*J. Phys. Chem. A* **1999,** *103,* 4553−4559

# Extension of the Neglect of Diatomic Differential Overlap Method to Spectroscopy. NDDO-G Parametrization and Results for Organic Molecules

**Alexander A. Voityuk,[†] Michael C. Zerner,[‡] and Notker Rösch*,[†]**

*Lehrstuhl für Theoretische Chemie, Technische Universität München, D-85747 Garching, Germany, and Quantum Theory Project, University of Florida, Gainesville, Florida 32611*

*Received: January 18, 1999; In Final Form: April 9, 1999*

A new semiempirical scheme, referred to as NDDO-G, for calculating geometries and spectroscopic properties of molecules, is described. The method is based on the NDDO (neglect of diatomic differential overlap) approximation. It uses the point-charge model and the Mataga−Nishimoto formula to evaluate two-center two-electron integrals. The NDDO-G model has been parametrized for the elements H, C, N, and O. Molecular geometries of organic molecules are well predicted by NDDO-G; for 60 molecules, the mean absolute error of bond lengths is 0.014 Å and that of bond angles is 1.9°. The spectroscopic variant of the NDDO-G scheme provides electronic excitations using configuration interaction of singly excited states (CIS). It has been applied to calculate absorption spectra (vertical transitions) of several dozens organic molecules and photoelectron spectra within Koopmans' approximation. These NDDO-G results are compared with experimental results and with results of high-level ab initio calculations. The mean absolute error of NDDO-G excitation energies is 0.13 eV $= 1050$ cm$^{-1}$ (196 comparisons). First and several higher ionization potentials are reproduced with a mean absolute error of 0.24 eV (123 comparisons). The proposed method may be used for studying structures of large organic and biological molecules and for interpreting and predicting their absorption and photoelectron spectra. As an example, we discuss the spectroscopy of free-base porphin.

## Introduction

Ultraviolet and visible spectroscopy is a very important tool for analyzing chemical and biochemical systems. However, the interpretation of electronic spectra of molecules has remained a challenge for both experiment and theory.[1] Currently, two approaches are used to calculate absorption spectra: high-level ab initio and semiempirical quantum chemical calculations.[2,3]

Although ab initio CIS theory (configuration interaction based on singly excited configurations) may be considered as an affordable technique for calculating molecular electronic excitation energies, in general this approach is not accurate and reliable enough; typical errors are greater than 1 eV.[2] To improve this situation, a more sophisticated treatment of electronic correlation is required. A few years ago the CASPT2 methodology (complete active space method supplemented by a second-order perturbation scheme) was proposed as a very promising high-level ab initio approach for calculating electronically excited states.[4,5] This method accounts for correlation effects associated with different types of excitations in a balanced way. It has been successfully applied to a number of small and medium-size molecules (see refs 6 and 7 and references therein).

The coupled-cluster method also provides a reliable treatment of excited states and has been employed for a number of molecular systems.[11−13] Recently, Bartlett and co-workers have presented a new variant of the method, STEOM-CC (similarity transformed equation-of-motion couple-cluster), which allows accurate and efficient calculations on excited states.[14] The performance of the approach was demonstrated by application to the electronic excitation spectrum of the free-base porphin.[15] Another quite similar high-level ab initio scheme for calculating excitation energies is provided by the SAC-CI (symmetry-adapted cluster-configuration interaction) method.[16]

However, the bottleneck due to molecular size so far remains despite the considerable increase in available computing power. At best, a molecular system with up to 25−30 first-row atoms can be studied with high-level ab initio schemes, and in these cases only where there is high symmetry. Therefore, many important systems, especially those of biological importance that generally have no symmetry and even reasonable models of such systems, lie far beyond these approaches.

Semiempirical procedures are computationally much less demanding than ab initio methods, and therefore, they remain very useful for large systems.[3,17] While theoretically inferior to high-level methods in the sense that there are no systematic ways to improve them, semiempirical approaches are able to reveal the main features and trends of a given class of systems, comparable to high-level methods but at a significantly reduced cost. The INDO/S method (intermediate neglect of differential overlap/spectroscopic parametrization) parametrized at the CIS level of theory has proven to be a useful tool for calculating spectroscopic properties of organic 3d and 4d transition metal compounds.[18−21] After augmentation of the INDO/S scheme by a treatment of the spin−orbit interaction, this model has also been extended to excited states of lanthanide molecules.[22]

Since the NDDO approximation[23] (neglect of differential diatomic overlap) contains more of the two-electron integrals, it should be more accurate than the INDO scheme at small cost in computing time. One might therefore expect to gain more reliable spectroscopic predictions within this scheme, especially for transition metal compounds. Furthermore, for d-metal complexes the INDO/S approximation exhibits a fundamental deficiency; it can be shown that the extra integrals of the NDDO approximation are quite important in determining the ligand field

---

[†] Technische Universität München.

[‡] University of Florida.

10.1021/jp9902001 CCC: $18.00

splitting. Nowadays semiempirical methods based on the NDDO approximation (MNDO,[24] AM1,[25] PM3,[26] and MNDO/d[27,28]) are widely applied to studies of the structure and thermochemistry of various molecular systems.[29] However, these methods fail to predict spectroscopic parameters accurately enough.[17] The purpose of the present work is to parametrize the NDDO model for spectroscopy and molecular geometry and to study its performance. The resulting unified approach to ground-state properties and excitation energies will be referred to as the NDDO-G method (neglect of differential overlap, Garching−Gainesville parametrization).

## Method and Parametrization

For orthonormalized atomic orbitals (AO) the Fock equation of a closed-shell system can be written as

$$\mathbf{FC} = \mathbf{CE}$$

The diagonal matrix **E** contains the energy eigenvalues of the molecular orbitals (MO) that can be used to estimate the ionization potentials according to Koopmans' theorem, at least for the delocalized MOs of this study. The coefficients of each molecular orbital are collected in columns of the square matrix **C**. The matrix elements of the Fock operator $F$ are specified by[3]

$$F_{\mu\mu} = U_{\mu\mu} + \sum_{B} V_{\mu\mu,B} + \sum_{\nu} P_{\nu\nu}(g_{\mu\nu} - {}^{1}/_{2} h_{\mu\nu}) + \sum_{B}\sum_{\kappa\lambda} P_{\kappa\lambda}(\mu\mu,\kappa\lambda)$$

$$F_{\mu\nu} = \sum_{B} V_{\mu\nu,B} + {}^{1}/_{2} P_{\mu\nu}(3h_{\mu\nu} - g_{\mu\nu}) + \sum_{B}\sum_{\kappa\lambda} P_{\kappa\lambda}(\mu\nu,\kappa\lambda)$$

$$F_{\mu\kappa} = \beta_{\mu\kappa} - {}^{1}/_{2} \sum_{\nu\lambda} P_{\nu\lambda}(\mu\nu,\kappa\lambda)$$

The indices $\mu$ and $\nu$ in these expressions designate atomic orbitals of center A, while the indices $\kappa$ and $\lambda$ belong to center B. The following terms are included in the Fock matrix.

