (12) he considered a modified relation of the OS to the integral of the attenuation coefficient that included the factor n^2 in the denominator.

The value of n to use in the corrections above is also subject to debate. Some authors recommend using the refractive index of the solvent for the frequency of the transition, $n(\tilde{\nu})$, while other authors recommend $n(\infty)$ (the refractive index of a material for a field in the

limit of infinite frequency). This latter value has being said to be well approximated by the refractive index at the sodium D line (n_D).[41] Warner and Wolfsberg,[55] in their study of spectra in condensed phases, used the Lorentz field to derive the correction factor $\frac{1}{n_b}\left(\frac{2+n_b^2}{3}\right)^2$ where n_b is the "slowly varying background contribution to the refractive index" that depends on the off-resonance polarizability of the solute molecules. They obtained n_b values by fitting a model for $n(\tilde{\nu})$ presented in their paper to data from reflection experiments. For benzene, chloroform, and methyl iodide, they found n_b values close to, but consistently lower than, the corresponding n_D values.[55]

It is worth mentioning that the dependence on n in the f''/f correction factors comes not only from the local/cavity field correction. This may be best understood in terms of the absorption cross section (see, for example, the derivations in Ref. 56), which is defined as *energy (per unit time) absorbed by the molecule* divided by the *energy flux of the radiation field*. Macroscopically, the flux of energy of the incident electromagnetic wave in a non-magnetic medium of refractive index n is given by:[20,57]

$$S = \frac{c}{8\pi} |E|^2 n. \tag{31}$$

While the local/cavity field is considered for the *energy (per unit time) absorbed by the molecule*, equation (31) must be taken into account for the *energy flux of the radiation field*. Comparison of Chako's correction (equation 26) to the Lorentz local field (equation 24) shows that n in the denominator of (26) comes from (31). The same applies to Shibuya's correction (expression 27).

The computational description of the reaction field (responsible for effect ii) using implicit solvation models typically accounts for solute–solvent electrostatic interactions using apparent charges at the surface of a cavity constructed around the molecule. Details of how the cavity is constructed, how the apparent charge on the cavity's surface is discretized, or how to treat non-equilibrium effects, have being extensively researched. More recent efforts have being made to describe the cavity field in terms of additional surface charges. See, for

example, Refs. 56,58,59.

Methods

Computed absorption transitions

The coordinates of the 100 molecules optimized at the B3LYP/6-31+G* level of theory were obtained from the supporting information of Ref. 12. Here, we refined the structures at the B3LYP/6-311++G** level of theory.[60–62] Frequency calculations were carried out at the same level of theory to ensure that all positive frequencies were obtained. The updated geometries are provided in the Supporting Information (SI) as xyz coordinate files.

In this study, we will focus our analysis on 85 transitions categorized as very high, high, or medium confidence in Ref. 12. This subset of data will be referred to as VHHM throughout this work (where VHHM = VH ∪ H ∪ M). Since more than one transition per molecule is sometimes included, the 85 transitions come from 69 molecules. Table S1 of the SI document lists the molecules and transitions included in the subset.

From the optimized geometries, the energies and OSs of the lowest 30 singlet exited states were computed using single-point TD-DFT calculations. Nine different functionals were tested: One pure functional (SVWN[63,64]), five hybrid functionals (B3P86,[60,65] O3LYP,[61,66] mPW1PW91,[67] M05,[68] and B3LYP[60,61]), and three long-range corrected hybrid functionals (CAM-B3LYP,[69] LC-wHPBE,[70,71] and wB97XD[72]). The 6-311++G** basis set[62] was also used for all TD-DFT calculations. The solvent effect was included, in both geometry optimizations and single-point calculations, through PCM using the integral equation formalism (IEFPCM).[73]

Excitation energies and OSs for all nine functionals were computed both with and without the Tamm-Dancoff approximation (TDA)[74] using the same basis set and solvation method. We also recomputed the transition energies and OSs for TD-B3LYP with multiple basis sets: STO-3G, 3-21G, 6-31G*, 6-31++G**, cc-pVDZ, aug-cc-pVDZ, and aug-cc-pVTZ.[62,75–77]

The calculations above were all performed using Gaussian 16 version C.01.[78] In addition, using the PCM/6-31+G* optimized geometries reported by Tarleton et al.,[12] we carried out additional calculations using the 6-31+G* basis set with Q-Chem 5.3.[79] Specifically, we ran the calculation in vacuo and also using two additional solvation models; the conductor-like PCM (CPCM) model and COSMO. Those were compared to the B3LYP/6-31+G* calculations using IEFPCM solvation from Gaussian. Differences in the computed OSs using the three solvation models and two different software were negligible, indicating that the strengths are relatively insensitive to the details of the solvent model implementations tested. The gas phase calculations are discussed further in the Results and Discussion Section.

In addition to TD-DFT calculations, we carried out single-point excited state energy calculations using three *ab initio* methods: time-dependent Hartree-Fock (TD-HF), configuration interaction singles (CIS) and equation of motion coupled cluster with singles and doubles (EOM-CCSD). In the case of EOM-CCSD, we compute OSs from linear response transition densities (LR-CCSD) in addition to the unrelaxed EOM ones.[80–82] EOM-CCSD applies excitation operators to a CCSD ground state reference and includes doubly excited configurations.[83,84] These wave funcion method calculations were carried out for a smaller subset of 35 transitions from 26 molecules for which EOM-CCSD calculations were tractable. Those molecules and transitions are listed in Table S1 of the SI document. Furthermore, for the EOM-CCSD calculations, only 15 excited states were requested instead of 30. The EOM-CCSD calculations were carried out using the double-ζ aug-cc-pVDZ basis set.[77]

Statistical analysis

When running, for instance, a TD-DFT calculation for 30 excited states of a given molecule, 30 OSs are obtained in each gauge (i.e., 30 of each of f_{nm}^{lg}, f_{nm}^{vg}, and f_{nm}^{mx}). Here, we will drop the superscripts related to the gauge, as the same discussion will apply to all three gauges. The individual state-specific OSs are denoted f_{n0}, where 0 is the index for the ground state and n represents the excited state index (e.g., 1 for the first singlet excited

state, 2 for the second singlet excited state, etc.). As a first approximation, a code assigns the computed f_{n0} to an experimental band of the molecule if its corresponding energy is within the energy limits of the band ($\varepsilon(\tilde{\nu})$ minima in the experimental spectra). Often, more than one transition contributes to a band. When that is the case, we use the sum of the corresponding f_{n0} values to find the total OS of the band. The OS computed in this way for a specific band, k, is referred to as $f_{comp,k}$ and can be compared to the corresponding $f_{exp,k}$.

Our benchmark for OSs faces the inconvenience that not only are the individual f_{n0} values dependent on energy, but also that the set of f_{n0} values that contribute to a given $f_{comp,k}$ is affected by the accuracy of the computed (vertical) electronic transition energies. Judging whether or not a computed transition belongs in a band may not be straightforward, even more so if we are not sure whether f_{comp} values should reproduce f_{exp} values as given by (12), or f_{exp}/n as proposed in Refs. 15,18, or one of the other solvent effect corrections proposed. Therefore, an algorithm has been used that actively maximizes the agreement between f_{comp} and f_{exp} (or f_{exp}/n, nf_{exp}, etc...) by modifying which f_{n0}s contribute to a band. The algorithm does this by shifting the band limits to include or exclude computed transitions to minimize $|f_{comp} - f_{exp}|$ (or equivalently $|f_{comp} - f_{exp}/n|$, $|f_{comp} - nf_{exp}|$, etc). As an initial guess, the band limits are given by the experimental $\varepsilon(\tilde{\nu})$ minima.

The implementation avoids having a specific transition f_{n0} double-counted towards two different bands. This should give a "best-case scenario" where the computed OSs are as close as possible to the experimental ones. Note that in case of an incorrect energy ordering of excited states, the algorithm will not be able to repair the issue. While we recognize the shortcomings of this approach, automation was necessary given the large number of computations in the present benchmark.

For each set of computations, two comparisons with experimental data are presented. These are labeled "Exact Band Limits" and "Improved Fit". In the first case computed f_{n0} are assigned to an experimental band if the computed energy lies strictly within the energy limits of the band. The second case corresponds to the application of the algorithm

described above (labeled "Improved Fit Algorithm" as well). We emphasize that with the use of the Improved Fit algorithm, it becomes only possible to discuss an upper limit to the accuracy of a method's f values. On the other hand, within the Exact band limits framework, large computed energy discrepancies with experimental energies will severely affect the OS comparisons.

