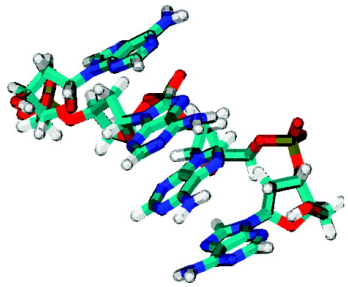
Article

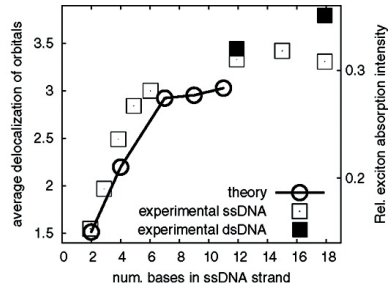
# Electronic Excitations and Spectra in Single-Stranded DNA

Stefano Tonzani, and George C. Schatz

*J. Am. Chem. Soc.*, **2008**, 130 (24), 7607-7612 • DOI: 10.1021/ja7103894 • Publication Date (Web): 21 May 2008








## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 2 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES

Published on Web 05/21/2008

# Electronic Excitations and Spectra in Single-Stranded DNA

Stefano Tonzani* and George C. Schatz

*Department of Chemistry, Northwestern University, Evanston, Illinois 60208-3113*

Received November 16, 2007; E-mail: tonzani@northwestern.edu

***Abstract:*** Using density functional theory and molecular dynamics simulations, we show that delocalized states extending over three bases can be directly excited in single-stranded poly(A) DNA. The results are in semiquantitative agreement with recent experimental results for the delocalization length of these states in single- and double-stranded DNA. The structures used in these molecular dynamics calculations are validated by comparing calculated circular dichroic spectra for $d(A)_2$ and $d(A)_4$ with experiment. These spectra, which arise from highly stacked structures, are in good agreement with experiment, suggesting that the short delocalization in ssDNA arises in spite of strong stacking.

## 1. Introduction

Interest in the nature of excited singlet states in DNA strands has been strong since the 1960s,[1] as a result of their importance in chemical reactions and ultimately damage and mutations caused by UV radiation.[2] In recent years, with the advent of sophisticated experimental tools such as ultrafast lasers and accurate theoretical approaches, the subject has become quite controversial. Much of this controversy stems from the amount of delocalization of the excited states in the $\pi$ stack, with different pieces of evidence pointing to different answers. For example, circular dichroic (CD) spectra for modified DNA oligomers show strong coupling between nearest neighbors,[3] while the similarity of UV spectra to those of the monomer has led to speculation that the excited states are actually localized on a single base.[4]

In recent years, experimental evidence for the evolution of electronic excitations has led to a qualitative distinction between the monomer bases and stacked single/double-stranded DNAs, in that monomers have only an ultrafast relaxation component, while polymers also possess a component on the order of 100 ps. This, combined with the fact that concentrated solutions of nucleosides at low temperature exhibit a red-shifted emission,[5] has resulted in different hypotheses concerning the nature of the excited states. One school of thought claims that the excitation is localized, and then the interaction between one excited monomer and a ground-state neighboring subunit leads to the formation of excimers,[6] therefore (at least in the classical sense) with delocalization shared over only two adjacent bases and formed only several picoseconds after the excitation. The second school of thought stresses instead that the excitation is delocalized from the outset,[7] its spatial extent is several bases,[7,8] and subsequent electronic relaxation does not change this fundamental characteristic, at least on a time scale of a few picoseconds.[8]

A further complication is added by the possible relaxation pathways, which can be intrastrand but also interstrand (e.g., proton transfer), as has been observed in gas-phase oligomers in experiments and calculations.[9] It has recently been suggested that base stacking is mostly responsible for decay mechanisms,[6] and the similarity in decay rates between single- and double-stranded oligomers would also imply that the structure of single-stranded DNA (ssDNA) is actually rather similar to that of double-stranded DNA (dsDNA). CD and NMR spectra[10–12] suggest the same evidence. On the other hand, theoretical modeling of ssDNA with empirical force fields is not, to our knowledge, well developed[13–15] compared to modeling of its double-stranded counterpart, and especially missing is a connection with the spectra cited above and with detailed NMR spectra from ref 16. Such a connection might help resolve the performance of force fields concerning the structure of ssDNA and allow them to be used to model complex systems in which ssDNA is present. In fact, an atomistic study of the detailed structure of ssDNA would be quite important, especially in view of its involvement in biodetection schemes,[17] materials technology,[18] elaboration of new probes for biodetection, such as nonlinear optical spectroscopy,[19] and efforts in DNA sequencing using nanopores.[20] Short oligomers of single-stranded poly-

