PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

A fast but accurate excitonic simulation of the Electronic Circular Dichroism of Nucleic Acids: how can it be achieved? †

Daniele Loco[a], Sandro Jurinovich[a], Lorenzo Di Bari[a] and Benedetta Mennucci[a]*

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

We present and discuss a simple and fast computational approach to the calculation of electronic circular dichroism spectra of nucleic acids. It is based on a exciton model in which the couplings are obtained in terms of the full transition-charge distributions, as resulting from TDDFT methods applied on the individual nucleobases. We validated the method on two systems, a DNA G-quadruplex and a RNA β-hairpin whose solution structure has been accurately determined by means of NMR. We have shown that the different characteristics of composition and structure of the two systems can lead to quite important differences in the dependence of the accuracy of the simulation on the excitonic parameters. The accurate reproduction of the CD spectra together with their interpretation in terms of the excitonic composition suggest that this method may lend itself as a general computational tool to both predict the spectra of hypothetic structures and define clear relationships between structural and ECD properties.

1 Introduction

Electronic circular dichroism (ECD) is one of the most powerful and widely used techniques for investigating the structure of biomacromolecules in solution.[1,2] The great advantages of this technique include: a) overall sensitivity, which permits the investigation of minute samples; b) the possibility to use a variety of solvent compositions/conditions; c) the marked dependence of spectral features on molecular conformations.

Nucleic acids (NAs) are most suitable to ECD, because nucleobases are endowed with transitions in the near UV region. As a consequence, ECD spectra of NAs can be measured without major interferences from most salts and buffers, usually absorbing deeper in the UV; finally, the mere appearance of the ECD spectrum can be regarded as a NA structural fingerprint.[3] The intrinsic ECD of isolated nucleotides is in fact small compared to the entity of the Cotton effects found in oligo- or polymeric sequences, which may be attributed primarily to through-space interactions between the various bases.[4–6] A signature of this mechanism is the fact that NA ECD are often conservative, i.e. the integral between 180 and 300 nm is practically vanishing.

Due to the large interest in finding relationship between structural and ECD properties, the attempt to predict the ECD of NA by means of computational methods has a long history. The nucleobases are well-known chromophores with electronic absorption bands presenting an electric dipole character; for this reason, since the very beginning of the ECD studies, a coupled oscillators method appeared reasonably successful to account for the spectrum of various models of helices.[5,7] In much more recent times, the same approach was used to qualitatively account for the spectra of complex quadruplex structures.[8] Moreover, chemometric methods provide a simple and fast way for efficiently classifying NAs, based on datasets of ECD spectra corresponding to widely different NAs structures.[9]

The simulation of NA ECD has also been attempted by means of the so-called Matrix Method.[10,11] Within this framework, the excited states of the system (which can be seen as a set of interacting but independent units) are defined in terms of a basis of "artificial" states localized in the single units. The resulting excitonic transition energies will thus be obtained by solving an excitonic Hamiltonian built in terms of the localized transition energies, also called "site energies", and their electronic couplings.

Recently, an exciton modeling of the ECD spectra of DNA model systems has been presented[12] combining quantum-mechanical (QM) calculations with classical molecular dynamics (MD) simulations to account for the effect of structural fluctuations. A similar strategy has been used to simulate double helix B-DNA sequences.[13] In that study a hybrid time-dependent density functional theory (TDDFT)/molecular mechanics (MM) description was used to account for environment effects. In both cases, the calculation of the electronic couplings was based on a simpli-

[a] Dipartimento di Chimica e Chimica Industriale, University of Pisa Via G. Moruzzi 13, 56124 Pisa, Italy

* Corresponding author. E-mail benedetta.mennucci@unipi.it

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

fied formulation. In the former atomic (transition) charges obtained using singly excited configuration interaction (CIS) calculations based on a semiempirical wave function were used while in the latter case the point dipole approximation (PDA) was applied.[14,15]

These simplified formulations however do not seem to be easily and safely generalized to any NA structure. In particular, the placement of a point dipole on a chromophore devoid of inversion centre (such as the DNA bases) is not straightforward.[16] Moreover, this approximation is known to be inappropriate when the interchromophore distance is of the same order of the chromophores dimension, which is exactly what happen in NA structures. In spite of this potentially dangerous weakness of the approaches used so far to predict NA ECD spectra, a detailed analysis of the possible consequences of an approximated description of the couplings has never been presented so far whereas similar analyses have already appeared in other contexts, such as the simulation of fluorescence-detected resonance energy transfer (FRET).[17]

Another related aspect which has never been analyzed in the context of NA CD is the role played by the electronic couplings with respect to the site energies in determining the accuracy of the final excitonic spectra. This aspect has been instead deeply analyzed for other supramolecular structures such as artificial and natural light-harvesting complexes.[18–22] In those systems, however the picture can be rather different with respect to NA structures as the site energies involved are generally the same for all chromophoric units (which in most cases are identical). In NAs instead, the CD spectra comes from the interactions of different types of excitations in different nucleobases: this heterogeneity makes the computational simulation much harder as the number of excitonic parameters to be determined largely increases.

Here, we analyze all these problems that QM based methods necessarily encounter in the simulation of NA ECD spectra and we propose an approximated but effective protocol to make them numerically solvable within the limits of the approximations introduced. The protocol uses a QM exciton model (QM-EC) in which electronic couplings are obtained by explicitly calculating the Coulomb interaction of the full transition densities of the interacting units. These latter are obtained by solving the TDDDFT equations of each unit eventually taking into account the effects of the other units (and the solvent) through a classical polarizable MM embedding.[23] The accuracy and robustness of the protocol is evaluated by a direct comparison to experimental CD spectra of a RNA β-hairpin,[24] and a G-quadruplex.[25] The two selected applications have the merit of being very well structurally characterized in solution and as such they constitute an ideal playground to test computational methods.

2 Methods and computational details

In the present formulation of the exciton model, the excitonic states of the NA system are expressed as linear combinations of excitations localized on the single chromophoric units, namely

$$|\Psi_k^{Exc}\rangle = \sum_m^{N_{mol}} \sum_i^{n_{state}} C_{mi}^{(k)} |\phi_{mi}\rangle \qquad (1)$$

where $|\phi_{mi}\rangle$ is the product of the molecular electronic wave functions on the individual nucleobases, where nucleobase m is in the ith excited state and all other nucleobases are in the ground state; $C_{mi}^{(k)}$ is the corresponding coefficient. N_{mol} is the total number of nucleobases in the NA and n_{state} the number of local excitations for each nucleobase. The corresponding Hamiltonian is written as:

$$H^{Exc} = \sum_m^{N_{mol}} \sum_i^{n_{state}} \varepsilon_{mi} |\phi_{mi}\rangle \langle\phi_{mi}| + \sum_{m\neq n}^{N_{mol}} \sum_{i,j}^{n_{state}} V_{mi,nj} |\phi_{mi}\rangle \langle\phi_{nj}| \qquad (2)$$

where ε_{mi} are site energies of the ith excitation of the nucleobase m and $V_{mi,nj}$ the electronic couplings between to excitations i and j in nucleobase m and n respectively. The eigenvalues of the Hamiltonian correspond to the excitonic energies. The analysis of the eigenvector matrix also gives useful information about the degree of localization of the excitons by calculating the inverse participation ratio, also called exciton delocalization length L_k, reported in Eq. 3.

$$L_k = \frac{1}{\sum_m^{N_{mol}} \sum_i^{n_{state}} (C_{mi}^{(k)})^4} \qquad (3)$$

This parameter indicates the number of coherently coupled chromophores for the k-th excitonic state. For a completely delocalized state L_k corresponds to the number of exciton states whereas for completely localized states L_k is one.