To quantify the agreement of a method with the experimental data, two sets of metrics are employed: The first metric is the mean absolute error, MAE, calculated as:

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{transitions}}} \sum_k \left| f_{exp,k} - f_{comp,k} \right| , \tag{32}$$

where $N_{\mathrm{transitions}}$ is the total number of transitions. The second set of metrics is obtained from linear regression analysis of the (f_{exp}, f_{comp}) pairs. A small MAE, a linear fit close to $y = x + 0$, and an R^2 value close to one are indicators of a good agreement between the set of f_{comp} obtained with a given method and the corresponding set of experimental values f_{exp}.

Since $f_{comp,k}$ values depend on the computed energy (with f_{n0} explicitly dependent in the position and momentum gauges) we must also pay attention to how the computed transition energies compare to the experimental ones. To describe the experimental transition energies, an average transition energy is obtained for a band k as

$$E_{exp,k} = \frac{\int_{band} \tilde{\nu} \varepsilon(\tilde{\nu})\, d\tilde{\nu}}{\int_{band} \varepsilon(\tilde{\nu})\, d\tilde{\nu}}, \tag{33}$$

using the data from Ref. 85 and digitized in Ref. 12 for $\varepsilon(\tilde{\nu})$.

Analogously, computed energies are obtained as oscillator strength-weighted transition energy averages:

$$E_{comp,k} = \frac{\sum_{n \in band} \tilde{\nu}_n \cdot f_{n0}}{\sum_{n \in band} f_{n0}} \tag{34}$$

Three metrics are employed to monitor how computations reproduce experimental energies:

i) The mean absolute error (in energy) computed as:

$$\langle|\Delta E|\rangle = \frac{1}{N_{\text{transitions}}} \sum_k \left|E_{comp,k} - E_{exp,k}\right|. \tag{35}$$

ii) The mean error (in energy) obtained as:

$$\langle\Delta E\rangle = \frac{1}{N_{\text{transitions}}} \sum_k E_{comp,k} - E_{exp,k}. \tag{36}$$

iii) The mean ratio of computed to experimental energies, given by:

$$\left\langle \frac{E_{comp}}{E_{exp}} \right\rangle = \frac{1}{N_{\text{transitions}}} \sum_k \frac{E_{comp,k}}{E_{exp,k}}. \tag{37}$$

As mentioned earlier, solvent effects need to be accounted for when comparing computed vs. experimental OSs. In the experimental spectra, the solvent effects are intrinsic to the measured attenuation coefficient leading to an apparent f_{exp} derived using equation (12). On the other hand f_{comp} values computed with PCM only account for the reaction field component of the local field while computations without any solvent model make no account at all. That is why, initially, we start by comparing f_{comp} not only to f_{exp}, but also to f_{exp}/n and nf_{exp}. We also test some of the cavity field corrections mentioned before, noticing the equivalence of f_{exp} and f_{comp} to the notation used in the cavity field corrections subsection: $f'' \rightarrow f_{exp}$ and $f \rightarrow f_{comp}$. Making that substitution, for example, on Chako's correction (equation 26) we obtain:

$$f_{comp} = \frac{9n}{(n^2+2)^2} f_{exp} \tag{38}$$

$$f_{comp} = C_{\text{Chako}}(n) f_{exp}, \tag{39}$$

where $C_{\text{Chako}}(n) = 9n/(n^2+2)^2$. We can obtain equivalently $C_{\text{Shibuya},k}(n) = n/[s_k(n^2-1)+1]^2$, $C_{\text{AbeL}}(n) = 9n^2/(n^2+2)^2$, $C_{\text{AbeO}}(n) = 9n^2/(2n+1)^2$, and $C_{\text{Schuyer}}(n) = [(2n^2+1)(n^2+1)]/(6n)$, from equations (27), (28), (29), and (30), respectively. The agreement of f_{comp} to

$C_{\text{Chako}} \cdot f_{exp}$, $C_{\text{AbeL}} \cdot f_{exp}$, $C_{\text{AbeO}} \cdot f_{exp}$, and $C_{\text{Schuyer}} \cdot f_{exp}$ is also tested in this work.

We calculate the expressions above using $n(\tilde{\nu})$ evaluated at the corresponding frequency of each transition. We used the dispersion formulas reported in the litearture for water,[86] ethanol,[87] CCl_4,[88] dioxane,[88] acetonitrile,[89] methanol,[89] cyclohexane,[89] hexane,[90] and heptane.[90] Those expressions are collected in Refs. 91,92. We also evaluated the correction factors using the value at the sodium D-line (n_D) obtained from Ref. 93.

An alternative approach to using these cavity field corrections is to find the optimal scaling factor that relates the computed and experimental OSs. Consider Algorithm 1, shown below. It computes C constants that reflect the slope between (f_{exp} and f_{comp}) accounting for the Improved Fit algorithm. The value of C often converges after a few iterations. However, in a few cases, the algorithm is sensitive to the initial guess for C. This is discussed further in the Results and Discussion Section.

Algorithm 1 Iterative approach to find a scaling factor C

1: Program Start
2: Obtain f_{comp} within the "Exact Band Limits"
3: Initialize constant $C = 1$
4: **loop**
5: Read f_{exp} set
6: Transform the experimental set $f_{exp} = C f_{exp}$
7: Transform f_{comp} set according to the band-matching "Improved Fit" algorithm to reduce MAE(f_{exp}, f_{comp})
8: Linear regression analysis of the pairs (f_{exp}, f_{comp})
9: c_1 = Slope of the linear regression through the origin
10: Update $C = c_1 \cdot C$
11: **end loop**

As discussed in the computed absorption transitions subsection, the majority of the statistical analysis focuses on the subset of VHHM transitions. A smaller subset of 35 transitions is used for comparison between wave function methods. We carry out further analysis by looking at subsets of VHHM prepared based on: 1) transition character ($\pi \rightarrow \pi^*$, charge transfer, or mixed character), 2) point group symmetry (C_1, C_s, or higher symmetry), 3) solvent (water, electrolyte solution, ethanol, methanol, heptane, hexane, and cyclohexane), and

4) Spectrophotometer used to measure the experimental UV-visible spectra (Zeiss PMQ II, MM12, Perkin Elmer 4000 A, Unicam SP 500, or other). In the case of solvents, we also prepared a subset that is a union of protic solvents (ethanol, methanol, water, electrolytes) and aprotic nonpolar solvents (heptane, hexane, CCl_4, petroleum ether, cyclohexane, dioxane).

Results and Discussion

In Ref. 12, we focused on benchmarking a single method, TD-B3LYP. In Figure 2, we extend the comparison to include wave function method calculations for 35 transitions that belong to the VHHM subset. Large molecules from that subset are excluded due to computational cost. These transitions are listed in Table S1. Six methods are tested: CIS, TD-HF, EOM-CCSD, LR-CCSD, Tamm-Dancoff approximated (TDA) DFT and TD-DFT.

The B3LYP functional is used for this plot. For the sake of simplicity, at this stage, we consider only three possible approximate solvent effect corrections: f_{exp}/n, f_{exp}, and $n \cdot f_{exp}$. We present the other pre-factors later in this Section.

In the center of Fig. 2, a plot shows the MAE between computed and experimental OSs, calculated using Eq. (32). The MAEs are shown for different methods (labeled at the bottom of the plot), for different refractive index pre-factors for the experimental strengths (f_{exp}/n, f_{exp}, and $n \cdot f_{exp}$, shown on the top of the plot), different gauges (represented using different symbols), and before and after the application of the Improved Fit band matching algorithm (red and green, respectively). The blue bar outline indicates the average values of f_{exp} multiplied by the respective refractive index pre-factor, and serves as a reference to allow comparison of the magnitude of the MAE relative to the average value of the experimental OS itself.