(1) The one-center one-electron energies $U_{\mu\mu}$ represent the sum of the kinetic energy of an electron in AO $\mu$ and its potential energy due to the attraction by the core. These energies may be derived from experimental valence state energies of the corresponding atomic configurations or may be treated as optimized parameters.[3]

(2) In an sp basis set only two types of one-center two-electron integrals differ from zero: the Coulomb integrals $g_{\mu\nu} = (\mu\mu,\nu\nu)$ and the exchange integrals $h_{\mu\nu} = (\mu\nu,\mu\nu)$. We use the spectroscopic values of these parameters.[30]

(3) The core−electron attraction terms $V_{\mu\nu,B}$ between the one-center electronic distribution $\mu\nu$ of atom A and the core of atom B is evaluated using the corresponding two-center two-electron integral: $V_{\mu\nu,B} = -Z_B(\mu\nu,s^B s^B)$ where $s^B$ is an s orbital and $Z_B$ is the effective nuclear charge of center B.

(4) The two-center one-electron resonance integrals $\beta_{\mu\kappa}$ are obtained in the molecular coordinate system by a rotational transformation of the corresponding terms evaluated in a local coordinate system with one axis oriented along the AB "diatomic" direction.[3] The quantities $\beta_{\mu\kappa}$ are taken to be proportional to the orbital overlap $S_{\mu\kappa}$:

$$\beta_{\mu\kappa} = (\beta_{\mu} + \beta_{\kappa})\, S_{\mu\kappa}$$

where the parameters $\beta_{\mu}$ and $\beta_{\kappa}$ are chosen empirically to yield the best agreement with experimental data. In the local coordinate system, the $\sigma$-resonance interaction is calculated using the parameters $\beta_s$ and $\beta_p$ while for the $\pi$-resonance interaction an additional parameter $\beta_{\pi}$ is employed. This weighting of $\pi$ components is similar to that used in CNDO/S and INDO/S.[3] The overlap integrals $S_{\mu\nu}$ are calculated using Slater-type orbitals with exponents $\zeta_s$ and $\zeta_p$.

**TABLE 1: Optimized NDDO-G Parameters**

| | H | C | N | O |
|---|---|---|---|---|
| $U_s$ (eV) | −13.38 | −50.74 | −71.62 | −92.23 |
| $U_p$ (eV) | | −39.58 | −58.87 | −79.94 |
| $\zeta_s$ (au) | 1.20 | 1.70 | 2.35 | 3.50 |
| $\zeta_p$ (au) | | 1.70 | 2.35 | 2.65 |
| $\beta_s$ (eV) | −10.32 | −20.20 | −22.92 | −36.00 |
| $\beta_p$ (eV) | | −11.90 | −27.78 | −40.00 |
| $\beta_{\pi}$ (eV) | | −10.19 | −27.28 | −40.00 |

(5) Two-center two-electron repulsion integrals $(\mu\nu,\kappa\lambda)$ are estimated using a scheme that ensures rotational invariance.[27] As established earlier for the INDO/S model,[20] it was also found for the NDDO-G scheme that the Mataga−Nishimoto formula[31] for the two-electron repulsion integrals allows a more accurate reproduction of excitation energies than the Ohno−Klopman formula, applied in MNDO-like methods.[24] The additive terms required for evaluating the integrals are chosen in such a way as to reproduce the corresponding one-center two-electron integrals at $R_{AB} = 0$ (see ref 27 for more details).

Several functions were tried for the core−core repulsion integrals. The expression

$$E^{\mathrm{core}}_{AB} = Z_A Z_B (s^A s^A, s^B s^B)[1 + 4\exp(-\alpha_{AB} R_{AB})]$$

with the adjustable bond-dependent parameter $\alpha_{AB}$ allowing the most accurate reproduction of molecular geometries and therefore was employed in the parametrization. For organic compounds composed of the H, C, N, and O compounds, 10 parameters $\alpha_{AB}$ are needed, one parameter per each type of bond. Note for comparison that the AM1 and PM3 schemes employ 40 and 28 parameters, respectively, to describe core−core repulsion of these four elements.

For each of the atoms C, N, and O, seven parameters were fitted using spectroscopic reference data for selected organic compounds (Table 1): the core energies $U_s$ and $U_p$, the exponents $\zeta_s$ and $\zeta_p$ required for estimating overlap integrals, and the resonance parameters $\beta_s$, $\beta_p$, $\beta_{\pi}$. For H, three parameters were optimized: $U_s$, $\zeta_s$, and $\beta_s$ (Table 1). The NDDO-G parameters for H, C, N, and O were adjusted to best reproduce the experimental characteristics using a least-squares fit to molecular geometries, excitation energies (also taking into account the corresponding oscillator strengths; Table 2), and vertical ionization potentials of about 40 molecules (see Supporting Information). The success of a semiempirical parametrization substantially depends on the choice of molecules in the training set. Taking into account that the NDDO-G method is intended for calculating rather extended organic molecules and biologically relevant models, large molecules should be included in the parametrization set. On the other hand, CI calculations of extended systems are time-consuming, and therefore, the number of compounds in the training set has to be limited. Thus, several parametrization runs were carried out starting from different parameter values and using different training sets. The initial parametrizations employed about 20 molecules and 100 reference properties. A nonlinear least-squares method was used to optimize the semiempirical parameters. The resulting optimized parameter set was tested in extensive survey calculations in order to choose the set that