In the computation of both site energies and couplings we eventually take into account the perturbation effects due to the environment in which the QM chromophores are embedded by using a polarizable Molecular Mechanics (MM) force field. In this MM-Pol approach the MM sites are described as a set of fixed point charges and polarizabilities. The polarization terms of the interactions are computed in terms of induced dipoles generated at each MM site a by the QM electron density and all other MM fixed charges and induced dipoles, namely:[26,27]

$$\vec{\mu}_a^{ind} = \alpha[\mathbf{E}^{QM} + \mathbf{E}^{chg} + \mathbf{E}^{\mu}(\{\mu\})] \qquad (4)$$

where each electric field $\mathbf{E}$ (due, respectively, to QM density, charges and the other induced dipoles) is evaluated at the MM-Pol site and α represents the polarisability of the classical site. This problem has to be solved in a *Self-Consistent* way through a matrix inversion or the numerical solution of the corresponding system of linear equations. The effects on the QM density due to the presence of the polarizable embedding are taken into account introducing appropriate interaction terms in the vacuum Hamiltonian operator, such as:

$$\hat{H}_{QM/MM}^{Pol} = \sum_m q_m^{MM} \hat{V}^{QM}(\vec{r}_m) - \sum_a \vec{\mu}_a^{ind} \cdot \hat{\mathbf{E}}^{QM}(\vec{r}_a) \qquad (5)$$

in which the first term represents the interaction of the QM electron density with the fixed charges and the second that with the induced dipoles. By adding the Hamiltonian reported in Eq. 5 to the in vacuum one and solving the resulting TDDFT equations,[26] we directly obtain site energies and couplings modified by the presence of a polarizable environment. More in details, the electronic coupling between the excitations localized on the chro-

mophores m and n can be calculated from the transition densities of the non-interacting chromophores plus an explicit MMPol-mediated term, namely:

$$V_{mn} = \int d\vec{r}' \int d\vec{r} \rho_m^T(\vec{r}') \frac{1}{|\vec{r}-\vec{r}'|} \rho_n^T(\vec{r}) + \\ + \sum_k \vec{\mu}_k^{ind} \left(\alpha; \rho_m^T\right) \int d\vec{r} \frac{\rho_n^T(\vec{r})(\vec{r}_k - \vec{r})}{|\vec{r}_k - \vec{r}|^3} \quad (6)$$

where we have assumed that the interactions between the transition densities are dominated by coupling terms (this is indeed the case for dipole-allowed excitations).

2.1 Rotational strength calculation

By definition the rotational strength of a CD electronic transition, namely the k←0 transition, is given by

$$R^k = Im[\langle 0|\hat{\mu}|k\rangle \cdot \langle k|\hat{m}|0\rangle] \quad (7)$$

where $\hat{\mu}$ and $\hat{m}$ are the electric and magnetic dipole operators respectively. To account for the problem of the origin-dependence of the magnetic dipole, we introduce the dipole-velocity formulation and exploit the correspondence of the magnetic and electric operators. Since the NB transitions have a predominant electric character, the simplified formulation of R^k, in which only the electric transition dipole moment operators appear, can be used, namely:

$$R^k = \frac{\pi}{2\lambda_k} \sum_{n,m=1}^{N_{mol}} \sum_{i,j=1}^{n_{trans}} c_{ni}^{(k)} c_{mj}^{(k)} \vec{r}_{nm} \cdot \vec{\mu}_{ni} \times \vec{\mu}_{mj} \quad (8)$$

where $\vec{r}_{nm}$ is the distance vector between the chromophores i and j and λ_k is the wavelength of the electronic transition k←0. The validity of this simplified formulation has been checked with respect to the full one using both the electric and the magnetic transition moments:[28] the comparison is reported in Fig. S1 of ESI†.

The coefficients $c_{ni}^{(k)}$ corresponds are the eigenvectors of the k-th excitonic state, coming from the diagonalization of the matrix in Eq. 2. This latter formula is the one we use in this work to compute the rotational strength of each excitonic state. At least, to obtain the CD spectra we apply a Gaussian broadening[29] to the excitonic rotational strength located at the corresponding energy in the spectrum so that

$$CD(a.u.) = \frac{1}{22.97} \sum_{k=1}^{N} \omega_k R^k g(\omega, k) \quad (9)$$

where ω_k is the excitonic energy relative to the k←0 transition and $g(\omega,k)$ the Gaussian distribution function, which we assume it is sufficient to take into account the broadening effects which occur during a real measurement.

2.2 Computational details

In the following analysis, all QM calculations will be performed at DFT level of theory using the hybrid meta-GGA functional M06-2X[30] functional with a development version of the Gaussian software.[31] This functional in fact has shown to give accurate excitations in organic molecules for both valence and charge transfer states.[32]

The excitonic analysis and the simulation of the all the spectra have been obtained by using the EXcitonic Analysis Tool (EXAT) developed in our group.[33]

As first step of the study we performed QM calculations for both the oligonucleotides on their NMR structures. Two different models have been tested, namely the VAC model where each chromophoric unit is assumed in gas-phase and an MMPol model where the same unit is surrounded by a molecular environment represented by the other units as fixed point charges and induced dipoles.

All QM calculations are performed on nucleobase geometries optimized for the isolated nucleobase at B3LYP/6-311G(d) level in C_s symmetry. When a specific experimental conformation is chosen, we replace the QM nucleobase projecting the optimized structure on the original one minimizing the RMSD between their corresponding atom position.

The atomic polarisabilites are taken from the work of Wang et al.[34] In the chosen polarisation model only the interaction from the 1-4 neighbour atoms forward are taken into account as contribution to the polarisation energy in a Thole smeared scheme[35,36] applying a linear screening function to the included interaction without any further scaling. A set of atomic charges, appropriate and consistent with the polarisation model, is computed following the Electrostatic Potential (ESP) method implemented in the Gaussian suite of programs and subsequently treated with a set of restrains by means of PolChat[37].

The atom charges are fitted to reproduce the electrostatic potential evaluated on a grid of points surrounding the molecule and due to its electronic density computed at DFT level. In addiction to the potential term due to the charge also the contribution of the point-dipoles induced by the charges is included in the fitting in order to take into account the polarisation effects,[38] so obtaining the Polarisable (P)-ESP by minimizing

$$J = \sum_i^m \left[V_i^{QM} - \sum_j^n \frac{q_j}{|\vec{r}_{ij}|} - \sum_j^n \frac{\vec{\mu}_j \vec{r}_{ij}}{|\vec{r}_{ij}|^3} \right]^2 + J_c, \quad (10)$$

where the i index runs over the m grid points while j runs over the n MM sites, with $|\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j|$ where $\mathbf{r}_i$ is the position vector of the ith grid point. The additive term J_c which we have to include in the minimizer of Eq. 10 arise from the need of partitioning the system in its chromophoric subunits. It takes into account the constrains which we have to impose to the system in order to retain physical essential properties for each subunit, like, and principally, the conservation of the negative (-1) charge of each nucleotide, due to the presence of the phosphate group, and the electroneutrality of nucleobases in each of them, in order to be consistent with the QM calculation which are performed on free nucleobase optimized structure. The expression for the total charge constrain is given by

$$J_c = \lambda \left(\sum_{k \in F}^{l} q_k - q_F^{tot} \right) \quad (11)$$

with q_F^{tot} the total charge of the fragment F, q_k the atom charge

of each atom of F and λ a Lagrangian Multiplier.

The complete the description of the computational protocol, we report some details computational and human costs of each of the four involved steps to go from the structure to the CD spectrum:

1 projection of the QM optimized structures on the experimental structures;

2 derivation of parameters for QM/MMPol calculations (if we want to include environmental effects);

3 QM or QM/MMPol calculations to obtain the excitonic parameters and electric transition dipole moments;

4 combination of the QM results to simulate the CD spectrum.

Steps 1-2 have to be performed once for each building block of NAs. This represents a big advantage in terms of CPU time because, once a reference database is obtained, the parameters can be easily transferred to every NA systems. In other words, the preparation steps do not depend on the dimension of the system. As it regards step 2, namely the parameterization of the polarizable force field for QM/MM calculations, the values of isotropic polarizabilities for the different atom types of interest are available in the literature, thus the only extra-parameters needed for the MMPol simulations are the atomic charges. From a computational point of view this step requires a single point calculation on each building block in order to obtain the electrostatic potential. After that, the charge derivation procedure is carried out by using the PolChat tool described above. On the contrary, the step 3 needs to be repeated for each configuration and it involves a CPU time which depends on the size of the system and the number of configurations. The bottleneck is represented by the electronic coupling calculations because, even if the computation of a single coupling is fast, many couplings have to be computed depending on the number of interacting chromophores and the electronic transition considered. Finally step 4 is rather fast as, once all the parameters are derived, the simulation of CD spectra is done by the EXAT tool with essentially no additional CPU costs.

