Assessment of semiempirical methods for the computation of charge transfer in DNA $\pi$-stacks

Alexander A. Voityuk *

Institució Catalana de Recerca i Estudis Avançats, Institute of Computational Chemistry, Universitat de Girona, 17071 Girona, Spain

Received 23 May 2006
Available online 4 July 2006

Abstract

Using CAS-PT2 and CASSCF charge transfer (CT) energies and electronic couplings as reference data, we assess the performance of several semiempirical methods in the calculation of hole transfer in $\pi$-stacks of nucleobases. INDO/S is the most reliable of the methods examined. It provides very good estimates for the key CT parameters. While the MNDO, AM1 and PM3 methods give reasonable CT energies, they fail to provide acceptable values of the coupling matrix elements in the stacks.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Charge transfer (CT) through organic $\pi$-stacks has attracted considerable attention because of its key importance for nanoelectronics and several biological processes. CT energies and donor–acceptor electronic couplings are key parameters which determine the efficiency of the charge transport [1–4]. Long-range CT in DNA is an important example of the charge migration mediated by $\pi$-stacks [5]. CT energies and especially electronic couplings are difficult to derive from experimental data, and therefore, quantum mechanical approaches are employed to estimate these parameter both in DNA [6–11] and other types of $\pi$-stacks [12–14].

The electronic coupling of the donor and acceptor diabatic states can be estimated on the basis of adiabatic CT excitation energies using an unitary transformation of adiabatic to diabatic states [2,15–18], and therefore, the accuracy of the calculated adiabatic energies will directly affect the coupling matrix elements. Because the CT energies of a $\pi$-stack with an excess charge (either hole or electron) are quite small, often only several tens of eV, accurate methods should be employed for the calculation. Single-reference $ab$ $initio$ and DFT approaches (CIS, TDDFT and related schemes) cannot be applied because of the known limitation of these methods with charge transfer excitations [19]. The CASSCF and CAS-PT2 methods are more adequate for such calculations, however, they are tractable only for relatively small systems [20]. Recently, based on the CASSCF and CAS-PT2 calculation of several models consisting of two and three nucleobases, it has been shown that reasonable estimates of electronic couplings in the stacks can be obtained using Koopmans’ approximation within the Hartree–Fock (HF) method [21]. However, even HF calculations of large systems are very time consuming. It has been found that individual nucleobases in a stack exhibit significant freedom of movement leading to large fluctuations of electronic couplings [22,23]. To take into account the influence of structural dynamics on the coupling matrix elements very efficient quantum chemical methods should be applied. As the semiempirical methods are by several orders of magnitude faster than $ab$ $initio$ and DFT methods, they may be very useful for treatment of CT in organic $\pi$-stacks. The semiempirical methods INDO/S [24], MNDO [25], AM1 [26] and PM3 [27] have been widely used to study charge transfer in $\pi$-stacks (see, for instance, Refs. [10–14]). However, the accuracy of these methods in the calculation of electronic couplings has not still been addressed.

* Fax: +34 972418356.
E-mail address: alexander.voityuk@icrea.es.
The purpose of this Letter is to assess the performance of the semiempirical methods comparing the calculated CT energies and electronic couplings with CAS-PT2 and CASSCF data published recently [21]. We will show that INDO/S provides good estimates of the CT parameters. While the MNDO, AM1 and PM3 methods give reasonable CT energies, they strongly underestimate the couplings matrix elements, and therefore, cannot be applied to model CT in π-stacks.

2. Method

The electronic couplings were calculated using the generalized Mulliken–Hush (GMH) method [16,17] and electronic functions obtained from the semiempirical techniques. The adiabatic energies and dipole matrix elements were computed using Koopmans’ approximation (see Refs. [6,8,9] for details). The following single-stranded stacks were considered: five dimers 1GG, 1GA, 1AG, 1GT and 1TG and two trimers 1GAG and 1GTG (nucleobases are given from the 5' end to the 3' end). The CAS-PT2 and CASSCF calculations of these systems have been recently performed [21]; the estimated CT energies and electronic couplings are used in the present study as reference data. Cartesian coordinates of the π-stacks are given in the supporting material of Ref. [21].

