JCTC Journal of Chemical Theory and Computation

Toward a Theoretical Quantitative Estimation of the λ_{max} of Anthraquinones-Based Dyes

Eric A. Perpète,*[,†,‡] Valerie Wathelet,[‡] Julien Preat,[‡] Christophe Lambert,[§] and Denis Jacquemin*[,†,‡]

Laboratoire de Chimie Théorique Appliquée, Facultés Universitaires Notre-Dame de la Paix, rue de Bruxelles, 61, B-5000 Namur, Belgium, and BioXpr, Centre technologique FUNDP, Rue du séminaire, 22, B-5000 Namur, Belgium

Received November 23, 2005

Abstract: We have computed the absorption spectra of a large series of anthraquinone dyes by using the time-dependent density functional theory (TD-DFT) for the excited-state calculations and the polarizable continuum model (PCM) for evaluating bulk solvent effects. On one hand, we compare the results obtained with the B3LYP and the PBE0 hybrid functionals, combined with different atomic basis sets. On the other hand, using multiple linear regression, we take advantage of the λ_{max} predicted by these two functionals in order to reach the best agreement between theoretical estimates and experimental measurements. It turns out that 1. PBE0 provides more accurate results than B3LYP; in addition the average errors provided by the former are less basis set dependent. 2. Multiple linear regression provides excited state spectra in better agreement with experiment than any simple linear fit that could be performed. 3. Using our best fitting procedure, we obtained a mean absolute error of 6 nm for a set of 66 anthraquinones, with no deviations exceeding 25 nm. The related standard deviation, useful for predictions, is only 8 nm, i.e., $\lambda_{max}^{theo} = \lambda_{max}^{exp} \pm 8$ nm (or ± 0.05 eV) for unknown anthraquinone compounds.

I. Introduction

Today, the molecular modelization techniques offer a competitive alternative for the interpretation of experimental data arising from both academic and industrial measurements. Schäfer recently stated that, in the majority of cases, IR or Raman spectra can be accurately computed with the help of quantum mechanical methods that could be found in many computational chemistry packages.[1] However, it is not the case for UV/VIS spectra of large conjugated molecules. For instance, semiempirical methods, though especially tailored for, are often found to be lucky either inaccurate when reproducing spectral patterns or trends.[1] One of the main difficulties in determining the color of organic compounds is the astonishing accuracy of the standard human eye, which can distinguish, in some parts of the visible spectra (typically in the green region), differences of coloration corresponding to less than 1 nm λ_{max} shifts. Nevertheless, in regard to practical industrial applications, the theoretical calculations could be regarded as serious competitors to experimental approaches for developing new dyes and/or pigments if they were to deliver an estimate of the λ_{max} values within a 5−15 nm accuracy (∼0.05 eV). Such a chemical accuracy for large conjugated molecules is still a tremendous challenge for the modelization approaches. On one hand, highly correlated methods, such as EOM-CC or MR-CI, are completely out of computational reach for molecules possessing several π-electrons and used in solution. On the other hand, as stated above, semiempirical methods are not able to consistently deliver quantitative λ_{max} of dyes. For instance, Adachi and Nakamura reported large errors and poor correlation coefficients with CNDO/s and INDO/s methods for a large set of dyes.[2] In fact, the most promising scheme for systematically evaluating the color of conjugated compounds is the

* Corresponding author e-mail: denis.jacquemin@fundp.ac.be; URL: http://perso.fundp.ac.be/∼jacquemd.

† Research Associate of the Belgian National Fund for Scientific Research.

‡ Facultés Universitaires Notre-Dame de la Paix.

§ Centre technologique FUNDP.

10.1021/ct050287w CCC: $33.50
Published on Web 02/07/2006

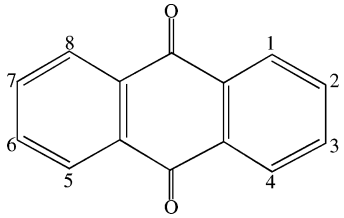

Figure 1. Sketch of anthraquinone with the numbering of substitution positions.

time-dependent density functional theory (TD-DFT).[3] Indeed, TD-DFT is often found robust and efficient for evaluating the low-lying excited spectra of conjugated molecules[4−7] and has been the subject of numerous applications.[8−19] Nevertheless, recent TD-DFT determinations of the λ_{max} of dyes or related conjugated molecules often report mean absolute errors (MAE) in the 0.1−0.4 eV range, i.e., at least twice as large as our target. Indeed, for a large set of sulfur-containing compounds, Fabian obtained MAE of 0.24 eV;[20] the same MAE can be determined for the first $\pi \rightarrow \pi^*$ singlet excitation of the 11 thiouracil for which measurements are given in ref 21. The typically reported TD-DFT error ranges from 0.2 to 0.5 eV for coumarins,[22] from 0.3 to 0.4 eV for uroanic-acid-based molecules,[23] from 0.1 and 0.3 eV for alkyl-amino-benzonitrile compounds,[24] and from 0.3 to 0.4 eV for transition-metal complexes.[25] In 2005, we have found only three published studies with (almost-)quantitative absorption spectra of several dyes. The first is due to Hommen-de-Mello and co-workers who obtained a MAE of 19 nm, for six cationic dyes in water using a PCM-ZINDO// PCM-B3LYP (PCM: Polarizable Continuum Model, see below) approach.[26] The second and third investigations, on diazonium salts and thioindigo, respectively, reported MAE of 6 and 10 nm, using PCM-PBE0 approaches combined with extended basis sets.[18,27] These comparisons highlight the fundamental importance of including solvent effects when simulating absorption spectra of organic dyes.