3 Results

The molecular structures of the two biological molecules are experimentally determined by means of in solution NMR spectroscopy and they are available in the Protein Data Bank (PDB): the 2KOC RNA β-hairpin[24] and the 143D DNA G-quadruplex.[39] For each system a different number of possible conformations have been detected. In the text we will refer to them as N-RNA or N-GQ to indicate the N-th structure in the PDB file of RNA β-hairpin or DNA G-quadruplex respectively. The RMSD of the different conformations for both the systems are collected in Table S1 of ESI†.

3.1 RNA β-hairpin

The system under investigation is a 14-mer RNA hairpin consisting of a 5 base pairs helix capped by an highly stable tetraloop constituted by U6-U7-C8-G9 sequence (see Fig. 1). In the simulation of the excitonic spectra we included the first 5 base pairs plus the U6-G9 pair of the tetraloop. Even if this pair does not correspond to a standard Watson and Crick (WC) pair, it plays a not negligible contribution to determine the spectral shape because of the strong π-stacking interactions with the other base pairs (see Fig. S2 of ESI†). On the other hand, the U7 and U8 nucleobasis are not interacting and they assume a random orientation in the structures.

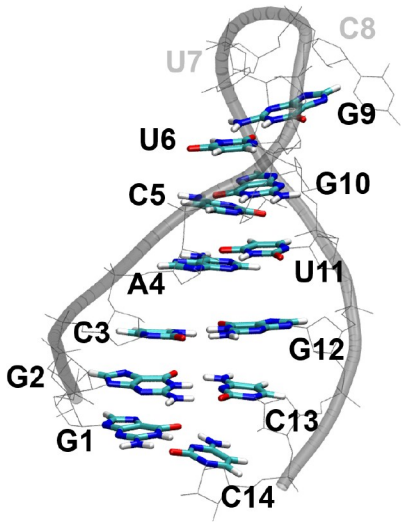

Fig. 1 Schematic representation of 2KOC RNA β-hairpin structure. The coloured nucleobases are those included in the exciton Hamiltonian. The rest of the system (other nucleobases and the RNA backbone) are represented with gray lines. The ribbon follows the backbone structure to point out the hairpin configuration.

In Fig. 2 we report two spectra calculated at M062X/6-311+G(d) level for the first NMR β-hairpin structure (1-RNA). The black spectrum is computed including up to 10 transitions of each nucleobase, while the other one, in red, with only the first $\pi-\pi^*$ transitions. This selection corresponds to the one reported in Table 1 while all the first ten excited states are collected in Tables S2-S5 of ESI†. In both cases a Gaussian broadening with a standard deviation of σ 0.21 eV has been used for all the excitonic lines (the same broadening will be used for all the following spectra).

From this comparison it is evident that the principal features of the CD spectrum are determined by the $\pi-\pi^*$ transitions. In light of this result, from now on, we will include only this restricted number of transitions in the excitonic Hamiltonian. This selection also allows us to simplify the following excitonic analysis and the data interpretation. Moreover, as this comparison shows that the Rydberg states do not significantly contribute to the CD spectra, it is well justified to try to explore the possibility of using a smaller basis set.

This analysis is reported in Table 1 where we compare the lowest $\pi-\pi^*$ transitions of each nucleobase computed with the previous basis set (6-311+G(d)) and with a different one using two instead of three basis functions for each valence atomic orbital and lacking the additional diffuse functions on the non-hydrogen atoms but still keeping the polarization ones, namely 6-31G(d).

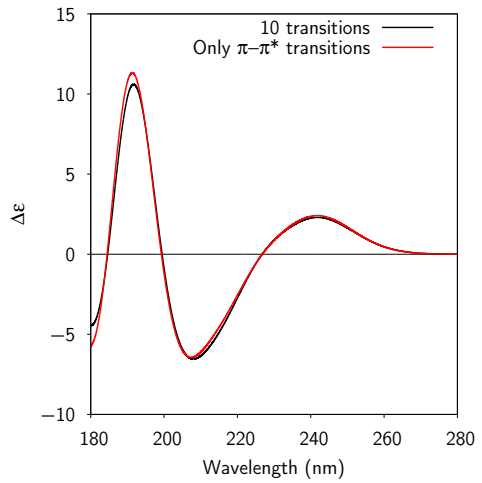

Fig. 2 Comparison between M062X/6-311+G(d) CD spectra computed on the 1-RNA structure, including up to 10 transitions of each nucleobase (black curve), and only the first $\pi - \pi^*$ transitions (red curve).

It is important to notice that the experimental values are measured in solution: therefore part of the discrepancy refers to solvent effects which are not taken into account in our TDDFT calculations. What comes out from Table 1 is that the differences between calculated and experimental energies strongly depend on the specific nucleobase. In particular they are larger for the two purines (Adenine and Guanine). However, the relative errors for the different nucleobases are similarly reproduced by the two basis sets; similar behaviours are shown by the oscillator strengths. This nucleobase-specificity of the error on the excitation energies can become extremely dangerous when an excitonic model is used, as the "site energies" used to in the excitonic Hamiltonian are not homogeneously shifted with respect to the experimental values and thus they introduce uncontrollable errors in the resulting excitonic states. To avoid this problem we can try to employ the experimental values as parametric site energies of the exciton matrix. Along this line, two alternative strategies have been tested: i) for each nucleobase we selected the transitions of interest and we assigned to them the corresponding experimental site energies (SS model); ii) a unique shift was applied to all the site energies of a specific nucleobases according to the corresponding mean error as reported in Table S6 of ESI† (ES model).

The two resulting CD spectra are reported in Fig. 3 for the smaller basis while for the larger basis set only the strategy i) is shown. All the calculated spectra are finally compared with the experimental ones.[45] Both the experimental and the calculated spectra have been normalized to the number of nucleobases considered in the exciton calculations.

We notice considerable changes with respect to the spectra of Fig. 2 showing the importance of the site energy accuracy. Looking at the large basis set spectrum and comparing it with the previous one and the experimental one, we clearly see that the overall spectral features have been largely improved. An almost negligible dependence of these features is instead found on the basis set. In fact, the two spectra obtained with the two basis

Table 1 TDM062X and experimental excitation energy (E, eV) and oscillator strength (f) for the $\pi - \pi^*$ lowest transitions of the four nucleobases. Excited states are labeled with the nucleobase first letter and the number in which the $\pi - \pi^*$ transition are ordered in the QM calculation outputs.

		Adenine			
State	**Exp**[a]	**6-31G(d)**		**6-311+G(d)**	
	E	E	f	E	f
A(1)	4.63	5.67	0.056	5.56	0.017
A(2)	4.77	5.58	0.248	5.39	0.298
A(3)	6.00	6.89	0.431	6.63	0.276
		Cytosine			
State	**Exp**[b]	**6-31G(d)**		**6-311+G(d)**	
	E	E	f	E	f
C(1)	4.6	5.13	0.066	5.02	0.074
C(2)	5.4	6.14	0.152	5.96	0.142
C(3)	6.3	7.02	0.437	6.59	0.413
		Guanine			
State	**Exp**[c]	**6-31G(d)**		**6-311+G(d)**	
	E	E	f	E	f
G(1)	4.45	5.34	0.172	5.18	0.176
G(2)	4.95	5.88	0.316	5.63	0.342
		Uracil			
State	**Exp**[d]	**6-31G(d)**		**6-311+G(d)**	
	E	E	f	E	f
U(1)	4.80	5.66	0.1868	5.51	0.202
U(2)	6.10	6.81	0.0402	6.69	0.034
U(3)	6.90	7.20	0.1568	6.88	0.185

[a] Values refer to aqueous solution.[40]
[b] Values refer to aqueous solution.[41]
[c] Averages values from different experimental data are reported.[42]
[d] Values refer to aqueous solution.[43,44]

sets are almost superimposable with only some slight differences in the relative intensities of the bands. This shows that the electronic couplings and transition dipoles are similarly described by the two basis sets. Evident differences are instead found when the small basis set spectrum is obtained applying the strategy ii) (ES model) for the site energy shifting: the use of an equal shift for all the transitions of each type of nucleobase significantly worsens the agreement with experiments by loosing almost completely the reproduction of the spectrum in low-energy region.