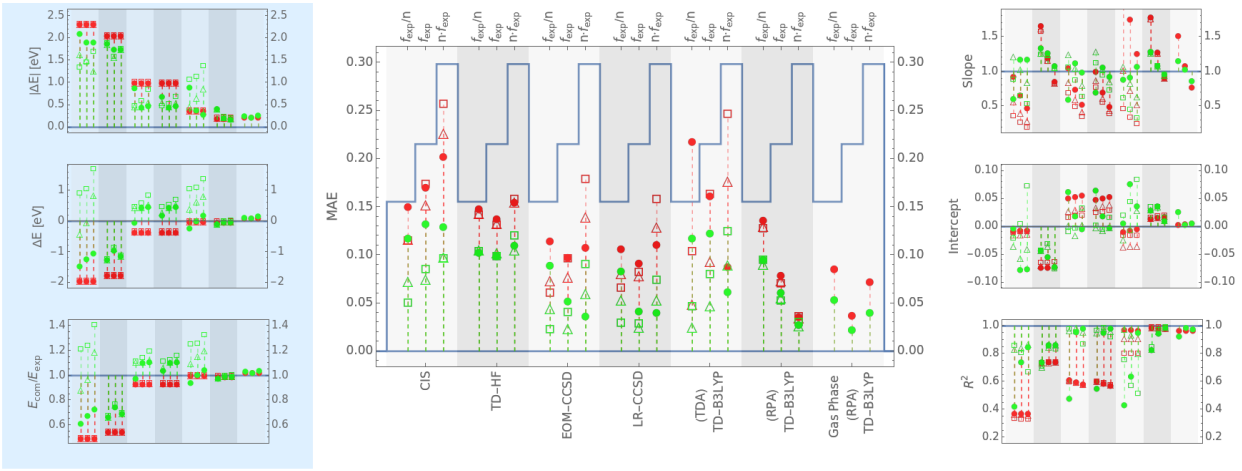

Figure 2: Comparison of f-values computed using CIS/6-311++G**, TD-HF/6-311++G**, EOM-EE-CCSD/aug-cc-pVDZ, LR-CCSD/aug-cc-pVDZ, TDA-B3LYP/6-311++G**, RPA TD-B3LYP/6-311++G**, and gas phase-RPA-TD-B3LYP/6-31+G* for a subset of 35 experimental transitions. For each method, the f_{comp} values are compared to f_{exp}/n (left), f_{exp} (center), and $n \cdot f_{exp}$ (right) as indicated by the labels on top of the central plot. The blue bar outline indicates the average values of f_{exp} multiplied by the respective refractive index pre-factor ($\langle f_{exp}/n \rangle = 0.155384$, $\langle f_{exp} \rangle = 0.215319$, and $\langle n \cdot f_{exp} \rangle = 0.298506$). A full circle corresponds to the data obtained with the length gauge, an empty square corresponds to the velocity gauge, and an empty triangle corresponds to the mixed gauge. Markers in red correspond to transitions assigned using the Exact Band Limits, while markers in green correspond to transitions assigned using the Improved Fit algorithm. The data displayed can be found in Tables S2 to S19 of the SI document. See the text regarding the interpretation of these plots.

Six additional panels are shown in all figures presented in this section. The panels on the left and right sides are set up in the same way as the central one but with x-axis labels excluded. The results for different methods are shown with alternating background shades to help correlate with the labels in the central panel.

The three panels on the left indicate the agreement between experimental and computed excitation energies, obtained from Eqs. (33) and (34), respectively. $|\Delta E|$, ΔE, and $\frac{E_{comp}}{E_{exp}}$ are calculated using equations (35), (36), and (37), respectively. Notably, if no state was found within the band limits, $E_{comp,k}$ was assigned a value of zero. Therefore, a negative value of ΔE does not necessarily mean that computed transitions are systematically red-shifted with respect to computed ones but instead indicates that many transitions may fall outside the

band limits. When this occurs, the associated $|\Delta E|$ will be large and $\frac{E_{comp}}{E_{exp}}$ will be smaller than 1.

The comparison between computed and experimental energies before (red symbols) and after (green symbols) the Improved Fit algorithm reflects what the algorithm did. The algorithm does not alter the computed energy associated with a given f_{n0}; it just assigns or unassigned computed transitions to each band. Therefore, when the red and green symbols are not equal, that means that the algorithm made changes to the band assignments. The values of $|\Delta E|$, ΔE, and $\frac{E_{comp}}{E_{exp}}$ after the Improved Fit algorithm are more representative of the errors stemming from the electronic structure method used, as those are assumed to have assigned computed transitions reasonably well to the experimental bands.

Fig. 2 indicates that CIS and TD-HF computed transitions often fall completely outside of the experimental absorption band limits. This is consistent with previously reported errors associated with CIS, often in the 0.5-2.0 eV range.[94] The Improved Fit algorithm partly resolves this issue, reducing the errors in the excitation energies, but CIS and TD-HF transitions may still not have been assigned correctly in all cases to the experimental bands. Therefore, those two methods will not be discussed extensively in this section.

Several EOM (or LR) CCSD transitions also fall outside of the experimental band limits. However, the center-left panel shows that the Improved Fit algorithm largely resolves the issue and that EOM-CCSD transitions are typically overestimating rather than underestimating relative to the experimental band energies. This is largely consistent with what is expected of EOM-CCSD with a double-ζ basis set;[95,96] better agreement between computed and experimental excitation energies would require the triples correction and/or a larger basis set, but those would not be tractable for the systems studied here.

TD (or TDA) DFT transitions mostly fall within the band limits, as reported also in the Supporting Information of Ref. 12. The Improved Fit algorithm makes a few changes in the band assignments, but those changes do not significantly affect the energetic error metrics.

The three panels on the right of Fig. 2 indicate the values of the slope, y-intercept,

and R^2 for the linear regression between computed and experimental OSs. An excellent agreement would yield values of 1.0, 0.0, and 1.0, respectively, so any deviations from those values indicate differences between the computed and experimental strengths. Together with the center panel, which presents the MAE, those metrics aid in quantifying the differences in the computed and experimental OSs.

CIS and TDA DFT exhibit a strong dependence of the OS on the gauge used; the length gauge (full circle) and velocity gauge (empty square) often give OSs that can vary significantly. The mixed gauge (empty triangle) is typically in between the other two gauges. This gauge dependence is almost eliminated when using TD-HF or TD-DFT, which follow the Thomas–Reiche–Kuhn sum rule ($\sum_i^N f_i = N$, where N is the number of electrons in the system)[97–99] unlike CIS and TDA.[94,100] Similarly, EOM-CCSD has a larger gauge-dependence than LR-CCSD, but the difference is not as pronounced.

For the remainder of this section, we focus our discussion on TD-DFT (without the Tamm-Dancoff approximation) and LR-CCSD.

In Ref. 12, we verified that (RPA) TD-B3LYP OSs with PCM solvation improved by almost all metrics when compared against $n \cdot f_{exp}$ instead of just f_{exp}. Here, we revisit this comparison focusing on only the subset of 35 VHHM transitions and applying the Improved Fit algorithm. We find, consistently with Tarleton et al.,[12] that TD-B3LYP f_{comp} are overestimated relative to f_{exp}, and are in much better agreement with $n \cdot f_{exp}$. This is reflected in each of the MAE, slope, y-intercept, and R^2 plots in Fig. 2. We note also that the relative error (compared to the average value of experimental OS) is significantly lower for $n \cdot f_{exp}$, as shown in Table 1.

Table 1: Relative MAE for TD-B3LYP compared to the average of the experimental reference. We use the average of the three gauges since TD-B3LYP does not exhibit strong gauge-dependence.

Framework	f_{exp}/n	f_{exp}	$n \cdot f_{exp}$
Exact Band Limits	84.3 %	34.4 %	11.8 %
Improved Fit	60.1 %	26.1 %	9.3 %

Within the Exact Band Limits framework, the metrics for MAE and slope are best for $n \cdot f_{exp}$ and worst for f_{exp}/n. R^2 and y-intercept are instead comparable for f_{exp}/n, f_{exp}, and $n \cdot f_{exp}$. However, application of the Improved Fit algorithm, which reduces the MAE, improves the agreements with $n \cdot f_{exp}$ by almost all metrics; it results in a slight improvement in R^2, minimizes the y-intercept, and reduces the absolute error in energy $|\Delta E|$. Meanwhile, applying the same Improved Fit algorithm when comparing to f_{exp} and f_{exp}/n yields a limited improvement or even a worst agreement (in terms of R^2, y-intercept, and $|\Delta E|$) compared to the Exact Band Limits framework.