Two classes of chromophoric unit are principally used as industrial dyes: the N═N chromophore of azo pigments and the C═O group. The latter is present under a variety of chemical forms: coumarins, naphthaquinones, quinacridones, perinones, indigoids ... The carbonyl dyes owe their success to their ability to provide a wide range of colors covering the entire visible spectrum and to their capacity to show long wavelength absorption bands when combined with relatively short π-conjugated systems. In particular, the 9,10-anthraquinones derivatives (Figure 1), in which the central ring bearing two carbonyl groups is fused to two fully aromatic six-member rings, can give rise to a complete range of shades (especially in the green/blue region), depending on the nature and relative position(s) of the auxochromic group(s) substituting hydrogen atom(s) on the outer rings.[28,29] Consequently, anthraquinoidic derivatives represent about 30% of today's world dye production.[28]

Following our first investigation,[30] we aim at setting up an approach able to accurately predict ab initio the λ_{max} of absorption of 9,10-anthraquinones. In addition, we want to assess the basis set effects as well as the relative accuracy of two selected functionals (see below). As experimental input, we have chosen the measurements of Labhart,[31] who obtained the electronic excitation spectra of a large number of anthraquinones, in dichloromethane, with absorption wavelengths almost covering the entire visible spectrum (from 325 to 645 nm). In the Labhart set, one finds a large variety of side groups: hydroxy, amino, nitro, chloro, ... This allows consistent comparisons and meaningful statistical treatment.

This paper is divided as follows. Section 2 gives a description of the quantum-chemical and statistical tools. In section 3.1, we compare the respective accuracy of B3LYP and PBE0 functionals for evaluating the λ_{max} of anthraquinones, whereas in section 3.2, we combine both to obtain an optimal accuracy.

II. Methodology

A. Quantum-Mechanical Calculations. We have chosen the Gaussian03[32] package of programs to perform the geometry optimizations, vibrational analysis, and excited-state evaluations.

The ground-state geometry of each molecule has been fully optimized until the RMS residual force is smaller than 1×10^{-5} au (TIGHT threshold in Gaussian). For anthraquinones, it turns out that the B3LYP[33] functional gives geometries in good agreement with second-order Møller−Plesset structures;[30] B3LYP has therefore been selected. In this functional the exchange is a combination of 20% HF exchange, Slater functional, and Becke's GGA correction,[34] whereas the correlation part combines VWN and LYP[35] functionals. Following each optimization, the vibrational spectrum has been determined at the same level of theory, and it has been systematically checked that all vibrational frequencies are real.

TD-DFT[3] methodology is then used to compute the low-lying excited states of anthraquinone derivatives. We have used two hybrid functionals: B3LYP and PBE0.[36,37] PBE0 is built on the Perdew-Burke-Ernzerhof pure functional,[38] in which the exchange is weighted (75% DFT/25% HF) accordingly to purely theoretical considerations.[39] As expected for this type of dye, the electronic excitation responsible for the color of anthraquinone presents a typical $\pi \rightarrow \pi^*$ character often associated with a large oscillator force. Our theoretical λ_{max} are always related to a transition toward the first singlet excited state, except for some of the dyes absorbing in the 320−350 nm region, for which the experimentally reported λ_{max} corresponds to a higher excited state (but the oscillator strength toward the lower-lying excited states is negligible in that case).

In ref 30, we show that the solvent effects on the ground-state geometry are negligible due to the rigidity of the anthraquinone core but are sizable for transition energies. Therefore, the bulk solvent effects are evaluated during the TD-DFT calculations by means of the standard Polarizable Continuum Model (IEF-PCM).[40,41] In PCM, one divides the problem into a solute part (anthraquinone) lying inside a cavity and a solvent part (in this case, dichloromethane) represented as a structureless material, characterized by its macroscopic properties (dielectric constant, radius, density, molecular volume, ...). PCM is able to obtain a valid approximation of solvent effects as long as no specific interactions (such as hydrogen bonds, ion pairing, ...) link

the solute and the solvent molecules. Because we study UV/Vis spectra, we have selected the so-called nonequilibrium PCM solutions.[41] Indeed the absorption process presents a short characteristic time. Therefore, only the solvent electronic distribution can "adapt" to the new (excited) electronic structure of the solute, while molecular motions of the solvent are frozen during the process.[41]

We use two different atomic basis set combinations in this study. In the less demanding approach (**M-I**), only the Pople's polarized double-ζ basis set, 6-31G(d,p), is selected as our methodological study shows that this basis set could be sufficient.[30] This means that **M-I** corresponds to a PCM-TD-[B3LYP/PBE0]/6-31G(d,p)//B3LYP/6-31G(d,p) approach. In the second method (**M-II**), we use the atomic basis sets often recommended for UV/vis investigations with TD-DFT,[42] i.e., a triple-ζ basis set for the geometry, with additional diffuse functions for the excitation spectra, i.e., **M-II** is PCM-TD-[B3LYP/PBE0]/6-311++G(d,p)//B3LYP/6-311G(d,p).