To confirm that these findings are not specific of the selected NMR structure, we have repeated the calculations on other four available structures, namely 2-RNA, 4-RNA, 8-RNA and 10-RNA. In Fig. 4 we compare the five spectra with respect to the average one.

All the spectra corresponding to different structures are almost superimposed. The only visible difference is in the intensity of the negative band around 250 nm. This is in agreement with the small structural differences among the experimental structures (the average RMSD of the four explored structures results 0.51 Å). For the largest RMSD structure, namely the 2-RNA, we also observe a different intensity in the positive band at around 260 nm. The agreement between the experimental spectrum in Fig. 3 with the average spectra in Fig. 4 is good, in particular for that concerns the relative bands intensities. Going into details, we can compare the energy difference between the two most distant peaks detectable in the experiment, namely the high energy negative peak at 225 nm and the lowest energy positive peak at 268

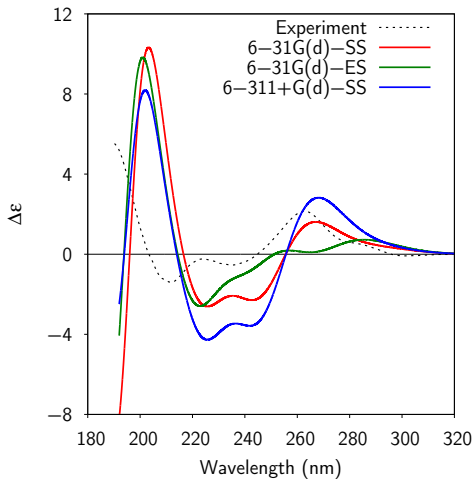

Fig. 3 Comparison between experimental ECD spectrum of 2KOC and calculated CD spectra with different basis sets. The label SS corresponds to the assignment of specific experimental site energy to each transition in each nucleobase; the label ES corresponds to the application of an equal site energy shift for all the transitions of each specific nucleobase. In all the simulated spectra a Gaussian broadening with a standard deviation of σ 0.21 eV has been used for all the excitonic lines.

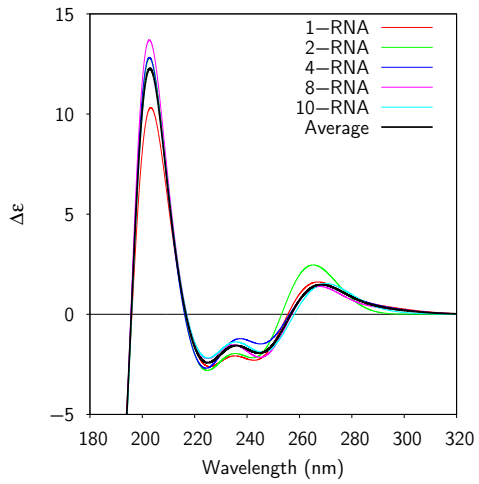

Fig. 4 Comparison of calculated CD spectra using 5 different NMR structures (thin coloured lines) and their average (black bold line). In all the simulated spectra a Gaussian broadening with a standard deviation of σ 0.21 eV has been used for all the excitonic lines.

nm with respect to the simulated spectra. Such energy difference is ∼0.9 eV in the computed average spectrum in contrast to 1.1 eV of the experiment.

As the correct choice of site energies represents a key point to obtain reliable results in the ECD spectra, we have further analyzed the effect of a different assignment of the first two $\pi-\pi^*$ transitions of adenine that are characterized by very different oscillator strengths. The energy order of these two transitions, commonly labelled as L_a (strong state) and L_b (weak state), is in fact still an open question.[40,46] In the spectra reported so far we used the assignment given by CCSD(T) calculations[47], namely the L_b state corresponds to the first peak in the experimental absorption spectrum (∼ 4.63 eV) and the L_a state to the second one (∼ 4.77 eV). However, we have also simulated the ECD spectra using the SS model for site energies with the other possibility (see Fig. S3 ESI) but no significant differences have been found. This low sensitivity of the present RNA system may be due to its nucleobases composition in which only one adenine is present.

The approach illustrated so far also gives us the possibility to understand the nature of the different bands in terms of the contribution of the different units. To achieve such an analysis we can look at the expansion coefficients of the excitonic wave function (see Eq. 1) obtained as eigenvectors of the Hamiltonian matrix. This analysis is summarized in Fig. 5 in which we report the excitonic signals for absorption and CD spectra represented as bars whose intensities corresponds to the square dipole moment and to the rotational strength, respectively. For this analysis we have used the 1-RNA structure, as representative of the average as just shown.

The excitonic states form eight distinct groups. Each color in Fig. 5 collects excitonic signals which are mainly determined by a specific electronic transition reported in the legend. For example, the group labeled with "G(2)" means that the corresponding signals mostly come from the second excitation of the guanines, following the notation used in Table 1.

The analysis of the different contributions indicates that the positive band in the CD spectrum at ∼260 nm is mainly determined by the two positive bands from excitonic states composed by the mixing of A(1)+C(1) (orange curve) and G(2) (green curve) transitions. The negative band at ∼250 nm is instead mainly due to excitonic states from the mixing of A(2)+U(1) transitions, whereas the negative band at ∼220 nm comes from the mixing of C(2) states. Finally, the strong intense positive peak at ∼200 nm is mainly determined by the mixing of C(3) states. The degree of localization of each exciton has been evaluated by computing the "exciton delocalization length", L_k, defined in Eq. 3. The obtained average L_k value of 2.1 (σ = 0.6) indicates that most of excitons are delocalized over two nucleobases. The minimum L_k value is 1.2 and it corresponds to an excitonic state strongly localized on the U6 nucleobase (blue stick in Fig. 5). The maximum L_k value of 3.6 corresponds instead to the excitonic state delocalized over the three π-stacked guanines (G9, G10, G12).

In general, the excitonic states are delocalized over the same units being the site energies separation of the nucleobases' transitions much larger than the electronic couplings. Within this picture, excitons mainly involve stacked units instead of the WC pairs and they therefore spread along the vertical axis of the macrostructure. The matrix collecting all the coefficients and the L_k values are reported in Table S7 of ESI†. It is important to add that, since this analysis has been done using an excitonic Hamiltonian with experimental site energies on its diagonal, and the experimental site energies are identical for equal nucleobases, the L_k values are surely overestimated. Moreover our calculations do not take into account any disorder effect which can play a role in the localization of the excitons.[12] However, we expect that the main findings here obtained about delocalization and excitonic

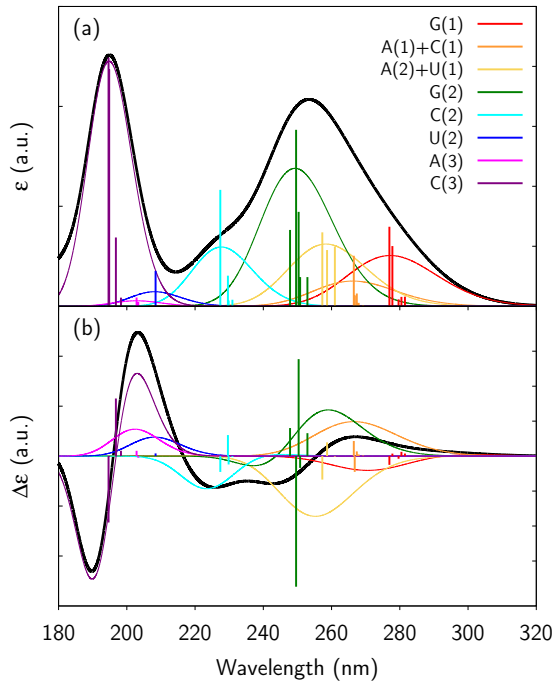

Fig. 5 Absorption (top) and CD (bottom) spectra of 1-RNA structure. Black curves represent the total spectra. Colored curves represent the contributions from specific groups of excitonic states identified by different colors. Vertical bars correspond to square dipole moments and rotational strengths.

contribution cannot largely change when adopting a more refined model, in fact the CD spectrum could not be so well reproduced if the excitonic picture was not only quantitatively but also qualitatively wrong.