Due to the stronger gauge-dependence of EOM-CCSD (which is only partially but not fully resolved with LR-CCSD), it is more difficult to draw conclusions about which solvent correction (f_{exp}/n, f_{exp}, or $n \cdot f_{exp}$) is in best agreement with the EOM-CCSD results. The MAE and other metrics in the length gauge are in best agreement with $n \cdot f_{exp}$. However, velocity gauge calculations give a better agreement with f_{exp}/n. The mixed gauge calculations appear to have a similar error with all three gauges but agree best with f_{exp}. Overall, we expect the results of the length gauge to be more reliable for the double-ζ basis set used.[101]

The fact that the length gauge OSs overestimate f_{exp}, while the momentum gauge OSs underestimate f_{exp} could be partially explained by the computed transition energies being systematically larger than the experimental energies (f_{nm}^{lg} is proportional to the transition energy while f_{nm}^{vg} is inversely proportional to it; see equations 21 and 22) From the E_{comp}/E_{exp} plot in Figure 2 (see Figure 3 as well) EOM(LR)-CCSD computed energies appear to be $\sim$10% larger than the experimental energies. That said, the difference between OSs computed in different gauges seems to be larger than what can be explained by the energy overestimation, as will also become apparent during the discussion of Table 2 below.

The gas phase OS calculations agree better with f_{exp} rather than $n \cdot f_{exp}$. In other words, the reaction field effect introduced by using PCM significantly increases the computed OS. This can be traced to the individual f_{n0}. For most transitions computed in this set, the f_{n0} were larger when computed using PCM. We believe that the agreement of gas-phase

calculations to f_{exp} result from a cancellation of errors.

Next, we optimize the scaling factor "C" that relates the computed and experimental OSs by following the approach in Algorithm 1. For the electronic structure methods presented in Table 2, three initial guesses were tested: $C_0 = 0.7$, 1.0, and 1.4. The converged C values resulting in the highest R^2 are presented in Table 2.

Table 2: C values obtained according to Algorithm 1. The background color indicates the convergence from three different starting points for C. Green indicates that the same C value is obtained from either $C_0 = 0.7$, 1.0, or 1.4. Red or blue indicate that the C value with the highest R^2 comes from $C_0 = 1.4$ or 1.0, respectively. Purple indicates that the same C value with the highest R^2 is obtained from either $C_0 = 1.4$ or 1.0.

Method	**X-gauge**	**P-gauge**	**XP-gauge**
CIS	1.49	0.55	0.88
TD-HF	1.20	1.14	1.17
EOM-CCSD	1.36	0.77	1.01
LR-CCSD	1.27	0.88	1.06
TDA-B3LYP	1.76	0.35	0.72
B3LYP	1.33	1.29	1.31
Gas B3LYP	1.11		

Most of the data in Table 2 has a green background which indicates that the algorithm converged to the same C value independent of the starting point used (0.7, 1.0, or 1.4). Even among the values colored with a red, blue, or purple background, most converged to similar values from the different starting points. For the few exceptions to this, there usually is a poor quality fit (e.g., for CIS) and the R^2 value between $C \cdot f_{exp}$ and f_{comp} gives a good indication for which value of C to trust.

The results in Table 2 again highlight the strong gauge dependence of CIS, EOM-CCSD, and TDA-B3LYP, and the slightly reduced gauge dependence of LR-CCSD. In all those cases, the length gauge gives a scaling factor larger than one and the velocity gauge gives a scaling factor smaller than one. On the other hand, the more gauge-independent TD-HF and TD-B3LYP methods systematically overestimate f_{exp} regardless of the gauge.

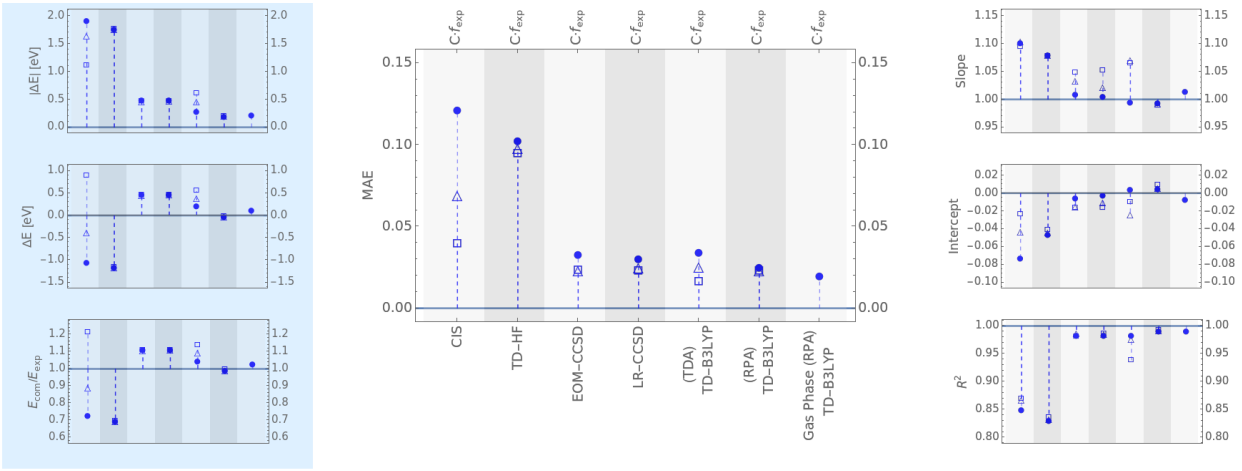

Figure 3: Comparison of f-values computed using CIS/6-311++G**, TD-HF/6-311++G**, EOM-EE-CCSD/aug-cc-pVDZ, LR-CCSD/aug-cc-pVDZ, TDA-B3LYP/6-311++G**, RPA TD-B3LYP/6-311++G**, and gas phase-RPA-TD-B3LYP/6-31+G* for a subset of 35 experimental transitions. For each method, the f_{comp} values are compared to $C \cdot f_{exp}$, where the constants C were obtained according to algorithm 1 and are displayed in Table 2. A full circle corresponds to the data obtained with the length gauge, an empty square corresponds to the velocity gauge, and an empty triangle corresponds to the mixed gauge. The data displayed can be found in Tables S20 to S28 of the SI document.

The statistics for the pairs $(C \cdot f_{exp}, f_{comp})$ are displayed in figure 3. These results represent the best possible fit using a simple scaling factor, C, in combination with the Improved Fit algorithm. We find that all of EOM-CCSD, LR-CCSD, TD-DFT, and TDA-DFT can be give a good agreement (MAE lower than 0.04) with experimental OS through selection of band assignments and appropriate scaling factor C. This algorithm also reduces the gauge dependence for EOM-CCSD and LR-CCSD by using an different scaling factor C for each gauge to better fit the experimental data. This is also true, to a lesser degree, for TDA-DFT, but not CIS, where the MAE still varies strongly with the gauge. Among these methods, TD-DFT with the B3LYP functional still results in the best metrics overall for the slope, y-intercept, R^2, and energy errors, compared to the other methods. Meanwhile, other methods can be fit to reproduce OSs with a similarly low MAE but they either result in a worse linear regression slope and intercept or require using transitions outside of the band limits, which manifests in larger errors in the energy metrics.

For gauge dependent methods, the scaling factor C is a more complicated function that captures multiple effects, including a correction for the gauge used. However, for gauge-independent methods like (RPA) TD-DFT, the origin of the scaling factor C may be largely attributed to the solvent effect. As discussed in the theoretical background section, the solvent impacts the absorption intensity in several ways. The reaction field (effect ii) is accounted for through PCM. However, the cavity field (effect iii) is missing. In Figure 4, we test some of the simple cavity field corrections proposed in the literature (see equations 26 to 30) using both the frequency-specific refractive index of the solvent (n) and the refractive index at the sodium D-line (n_D). We focus at this point on TD-B3LYP calculations. Similar figures for the gas phase and for LR-CCSD are shown in SI Figure S1.