**TABLE 2: Calculated and Observed Excitation Energies (eV)**

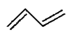
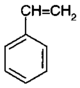
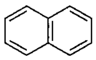
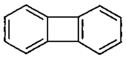
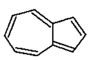
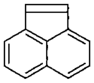
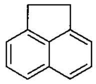
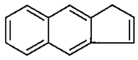
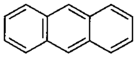
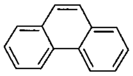
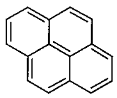
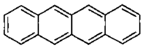
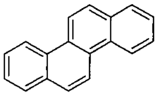
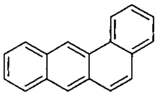
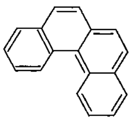
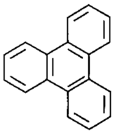
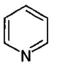
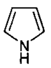
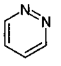
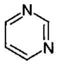
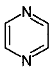
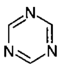
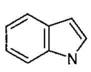
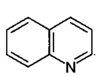
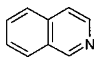
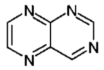
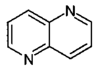
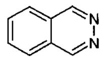
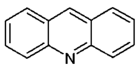
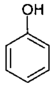
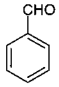
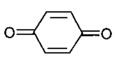
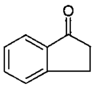
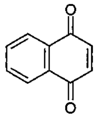
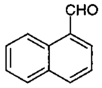
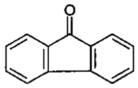
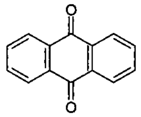
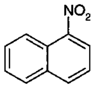
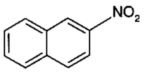
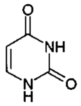
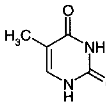
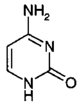
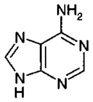
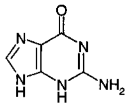
| Molecule | Exp.[a] | NDDO-G | INDO/S |
|---|---|---|---|
| Ethylene | 8.00[b] | 7.96 | 8.34 |
|  | | | |
| 1,3-trans-Butadiene | 5.92[b] | 5.74 | 5.84 |
|  | 6.66[b] | 6.69 | 6.56 |
| Benzene | 4.76[c] | 4.58 | 4.69 |
|  | 6.07[c] | 5.87 | 6.05 |
| | 6.73[c] | 6.68 | 6.88 |
| Styrene | 4.26[c] | 4.30 | 4.40 |
|  | 5.00[c] | 4.92 | 5.05 |
| Naphthalene | 3.99 | 3.82 | 3.90 |
|  | 4.34 | 4.39 | 4.49 |
| | 5.62 | 5.47 | 5.58 |
| Biphenylene | 3.11 | 3.08 | 3.13 |
|  | 3.45 | 3.61 | 3.63 |
| | 5.00 | 5.11 | 5.27 |
| Azulene | 1.76 | 2.00 | 2.05 |
|  | 3.51 | 3.35 | 3.45 |
| | 4.43 | 4.17 | 4.27 |
| | 4.57 | 4.71 | 4.82 |
| Acenaphtylene | 3.00 | 3.05 | 3.07 |
|  | 3.64 | 3.83 | 3.94 |
| | 3.84 | 3.88 | 3.97 |
| | 4.53 | 4.55 | 4.65 |
| | 5.42 | 5.29 | 5.38 |
| Acenaphtene | 3.86 | 3.77 | 3.84 |
|  | 4.12 | 4.23 | 4.27 |
| | 5.42 | 5.37 | 5.29 |
| | 5.86 | 5.84 | 5.37 |
| Benz[f]indene | 3.36 | 3.60 | 3.68 |
|  | 4.12 | 4.13 | 4.17 |
| | 4.84 | 4.74 | 4.79 |
| | 5.13 | 5.20 | 5.23 |
| | 5.89 | 5.83 | 5.73 |
| | 6.11 | 6.07 | 5.94 |
| | 6.46 | 6.35 | 6.16 |
| Anthracene | 3.31 | 3.36 | 3.42 |
|  | 3.66 | 3.48 | 3.53 |
| | 4.92 | 4.79 | 4.88 |
| | 5.67 | 5.59 | 5.71 |
| Phenanthrene | 3.58 | 3.56 | 3.64 |
|  | 4.24 | 4.14 | 4.24 |
| | 4.95 | 4.84 | 4.94 |
| | 5.08 | 4.96 | 5.04 |
| | 5.63 | 5.49 | 5.62 |
| | 5.72 | 5.65 | 5.75 |
| Pyrene[d] | 3.71 | 3.63 | 3.70 |
|  | 4.55 | 4.67 | 4.74 |
| | 5.14 | 5.10 | 5.20 |
| | 6.32 | 6.10 | 6.19 |
| Naphthacene[d] | 2.63 | 2.89 | 2.92 |
|  | 3.14 | 3.07 | 3.13 |
| | 4.51 | 4.34 | 4.43 |
| | 5.18 | 5.24 | 5.35 |
| | 5.47 | 5.45 | 5.59 |
| | 5.89 | 6.07 | 6.21 |
| Chrysen[d] | 3.43 | 3.38 | 3.45 |
|  | 3.89 | 3.78 | 3.86 |
| | 4.64 | 4.54 | 4.62 |
| | 5.13 | 4.93 | 5.00 |
| | 5.58 | 5.78 | 5.96 |
| | 6.20 | 6.17 | 6.37 |
| | 6.43 | 6.41 | 6.55 |
| 1,2-Benzantracene[d] | 3.22 | 3.27 | 3.34 |
|  | 3.46 | 3.58 | 3.85 |
| | 4.31 | 4.27 | 4.36 |
| | 4.62 | 4.53 | 4.63 |
| | 4.83 | 4.74 | 4.86 |
| | 4.98 | 4.87 | 4.97 |
| | 5.43 | 5.54 | 5.66 |
| | 6.15 | 6.12 | 6.46 |
| Benzophenanthrene[d] | 3.38 | 3.36 | 3.43 |
|  | 3.82 | 3.87 | 3.96 |
| | 4.41 | 4.31 | 4.40 |
| | 4.71 | 4.78 | 4.86 |
| | 5.44 | 5.44 | 5.54 |
| | 5.52 | 5.61 | 5.75 |
| Triphenylene[d] | 3.71 | 3.55 | 3.63 |
|  | 4.09 | 4.05 | 4.16 |
| | 4.36 | 4.25 | 4.34 |
| | 4.81 | 4.68 | 6.78 |
| | 5.20 | 5.13 | 5.69 |
| | 5.95 | 5.93 | 6.06 |
| | 6.27 | 6.17 | 6.38 |
| | 6.66 | 6.56 | 6.74 |
| Pyridine | 4.59[c] | 4.35 | 4.38 |
|  | 4.83[c] | 4.87 | 4.79 |
| | 6.20[c] | 6.07 | 6.17 |
| | 7.22[c] | 7.09 | 7.10 |
| Pyrrole | 5.92 | 5.47 | 5.41 |
|  | 6.00 | 5.63 | 5.50 |
| | 7.30 | 7.21 | 7.20 |
| 1,2-diazine | 3.60[c] | 3.94 | 3.72 |
|  | 5.00[c] | 5.17 | 4.98 |
| | 7.30[c] | 7.35 | 7.37 |
| 1,3-Diazine | 3.85[c] | 4.15 | 4.06 |
|  | 5.12[c] | 5.26 | 5.03 |
| | 6.70[c] | 6.45 | 6.39 |
| | 7.57[c] | 7.64 | 7.33 |
| 1,4-Diazine | 3.83[c] | 3.72 | 3.73 |
|  | 4.81[c] | 5.03 | 4.65 |
| | 6.38[c] | 6.08 | 6.19 |
| | 7.67[c] | 7.83 | 7.61 |
| 1,3,5-Triazine | 4.49[c] | 4.56 | 4.33 |
|  | 4.59[c] | 4.66 | 4.49 |
| | 7.76[c] | 8.00 | 7.53 |
| Indol | 4.36 | 4.08 | 4.12 |
|  | 4.82 | 4.61 | 4.68 |
| | 5.76 | 5.47 | 5.46 |
| | 6.32 | 5.98 | 6.03 |
| Quinoline | 3.95 | 3.95 | 3.97 |
|  | 4.05 | 3.96 | 4.03 |
| | 4.52 | 4.56 | 4.56 |
| | 5.40 | 5.38 | 5.38 |
| | 6.11 | 5.99 | 6.04 |
| Isoquinoline | 3.91 | 3.95 | 3.93 |
|  | 4.08 | 4.19 | 4.20 |
| | 4.66 | 4.50 | 4.56 |
| | 5.74 | 5.62 | 5.68 |
| Pteridine[d] | 3.17 | 3.01 | 3.21 |
|  | 4.10 | 4.33 | 3.99 |
| | 5.24 | 5.27 | 5.55 |
| | 5.90 | 5.98 | 5.96 |
| 1,5-Naphthyridine[d] | 4.10 | 4.09 | 3.99 |
|  | 4.82 | 4.93 | 4.84 |
| | 6.01 | 5.82 | 5.85 |
| Phthalazine[d] | 4.19 | 4.15 | 4.11 |
|  | 4.78 | 4.84 | 4.73 |
| | 5.76 | 5.79 | 5.83 |
| Acridine | 3.27 | 3.48 | 3.50 |
|  | 3.50 | 3.53 | 3.64 |
| | 3.66 | 3.65 | 3.71 |
| | 4.98 | 4.86 | 4.94 |
| Furan | 6.06 | 5.61 | 5.71 |
|  | 7.80 | 7.64 | 7.83 |
| Phenol | 4.46 | 4.33 | 4.37 |
|  | 5.70 | 5.44 | 5.51 |
| | 6.44 | 6.44 | 6.62 |
| Benzaldehyde | 3.50 | 3.52 | 3.31 |
|  | 4.31 | 4.42 | 4.55 |
| | 5.08 | 5.37 | 5.59 |
| | 6.20 | 6.25 | 6.43 |