To conclude the analysis on RNA, we analyze the role of the description of the electronic couplings. To do that we compare our approach based on the full transition densities with the point dipole approximation. The comparison between the two spectra obtained by keeping the same site energies and electric transition dipole moments used in the previous analysis while using either the PDA or the QM couplings are reported in Fig. 6.

As it can be seen from the comparison, the use of PDA changes the relative intensities of the bands, in particular in the region around 240 nm making the comparison with the experiments worse. These differences are mainly due to an overestimation of the inter-plane couplings in the PDA which changes the composition of the excitonic bands as shown in Fig. S4 of ESI†. This analysis shows that an accurate description of the excitonic states requires not only a correct estimate of the site energies but also of the electronic couplings. While the former are directly available from the experiments, the latter require some modellistic assumption and here it appears evident that the dipolar approximation is not accurate enough but a more complete estimation in terms of the Coulomb interaction of the full transition densities is required.

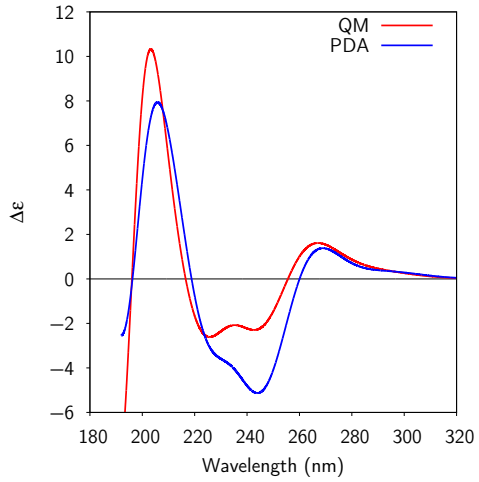

Fig. 6 Comparison between the CD spectra of 1-RNA structure computed using the QM electronic coupling (red curve) and the PDA model (blue curve).

3.2 DNA G-quadruplex

The same investigation presented before for β-hairpin is repeated here for a DNA G-quadruplex. G-quadruplexes are of many types but all of them are based on guanine tetrads. There are two basic types of the guanine quadruplexes with parallel and anti-parallel strand orientations, arising from different arrangements of *anti*/*syn* glycosidic angles. It has been observed that CD spectroscopy is able to discriminate between quadruplex topologies. In particular, the spectra of parallel quadruplexes have a dominant positive band at 260 nm, whereas the spectra of anti-parallel quadruplexes have a negative band at 260 nm and positive band at 290 nm.[3,48]

The G-quadruplex under investigation is rather larger than the previous RNA sequence (22 residues instead of 14) but there are 12 guanines organized in the so called quartets with an anti-parallel arrangement, as shown in Fig. 7. Within each plane (denoted by A, B and C) four nucleobases form a network of eight hydrogen bonds that hold them together in this characteristic planar configuration.

In addition to the guanine units, the overall structure also contains thymine and adenine nucleobases which can be neglected in the excitonic calculations because their contribution in the spectral region of interest is not significant. For this reason, in the following analysis, we will consider the guanine nucleobase only, for which we computed up to the 10th excited state at TDDFT level. As in the RNA case, we checked the difference in CD spectra by including all the computed transitions with respect to selecting only the two bright π-π^* ones (the results are collected in Fig. S5 of the ESI†). As expected, no significant differences have been observed between the two spectra thus we can limit the excitonic analysis to the π-π^* excitations of guanines.

In the case of RNA we have shown that the dependence of the accuracy of the calculated site energies on the specific transition and the specific nucleobase represents a delicate aspect in the application of the excitonic method. In the case of G-quadruplex, however the situation is much easier: we are in fact dealing with

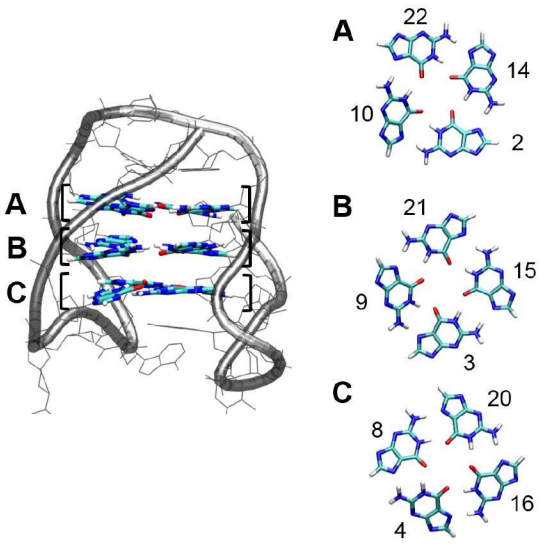

Fig. 7 Schematic representation of the 143D G-quadruplex structure (left). The three guanine tetrads included in the excitonic calculations appear in colored tube representation whereas the rest of nucleobaseis and the sugar backbone are represented with gray lines. The gray ribbon is also shown to point out the anti-parallel strand orientation. On the right we report the guanine nucleobases belonging to different tetrads observed from a top view. The guanine labels corresponds to those in the pdb structure.

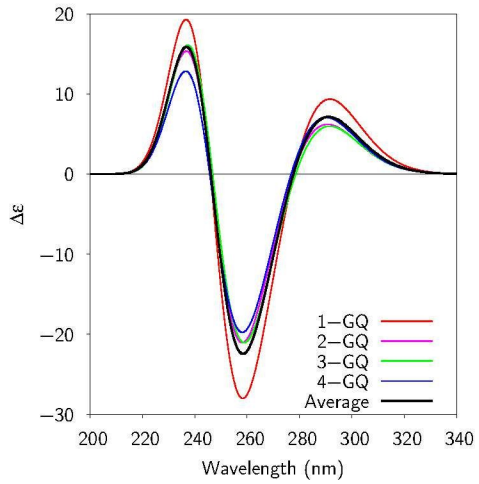

Fig. 8 Comparison between the calculated CD spectra of G-quadruplex using four different NMR structures (colored thin lines) and the average spectrum (bold black line). All the spectra have been shifted of -0.91 eV.

a multichromophoric system composed by a single type of nucleobases and, from Table 1, we observed that the errors for the two π-π^* excitations are almost the same. This allows us to use the site energies obtained from the TDDFT calculations, without including any external parameters or excitation-specific shift. The only required post-processing of the calculated excitonic spectrum is a global shift by the TDDFT error estimated for the guanine excitations so to have a fair comparison with the experimental spectrum.

As done before for β-hairpin also here we have repeated the excitonic simulation for four different NMR structures to investigate the possible structural freedom of the G-quadruplex. The simulated CD spectra for each structure are reported in Fig. 8 together with the average spectrum.

We note that the differences between the spectra of the selected structures and the average are larger than in the β-hairpin. This difference can be correlated with larger differences between the G-quadruplex conformers, as confirmed by an average RMSD of 0.80Å among the computed structures, instead of the 0.47 Å of the β-hairpin case. The larger difference in the spectra is observed between 1-GQ and 4-GQ that are characterized by the higher value of RMSD (0.80Å). The differences in the simulated spectra further confirm the capacity of our method to reproduce the expected sensitivity of the CD spectrum to even small structural differences.