B. Statistical Treatment. To reach the best agreement between theory and experiment, the results of different approaches can be advantageously combined. To obtain the most efficient combination, the Multiple Linear Regression (MLR),[43−45] which is based on the numerical technique of least-squares fitting and analyzes the relationship between one dependent variable (experimental value) and one or more independent variables (theoretical values), is a method of choice. MLR is a tool for determining the (experimental) property, *Y*, as a function of *p* independent (theoretical) variables (*x*):

$$Y = b_0 + b_1x_1 + b_2x_2 + ... + b_px_p + R \quad (1)$$

To test the significance of a regression curve, the total sum of squares (TSS) is split into two components, the model sum of squares (MSS) and the residual sum of squares (RSS)

$$\mathrm{TSS} = \mathrm{MSS} + \mathrm{RSS} \quad (2)$$

$$\sum_{i=1}^{n}[y_i - \bar{y}]^2 = \sum_{i=1}^{n}[y(x_i) - \bar{y}]^2 + \sum_{i=1}^{n}[y_i - y(x_i)]^2 \quad (3)$$

with y_i the experimental value, $y(x_i)$ the regression value, $\bar{y}$ the average, and *n* the number of points (number of dyes considered). The correlation coefficient reads

$$R^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{MSS}} \quad (4)$$

If the fitted curve passes through all the original data points, the MSS is equal to the TSS and the RSS is zero. In that case $R^2 = 1$. To test the significance of the regression, test calculations are carried out in a so-called analysis of the variance table (ANOVA), where the mean squares MMS and RMS are obtained by reporting MSS to the number of independent variables, and dividing RSS by *n*-*p*-1, respectively. Indeed, R^2 could be abnormally large if there are numerous descriptors (i.e. large *p*) but a few data points (i.e. small *n*). Therefore, one uses an adjusted coefficient

$$R^2_{\mathrm{adj}} = 1 - \frac{n-1}{n-p-1}\left(1 - \frac{\mathrm{RSS}}{\mathrm{MSS}}\right) \quad (5)$$

If the MMS/RMS ratio is significantly large, the null hypothesis may be rejected and the regression is meaningful. Confidence limits for the regression parameters b_i, measuring the adequacy of each independent variables in the model, are also determined. The ratio between b_i and the associated error could be compared to the critical values for which the probability (*P*-value) to obtain a regression coefficient by chance has been tabulated. If necessary, this treatment allows for eliminating step-by-step the less significant independent variables. MLR provides not only the usual mean absolute error (MAE) but also a standard deviation, d_R, computed as

$$d_R = \sqrt{\frac{\mathrm{RSS}}{n-p-1}} \quad (6)$$

d_R is useful for the prediction of the properties of compounds not included in the training set. In the present study, MLR has been performed with the Statgraphics Plus 5.1. program.[46]

III. Results and Discussion

A. Comparison between B3LYP and PBE0. The computed λ_{max} for 66 anthraquinones are reported (in nm) in Table 1 and are compared to the experimental data taken in ref 31. Before using multilinear regression, it is worth evaluating the performance of the two functionals using "raw" values directly extracted form TD-DFT calculations. When using the most accurate theoretical level (**M-II**), we obtain, for B3LYP a MAE of 20 nm (RMS of 25 nm), whereas for PBE0 the MAE is 14 nm (RMS of 18 nm). The corresponding MAE (RMS) in eV are 0.12 (0.16) and 0.08 (0.10) for B3LYP and PBE0, respectively. Thus, both functionals are quite efficient for anthraquinones: the MAE are clearly in the lower range of the expected TD-DFT deviations (see Introduction). This is well illustrated in Figure 2, where the qualitative and quantitative agreements between theory and experiment is striking. In general, B3LYP slightly overshoots the λ_{max} of dyes with large excitation energies, PBE0 presenting the (completely) opposite behavior.

In addition, one can state that PBE0 is statistically more efficient than B3LYP for evaluating the UV/Vis spectra of anthraquinones at a 99% confidence level. This assertion is confirmed by the extreme deviations that are smaller with PBE0 functional: +63/−32 nm for B3LYP and +38/-45 nm for PBE0. Using the less demanding computational scheme, **M-I**, the errors calculated with PBE0 are almost unchanged: MAE = 13 nm (0.08 eV) and RMS = 17 nm (0.09 eV) (in fact there is a 30% probability that the basis set modification does not statistically alter the results obtained with the PBE0 functional). The same is true for the largest deviations: +33/−49 nm. This illustrates that the PBE0 λ_{max} are already (almost) converged with 6-31G(d,p) basis set. Using **M-II** does not reduce the average errors and is therefore quite useless. Of course, this is a general conclusion, and an individual compound might be significantly affected by the basis set change (NMe_2 groups for instance). Nevertheless, the absolute average change when shifting from **M-I** to **M-II** is small: 4 nm (0.03 eV) with PBE0, i.e., significantly smaller than the theory-experiment discrepancies.