TABLE 2: (Continued)

| Molecule | Exp.[a] | NDDO-G | INDO/S |
|---|---|---|---|
| p-Benzochinon | 2.59 | 2.63 | 2.58 |
|  | 2.85 | 3.05 | 2.64 |
| | 4.46 | 4.49 | 4.86 |
| | 5.17 | 5.41 | 5.68 |
| Indanon | 3.73 | 3.41 | 3.28 |
|  | 4.32 | 4.39 | 4.45 |
| | 5.22 | 5.30 | 5.45 |
| | 6.07 | 6.08 | 6.09 |
| p-Naphthochinone | 3.73 | 4.12 | 4.30 |
|  | 4.96 | 5.21 | 5.42 |
| | 5.05 | 5.42 | 5.67 |
| 1-Naphthaldehyde | 3.35 | 3.42 | 3.21 |
|  | 3.77 | 3.77 | 3.86 |
| | 3.99 | 4.11 | 4.25 |
| | 5.00 | 5.10 | 5.24 |
| | 5.23 | 5.44 | 5.54 |
| Fluorenone[d] | 3.14 | 3.02 | 2.90 |
|  | 3.79 | 3.53 | 3.69 |
| | 4.13 | 4.08 | 4.20 |
| | 4.85 | 5.05 | 5.24 |
| | 5.00 | 5.44 | 5.62 |
| Anthrachinon[d] | 2.92 | 2.96 | 2.90 |
|  | 3.71 | 4.15 | 3.69 |
| | 4.57 | 5.06 | 4.20 |
| | 4.93 | 5.27 | 5.24 |
| 1-Nitronaphthalene[d] | 3.10 | 2.93 | 2.45 |
|  | 3.22 | 3.20 | 3.38 |
| | 3.45 | 3.33 | 3.74 |
| | 4.81 | 4.53 | 4.90 |
| | 5.88 | 5.87 | 5.55 |
| 2-Nitronaphthalene[d] | 3.10 | 3.00 | 2.83 |
|  | 3.47 | 3.37 | 3.71 |
| | 4.09 | 3.90 | 4.20 |
| | 4.77 | 4.59 | 4.85 |
| | 5.88 | 5.86 | 6.21 |
| Uracil | 4.50[e] | 4.66 | 4.02 |
|  | 5.10[e] | 5.18 | 5.31 |
| | 6.00[e] | 6.24 | 6.43 |
| | 6.60[e] | 6.62 | 6.60 |
| | 7.00[e] | 7.34 | 7.40 |
| Thymine | 5.10[e] | 4.98 | 5.06 |
|  | 6.00[e] | 5.93 | 6.09 |
| | 6.60[e] | 6.65 | 6.65 |
| Cytosine | 4.60[f] | 4.45 | 4.37 |
|  | 5.50[f] | 5.74 | 5.63 |
| | 6.30[f] | 6.13 | 6.17 |
| Adenine | 4.60[g] | 4.55 | 4.27 |
|  | 4.80[g] | 4.76 | 4.60 |
| | 6.00[g] | 5.82 | 5.87 |
| | 6.30[g] | 6.18 | 6.13 |
| | 6.80[g] | 6.50 | 7.39 |
| Guanine | 4.40 | 4.12 | 3.93 |
|  | 4.90 | 4.87 | 4.68 |
| | 5.70 | 5.50 | 4.98 |
| | 6.20 | 6.39 | 6.19 |
| | 6.60 | 6.77 | 6.47 |
| Mean error (196 comparisons) | | -0.01 | 0.03 |
| Mean absolute error | | 0.13 | 0.17 |
| Maximum deviation | | 0.49 | 0.72 |