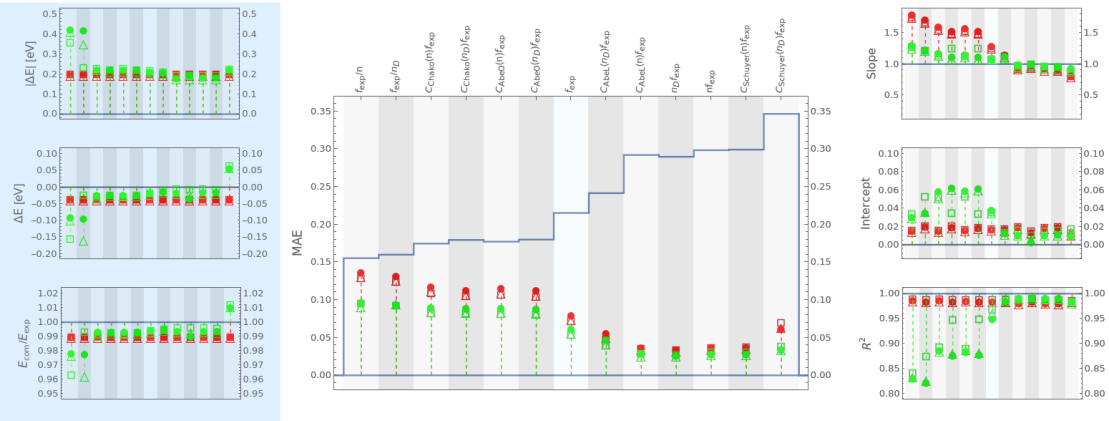

Figure 4: Comparison of f-values computed using TD-B3LYP/6-311++G**/PCM for a subset of 35 experimental transitions multiplied by different cavity field corrections. The blue bar outline indicates the average values of f_{exp} multiplied by the respective cavity field factor. A full circle corresponds to the data obtained with the length gauge, an empty square corresponds to the velocity gauge, and an empty triangle corresponds to the mixed gauge. Markers in red correspond to transitions assigned using the Exact Band Limits, while markers in green correspond to transitions assigned using the Improved Fit algorithm. The data displayed can be found in Tables S29 to S106 of the SI document.

All the cavity field corrections used in Fig. 4 assume a spherical cavity. For the molecules in this benchmark, we find that none of the corrections performed better than a simple multiplication by the refractive index (either n or n_D). The corrections displayed to the left side of f_{exp} give considerably worse agreement by almost all metrics.

As shown earlier in Eq. (31), the factor n that appears in $n \cdot f_{exp}$ arises from the energy flux of the radiation field in a dielectric and appears in early cavity field literature corrections.[48,49] It has been shown that more accurate cavity field corrections would need to account for the cavity shape beyond using a simple spherical approximation.[48,56] Such corrections will be tested in future work.

A similar analysis of different cavity field corrections was carried out for LR-CCSD calculations (see SI Fig. S1). The results in the length gauge, which also give an overestimation of the computed OS relative to the experimental one, largely follow the same trend as observed for TD-B3LYP.

Next, we expand the benchmark set to include all 85 VHHM transitions to compare f-values computed using different TDDFT functionals and basis sets. Hereon, we focus on comparing the computed transitions relative to only f_{exp} and $n \cdot f_{exp}$, and no longer consider other cavity field correction terms. The average experimental OS in the set, as given by equation (12), is $\langle f_{exp} \rangle = 0.3077$, while $\langle n \cdot f_{exp} \rangle$ is 0.4333.

Figure 5 presents the data for the OS calculations carried out using TD-B3LYP and different basis sets. Small basis sets exhibit a strong gauge-dependence, especially for STO-3G and 3-21G*, that is significantly reduced for larger basis sets. In general, the MAE and gauge-dependence continue to decrease with increasing basis set size above 6-31G* (see SI Fig. S2 for more detailed figure). For example, the range of MAEs for the different gauges decreases from 0.083(length)-0.089(velocity) for 6-31G* to 0.076(length)-0.083(velocity) for 6-31++G** to 0.076(length)-0.077(velocity) for 6-311++G** when not using the Improved Fit algorithm. Similarly, across the Dunning basis set series, 0.085(length)-0.088(velocity) for cc-pVDZ to 0.075(length)-0.077(velocity) for aug-cc-pVDZ to 0.076(length)-0.077(velocity) for aug-cc-pVTZ. The same trends are largely conserved when using the Improved Fit algorithm.

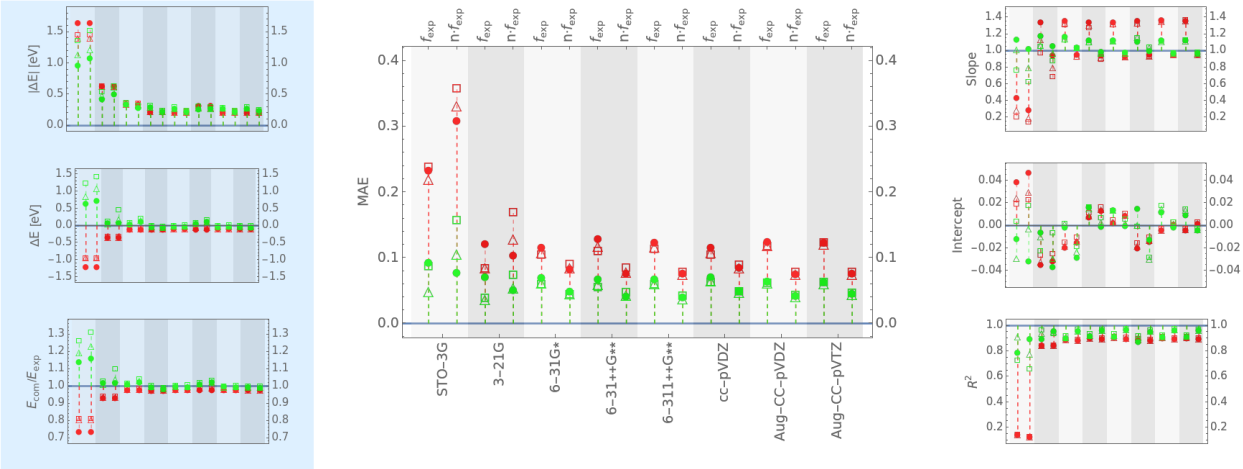

Figure 5: Comparison of f-values computed using B3LYP with different basis sets to a subset of 85 experimental transitions. For each basis set the f_{comp} values are compared to f_{exp} (left) and $n \cdot f_{exp}$ (right). For reference, the average value of the experimental f-values are $\langle f_{exp} \rangle = 0.307655$, and $\langle n \cdot f_{exp} \rangle = 0.433273$. A full circle corresponds to the data obtained with the length gauge, an empty square corresponds to the velocity gauge, and an empty triangle corresponds to the mixed gauge. Markers in red correspond to transitions assigned using the Exact Band Limits, while markers in green correspond to transitions assigned using the Improved Fit algorithm. The data displayed can be found in Tables S107 to S118 of the SI document.

Figure 6 compares the OS errors relative to f_{exp}, $n \cdot f_{exp}$, and $C \cdot f_{exp}$ for a series of 9 TD-DFT functionals (one pure, five hybrid, and three long-range corrected functionals). The values of C for each gauge and each functional are shown in Table 3. The green background in Table 3 indicates that in all cases, convergence was achieved regardless of the starting value of C (0.7, 1.0, or 1.4). All functionals behave consistently with TD-B3LYP and overestimate the OS relative to f_{exp}. A weak gauge-dependence is observed in all cases. On the other hand, the TDA equivalents, shown in Table S275 to S292 and Figure S5 of the supporting information, display much stronger gauge-dependence consistent with TDA-B3LYP.

Hybrid functionals without long-range corrections give an optimal C value in the range of 1.25-1.42. the pure functional SVWN gives an optimal C value in the range of 1.11-1.15. Long-range functionals give optimal C values in the range 1.47-1.58. Most of those factors, especially for the hybrid functionals, are close to the refractive index of solvents, so the agreement with $n \cdot f_{exp}$ is usually better than the agreement with just f_{exp}, with the

exception of SVWN. B3LYP still gives the best agreement with $n \cdot f_{exp}$, although a few other hybrid functionals are close.