For B3LYP, the changes induced by the basis set effects are larger. Indeed, with **M-I** one obtains a MAE of 16 nm

Table 1. Comparison between the Experimental and Theoretical λ_{max} for Anthraquinones in CH_2Cl_2[a]

compound		M-I			M-II			
substitution	no.	B3LYP	PBE0	MLR	B3LYP	PBE0	MLR	expt
2-F	13	327.7	316.4	324.1	334.7	322.1	321.7	325
	47	327.2	315.8	322.9	333.1	320.8	321.4	327
2-Cl	8	331.0	319.4	326.2	336.7	323.8	322.5	330
2,3-Cl	12	335.4	323.6	330.1	339.7	327.1	327.5	330
2,3-Br	15	337.2	324.9	329.3	343.3	329.5	325.6	330
2,6-Cl	72	334.1	321.8	325.7	341.1	327.5	324.1	330
2,7-Cl	73	322.7	311.5	318.9	328.0	316.0	317.0	330
1-NO_2,4-Cl	44	351.6	338.0	338.5	356.5	341.9	336.9	335
1-Cl	7	346.6	334.0	338.5	355.4	341.3	338.1	337
1,8-Cl	11	353.9	340.7	343.5	360.9	346.4	342.5	344
1,5-Cl	9	358.1	344.4	343.5	364.3	349.3	343.9	347
1,4-Cl	10	359.6	346.4	350.1	368.7	353.0	345.5	350
2-OMe	69	398.8	382.8	379.4	406.6	394.6	408.3	363
2-OH	68	395.5	379.7	376.7	400.1	383.6	377.9	365
1-OMe	6	393.9	379.7	384.0	402.1	386.5	384.8	380
1,8-OMe	35	405.6	390.1	390.1	412.5	395.6	390.3	385
1-OH	1	410.9	396.4	402.0	411.2	396.6	400.5	405
1-Cl,2-NH_2	61	449.9	430.8	420.7	458.5	438.2	426.5	405
2-NH_2,3-Br	63	446.3	426.6	413.1	454.6	433.4	417.3	406
1-NO_2,2-NH_2	55	462.3	439.9	416.0	470.9	446.0	417.5	410
1-NHCOMe	65	430.7	415.0	418.0	430.9	414.9	416.2	410
2-NH_2	17	455.1	435.3	422.7	464.8	443.4	428.2	410
2-NH_2,3-Cl	62	444.0	424.7	412.7	453.7	432.9	418.3	414
1-NHCOPh	48	434.1	417.2	415.1	436.2	418.8	415.3	415
1,2-OH	5	439.9	422.8	420.6	437.9	421.0	419.8	416
2-NH_2,3-NO2	56	433.5	417.3	418.4	442.6	425.0	421.7	420
1,5-OH	3	433.2	417.8	422.7	432.9	417.1	419.6	428
1,8-OH	4	435.4	420.2	426.4	437.7	420.1	416.0	430
1-SMe	36	449.8	431.0	422.3	469.0	446.2	426.0	438
2,3-NH_2	19	498.0	475.0	453.9	504.6	480.0	458.2	442
1-NH_2,4-NO_2	57	481.7	462.0	454.1	486.9	465.5	453.9	460
1-NH_2,5-OMe	59	478.3	461.1	464.5	481.2	463.2	464.5	460
1-NH_2	16	475.7	459.0	464.4	480.3	462.6	465.0	465
1-NH_2,2-Me	40	473.0	456.8	464.2	477.7	460.6	465.0	465
1-NH_2,4-Cl	37	475.9	460.0	469.2	479.4	463.0	470.6	466
1-NH_2,6-Cl	38	487.6	470.1	473.6	490.7	472.3	473.5	470
2-NMe_2	74	495.8	474.8	462.9	513.8	489.1	468.4	470
1-NH_2,2-Me,4-Br	46	476.1	460.1	468.9	478.2	461.8	469.2	473
1-NH_2,2-NHCOPh	64	476.7	459.1	460.4	480.6	462.4	462.8	475
1,4-OH	2	472.8	459.2	478.9	467.2	454.1	473.2	476
1-NH_2,6,7-Cl	39	493.1	475.6	480.0	493.9	475.8	478.8	477
1,2-NH_2	21	487.2	469.6	472.5	497.7	475.0	459.8	480
1,5-NH_2	20	485.4	468.6	475.1	490.1	472.2	475.4	480
1,4-NHCOPh	49	507.4	489.4	493.6	507.4	489.2	494.0	490
1,8-NH_2	22	512.6	493.8	495.1	517.7	497.5	495.8	492
1-NH_2,4-OMe	41	505.7	491.0	510.7	508.8	493.6	510.9	500
1-NMe_2	24	516.6	497.8	499.7	531.3	509.5	503.5	504
1-NHMe	23	500.9	483.4	489.0	509.6	490.4	491.4	508
1-NHPh	26	545.3	519.3	491.5	550.2	522.6	495.9	508
1-NHMe,4-Br	66	504.2	487.2	495.7	509.8	491.6	496.8	510
1-NO_2,4,5,8-OH	33	507.9	491.8	505.2	501.6	485.6	498.4	510
1-OH,4-NH_2	25	520.4	506.5	532.2	517.6	503.8	528.2	520
1-OH,2,4-NH_2	43	529.2	513.0	529.3	527.7	511.6	528.2	530
1-NH_2,4-NHCOPh	50	529.6	513.8	532.0	529.6	513.6	531.0	532
1-NHMe,4-OMe	32	530.1	515.1	537.2	537.1	521.2	540.2	540
1,4-NH_2	18	539.8	527.2	562.3	539.0	526.3	558.7	550
1,4-NH_2,2-OMe	60	532.2	519.0	550.0	532.5	519.2	548.1	550
1-OH,4-NHPh	28	581.2	560.4	563.0	580.6	559.1	562.3	566
1-NH_2,4-NHPh	29	585.2	568.3	590.1	586.3	569.0	590.3	590
1,5-NH_2,4,8-OH	45	577.8	563.1	594.2	574.2	559.1	587.4	590
1-NH_2,4-NHMe	31	561.6	548.9	587.0	563.7	550.7	585.9	590
1,4,5,8-NH_2	67	602.8	587.8	621.4	602.4	586.7	617.1	610
1,4-NHMe	34	583.4	570.7	612.2	588.5	575.3	613.7	620
1,4-NHPh	27	621.5	602.0	617.2	623.2	603.1	618.9	620
1-NHMe,4-NHPh	30	606.1	589.2	614.3	609.7	592.3	617.0	625
1,4-NH_2,2-NO_2	58	682.9	657.5	654.3	692.0	662.6	651.6	645