[a] Reference 35 unless stated otherwise. [b] Reference 13. [c] Reference 12. [d] Molecules not included in the training set of the parametrization. [e] Reference 9. [f] Reference 10. [g] Reference 7.

yields the most balanced results. Molecules that had large differences between calculated and experimental properties were included in the training set and the parameters reoptimized. In this way, we used increasingly larger molecular sets of up to 34 species and 180 reference values (see Table 2). This iterative refinement procedure was repeated eight times.

The electronic spectra were calculated using configuration interaction of selected singly excited states (CIS). The active space was defined to include all orbitals from the lowest occupied $\pi$ MO to the highest unoccupied $\pi$ MO; all possible "single" excitations that can be generated in this active space were taken into account. Note that the lone pairs of N and O responsible for n $\rightarrow$ $\pi^*$ transitions are included in the active space determined in this fashion. One may expect that using the same selection procedure for the CI space in both parametrization and application will lead to the most accurate NDDO-G predictions of spectroscopic properties. Vertical ionization potentials were calculated from the corresponding orbital energies using the frozen orbital approximation as suggested by Koopmans' theorem. The calculations were carried out with the programs SIBIQ[32] and ZINDO.[33]

During the optimization, the orbital exponents $\zeta_s$ and $\zeta_p$ of C and N converged to similar values and thus were set to be equal. The final parameters are listed in Table 1; the values found by fitting were rounded off so that there is no effect on the accuracy of the method. The parametrization yielded the following core−core repulsion parameters $\alpha_{AB}$ in (in Å$^{-1}$): 2.24 (H−H), 2.81 (H−C), 3.12 (H−N), 3.35 (H−O), 2.78 (C−C), 2.87 (C−N), 3.10 (C−O), 2.91 (N−N), 3.10 (N−O), and 3.18 (O−O). Note that these $\alpha_{AB}$ parameters are required only for geometry optimizations; they do not directly affect the computation of vertical excitation or ionization processes, since the nuclear repulsion energy is a constant under the Franck−Condon fixed-geometry approximation.

TABLE 3: Mean Absolute Errors for the NDDO-G, INDO/S, AM1, and PM3 Models

| | $M^a$ | $N^b$ | NDDO-G | INDO/S | AM1 | PM3 |
|---|---|---|---|---|---|---|
| bond lengths, Å | 60 | 176 | 0.014 | | 0.019 | 0.011 |
| bond angles, deg | 60 | 90 | 1.8 | | 2.0 | 2.2 |
| excitation energies, eV | | | | | | |
| CH compounds | 18 | 83 | 0.10 | 0.12 | | |
| CHN compounds | 13 | 48 | 0.15 | 0.15 | | |
| CHNO compounds | 16 | 65 | 0.17 | 0.27 | | |
| | 47 | 196 | 0.13 | 0.17 | | |
| Ionization potentials,[c] eV | 46 | 123 | 0.24 | 0.40 | 0.60 | 0.64 |

[a] Number of molecules. [b] Number of comparisons. [c] Vertical ionization processes.

## Results and Discussion

**Molecular Geometries**. Experimental structural parameters for 60 molecules in the gas phase were taken from standard compilations[34] and occasionally from previous semiempirical evaluations.[24] Comparison of experimental and calculated structures showed that NDDO-G geometries are of similar accuracy as AM1 and PM3 data (for details, see Supporting Information). The mean absolute error,

$$N^{-1}\sum_{i=1}^{N}|F_i^{\mathrm{calc}} - F_i^{\mathrm{exp}}|$$

estimated for 176 bond lengths amounts to 0.014 Å in NDDO-G, 0.019 Å in AM1, and 0.011 Å in PM3 (Table 3). Bond angles are reproduced with very similar accuracy; the mean absolute errors found for 90 bond angles are 1.8° (NDDO-G), 2.0° (AM1), and 2.2° (PM3) (Table 3).

Because these three semiempirical methods provide rather reliable structures close to experimental results, the vertical transition energies as well as ionization potentials calculated within NDDO-G are often not very sensitive to the method employed for generating the molecular structure. However, from a practical point of view it is convenient to use the same method for estimating both geometric and spectroscopic parameters. In some models using the experimental geometry yields the most accurate spectroscopic predictions, as the spectroscopy seems more sensitive to geometry than is the total energy of the ground state.[4] On the other hand, in many systems of biological interest, an accurate experimental geometry is not available. Thus, for this study excitation energies and ionization potentials were evaluated at the NDDO-G geometries.

**Excitation Energies**. Table 2 lists excitation energies calculated for 47 organic molecules together with available experimental data.[35] In general, good agreement between experimental values and those calculated with NDDO-G is found. For 196 excitation energies the mean absolute error is 0.13 eV or 1050 cm$^{-1}$ (Table 3). The INDO/S method exhibits also very good performance with a mean absolute error of 0.17 eV (Table 3). The most accurate estimates of excitation energies may be obtained for $\pi$-hydrocarbon molecules; the values of the mean absolute error for 18 molecules (83 comparisons) are 0.10 and 0.12 eV for NDDO-G and INDO/S, respectively (Table 3). For nitrogen-containing compounds both methods are equally accurate; for 13 molecules the mean absolute error is estimated to be about 0.15 eV based on 48 excitation energies. However, the two semiempirical schemes differ substantially for oxygen-containing molecules (CHO and CHNO compounds; Table 3). On the basis of 65 excitation energies of 16 molecules, the corresponding mean absolute error of INDO/S, 0.27 eV, reduces to 0.17 eV for NDDO-G (Table 3). Thus, in the present study, the main improvement by the NDDO-G scheme is found for oxygen-containing molecules. The correlation between results calculated with NDDO-G and INDO/S and experimental excitation energies (196 comparisons) is illustrated in Figure 1.