Table 3: C values obtained according to Algorithm 1 for the nine functionals considered

Method	**X-gauge**	**P-gauge**	**XP-gauge**
B3P86	1.34	1.29	1.31
CAM-B3LYP	1.54	1.49	1.52
LC-wHPBE	1.55	1.47	1.54
M05	1.39	1.38	1.38
mPW1PW91	1.42	1.37	1.39
O3LYP	1.27	1.25	1.27
SVWN	1.15	1.11	1.14
wB97XD	1.58	1.53	1.56
B3LYP	1.34	1.30	1.31

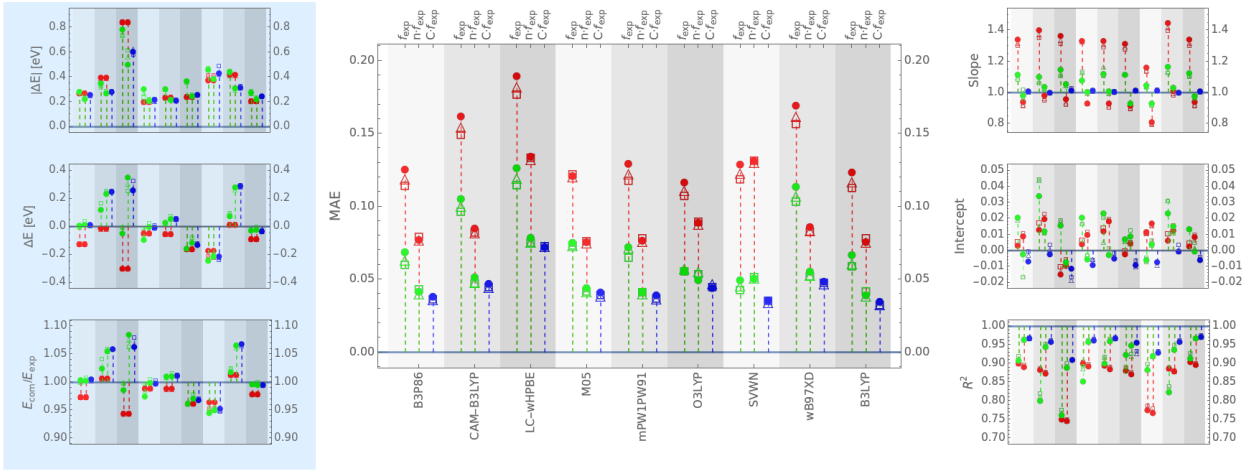

Figure 6: Comparison of f-values computed using different density functionals to a subset of 85 experimental transitions. For each method the f_{comp} values are compared to f_{exp} (left), $n \cdot f_{exp}$ (center), and $C \cdot f_{exp}$ (right). For reference, the average value of the experimental f-values are $\langle f_{exp} \rangle = 0.307655$, and $\langle n \cdot f_{exp} \rangle = 0.433273$. A full circle corresponds to the data obtained with the length gauge, an empty square corresponds to the velocity gauge, and an empty triangle corresponds to the mixed gauge. Markers in red correspond to transitions assigned using the Exact Band Limits, while markers in green correspond to transitions assigned using the Improved Fit algorithm. Blue markers are used for the iterative comparison to $C \cdot f_{exp}$ The data displayed can be found in Tables S119 to S133 of the SI document.

Out of the 85 transitions that belong to the VHHM subset, 43 come from molecules whose ground state symmetry point group is C_1, 30 come from molecules of point group C_s, and 12 come from molecules of point groups of higher symmetry (one from D_{2h}, two from

C_{2h}, two from D_2, four from C_{2v}, and three from C_2). Of the 85 B3LYP transitions, 54 have $\pi\pi^*$ character, 7 are $\pi\pi^*$ with significant charge transfer character, and 21 are mixed, containing Rydberg or diffuse character in addition $\pi\pi^*$. In the reference experimental data, 32 of the 85 transitions were measure in non-polar solvent, 50 in protic solvents, and 3 in polar aprotic solvents. We carry out further statistical analyses on these subsets of data in Supporting Information Figures S3 and S4 and find that the trends observed overall for the 85 transitions are largely reproduced by all the subsets if they have a large enough sample size. In other words, we do not identify significantly different trends for molecules of different symmetry, excitation character, or solvent polarity.

Conclusions

In a previous study, Tarleton et al. derived experimental oscillator strengths from well defined UV-visible absorption spectral peaks of 100 molecules in solution.[12] Here, we use a subset of transitions identified as having reliable experimental strengths, based on the reproducibility and quality of their deconvolution and having little overlap with other peaks, to further benchmark several wave function methods, density functionals, basis sets, transition dipole gauges (length, velocity, and mixed), and solvent corrections. A band-matching algorithm is used to assign computed transitions to experimental peaks.

Large errors and gauge-dependence were observed and quantified in oscillator strengths computed with CIS or TD-DFT paired with the Tamm-Dancoff approximation (TDA). These theories, which do not satisfy the Thomas–Reiche–Kuhn sum rule, gave oscillator strengths that do not match well with the experimental data. Linear response methods like TD-DFT and TD-HF (RPA) showed much smaller gauge dependence. TD-DFT calculations resulted in mean errors that are less than half of those observed in the best TDA cases.

The size of the molecules (average molecule weight = 160 g/mol for the molecules included in the 85 VHHM transitions) made systematic calculations using high-level wave function

methods and large basis sets intractable. Instead, we opted to run EOM-CCSD and LR-CCSD calculations with the aug-cc-pVDZ basis set on a subset of 35 transitions. Overall, EOM-CCSD calculations also exhibited a strong gauge-dependence which was only slightly reduced with LR-CCSD.

In general, we find that an increase in the size of the basis set resulted both in smaller gauge-dependence and smaller mean errors relative to the experimental data.

Several functionals were benchmarked in addition to TD-B3LYP. In all cases, the oscillator strengths were overestimated relative to the experiments, but the degree of this overestimation depends on the class of functional used. A pure functional only overestimated f_{exp} by a factor of around 1.1, while hybrid functionals had a larger factor ranging from 1.25 to 1.4. Long range corrected functionals gave the largest factor relative to f_{exp}, up to 1.58. The EOM-CCSD and LR-CCSD in the length gauge also overestimate the data by a similar factor as hybrid functionals, close to 1.3.

The systematic overestimation of most computational methods compared to f_{exp} is consistent with the refractive index factor that appears in the denominator of several theoretical solvent effect corrections. This factor arises from the energy flux of the radiation field in a dielectric (eq. 31). Through testing several simple cavity field corrections, we find that factors derived using a spherical cavity do not improve the agreement between computations and experiments. As highlighted by several studies,[48,56] a suitable cavity field correction can be obtained by using a cavity shaped to the dimensions of the molecule and that considers the direction of the transition dipole moment relative to that cavity. While these effects can be explored further, in the meantime, we find that simply multiplying the experimental oscillator strength by the solvent refractive index, which is equivalent to assuming that the cavity field acting on the molecules is equal to the macroscopic (averaged) field, gives a reasonably good agreement with computed oscillator strengths for TD-DFT/PCM methods, especially when using a hybrid functional. For example, in the case of TD-B3LYP, the error when comparing f_{comp} and $n \cdot f_{exp}$ is on the order of 0.02, which is near 9-12 %, depending

on how band assignments are made, of the actual magnitude of the experimental strength.

Acknowledgement

This material is based upon work supported by the National Science Foundation (NSF) under Grant CHE-2047667 (S.G.). J.C.G.A. acknowledges a fellowship from the Molecular Basis of Disease Program at Georgia State University. This work used Expanse at SDSC through allocation CHE180027 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. We also acknowledge the use of Advanced Research Computing Technology and Innovation Core (ARCTIC) resources at Georgia State University's Research Solutions, made available by the NSF Major Research Instrumentation (MRI) grant number CNS-1920024.

Supporting Information Available

Tables with the numerical values for the figures shown in the Results and Discussion section of this work; Table and Figure of Cavity Field Correction results; Table of results for all 164 transitions with 9 functionals and 6-311++G** ; Table and Figure with TDA-DFT calculations for 9 functionals with 6-311++G** basis set ; Tables and Figures summarizing the results of oscillator strength calculations for different subsets of the data related to symmetry, transition character, solvent, and spectrophotometer used (PDF)

Structures of the 100 molecules from Tarleton et al. re-optimized with the 6-311++G** basis set (ZIP)

References

(1) Laurent, A. D.; Jacquemin, D. TD-DFT benchmarks: A review. *International Journal of Quantum Chemistry* **2013**, *113*, 2019–2039.