[a] Experimental λ_{max} and anthraquinone reference numbers have been taken from ref 31. The MLR λ_{max} are calculated with eqs 15 and 16. All values are in nm.

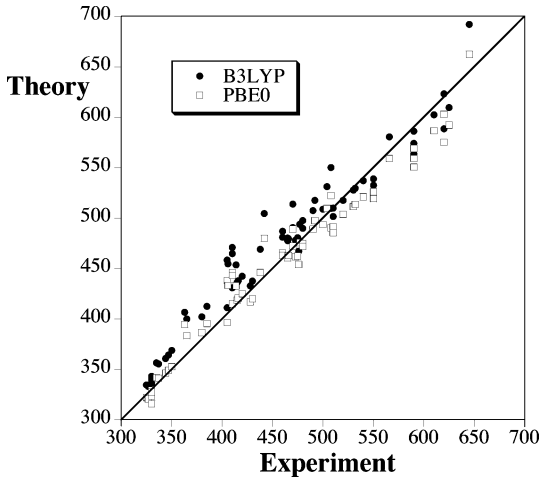

Figure 2. Comparison between the experimental and theoretical (**M-II**) λ_{max}. All values are in nm.

(0.09 eV) and a RMS of 20 nm (0.13 eV), surprisingly significantly (probability > 99%) smaller than with **M-II**. This means that, on one hand, the convergence with basis set size is slower with B3LYP than with PBE0 and, on the other hand, that the use of **M-I** leads to a "lucky" agreement in the case of B3LYP, i.e., there is some functional/basis set error compensation. If this is quite disappointing from the computational chemist point of view, this is useful in practice, as one could select less demanding methods ... to obtain more accurate λ_{max} for the average anthraquinone structure. In section 3.2, we demonstrate that **M-I** is also preferable after statistical treatment.

The substituents leading to significant auxochromic shifts belong to various chemical classes, but, in practice, the major groups used for anthraquinoidic dyes are the hydroxys (and the corresponding OR) present in naturally occurring quinones, and the amines (and NHMe, NMe_2, NHPh, NHCOPh, NHCOMe, ...) generally found in synthesized structures. Both groups are strongly electroactive and often (positions 1, 4, 5, and 8) affect the carbonyl chromophore by internal hydrogen bonds, especially strong with the NHPh groups. For the hydroxy group,[47] selecting **M-I** (**M-II**) we obtained a MAE of 10 (11) nm with PBE0, the corresponding B3LYP figures being 17 nm (22 nm). For the amino auxochroms,[47] the **M-I** (**M-II**) MAE are 18 nm (18 nm) with PBE0 and 18 nm (21) nm with B3LYP. This confirms the better behavior of the PBE0 functional with respect to basis set size. In addition, one clearly sees that the difference between the two functionals is larger when OH and OR groups are grafted to the anthraquinone core. In that case, PBE0 is clearly preferable.

B. Statistical Treatment. If one looks for a predictive tool for determining the color of anthraquinone dyes, the "raw" estimates of TD-DFT can be improved by using statistical treatment, either simple linear regression (SLR) if one uses only one functional, or MLR if the excitation energies obtained with both functionals are combined. The following SLR equations are obtained with nm units

$$\lambda_{max,nm} = -28.80 + 1.040\,\lambda_{max,nm}^{B3LYP-\mathbf{M-I}} \quad (7)$$

$$\lambda_{max,nm} = -24.18 + 1.067\,\lambda_{max,nm}^{PBE0-\mathbf{M-I}} \quad (8)$$

$$\lambda_{max,nm} = -37.64 + 1.049\,\lambda_{max,nm}^{B3LYP-\mathbf{M-II}} \quad (9)$$

$$\lambda_{max,nm} = -33.67 + 1.080\,\lambda_{max,nm}^{PBE0-\mathbf{M-II}} \quad (10)$$

The corresponding relationships in eV read

$$\lambda_{max,eV} = -0.051 + 1.045\,\lambda_{max,eV}^{B3LYP-\mathbf{M-I}} \quad (11)$$

$$\lambda_{max,eV} = -0.036 + 1.003\,\lambda_{max,eV}^{PBE0-\mathbf{M-I}} \quad (12)$$

$$\lambda_{max,eV} = -0.116 + 1.081\,\lambda_{max,eV}^{B3LYP-\mathbf{M-II}} \quad (13)$$

$$\lambda_{max,eV} = -0.098 + 1.034\,\lambda_{max,eV}^{PBE0-\mathbf{M-II}} \quad (14)$$