3. Results and discussion

3.1. CT energies

It has been shown that good estimates of the hole transfer energy in DNA stacks can be obtained at the HF level using Koopmans’ approximation [21]. Within this model the CT energy in a radical cation is calculated as the difference of the one-electron energies of the two highest occupied molecular orbitals HOMO and HOMO−1 in the corresponding neutral system. Then, for a donor–bridge–acceptor (d–b–a) system, e.g. 1GTG and 1GAG, the energy of hole transfer between the donor and bridge sites can be estimated as the difference of the HOMO and HOMO−2 orbital energies. Table 1 compares the semiempirical CT energies derived invoking the Koopmans approximation with the CAS-PT2 and CASSCF excitation energies of the radical cation. The INDO/S scheme is found to be the most reliable of the methods examined (cf. the standard deviation (SD) values listed in Table 1). Note that though the INDO/S method was parameterized at the CIS level of theory for calculating excitation energies of organic molecules, CIS calculations of the ground and excited states of the radical cations do not provide reliable values of the excitation energies. Thus, the computationally much more efficient approach based on the Koopmans approximation (as compared to CIS) allows one to obtain much better estimates of CT energies. The MNDO energies are also in good agreement with the reference data (Table 1). The largest deviations are found for the AM1 and PM3 methods. In some cases these schemes strongly underestimate the CT energies (see, for instance, AM1 and PM3 results for 1AG and 1GA, respectively).

3.2. Electronic couplings

First, we consider the results obtained for the dimers. Because in this case only two states (those of the donor and acceptor) determine the coupling matrix element, the two-state GMH treatment [16,17] was employed. Fig. 1 compares the calculated couplings of nucleobases in dimers. As can be seen, \( V \) is systematically overestimated by the HF method. Very good agreement is found for INDO/S. This method well reproduces the coupling magnitudes in 1GG, 1GA, and 1AG; however, it overestimates the strength of the coupling in 1GT (the predicted value is close to the HF result) and gives a somewhat smaller value for 1TG. As can be seen, the MNDO method considerably underestimates the couplings in all dimers. The AM1 and PM3 schemes (the data are not shown in Fig. 1) share this limitation with MNDO. Statistical evaluation of the performance of the methods is presented in Table 2. Comparison of the standard deviations and mean errors suggests that INDO/S performs the best of the methods examined. However, these criteria do not clearly reflect the difference in the performance of the semiempirical approaches. Actually the ratio of estimated and reference couplings characterizes the performance of a computational method \( M \). So we use a measure \( R(M) \) defined as

\[
R(M) = \frac{\sum_{i,j} |C_{ij}^{\text{est}} - C_{ij}^{\text{ref}}|}{\sum_{i,j} |C_{ij}^{\text{ref}}|}
\]

where \( C_{ij}^{\text{est}} \) and \( C_{ij}^{\text{ref}} \) are the estimated and reference couplings, respectively.

Table 1
CT excitation energies (in eV) calculated with different methods

<table>
<thead>
<tr>
<th>Method</th>
<th>1GG</th>
<th>1GA</th>
<th>1AG</th>
<th>1GT</th>
<th>1TG</th>
<th>1GTG</th>
<th>1GAG</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>d → a</td>
<td>d → b</td>
<td>d → a</td>
<td>d → b</td>
<td>d → a</td>
<td>d → b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ref.</td>
<td>0.392</td>
<td>0.560</td>
<td>0.340</td>
<td>1.175</td>
<td>0.797</td>
<td>0.178</td>
<td>1.239</td>
<td>0.118</td>
</tr>
<tr>
<td>HF</td>
<td>0.472</td>
<td>0.551</td>
<td>0.185</td>
<td>1.574</td>
<td>1.153</td>
<td>0.188</td>
<td>1.247</td>
<td>0.185</td>
</tr>
<tr>
<td>INDO/S</td>
<td>0.273</td>
<td>0.475</td>
<td>0.292</td>
<td>1.402</td>
<td>1.120</td>
<td>0.095</td>
<td>1.193</td>
<td>0.050</td>
</tr>
<tr>
<td>MNDO</td>
<td>0.337</td>
<td>0.365</td>
<td>0.103</td>
<td>1.196</td>
<td>0.830</td>
<td>0.115</td>
<td>0.905</td>
<td>0.082</td>
</tr>
<tr>
<td>AM1</td>
<td>0.343</td>
<td>0.277</td>
<td>0.036</td>
<td>1.039</td>
<td>0.682</td>
<td>0.127</td>
<td>0.738</td>
<td>0.073</td>
</tr>
<tr>
<td>PM3</td>
<td>0.366</td>
<td>0.081</td>
<td>0.151</td>
<td>0.956</td>
<td>0.568</td>
<td>0.104</td>
<td>0.664</td>
<td>0.074</td>
</tr>
</tbody>
</table>

a The CT energies for the dimer and trimer systems were computed with CAS-PT2 and CASSCF, respectively [21].
b The standard deviation of the estimated CT energies.
to distinguish the calculations. If the method $M$ overestimates (or underestimate) the coupling by a factor $f$, $R = \sqrt[f]{1 - \frac{1}{f}}$. The $R$ values are listed in Table 2. For instance, $R = 0.29$ found for INDO/S suggests that the mean z-value of $f$ is about 1.33, thus, a predicted coupling should deviate from the reference value by ~33%. The MNDO and related schemes provide considerably less accurate results. While the AM1 method provides more accurate results than MNDO and PM3 (Table 2) this method is essentially inferior to INDO/S. For AM1, $R = 1.22$ corresponds to $f \approx 3.2$. It means that the AM1 couplings are on the average three times smaller than the reference values. This estimation is in good agreement with Fig. 1.