One of the most appropriate fields for applications of the NDDO-G model involves biological systems in which UV−visible spectra are the only analytical tools available to follow activity. As shown in Table 2, NDDO-G reproduces the observed spectrum of indol well. This chromophore is a part of the amino acid tryptophan that is responsible for the UV absorption of proteins. Comparison of experimental and calculated spectra of nucleic bases uracil, thymine, cytosine, adenine, and guanine reveals good agreement between observed and predicted data (Table 2). The mean absolute error of excitation energies of nucleic bases reduces from 0.25 eV in INDO/S to 0.15 eV in NDDO-G. This good agreement suggests the usage of the NDDO-G scheme to interpret and predict spectroscopic properties of polynucleotides and related systems.

**Application to Porphin.** To illustrate the performance of NDDO-G for a system of biological relevance, we now turn to a discussion of free-base porphin (see Figure 2). For further assessment of the method some additional comparisons with high-level ab initio results on this system are presented. The electronic spectrum of free-base porphin has attracted much attention, since this molecule is a basic unit of porphyrins.[15,36−39] Because the geometry of porphin considerably affects its spectroscopic properties, we start by discussing the molecular structure.

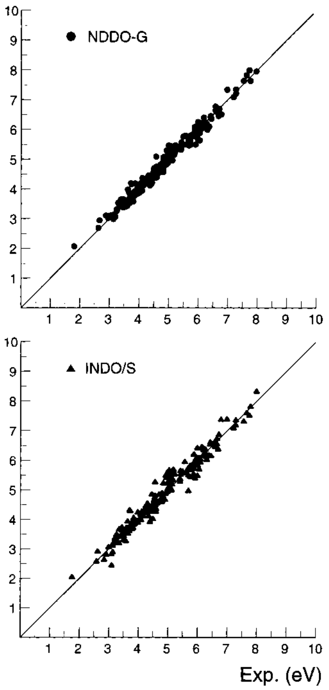


**Figure 1.** Correlation between calculated and experimental excitation energies for 47 molecules (in eV).

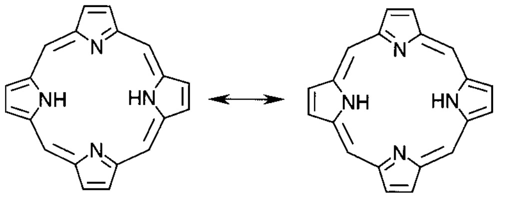


**Figure 2.** Two resonance $C_{2v}$ structures of porphin. The $D_{2h}$ structure is obtained by averaging.

Actually, various quantum chemical methods yield significantly differing structures of free-base porphin with noticeable differences in the calculated excitation spectra. At the ab initio[40] and semiempirical AM1[41] levels of theory it was found that the restricted Hartree−Fock approximation artificially favors a structure of free-base porphin with alternating bonds and relatively low symmetry, $C_{2v}$. This is at variance with experimental results[42] as well as with the results of quantum chemical methods that take electron correlation into account, like MP2 and the local density approximation,[40] which all yield structures with almost equivalent C−C bonds in the porphin ring that features approximate $D_{2h}$ structure. Even the unrestricted Hartree−Fock (UHF) method leads to such a structure with conjugated bonds.[40,41] Unlike all other semiempirical approaches, MNDO, AM1, and PM3, which predict bond alterations of $\Delta R = 0.07-0.08$ Å, the spin-restricted NDDO-G method gives almost equivalent bonds, $\Delta R = 0.01-0.02$ Å, in good agreement with the experimental value of $\Delta R = 0.025$ Å.[42] A UHF procedure improves the geometry calculated with MNDO-like methods. However, the UHF method has the distinct drawback of yielding a poorly defined mixture of spin states and an unphysical alternation of $\alpha$ and $\beta$ spin densities at adjacent carbon centers of porphin.

The experimental spectrum of free-base porphin consists of three region: (1) four rather weak bands in the visible, Q$_x$(0−0), Q$_x$(1−0), Q$_y$(0−0), and Q$_y$(1−0), (2) the very intense peak B (Soret band) with the shoulder N, and (3) the higher UV bands L and M. As suggested in the notation, the vibrational structure is quite apparant in the Q-bands, and our calculations refer only

**TABLE 4: Calculated and Observed Absorption Spectrum of Free-Base Porphine (in eV)**[a]

| band | state | expt[b,c] | NDDO-G[d] | INDO/S[d] | STEOM-CCSD[b] | CASPT2[d] | SAC−CI[e] |
|---|---|---|---|---|---|---|---|
| $Q_x$ | $1^1B_{3u}$ | 1.98 (0.01) | 1.75 (0.0001) | 1.72 (0.01) | 1.75 (0.0007) | 1.70 (0.001) | 1.75 (0.0001) |
| $Q_y$ | $1^1B_{2u}$ | 2.42 (0.06) | 2.07 (0.02) | 2.05 (0.04) | 2.40 (0.013) | 2.26 (0.016) | 2.23 (0.0006) |
| B | $2^1B_{3u}$ | 3.33 (1.15) | 3.25 (1.49) | 3.27 (1.30) | 3.47 (0.693) | 2.91 (1.66) | 3.56 (1.03) |
|  | $2^1B_{2u}$ |  | 3.46 (2.51) | 3.51 (2.29) | 3.62 (1.20) | 3.04 (1.54) | 3.75 (1.73) |
| N | $3^1B_{3u}$ | 3.65 (<0.1) | 3.91 (1.43) | 3.97 (1.47) | 4.06 (0.93) |  | 4.24 (0.98) |
| L | $3^1B_{2u}$ | 4.25 (∼0.1) | 4.37 (0.25) | 4.35 (0.32) | 4.35 (0.42) |  | 4.52 (0.35) |
| L | $4^1B_{2u}$ | 4.67 (∼0.1) | 4.77 (0.10) | 4.82 (0.13) | 5.00 (0.15) |  | 5.31 (0.28) |
|  | $4^1B_{3u}$ |  | 4.86 (0.12) | 5.15 (0.08) |  |  |  |
| M | $5^1B_{3u}$ | 5.50 (∼0.3) | 5.09 (0.17) | 5.07 (0.03) | $4^1B_{3u}$ 5.17 (0.27) |  | 5.45 (0.35) |
|  | $5^1B_{2u}$ |  | 5.20 (0.15) | 5.16 (0.07) | $5^1B_{3u}$ 6.07 (0.18) |  |  |
|  |  |  |  |  | $5^1B_{2u}$ 6.57 (0.04) |  |  |

[a] Oscillator strengths given in parentheses. [b] Reference 15. [c] Reference 39. [d] Present work. [e] Reference 38.

to $\lambda_{max}$ from the Franck−Condon overlap, one for $Q_x$ and one for $Q_y$. Table 4 compares the experimental spectrum of porphin to calculated results of various methods. The molecule is placed in the *xy* plane with the two internal hydrogen atoms along the *x* axis; the structure has approximately $D_{2h}$ symmetry (Figure 2).