A statistical analysis of the results obtained with these equations is given in Table 2. PBE0 equations are always more efficient that their B3LYP counterparts, systematically giving larger R^2 and smaller MAE and d_R, as well as a slightly smaller number of extreme deviations. In addition, for equations in the more physical energetic scale, the *a* and *b* [eqs (12) and (14)] are closer to zero and one, respectively [than the corresponding (11) and (13)]. From Table 2, one directly concludes that **M-I** is more appropriate than **M-II** for estimating the λ_{max} of anthraquinone using a simple linear fit. Therefore the SLR results confirm the conclusions of section 3.1: PBE0 has to be selected for TD-DFT calculations on anthraquinone dyes. Using the 6-31G(d,p) basis set, this functional leads to a standard deviation of 15 nm, i.e., the absorption energy of dyes of the same family (but not included in our set) could be predicted with an accuracy of ±15 nm.

Using MLR, the following equations have been obtained:

$$\lambda_{max,nm} = 9.54 - 4.604\,\lambda_{max,nm}^{B3LYP-\mathbf{M-I}} + 5.762\,\lambda_{max,nm}^{PBE0-\mathbf{M-I}} \quad (15)$$

$$\lambda_{max,nm} = -3.29 - 3.922\,\lambda_{max,nm}^{B3LYP-\mathbf{M-II}} + 5.084\,\lambda_{max,nm}^{PBE0-\mathbf{M-II}} \quad (16)$$

$$\lambda_{max,eV} = 0.112 - 5.599\,\lambda_{max,eV}^{B3LYP-\mathbf{M-I}} + 6.350\,\lambda_{max,\,eV}^{PBE0-\mathbf{M-I}} \quad (17)$$

$$\lambda_{max,eV} = 0.036 - 4.260\,\lambda_{max,eV}^{B3LYP-\mathbf{M-II}} + 5.087\,\lambda_{max,eV}^{PBE0-\mathbf{M-II}} \quad (18)$$

In all these equations the *P*-value analysis allows for stating that all DFT coefficients are statistically significant at the 99% confidence level. Figure 3 provides a comparison between the λ_{max} computed with eqs 15 and 16 and experimental results. As can be seen, the agreement is significantly better than in Figure 2, highlighting the interest of such post-treatment of quantum-chemical results. In Table 2, the MLR statistical data are compared to the SLR. With **M-I**, the MLR R^2_{adj} are larger than 99% indicating a nearly perfect fit, the MAE is limited to 6 nm (0.04 eV), the d_R is only 8 nm (0.05 eV), and the maximal deviations are essentially half of those obtained with SLR. It is also striking that none of the 66 estimates exceeds a 25 nm deviation, whereas only 9% present errors larger than 0.1 eV. Using **M-II** leads to slightly poorer R^2_{adj}, MAE, and d_R but to very large errors for 2-OMe, which is clearly a problematic

Table 2. Comparison of the Statistical Parameters Obtained By SLR-B3LYP, SLR-PBE0, and MLR[a]

	M-I		
property	SLR-B3LYP	SLR-PBE0	MLR
R^2 in %	96.0 [96.4]	97.0 [97.2]	99.2 [99.2]
R^2_{adj} in %	96.0 [96.4]	97.0 [97.2]	99.2 [99.1]
MAE in nm [eV]	13.4 [0.079]	11.5 [0.069]	5.8 [0.037]
d_R in nm [eV]	17.7 [0.105]	15.1 [0.092]	7.9 [0.051]
largest positive deviation in nm	47.0 (2,3-NH_2)	40.7 (2,3-NH_2)	16.4 (2-OMe)
largest negative deviation in nm	−42.2 (1,4-NHMe)	−35.2 (1,4-NHMe)	−19.0 (1-NHMe)
largest positive deviation in eV	0.21 (2,7-Cl)	0.20 (2,7-Cl)	0.12 (2,7-Cl)
largest negative deviation in eV	−0.28 (1-NO_2,2-NH_2)	−0.23 (1-NO_2,2-NH_2)	−0.14 (2-OMe)
cases with abs. deviations > 10 nm	34 (52%)	28 (42%)	16 (24%)
cases with abs. deviations > 25 nm	10 (15%)	7 (11%)	0 (0%)
cases with abs. deviations > 0.1 eV	15 (23%)	12 (18%)	6 (9%)
	M-II		
property	SLR-B3LYP	SLR-PBE0	MLR
R^2 in %	94.9 [95.7]	96.3 [96.6]	98.7 [98.2]
R^2_{adj} in %	94.8 [95.6]	96.3 [96.6]	98.7 [98.2]
MAE in nm [eV]	15.2 [0.088]	12.9 [0.077]	6.7 [0.046]
d_R in nm [eV]	19.9 [0.115]	16.9 [0.101]	10.0 [0.074]
largest positive deviation in nm	49.4 (2,3-NH_2)	42.6 (2,3-NH_2)	45.3 (2-OMe)
largest negative deviation in nm	−40.6 (1,4-NHMe)	−32.5 (1,4-NHMe)	−20.3 (1,2-NH_2)
largest positive deviation in eV	0.21 (2,7-Cl)	0.20 (2,7-Cl)	0.14 (2,7-Cl)
largest negative deviation in eV	−0.29 (1-NO_2,2-NH_2)	−0.26 (2-OMe)	−0.39 (2-OMe)
cases with abs. deviations > 10 nm	38 (58%)	38 (58%)	15 (23%)
Cases with Abs. Deviations > 25 nm	13 (20%)	9 (14%)	1 (2%)
Cases with Abs. Deviations > 0.1 eV	20 (30%)	17 (26%)	6 (9%)

[a] These values are obtained with fittings based on the λ_{max} computed in nm and eV.