To complete the description, in analogy to Fig. 5 we conducted an excitonic analysis on a selected experimental NMR structure (1-GQ). We recall that the two π-π^* transitions we have included in the model for each guanine are characterized by two transition dipoles directly almost along the short and the long axis of the molecule (see Fig. 9 and Table S4 of ESI†) and they are largely separated in energy (0.55 eV); due to this energy gap they do not mix in the excitonic states.

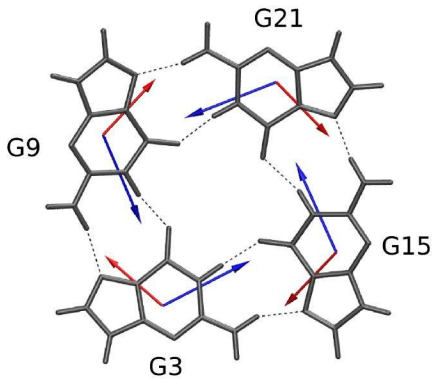

Fig. 9 Plane B of 1-GQ structure. Red and blue vectors indicates the electric transition dipole moments of G(1) and G(2) excited states respectively. The network of H-bond between adjacent guanines is also showed.

In the excitonic absorption spectrum shown in Fig. 10a we observe two distinct groups of excitonic transitions related to the G(1) and G(2) states. As expected, the spectrum of the aggregate differs from that of the single guanine. In particular the bright excitonic states resulted blue-shifted with respect to the monomer absorption, suggesting that the inter-plane interactions win on the intra-plane ones, as it happens in a typical spectrum of H-like aggregates. The CD spectrum reported in Fig. 10b shows the typical shape of the anti-parallel G-quadruplex arrangement: it is the result of a convolution of two asymmetric couplets corresponding to G(1) and G(2) excitations. In particular, the characteristic strong negative signal around 260 nm is dominated by the G(2) excitations.

Contrary to what we observed in the β-hairpin case, here, the

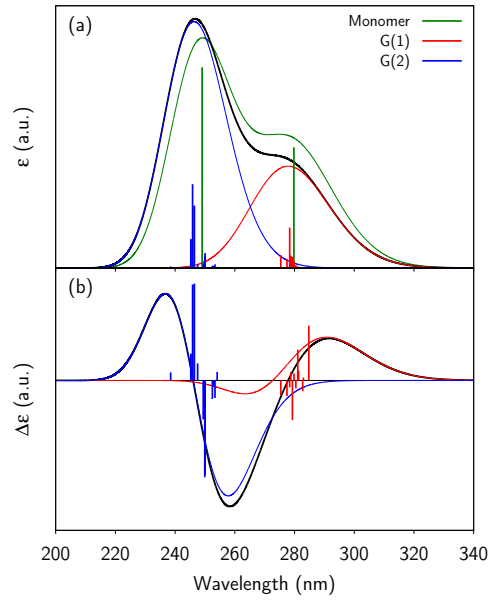

Fig. 10 Simulated Absorption (a) and CD (b) spectra of the 1-GQ structure. The black curve represents the total spectrum, whereas the red and blue curves are obtained as the sum of the exctionic states from the mixing of G(1) and G(2) excited states respectively. Square dipole moments and rotational strengths are also shown as vertical bars. For the absorption, the green curve corresponds to the computed guanine monomer spectrum. All the spectra have been shifted of -0.91 eV.

exciton delocalization length is totally determined by the couplings (the site energies are in fact degenerate within each of the two subsets). The average value of L_k is 5.3 with a maximum value of L_k is 10.8 for the highest excitonic state. Moreover the two sets of excitonic states present a similar delocalization scheme. The entire coefficient matrix and the L_k values are reported in Table S8 in the ESI†.

The network of interactions within each plane is almost the same because of similar inter-guanine distances and relative orientations. Each guanine strongly interacts with the two H-bonded units (i.e. the G3-G9 and G3-G10 in Fig. 9) with an average coupling of 131 cm^{-1} for G(1) transitions and 122 cm^{-1} for G(2) transitions. The average coupling between the opposite pairs within the planes (i.e. G3-G21) is much smaller for G(1) (25 cm^{-1}), while it is of the same order of those between the "adjacent" units for G(2) transitions (120 cm^{-1}) . On the contrary, the electronic couplings between guanines belonging to different planes are sensitive to the supramolecular arrangement, in particular to the relative twist of the guanine tetrads planes. For example, if we consider the coupling between the G(1) transition of G3, located in the central plane B, with the corresponding bases above (G2) and below (G4), the G3-G2 coupling is 181 cm^{-1}, but the G3-G4 couplings is 67 cm^{-1}. In the first case the transition dipole moments of the two basis are almost parallel, but in the latter case they form an angle of about 45° leading to a weaker Coulomb interaction.

As done before in the β-hairpin system, also here we have tested the performance of the PDA model. Thanks to the higher symmetry of the G-quadruplex we are able to investigate more in details the effects of the intra- and inter-plane interactions. In Fig. 11 we report the correlation between the PDA and QM couplings for the the two transitions in different pairs.

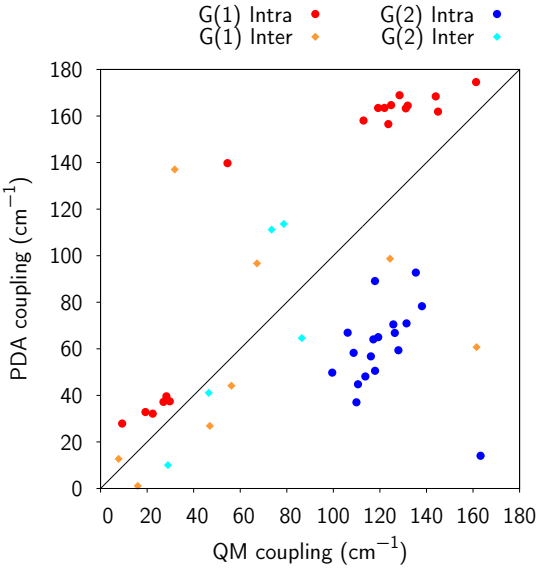

Fig. 11 Correlations between QM and PDA electronic couplings for the G(1) and G(2) states of guanines. Intra- label indicates couplings between guanines belonging to the same plane; Inter- label indicates couplings between consecutive guanine belonging to different planes (for example: G2, G3, G4).

From this graph it is evident that the intra-plane interactions (red and blue points) are qualitatively well described at PDA level even if for the G(1) transition a general overestimation is found while the opposite applies to the G(2) transition. In the case of inter-plane couplings instead the correlation is worse with PDA giving a general overestimation. These findings are not unexpected as the coupling in π-stacked pairs cannot be accurately described by a dipole-dipole approximation but they require at least a multipolar description.

What is less expected is instead the large similarity of the resulting CD spectra obtained with the QM or PDA models as reported in Fig. 12. This shows that the overall CD shape is not so sensitive to the details of the electronic couplings when we assume an exact degeneracy among the units: this corresponds to neglecting any disorder induced by the different local molecular environment.

To try to quantify this effect we have repeated the excitonic simulation taking into account the fact that each guanine units is embedded in the electrostatic field of the others. To do that we selected an NMR structure and applied a QM/MMPol approach in which each QM unit interacts with all the other units represented by atomic point charges and induced dipoles (see computational details). In Fig. 13, the resulting spectrum is compared with the one obtained for the isolated nucleobases and with the experimental one.

The QM/MMPol description modifies the spectrum with respect

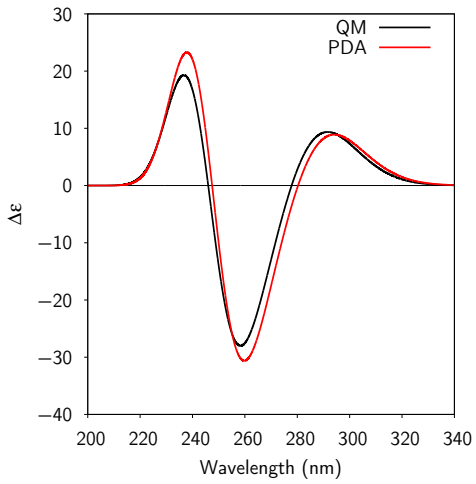

Fig. 12 Comparison between CD spectra of 1-GQ structure computed with QM (black curve) and PDA (red curve) cupling models. All the spectra are shifted by -0.91 eV.