(2) Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *Journal of Chemical Theory and Computation* **2009**, *5*, 2420–2435, PMID: 26616623.

(3) Loos, P.-F.; Galland, N.; Jacquemin, D. Theoretical 0–0 Energies with Chemical Accuracy. *The Journal of Physical Chemistry Letters* **2018**, *9*, 4646–4651, PMID: 30063359.

(4) Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction. *The Journal of Chemical Physics* **2008**, *129*, 104103.

(5) Schreiber, M.; Silva-Junior, M. R.; Sauer, S.; Thiel, W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *The Journal of chemical physics* **2008**, *128*.

(6) Caricato, M.; Trucks, G. W.; Frisch, M. J.; Wiberg, K. B. Oscillator Strength: How Does TDDFT Compare to EOM-CCSD? *Journal of Chemical Theory and Computation* **2011**, *7*, 456–466, PMID: 26596165.

(7) Chrayteh, A.; Blondel, A.; Loos, P.-F.; Jacquemin, D. Mountaineering Strategy to Excited States: Highly Accurate Oscillator Strengths and Dipole Moments of Small Molecules. *Journal of Chemical Theory and Computation* **2021**, *17*, 416–438, PMID: 33256412.

(8) Jacquemin, D.; Duchemin, I.; Blondel, A.; Blase, X. Assessment of the Accuracy of

the Bethe–Salpeter (BSE/GW) Oscillator Strengths. *Journal of Chemical Theory and Computation* **2016**, *12*, 3969–3981, PMID: 27403612.

(9) Labhart, H. Zur quantitativen beschreibung des einflusses von substituenten auf das absorptionsspektrum ebener molekeln. Anwendung auf anthrachinon. *Helvetica Chimica Acta* **1957**, *40*, 1410–1420.

(10) Tawada, Y.; Tsuneda, T.; Yanagisawa, S.; Yanai, T.; Hirao, K. A long-range-corrected time-dependent density functional theory. *The Journal of Chemical Physics* **2004**, *120*, 8425–8433.

(11) Miura, M.; Aoki, Y.; Champagne, B. Assessment of time-dependent density functional schemes for computing the oscillator strengths of benzene, phenol, aniline, and fluorobenzene. *The Journal of Chemical Physics* **2007**, *127*, 084103.

(12) Tarleton, A. S.; Garcia-Alvarez, J. C.; Wynn, A.; Awbrey, C. M.; Roberts, T. P.; Gozem, S. OS100: A Benchmark Set of 100 Digitized UV–Visible Spectra and Derived Experimental Oscillator Strengths. *The Journal of Physical Chemistry A* **2022**, *126*, 435–443, PMID: 35015532.

(13) Chan, W. F.; Cooper, G.; Brion, C. E. Absolute optical oscillator strengths for the electronic excitation of atoms at high resolution: Experimental methods and measurements for helium. *Phys. Rev. A* **1991**, *44*, 186–204.

(14) Pastore, M.; Mosconi, E.; De Angelis, F.; Grätzel, M. A Computational Investigation of Organic Dyes for Dye-Sensitized Solar Cells: Benchmark, Strategies, and Open Issues. *The Journal of Physical Chemistry C* **2010**, *114*, 7205–7212.

(15) Michl, J.; Thulstrup, E. W. *Spectroscopy with polarized light : solute alignment by photoselection, in liquid crystals, polymers, and membranes*; VCH: Deerfield Beach, FL, USA, 1986.

(16) *Molecular Fluorescence*; John Wiley & Sons, Ltd, 2012; Chapter 2, pp 31–51.

(17) Hills, M. E.; Olsen, A. L.; Nichols, L. W. Polarization in Cary Model 14 Spectrophotometers and Its Effect onTransmittance Measurements of Anisotropic Materials. *Appl. Opt.* **1968**, *7*, 1437–1441.

(18) Braslavsky, S. E. Glossary of terms used in photochemistry, 3rd edition (IUPAC Recommendations 2006). *Pure and Applied Chemistry* **2007**, *79*, 293–465.

(19) Berberan-Santos, M. N. Beer's law revisited. *Journal of Chemical Education* **1990**, *67*, 757.

(20) Jackson, J. D. *Classical Electrodynamics, 3rd Edition*; 1998.

(21) Ref. 20, Appendix.

(22) Foster, E. W. The measurement of oscillator strengths in atomic spectra. *Reports on Progress in Physics* **1964**, *27*, 469–551.

(23) Marlow, W. C. Hakenmethode. *Appl. Opt.* **1967**, *6*, 1715–1724.

(24) Huber, M. C. E.; Sandeman, R. J. The measurement of oscillator strengths. *Reports on Progress in Physics* **1986**, *49*, 397–490.

(25) Ref. 20, Chapter 4.

(26) Ref. 20, Chapter 7.

(27) Ladenburg, R. Die quantentheoretische Deutung der Zahl der Dispersionselektronen. *Zeitschrift für Physik* **1921**, *4*, 451–468.

(28) Tatum, J. B. The Interpretation of Intensities in Diatomic Molecular Spectra. *Astrophysical Journal Supplement* **1967**, *14*, 21.

(29) Thorne, A. P.; Litzén, U.; Johansson, S. S. *Spectrophysics : principles and applications*; Springer: Berlin ;, 1999.

(30) Cohen-Tannoudji, C.; Laloë, F.; Diu, B. *Quantum Mechanics.*; Quantum Mechanics Vol. 2; John Wiley & Sons, Inc. [US], 1977; pp 1303–1368.

(31) DAVYDOV, A. In *Quantum Mechanics (Second Edition)*, second edition ed.; DAVYDOV, A., Ed.; International Series in Natural Philosophy; Pergamon, 1965; Vol. 1; pp 388–434.

(32) Sakurai, J. J. *Modern quantum mechanics; rev. ed.*; Addison-Wesley: Reading, MA, 1994.

(33) Peleg, Y.; Pnini, R.; Zaarur, E.; Hecht, E. *Schaum's Outline of Quantum Mechanics, Second Edition*, 2nd ed.; McGraw-Hill Education: New York, 2010.

(34) Andrews, S. S. Using Rotational Averaging To Calculate the Bulk Response of Isotropic and Anisotropic Samples from Molecular Parameters. *Journal of Chemical Education* **2004**, *81*, 877.

(35) Crossley, R. In *The Calculation of Atomic Transition Probabilities*; Bates, D., Estermann, I., Eds.; Advances in Atomic and Molecular Physics; Academic Press, 1969; Vol. 5; pp 237–296.

(36) Hansen, A. E.; Bouman, T. D. *Advances in Chemical Physics*; John Wiley & Sons, Ltd, 1980; pp 545–644.

(37) Open-Shell and Excited-State Methods, Q-Chem 6.2 User's Manual. `https://manual.q-chem.com/latest/sec_EOMGRAD.html`.

(38) Weigang, J., Oscar E. Solvent Field Corrections for Electric Dipole and Rotatory Strengths. *The Journal of Chemical Physics* **1964**, *41*, 1435–1441.

(39) Abe, T. Comments on "The effect of solvent environment on molecular electronic oscillator strengths". *The Journal of Chemical Physics* **1982**, *77*, 1074–1074.

(40) Onsager, L. Electric Moments of Molecules in Liquids. *Journal of the American Chemical Society* **1936**, *58*, 1486–1493.

(41) Herbert, J. M. Dielectric continuum methods for quantum chemistry. *WIREs Computational Molecular Science* **2021**, *11*, e1519.

(42) Mennucci, B. Polarizable continuum model. *WIREs Computational Molecular Science* **2012**, *2*, 386–404.

(43) Lorentz, H. *The theory of electrons*; Teubner, 1909.

(44) Fröhlich, H. *Theory of dielectrics; dielectric constant and dielectric loss*; Monographs on the physics and chemistry of materials; Clarendon Press Oxford: Oxford, 1949.

(45) Kirkwood, J. G. The Dielectric Polarization of Polar Liquids. *The Journal of Chemical Physics* **1939**, *7*, 911–919.

(46) Kirkwood, J. G. The influence of hindered molecular rotation on the dielectric polarisation of polar liquids. *Trans. Faraday Soc.* **1946**, *42*, A007–A012.

(47) Scholte, T. A contribution to the theory of the dielectric constant of polar liquids. *Physica* **1949**, *15*, 437–449.