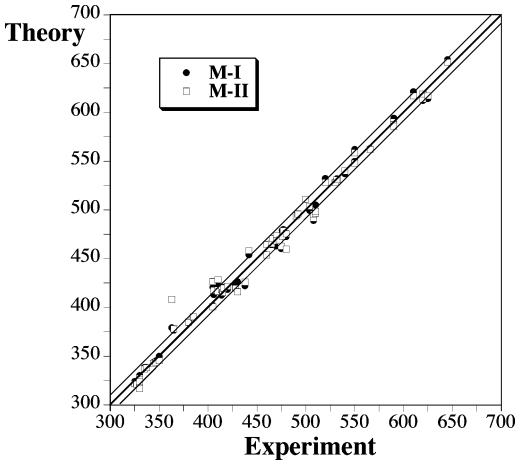

Figure 3. Comparison between the experimental and theoretical λ_{max} obtained by MLR (eqs 15 and 16). The central line indicates a perfect match, whereas the two side lines are borders for ±10 nm discrepancies. All values are in nm.

substitution for our approaches that systematically undershoot the related excitation energy.

IV. Conclusions and Outlook

The absorption spectra of 66 anthraquinone dyes have been computed with a PCM-TD-DFT approach using two hybrid functionals (B3LYP and PBE0) and two basis set combinations. The present study points out that, although both functionals provide at least satisfactory results, PBE0 is more adequate than B3LYP for evaluating the λ_{max} of anthraquinones. In addition, it turns out that the 6-31G(d,p) basis set provides converged transition energies with the PBE0 functional; a further extension of the basis does not improve (and sometimes slightly decreases) the average quality of the theoretical prediction. This means that the absorption spectra of substituted anthraquinones can be accurately evaluated at a relatively small computational cost. We have used a three-step procedure for comparing experimental and theoretical λ_{max}: 1. excitation energies directly taken from TD-DFT calculations, 2. absorption maxima evaluated by SLR, and 3. λ_{max} optimized with MLR using the results of B3LYP and PBE0 functionals. At each step, the accuracy is improved. However, as TD-DFT nicely reproduces the change in λ_{max} resulting from strong auxochromic substitution, the MAE is almost unchanged when using a SLR instead of the "raw" data. For instance, with PBE0, it goes from 13 to 12 nm. The improvement with MLR is more drastic with a MAE limited to 6 nm and a much smaller number of large deviations. More impressively the predicting power of the MLR equations is such that the blind tests for anthraquinones not included in our training set can be estimated with a standard deviation of ±0.05 eV (±8 nm). Although we use a wide panel of substituents, almost covering the entire visible spectrum, the errors reported in this study are much smaller than in most of the recent TD-DFT investigations. This is probably due, in parts, to the explicit consideration of medium effects in our model.

Acknowledgment. E.P. and D.J. thank the Belgian National Fund for Scientific Research for their research associate positions. J.P. acknowledges the FRIA (Belgian "Fonds pour la formation à la Recherche dans l'Industrie et dans l'Agriculture") for his Ph.D. grant. Most calculations

have been performed on the Interuniversity Scientific Computing Facility (ISCF), installed at the Facultés Universitaires Notre-Dame de la Paix (Namur, Belgium), for which the authors gratefully acknowledge the financial support of the FNRS-FRFC and the "Loterie Nationale" for the convention number 2.4578.02 and of the FUNDP.

References

(1) Schäfer, A. *Modern Methods and Algorithms of Quantum Chemistry*, 2nd ed.; volume 3 of *NIC* Johnvon Neumann Institute for Computing: Jülich, 2000.

(2) Adachi, M.; Nakamura, S. *Dyes Pigm.* **1991**, *17*, 287−296.

(3) Runge, E.; Gross, E. K. U. *Phys. Rev. Lett.* **1984**, *52*, 997−1000.

(4) Casida, M. E. In *Accurate Description of Low-Lying Molecular States and Potential Energy Surfaces*; Hoffmann, M. R., Dyall, K. G., Eds.; American Chemical Society: Washington, DC, 2002; Vol. 828.

(5) Onida, G.; Reining, L.; Rubio, A. *Rev. Mod. Phys.* **2002**, *74*, 601−659.

(6) Baerends, E. J.; Ricciardi, G.; Rosa, A.; van Gisbergen, S. J. A. *Coord. Chem. Rev.* **2002**, *230*, 5−27.

(7) Maitra, N. T.; Wasserman, A.; Burke, K. In *Electron Correlations and Materials properties 2*; Gonis, N. K., Ciftan, M., Eds.; Kluwer: Dordrecht, 2003.

(8) Jamorski-Jödicke, C.; Lüthi, H. P. *J. Am. Chem. Soc.* **2002**, *125*, 252−264.