In the stacks 1GTG and 1GAG consisting of three nucleobases, the hole donor (d) and acceptor (a) sites (guanine bases) are separated by the thymine and adenine bridges (b), respectively. Thus, the d–a coupling is mainly mediated due to the d–b and b–a electronic interactions. Recently it was shown that two-state approach fails to provide accurate values of $V_{d-a}$ for charge transfer in d–b–a $\pi$-stacks because of multistate effects and several adiabatic states must be accounted for [28,29]. However, the calculated values of $V$ do not converged with respect to the number of states included in the GMH model. It has been shown that reliable estimates of couplings are obtained when one adiabatic state per nucleobase is accounted for [30]. Thus, when estimating the couplings in 1GTG and 1GAG, three adiabatic states of the lowest energy were taken into account. Note, that the three-state GMH scheme was also employed to derive the reference CASSCF data [21].

The $R$ parameter listed in Table 1 suggests that the calculations provide less accurate coupling values in 1GTG and 1GAG than in the dimers. Fig. 2 compares the calculated results. The $V_{GG}$ couplings in the trimer stacks are essentially smaller than those found in 1GG (Fig. 1). INDO/S provides a very good estimate of $V_{GG}$ for 1GTG while remarkably underestimates this matrix element in 1GAG. This too small $V_{GG}$ value appears to be rather unexpected because the d–b and b–a interactions, $V_{GA}$ and $V_{AG}$, are well reproduced by INDO/S. Moreover, the donor–bridge CT energy calculated with INDO/S and CASSCF are found to be very similar, 0.317 and 0.345 eV, respectively (Table 1). Within the superexchange model, the d–a coupling mediated by a bridge can be expressed as [2]

$$V_{SE}^{da} \approx V_{db} V_{ba} / \Delta.$$
Because the quantities $V_{GA}$, $V_{AG}$ and $\Delta$ are well reproduced by INDO/S, one would also expect a good estimate for $V_{GG}$ in 1GAG. The remarkable difference of the calculated and reference couplings (Fig. 2) indicates that (1) the superexchange model becomes less accurate with decreasing energy gap (d–b CT energy) [2] and (2) the multistate effects considerably influence the donor–acceptor coupling [29,30]. Comparison of the INDO/S with the CASSCF reference data shows that INDO/S underestimates the relative energy of the d–a CT state in 1GAG; the INDO/S and CASSCF values are 0.050 and 0.118 eV, respectively (Table 1). As seen from Fig. 2, the MNDO couplings in 1GTG and 1GAG are much too weak. Table 2 suggests that also AM1 and PM3 estimates of the electronic couplings are essentially underestimated. The $R$ values calculated for the trimers are larger than the corresponding values for the dimer systems. This points to larger deviations between the semiempirical and reference data for 1GTG and 1GAG. It is possible that accounting for the effects of the dynamic electron correlation (PT2 corrections to CT excitations in the trimers) will decrease these deviations.

4. Concluding remarks

Based on the CAS-PT2 and CASSCF results [21] we assess the performance of the standard semiempirical methods in the calculation of CT energies and electronic couplings in DNA $\pi$-stacks. It has been shown that the INDO/S results obtained within the Koopmans approximation agree surprisingly well with the reference data from the high-level calculations. The accuracy of INDO/S is found to be better than that of the Hartee–Fock method. Good performance of INDO/S may also be expected in calculations of other organic $\pi$-stacks. Because the INDO/S method is computationally very efficient its combination with molecular dynamic simulations provides a quite reliable technique for computational modeling the charge movement in $\pi$-stacked materials. Important steps in this direction have been already done [12–14].

By contrast, the MNDO and related methods, AM1 and PM3, fail to provide reasonable values of the coupling matrix elements, and therefore, the application of these schemes to model the CT properties of $\pi$-stacks appears to be quite limited.

Acknowledgement

This work has been supported by the Spanish Ministerio de Educación y Ciencia, Project No. CTQ2005-04563.

References