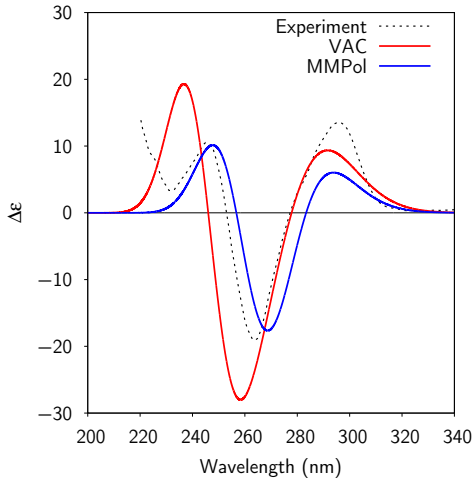

Fig. 13 (Experimental CD spectrum of 143D compared with simulated CD spectra of 1-GQ structure in vacuum and using the QM/MMPol model.

to the VAC one; in particular the intensity of the positive band at ∼230 nm is almost one half of the corresponding one obtained for the VAC spectrum; also the negative band at ∼260 nm is less intense in the MMPol simulation. The same peaks are also shifted towards lower energies in the MMPol simulation. The positive large band at ∼290 nm is almost unaffected, except for a smaller broadening in the QM/MMPol case. To have a more quantitative evaluation of the differences, in Table 2 we report the position of the three main bands of the CD spectrum, their relative energy differences, the peak intensities and their ratio.

From Fig. 13 and Table 2, it is evident both calculated spectra correctly reproduce the experimental shape. However, the MMPol spectrum seems to behave better: the relative intensity of the positive band around 250 nm which in the VAC spectrum is too intense with respect to the other two bands, is here correctly predicted as seen from the ratios I/III. Also the reproduction of I/II intensity ratio improves when the MMPol description is used. To

Table 2 λ_{max} of the first three main bands of CD spectrum and their relative energy difference (ΔE) (left part); CD band intensities and their absolute ratio (right part).

	λ_{max} (nm)			CD peak Δε		
	I	II	III	I	II	III
Exp	296	264	246	14	-19	11
MMPol	293	268	247	7	-20	10
VAC	294	261	239	9	-28	19
	ΔE (cm^{-1})			Δε ratio		
	II-I	III-I	III-II	I/II	I/III	II/III
Exp	4138	6949	2811	0.7	1.3	1.7
MMPol	3216	6374	3158	0.4	0.7	1.9
VAC	4414	7952	3538	0.3	0.5	1.5

investigate the origin of these differences we have dissected the MMPol effects on of three main sets of parameters, namely site energies, transition dipoles and couplings. This analysis is summarized in Fig. 14 where we report the spectra obtained with three "artificial" models in which each VAC set of parameters has been changed into the MMPol analog while the other two sets are kept unchanged.

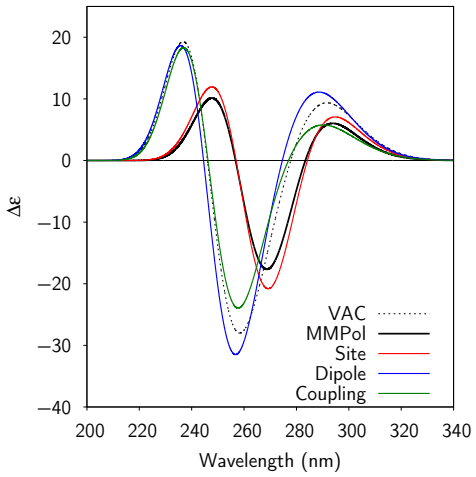

Fig. 14 Comparison of excitonic spectra obtained: dotted black) VAC, black) MMPol, red) VAC with MMPol site energies, blue) VAC with MMPol transition dipoles, green) VAC with MMPol couplings.

From the comparison we can conclude that:

- passing from VAC to the MMPol descriptions the transition dipole moments change both in orientation and in magnitude. Following the convention illustrated in Table S4 of ESI†, the variation of the angle between the transition dipole and the molecular long axis ($\Delta\varphi$) is $+25°$ and $+14°$ for G(1) and G(2) transitions, respectively. The corresponding changes in the transition dipole lengths are $\Delta\mu = -8\%$ and $\Delta\mu = +12\%$. Although these changes due to the local environment are significant, the net effect on the spectrum is negligible.

- the electronic couplings are modified by passing from VAC to MMPol because of the modification of transition densities and by the screening effect: as a result the intensities of

bands I and II decrease, but the overall effect on the spectrum remains small.

- in the VAC model site energies are degenerate for each transition while in the MMPol model the environmental effects induce site energy shifts that are unit dependent. This is reflected in a significant effect on the CD spectrum towards a close agreement to the full QM/MMPol simulation.

This analysis clearly shows that in a symmetric system like G-quadruplex lowering the degeneracy of site energies represents the most significant perturbation of the spectrum while changes in transition dipoles and also in electronic couplings have much less effect. This behaviour is rather different with respect to what observed for the β-hairpin where changes of the sole couplings (as those due to the PDA description) induced changes in the overall shape of the spectrum exactly as changes in the site energies. These results suggest that effects of possible disorder in both site energies and couplings have to be carefully taken into account for a correct simulation of the CD spectra of DNA structures.

4 Conclusion

Electronic CD spectroscopy is a fast, sensitive and selective probe of the structure of nucleic acids and it has been used for decades to discriminate between folds. Currently, solution structures of these macromolecules must be obtained using rather sophisticated NMR experiments, which are moreover limited by specific requirement as to sample quantity, isotope labeling, solvent or more in general medium. CD does not suffer from any of these problems, but on the other hand (unlike NMR) it does not provide measurements of interatomic distances or dihedrals and must totally rely on comparison between spectra. A general and reliable tool to predict a CD spectrum from a 3D structure, based on a full quantum chemical description is made difficult by several problems: the quality of the excited state description; the correct simulation of structural solvent effects; the overwhelming computational effort, owing to the large number of atoms of even the smallest systems of practical relevance.

We proposed a simple and fast approach based on a TDDFT estimation of the excitonic parameters. The idea is to limit at most the complexity of the excited state calculations using only $\pi-\pi^*$ excitations and neglecting any internal geometrical deformation for each nucleobase. Possible electrostatic and polarization effects due the other nucleobases are instead taken into account in terms of a polarizable embedding based on induced dipoles. For each $\pi-\pi^*$ excitation of a given unit we calculate the couplings with those of the other units by explicitly solving the Coulomb integrals between the corresponding TDDFT transition densities. By so doing we avoid the point dipole approximation, which becomes particularly critical, when the dimension of the chromophores is comparable to their spacing, as it naturally occurs in nucleic acids.

We showed that one must fine-tune the site energies of individual chromophore using the experimental data for the selected transitions of the nucleobases. On the contrary, for that concerns transition dipoles and electronic couplings, in our model we do not need to employ any external parameters and in principle the model can be applied to any DFT functional or even beyond: extensions to more correlated methods such as EOM-CC or multireference approaches are possible if the method gives access to transition densities. In any case, the computational cost remains limited to QM/MM calculations on the single nucleobases.

We applied our method to two very different systems (a β-hairpin and a G-quadruplex), whose solution structure is accurately known and can be considered two benchmarks for its validation. We have shown that the different characteristics of composition and structure of the two systems can lead to quite important differences in the dependence of the accuracy of the simulation on the excitonic parameters. In a symmetric system like G-quadruplex lowering the degeneracy of site energies represents the most significant perturbation of the spectrum while changes in transition dipoles and also in electronic couplings have much less effect. In the β-hairpin instead changes of the couplings (as those due to the PDA description) induces changes in the overall shape of the spectrum exactly as changes in the site energies.