(9) Bartholomew, G. P.; Rumi, M.; Pond, S. J. K.; Perry, J. W.; Tretiak, S.; Bazan, G. C. *J. Am. Chem. Soc.* **2004**, *126*, 11529−11542.

(10) Besley, N. A.; Oakley, M. T.; Cowan, A. J.; Hirst, J. D. *J. Am. Chem. Soc.* **2004**, *126*, 13502−13522.

(11) Ciofini, I.; Lainé, P. P.; Bedioui, F.; Adamo, C. *J. Am. Chem. Soc.* **2004**, *126*, 10763−10777.

(12) Improta, R.; Barone, V. *J. Am. Chem. Soc.* **2004**, *126*, 14320−14321.

(13) Rappoport, D.; Furche, F. *J. Am. Chem. Soc.* **2004**, *126*, 1277−1284.

(14) Stich, T. A.; Buan, N. R.; Brunold, T. C. *J. Am. Chem. Soc.* **2004**, *126*, 9735−9749.

(15) Jacquemin, D.; Preat, J.; Wathelet, V.; André, J. M.; Perpète, E. A. *Chem. Phys. Lett.* **2005**, *405*, 429−433.

(16) Chisholm, M. H.; D'Acchioli, J. S.; Pate, B. D.; Patmore, N. J.; Dala, N. S.; Zipse, D. J. *Inorg. Chem.* **2005**, *44*, 1061−1067.

(17) Masternak, A.; Wenska, G.; Milecki, J.; Skalski, B.; Franzen, S. *J. Phys. Chem. A* **2005**, *109*, 759−766.

(18) Jacquemin, D.; Preat, J.; Perpète, E. A. *Chem. Phys. Lett.* **2005**, *410*, 254−259.

(19) Jorge, F. E.; Autschbach, J.; Ziegler, T. *J. Am. Chem. Soc.* **2005**, *127*, 975−985.

(20) Fabian, J. *Theor. Chem. Acc.* **2001**, *106*, 199−217.

(21) Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2004**, *108*, 10367−10375.

(22) Cave, R. J.; Castner, E. W., Jr. *J. Phys. Chem. A* **2002**, *106*, 12117−12123.

(23) Danielsson, J.; Ulicny, J.; Laaksonen, A. *J. Am. Chem. Soc.* **2001**, *123*, 9817−9821.

(24) Jamorski-Jödicke, C.; Casida, M. E. *J. Phys. Chem. B* **2004**, *108*, 7132−7141.

(25) Petit, L.; Maldivi, P.; Adamo, C. *J. Chem. Theory Comput.* **2005**, *1*, 953−962.

(26) Hommen de Mello, P.; Mennucci, B.; Tomasi, J.; da Silva, A. B. F. *Theor. Chem. Acc.* **2005**, *113*, 274−280.

(27) Jacquemin, D.; Preat, J.; Wathelet, V.; Perpète, E. A. *J. Mol. Struct. (THEOCHEM)* **2005**, *731*, 67−72.

(28) Green, F. J. *The Sigma-Aldrich Handbook of Stains, Dyes and Indicators*; Aldrich Chemical Company, Inc.: Milwaukee, WI, 1990.

(29) Thomson, R. H. *Naturally Occurring Quinones*, 2nd ed.; Academic Press: London, 1971.

(30) Jacquemin, D.; Preat, J.; Charlot, M.; Wathelet, V.; André, J. M.; Perpète, E. A. *J. Chem. Phys.* **2004**, *121*, 1736−1743.

(31) Labhart, H. *Helv. Chim. Acta* **1957**, *152*, 1410−1421.

(32) Frisch, M. J. et al. *Gaussian 03, Revision B.04*; Gaussian, Inc.: Wallingford, CT, 2004.

(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(34) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(35) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.

(36) Adamo, C.; Barone, V. *J. Chem. Phys*. **1999**, *110*, 6158−6170.

(37) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029−5036.

(38) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(39) Perdew, J. P.; Ernzerhof, M.; Burke, K. *J. Chem. Phys.* **1996**, *105*, 9982−9985.

(40) Amovilli, C.; Barone, V.; Cammi, R.; Cancès, E.; Cossi, M.; Mennucci, B.; Pomelli, C. S.; Tomasi, J. *Adv. Quantum Chem.* **1998**, *32*, 227−261.

(41) Cossi, M.; Barone, V. J. *Chem. Phys.* **2001**, *115*, 4708−4717.

(42) Wiberg, K. B.; Stratmann, R. E.; Frisch, M. *J. Chem. Phys. Lett.* **1998**, *297*, 60−64.

(43) Dagnelie, P. *Statistique théorique et appliquée. Tome 1. Statistique descriptive et bases de l'inférence statistique*; De Boeck and Larcier: Bruxelles and Paris, 1998.

(44) Dagnelie, P. *Statistique théorique et appliquée. Tome 2. Inférence statistique à une et deux dimensions*; De Boeck and Larcier: Bruxelles and Paris, 1998.

(45) Pollard, J. *A Handbook of Numerical and Statistical Techniques*; Cambridge University Press: Cambridge, U.K., 1979.

(46) *Statgraphics Plus 5.1*; Manugistics Inc.: Herndon, VA, U.S.A., 2000.

(47) For these comparisons, we have considered anthraquinones substituted only by the groups of the given family, i.e., mixed substitution (as in anthraquinone 60) have not been considered.

CT050287W