(48) Shibuya, T. A dielectric model for the solvent effect on the intensity of light absorption. *The Journal of Chemical Physics* **1983**, *78*, 5175–5182.

(49) Chako, N. Q. Absorption of Light in Organic Compounds. *The Journal of Chemical Physics* **1934**, *2*, 644–653.

(50) Myers, A. B.; Birge, R. R. The effect of solvent environment on molecular electronic oscillator strengths. *The Journal of Chemical Physics* **1980**, *73*, 5314–5321.

(51) Böttcher, C. Zur Theorie Der Inneren Elektrischen Feldstärke. *Physica* **1942**, *9*, 937–944.

(52) Schuyer, J. The influence of the refractive index on the absorption of light by solutions. *Recueil des Travaux Chimiques des Pays-Bas* **1953**, *72*, 933–949.

(53) Osborn, J. A. Demagnetizing Factors of the General Ellipsoid. *Phys. Rev.* **1945**, *67*, 351–357.

(54) Abe, T. Theory of solvent effects on oscillator strengths for molecular electronic transitions. *Bulletin of the Chemical Society of Japan* **1970**, *43*, 625–628.

(55) Warner, J. W.; Wolfsberg, M. Dielectric effects on the spectra of condensed phases. *The Journal of Chemical Physics* **1983**, *78*, 1722–1730.

(56) Gil, G.; Pipolo, S.; Delgado, A.; Rozzi, C. A.; Corni, S. Nonequilibrium Solvent Polarization Effects in Real-Time Electronic Dynamics of Solute Molecules Subject to Time-Dependent Electric Fields: A New Feature of the Polarizable Continuum Model. *Journal of Chemical Theory and Computation* **2019**, *15*, 2306–2319, PMID: 30860829.

(57) Novotny, L.; Hecht, B. *Principles of Nano-Optics*, 2nd ed.; Cambridge University Press, 2012.

(58) Cammi, R.; Mennucci, B.; Tomasi, J. On the Calculation of Local Field Factors for Microscopic Static Hyperpolarizabilities of Molecules in Solution with the Aid of Quantum-Mechanical Methods. *The Journal of Physical Chemistry A* **1998**, *102*, 870–875.

(59) Pipolo, S.; Corni, S.; Cammi, R. The cavity electromagnetic field within the polarizable continuum model of solvation. *The Journal of Chemical Physics* **2014**, *140*, 164114.

(60) Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics* **1993**, *98*, 1372–1377.

(61) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **1988**, *37*, 785.

(62) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics* **1972**, *56*, 2257–2261.

(63) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **1965**, *140*, A1133.

(64) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of physics* **1980**, *58*, 1200–1211.

(65) Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical review B* **1986**, *33*, 8822.

(66) Cohen, A. J.; Handy, N. C. Dynamic correlation. *Molecular Physics* **2001**, *99*, 607–615.

(67) Adamo, C.; Barone, V. Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The m PW and m PW1PW models. *The Journal of chemical physics* **1998**, *108*, 664–675.

(68) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Exchange-correlation functional with broad accuracy for metallic and nonmetallic compounds, kinetics, and noncovalent interactions. *The Journal of chemical physics* **2005**, *123*.

(69) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical physics letters* **2004**, *393*, 51–57.

(70) Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *The Journal of chemical physics* **2006**, *125*.

(71) Henderson, T. M.; Izmaylov, A. F.; Scalmani, G.; Scuseria, G. E. Can short-range hybrids describe long-range-dependent properties? *The Journal of chemical physics* **2009**, *131*.

(72) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(73) Scalmani, G.; Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *The Journal of chemical physics* **2010**, *132*.

(74) Hirata, S.; Head-Gordon, M. Time-dependent density functional theory within the Tamm–Dancoff approximation. *Chemical Physics Letters* **1999**, *314*, 291–299.

(75) Dunning, J., Thom H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **1989**, *90*, 1007–1023.

(76) Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *The Journal of Chemical Physics* **1969**, *51*, 2657–2664.

(77) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of chemical physics* **1992**, *96*, 6796–6806.

(78) Frisch, M. J. et al. Gaussian˜16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(79) Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Molecular Physics* **2015**, *113*, 184–215.

(80) Koch, H.; Jo/rgensen, P. Coupled cluster response functions. *The Journal of Chemical Physics* **1990**, *93*, 3333–3344.

(81) Koch, H.; Kobayashi, R.; Sanchez de Merás, A.; Jo/rgensen, P. Calculation of size-intensive transition moments from the coupled cluster singles and doubles linear response function. *The Journal of Chemical Physics* **1994**, *100*, 4393–4400.

(82) Kállay, M.; Gauss, J. Calculation of excited-state properties using general coupled-cluster and configuration-interaction models. *The Journal of Chemical Physics* **2004**, *121*, 9257–9269.

(83) Stanton, J. F.; Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *The Journal of chemical physics* **1993**, *98*, 7029–7039.

(84) Krylov, A. I. Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker's guide to Fock space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.

(85) Staff, P. S. G. E. *UV Atlas of Organic Compounds*; Springer US, 1967.

(86) Daimon, M.; Masumura, A. Measurement of the refractive index of distilled water from the near-infrared region to the ultraviolet region. *Appl. Opt.* **2007**, *46*, 3811–3820.

(87) Rheims, J.; Köser, J.; Wriedt, T. Refractive-index measurements in the near-IR using an Abbe refractometer. *Measurement Science and Technology* **1997**, *8*, 601.

(88) Moutzouris, K.; Papamichael, M.; Betsis, S. C.; Stavrakas, I.; Hloupis, G.; Triantis, D. Refractive, dispersive and thermo-optic properties of twelve organic solvents in the visible and near-infrared. *Applied Physics B* **2014**, *116*, 617–622.

(89) Kozma, I. Z.; Krok, P.; Riedle, E. Direct measurement of the group-velocity mismatch and derivation of the refractive-index dispersion for a variety of solvents in the ultraviolet. *J. Opt. Soc. Am. B* **2005**, *22*, 1479–1485.

(90) Kerl, K.; Varchmin, H. Refractive index dispersion (RID) of some liquids in the UV/VIS between 20°C and 60°C. *Journal of Molecular Structure* **1995**, *349*, 257–260, Molecular Spectroscopy and Molecular Structure 1994.

(91) Polyanskiy, M. N. Refractiveindex.info database of optical constants. *Scientific Data* **2024**, *11*, 94.

(92) Polyanskiy, M. RefractiveIndex.INFO - Refractive index database — refractiveindex.info. `https://refractiveindex.info`, [Accessed 22-03-2024].

(93) Winget, P.; Dolney, D. M.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Minnesota solvent descriptor database. *Minneapolis, MN: Department of Chemistry and Supercomputer Institute* **1999**,

(94) Dreuw, A.; Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chemical Reviews* **2005**, *105*, 4009–4037, PMID: 16277369.

(95) Rishi, V.; Perera, A.; Nooijen, M.; Bartlett, R. J. Excited states from modified coupled cluster methods: Are they any better than EOM CCSD? *The Journal of Chemical Physics* **2017**, *146*.

(96) Acharya, A.; Chaudhuri, S.; Batista, V. S. Can TDDFT describe excited electronic states of naphthol photoacids? A closer look with EOM-CCSD. *Journal of Chemical Theory and Computation* **2018**, *14*, 867–876.

(97) Thomas, W. Über die Zahl der Dispersionselektronen, die einem stationären Zustande zugeordnet sind.(Vorläufige Mitteilung). *Naturwissenschaften* **1925**, *13*, 627–627.

(98) Reiche, F.; Thomas, W. Über die Zahl der Dispersionselektronen, die einem stationären Zustand zugeordnet sind. *Zeitschrift für Physik* **1925**, *34*, 510–525.

(99) Kuhn, W. Über die Gesamtstärke der von einem Zustande ausgehenden Absorptionslinien. *Zeitschrift für Physik* **1925**, *33*, 408–412.

(100) Casida, M. E.; Huix-Rotllant, M. Progress in time-dependent density-functional theory. *Annual review of physical chemistry* **2012**, *63*, 287–323.

(101) Bauschlicher, C. W.; Langhoff, S. R. Computation of electronic transition moments: the length versus the velocity representation. *Theoretica chimica acta* **1991**, *79*, 93–103.

TOC Graphic