Overall, we find good agreement between calculated NDDO-G and experimental excitation energies of free-base porphin (Table 4). As just mentioned, all methods, ab initio[15,38,39] and semiempirical, predict the excitation energy for the $Q_x$ band, which belongs to a $^1B_{3u}$ state, which is too low. The experimental splitting $Q_x-Q_y$ is well reproduced by the semiempirical methods and not by the ab initio methods. The NDDO-G and INDO/S results for the B band are in very good agreement with the observed spectrum. In accord with the STEOM-CCSD[15] and CASPT2[38] calculations, the B band is assigned to the states $2^1B_{3u}$ and $2^1B_{3u}$. While NDDO-G predicts the excited state $3^1B_{3u}$ to be 0.26 eV higher than the experimental value of the N peak, the assignment is identical to the traditional interpretation, which is also supported by the CCSD calculation.[15] All methods correctly predict the ordering of the Q and B bands deduced from polarization experiments.[43] The first and second L peaks may be assigned to states $3^1B_{2u}$ and $4^1B_{2u}$, respectively. Although there is a difference of about 0.2 eV between the NDDO-G and CCSD results for the $4^1B_{3u}$ excited state, we assign this transition to the second L peak whereas Gwaltney and Bartlett[15] suggest assignment to the M band (Table 4). The excitation energies for the states $5^1B_{3u}$ and $5^1B_{3u}$ calculated semiempirically are in good agreement with the observed M band.

By use of the NDDO-G scheme, we found that the $Q_x$ and $Q_y$ bands are very sensitive to small changes in the structure. In fact, for the PM3 structure with bond alternation, we calculated the $Q_x$ and $Q_y$ peaks at 2.31 and 2.47 eV, respectively; these transitions are 0.56 and 0.40 eV higher than those calculated for the NDDO-G structure with approximate $D_{2h}$ symmetry. Note that NDDO-G calculations with $D_{2h}$ constraints result in excitation energies of 1.68 and 2.04 eV, which are very similar to the values obtained at optimized structure (see Table 4). On the basis of this finding, one may speculate whether the consistent underestimation of the $Q_x$ peak energy in high-level approaches (Table 4) may in part be due to the idealized $D_{2h}$ structure assumed in these ab initio calculations.[15,38,39]

**Ionization Potentials.** Predicted and observed[44] ionization potentials of 46 organic molecules agree satisfactorily (Table 3 and Supporting Information). For NDDO-G the agreement with experimental results is noticeably better than for INDO/S; the mean absolute error of INDO/S, 0.40 eV, is reduced to 0.24 eV in NDDO-G. The results obtained (see Supporting Information) show that the both semiempirical schemes reproduce several higher ionization potentials with roughly the same accuracy as the first ionization energies.

**TABLE 5: Triplet Excitation Energies (in eV)**

|  | exptl[a] | NDDO-G | NDDO-G[b] |
|---|---|---|---|
| ethylene | 4.40 | 2.46 | 4.22 |
| 1,3-*trans*-butadiene | 3.20 | 1.58 | 2.81 |
|  | 4.95 | 2.92 | 4.77 |
| benzene | 3.95 | 2 22 | 3.80 |
|  | 4.76 | 3.68 | 5.17 |
|  | 5.60 | 4.68 | 6.06 |
| naphthalene | 2.98 | 1.63 | 2.81 |
|  | 3.82 | 2.68 | 4.21 |
| pyridine | 4.1 | 2.45 | 4.08 |
|  | 4.1 | 3.67 | 4.52 |
| furan | 3.99 | 2.95 | 3.77 |
|  | 5.22[g] | 3.17 | 4.84 |

[a] References 2, 6, 7, 8. [b] Parameters $\beta_\pi$ multiplied by a factor of 1.35.

**Limitations.** There is an important limitation that NDDO-G shares with INDO/S. Both methods predict triplet excited states significantly too low (Table 5). So far, we were not successful in developing a computational scheme that simultaneously reproduces singlet and triplet excitation energies with sufficient accuracy. Ridley and Zerner suggested using separate parameter sets for singlets and triplets.[45] Within NDDO-G, the results for triplet excitation energies can be considerably improved when the parameters $\beta_\pi$ are enlarged by a factor of 1.35 (Table 5).

## Conclusions

On the basis of the NDDO approximation, we have presented a unified semiempirical model, termed NDDO-G, for calculating structures and electronic spectra of molecules. The method was parametrized for the elements H, C, N, and O. Employing bond-type parameters in the NDDO-G expression of the core−core repulsion allows ground-state geometry optimizations. Molecular structures are predicted with an accuracy comparable to that of AM1 and PM3. Electronic absorption spectra are calculated via configuration interaction of singly excited states. The main difference between the NDDO-G and INDO/S models is the treatment of two-center two-electron integrals. The more accurate approach used in the NDDO-G procedure yields an improved performance for semiempirical calculations of excitation energies. Overall, the NDDO-G scheme seems to be accurate enough for a unified approach that simultaneously allows structure determination of organic molecules as well as conclusive assignments and semiquantitative predictions of vertical transitions in UV−vis and photoelectron spectra. There is an obvious direction for further work—to incorporate other main-group elements and transition metals into the NDDO-G scheme.

It may be remarked that the NDDO-G model discussed here seems little more accurate than is the INDO/S model in its predictions of electronic spectroscopy except for compounds containing oxygen where the NDDO-G model shows significant improvements. However, a recent reparametrization of oxygen by Li, Cramer, and Truhlar[46] improved the INDO/S results considerably. We do, however, expect NDDO-G results to be far superior in the case of ionic transition metal complexes, where it can be shown that the extra integrals are quite important in determining the ligand field. In addition, it should be emphasized that NDDO-G is a model for both spectroscopy and structure and as such has the potential to become a powerful tool in the study of photochemistry, whereas the present INDO/S model cannot be used for calculating structures. One might further remark in concluding, however, that the ability to calculate the geometry from the NDDO-G model seems more a consequence of the scaling of the nuclear repulsion energy than of the differences between the NDDO and INDO approximations themselves. This would be consistent with the findings of Jug and co-workers in the application of their SINDO model that claims an accuracy in geometry and thermochemistry equal to that of the NDDO models AM1 and PM3.[47,48]

**Acknowledgment.** This work has been supported by the Deutsche Forschungsgemeinschaft, the DAAD/NSF exchange program, and the Fond der Chemischen Industrie.