The two systems instead behave similarly for that concerns the role of charge-transfer contributions to the excitations: in both cases it seems that the main features of the CD spectra are correctly reproduced without these effects. However, their inclusion is always possible by extending the definition of the minimal unit to at least a dimer of nucleobases. This dimeric description has been already presented and shown to accurately describe the CD spectra of different DNA systems.[49,50]

Acknowledgments

The European Research Council (ERC) through the Starting Grant proposal n. 277755 (EnLight) is acknowledged by S.J. and B.M..

References

1 G. D. Fasman, *Circular Dichroism and the Conformational Analysis of Biomolecules*, Springer-Verlag, New York, NJ, USA, 2001.

2 J. Kypr, I. Kejnovska, K. Bednarova and M. Vorlìčková, *Circular Dichroism Spectroscopy of Nucleic Acids in Comprehensive Chiroptical Spectroscopy*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012.

3 J. Kypr, I. Kejnovská, D. Renčiuk and M. Vorlíčková, *Nucleic Acids Res.*, 2009, **37**, 1713–1725.

4 W. C. Johnson and I. Tinoco, *Biopolymers*, 1969, **7**, 727–749.

5 C. L. Cech, W. Hug and I. Tinoco, *Biopolymers*, 1976, **15**, 131–152.

6 S. Masiero, R. Trotta, S. Pieraccini, S. De Tito, R. Perone, A. Randazzo and G. P. Spada, *Org. Biomol. Chem.*, 2010, **8**, 2683–2692.

7 W. C. Johnson and I. Tinoco, *Biopolymers*, 1969, **7**, 727–749.

8 A. I. Karsisiotis, N. M. A. Hessari, E. Novellino, G. P. Spada, A. Randazzo and M. Webba Da Silva, *Angew. Chem. Int. Ed.*, 2011, **50**, 10645–10648.

9 J. Jaumot, R. Eritja, S. Navea and R. Gargallo, *Anal. Chim. Acta*, 2009, **642**, 117–126.

10 V. Rizzo and J. A. Schellman, *Biopolymers*, 1984, **23**, 435–470.

11 D. S. Moore and A. L. Williams, *Biopolymers*, 1986, **25**, 1461–1491.

12 S. Patwardhan, S. Tonzani, F. D. Lewis, L. D. A. Siebbeles, G. C. Schatz and F. C. Grozema, *J. Phys. Chem. B*, 2012, **116**, 11447–11458.

13 H. Gattuso, X. Assfeld and A. Monari, *Theo. Chim. Acta*, 2015, **134**, 36–8.

14 D. M. Gray, J.-D. Wen, C. W. Gray, R. Repges, C. Repges, G. Raabe and J. Fleischhauer, *Chirality*, 2008, **20**, 431–440.

15 R. Repges, C. Beuck, E. Weinhold, G. Raabe and J. Fleischhauer, *Chirality*, 2008, **20**, 978–984.

16 B. Bouvier, T. Gustavsson, D. Markovitsi and P. Millié, *Chem. Phys.*, 2002, **275**, 75 – 92.

17 A. Munoz-Losa, C. Curutchet, B. P. Krueger, L. R. Hartsell and B. Mennucci, *Biophys J*, 2009, **96**, 4779–4788.

18 T. Renger and F. Müh, *Photosynth Res*, 2012, **111**, 47–52.

19 F. Müh and T. Renger, *Biochim. Biophys. Acta*, 2012, **1817**, 1446–1460.

20 Z.-Q. You and C.-P. Hsu, *Int J Quantum Chem*, 2013, **114**, 102–115.

21 S. Jurinovich, L. Viani, I. G. Prandi, T. Renger and B. Mennucci, *Phys Chem Chem Phys*, 2015, **17**, 14405–14416.

22 S. Jurinovich, L. Viani, C. Curutchet and B. Mennucci, *Phys. Chem. Chem. Phys.*, 2015, doi=10.1039/C5CP00986C.

23 S. Jurinovich, G. Pescitelli, L. DiBari and B. Mennucci, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16407–16418.

24 S. Nozinovic, B. Fürtig, H. R. a. Jonker, C. Richter and H. Schwalbe, *Nucleic Acids Res.*, 2010, **38**, 683–94.

25 L. Martino, A. Virno, A. Randazzo, A. Virgilio, V. Esposito, C. Giancola, M. Bucci, G. Cirino and L. Mayol, *Nucleic Acids Res.*, 2006, **34**, 6653–6662.

26 C. Curutchet, A. Munoz-Losa, S. Monti, J. Kongsted, G. D. Scholes and B. Mennucci, *J. Chem. Theo. Comp.*, 2009, **5**, 1838–1848.

27 S. Caprasecca, S. Jurinovich, L. Viani, C. Curutchet and B. Mennucci, *J. Chem. Theo. Comp.*, 2014, **10**, 1588–1598.

28 S. Jurinovich, C. a. Guido, T. Bruhn, G. Pescitelli and B. Mennucci, *Chem. Commun.*, 2015, **51**, 10498–10501.

29 T. Bruhn, A. Schaumlöffel, Y. Hemberger and G. Bringmann, *Chirality*, 2013, **25**, 243–249.

30 Y. Zhao and D. G. Truhlar, *Theo. Chem. Acc.*, 2008, **120**, 215–241.

31 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ã. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian09 Revision D.01*, Gaussian Inc. Wallingford, CT, 2009.

32 A. D. Laurent and D. Jacquemin, *Int. J. Quantum Chem.*, 2013, **113**, 2019–2039.

33 S. Jurinovich, C. A. Guido, L. Cupellini and B. Mennucci, *EXAT - EXcitonic Analysis Tool*, 2014, Department of Chemistry, University of Pisa, Italy. URL: www.dcci.unipi.it/molecolab/tools.

34 J. Wang, P. , J. Li, T. Hou, R. Luo and Y. Duan, *J. Phys. Chem. B*, 2011, **115**, 3091–3099.

35 B. Thole, *Chem. Phys.*, 1981, **59**, 341–350.

36 P. T. van Duijnen and M. Swart, *J. Phys. Chem. A*, 1998, **102**, 2399–2407.

37 S. Caprasecca, C. Curutchet, S. Jurinovich and B. Mennucci, *POLCHAT - A polarisation-consistent charge-fitting tool, Department of Chemistry, University of Pisa*, 2014, www.dcci.unipi.it/molecolab/tools.

38 P. Cieplak, J. Caldwell and P. Kollman, *J. Comput. Chem.*, 2001, **22**, 1048–1057.

39 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.

40 W. Voelter, R. Records, E. Bunnenberg and C. Djerassi, *J. Am. Chem. Soc.*, 1968, **90**, 6163–6170.

41 D. Voet, W. B. Gratzer, R. A. Cox and P. Doty, *Biopolymers*, 1963, **1**, 193–208.

42 M. P. Fuelsche, L. Serrano-Andres and B. O. Roos, *J. Am. Chem. Soc.*, 1997, **119**, 6168–6176.

43 L. B. Clark and I. Tinoco, *J. Am. Chem. Soc.*, 1965, **87**, 11–15.

44 L. B. Clark, G. G. Peschel and I. Tinoco, *J. Phys. Chem.*, 1965, **69**, 3615–3618.

45 M. Felletti, *Solid-state NMR of non-crystalline recombinant viral particles*, MSc thesis, University of Pisa (Italy), https://etd.adm.unipi.it/t/etd-09062011-183641, 2011.

46 F. Santoro, R. Improta, T. Fahleson, J. Kauczor, P. Norman and S. Coriani, *J. Phys. Chem. Lett.*, 2014, **5**, 1806–1811.

47 P. G. Szalay, T. Watson, A. Perera, V. Lotrich, G. Fogarasi and R. J. Bartlett, *J. Phys. Chem. A*, 2012, **116**, 8851–60.

48 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–5415.

49 F. Di Meo, M. N. Pedersen, J. Rubio-Magnieto, M. Surin, M. Linares and P. Norman, *J. Phys. Chem. Lett.*, 2015, **6**, 355–359.

50 P. Norman, J. Parello, P. L. Polavarapu and M. Linares, *Phys. Chem. Chem. Phys.*, 2015, **17**, 21866–21879.