**Supporting Information Available:** Two tables (Tables 6 and 7) listing bond lengths and bond angles of the optimized structures and the calculated and observed vertical ionization potentials. This material is available free of charge via the Internet at http://pubs.acs.org.

## References and Notes

(1) Klessinger, M.; Michl, J. *Excited States and Photochemistry of Organic Molecules*; VCH: Weinheim, Germany, 1995.

(2) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frish, M. J. *J. Phys. Chem.* **1992**, *96*, 135.

(3) Zerner, M. C. In *Review of Computational Chemistry*; Lipkowitz K. B., Boyd, D. B., Eds.; VCH: New York, 1991; Vol. 2, p 313.

(4) Anderson, K.; Malmqvist, P.-A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483.

(5) Anderson, K.; Malmqvist, P.-A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218.

(6) Serrano-Andres, L.; Merchan, M.; Nebot-Gil, I.; Lindh, R.; Roos, B. O. *J. Chem. Phys.* **1993**, *98*, 3151.

(7) Fülscher, M. P.; Serrano-Andres, L.; Roos, B. O. *J. Am. Chem. Soc.* **1997**, *119*, 6168.

(8) Roos, B. O.; Anderson, K.; Fülscher, M. P. *Chem. Phys. Lett.* **1992**, *192*, 5.

(9) Lorentzon, J.; Fülscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 9265.

(10) Fülscher, M. P.; Roos, B. O. *J. Am. Chem. Soc.* **1995**, *117*, 2089.

(11) Stanton, J. F.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 7029.

(12) Del Bene, J. E.; Watts, J. D.; Bartlett, R. J. *J. Chem. Phys.* **1997**, *106*, 6051.

(13) Watts, J. D.; Gwaltney, S. R.; Bartlett, R. J. *J. Chem. Phys.* **1996**, *105*, 6979.

(14) Nooijen, M.; Bartlett, R. J. *J. Chem. Phys.* **1997**, *106*, 6441.

(15) Gwaltney, S. R.; Bartlett, R. J. *J. Chem. Phys.* **1998**, *108*, 6790.

(16) Nakatsuji, H. *Chem. Phys. Lett.* **1979**, *67*, 329.

(17) Thiel, W. *Adv. Chem. Phys.* **1996**, *93*, 703.

(18) Ridley, J. E.; Zerner, M. C. *Theor. Chim. Acta* **1973,** *32*, 134.

(19) Bacon, A. D.; Zerner, M. C. *Theor. Chim. Acta* **1979**, *53*, 21.

(20) Zerner, M. C.; Loew, G. H.; Kirchner, R. F.; Müller-Westenhof, U. T. *J. Am. Chem. Soc.* **1980**, *102*, 589.

(21) Anderson, W. P.; Edwards, W. D.; Zerner, M. C. *Inorg. Chem.* **1986**, *25*, 2728.

(22) Kotzian, M.; Rösch, N.; Zerner, M. C. *Theor. Chim. Acta* **1992**, *81*, 201.

(23) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.

(24) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(25) Dewar, M. J. S.; Zoebish, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1977**, *99*, 4899.

(26) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

(27) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391.

(28) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616.

(29) Thiel, W. In *Computational Thermochemistry*; Ikakura, K. K., Frurip, D. J., Eds.; ACS Symposium Series 677; American Chemical Society: Washington, DC, 1998; pp 142−161.

(30) Oleari, L.; DiSipio, L.; DeMichelis, G. *Mol. Phys.* **1966**, *10*, 97.

(31) Mataga, N.; Nishimoto, K. *Z. Phys. Chem.* **1957**, *13*, 140.

(32) Voityuk, A. A. SIBIQ 3.0, a program for semiempirical calculations, 1996.

(33) Zerner, M. C. ZINDO, a comprehensive semiempirical SCF/CI package, 1995.

(34) Hirota, E.; Iijima, T.; Kuchitsu, K.; Lafferty, W. J.; Ramsay, D. A. *Structure Data of Free Polyatomic Molecules*; Landolt-Boernstein, New Series; Springer: Berlin, 1992; Vols. II/21, II/15, II/7.

(35) Perkampus, H.-H. *UV−Vis Atlas of Organic Compounds*; VCH: Weinheim, 1992; Parts 1 and 2.

(36) Gouterman, M. In *The Porphyrines*; Dolphone, D., Ed.; Academic Press: New York, 1978; Part A.

(37) Zerner, M. C.; Baker, J. D. *Chem. Phys. Lett.* **1990**, *175*, 192.

(38) Merchan, M.; Orti, E.; Roos, B. O. *Chem. Phys. Lett.* **1994**, *226*, 27.

(39) Nagashima, U.; Takada, T.; Ohno, K. *J. Chem. Phys.* **1986**, *85*, 4524.

(40) Almlöf, J.; Fisher, T. H.; Gassman, P. G.; Ghosh, A.; Häser, M. *J. Phys. Chem.* **1993**, *97*, 10964.

(41) Reynolds, C. H. *J. Org. Chem.* **1988**, *53*, 6061.

(42) Chen, B. M.; Tulinsky, A. *J. Am. Chem. Soc.* **1972**, *94*, 4144.

(43) Anex, B. G.; Umans, R. S. *J. Am. Chem. Soc.* **1964**, *86*, 5026.

(44) Robinson, J. W., Ed. *Handbook of Spectroscopy*; CRC Press: Boca Raton, FL, 1980; Vol. 1.

(45) Ridley, J. E.; Zerner, M. C. *Theor. Chim. Acta* **1976**, *42*, 223.

(46) Li, J.; Williams, B.; Cramer, C. J.; Truhlar, D. G. To be published. We thank these authors for a preprint of this paper.

(47) Jug, K. Nanda, D. N. *Theor. Chim. Acta* **1980**, *57*, 107, 131.

(48) Jug, K.; Bredow, In *Computational Chemistry METECC-95*; Clementi, E., Corongiu, G., Eds.; STEF, Cagliari, 1995; p 89.