(1) Tinoco, I., Jr *J. Am. Chem. Soc.* **1960**, *82*, 4785.
(2) Cadet, J.; Sageb, E.; Douki, T. *Mutat. Res.* **2005**, *571*, 3.
(3) Tuma, J.; Tonzani, S.; Schatz, G. C.; Karaba, A. H.; Lewis, F. D. *J. Phys. Chem. B* **2007**, *111*, 13101.
(4) Eisinger, J.; Shulman, R. G. *Science* **1968**, *161*, 1311.
(5) Eisinger, J.; Gueron, M.; Shulman, R. G.; Yamane, T. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 1015.
(6) Crespo-Hernandez, C. E.; Cohen, B.; Kohler, B. *Nature* **2005**, *436*, 1141.
(7) Bouvier, B.; Dognon, J.-P.; Lavery, R.; Markovitsi, D.; Millie, P.; Onidas, D.; Zakrzewska, K. *J. Phys. Chem. B* **2003**, *107*, 13512.
(8) Buchvarov, I.; Wang, Q.; Raytchev, M.; Trifonov, A.; Fiebig, T. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4794.
(9) Schultz, T.; Samoylova, E.; Radloff, W.; Hertel, I. V.; Sobolewski, A. L.; Domcke, W. *Science* **2004**, *306*, 1765.
(10) Olsthoorn, C. S. M.; Bostelaar, L. J.; De Rooij, J. F. M.; Van Boom, J. H.; Altona, C. *Eur. J. Biochem.* **1981**, *115*, 309.
(11) Zimmer, C.; Birch-Hirschfeld, E.; Weiss, R. *Nucleic Acids Res.* **1974**, *1*, 1017.
(12) Greve, J.; Mestre, N. F.; Levin, A. *Biopolymers* **1977**, *16*, 1489.
(13) Cheatham, T. E., III; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435.
(14) Martinez, J. M.; Elmroth, S. K. C.; Kloo, L. *J. Am. Chem. Soc.* **2001**, *123*, 12279.
(15) Sen, S.; Nilsson, L. *J. Am. Chem. Soc.* **2001**, *123*, 7414.
(17) Kohli, P.; Harrell, C. C.; Cao, Z.; Gasparac, R.; Tan, W.; Martin, C. R. *Science* **2004**, *305*, 984.

nucleotides are also important in biology (for example, in microRNA[21] and to understand the numerous ssDNA-binding proteins[22,23] important in DNA replication), so finding the right theoretical tools to model them is very useful.

We therefore set out a two-pronged effort, to understand better and model at least semiquantitatively the delocalization of excited states in single- and double-stranded DNA and in the process to take a first step in modeling more precisely the structure of ssDNA with empirical atomistic force fields. At the same time we will also demonstrate the usefulness of calculating circular dichroic spectra from molecular dynamics (MD) simulations, both to understand system structure and to validate force fields, as has been done recently with neutron scattering.[24] In the present case, the added value of CD spectra is their dependence on both geometry and excited states. Therefore, correct modeling of the CD spectra is very important in the present context of calculating delocalization of excited states.

## 2. Theory

To achieve our goals we use molecular dynamics simulations combined with density functional theory (DFT) calculations to capture the most important characteristics of DNA, including on one hand the structure fluctuations, which we have shown in previous works to be fundamental for both spectra[3] and charge transfer[25] in DNA, and on the other hand the excitations, which are the main focus of this work, that can be reliably calculated for DNA strands by use of DFT.[3]

We have performed molecular dynamics simulations for various ssDNA strand lengths (2, 4, 7, 9, and 11 bases) using Amber (specifically Amber99[26]), which is known to give good results for double helices; we have examined the predicted structures from this force field and, in the case of the 2- and 4-mers, also the CD spectra (see Figures 2 and 3), taking as a reference the ideal B-form of DNA and the experimental ssDNA CD spectra.[10,11] The Amber simulations are similar to the ones described in ref . We started from an Arnott B-DNA single strand obtained from the nucgen package of Amber, which we solvated with a TIP3P water box and neutralized with $Na^+$ ions. Then we minimized the structure and ran 2 ns simulations at 300 K (with a 2 fs time step). The last 1.5 ns was used to produce regularly spaced snapshots to calculate spectra.

The spectra have been calculated much as in ref 3, so we will just briefly summarize the procedure here. We eliminate the water and DNA backbone (and cap the bases with hydrogens) and use the package ADF[27] to calculate orbitals, and subsequently absorption and CD spectra,[28] via time-dependent density functional theory (TDDFT). By this approach we are able to calculate hundreds of excited states for each snapshot, which in turn makes it feasible to average the spectra over the fluctuations. The calculated rotatory strength $R_i$ for each transition was then converted to CD spectra via the procedure delineated in ref 28 and references therein, by use of the equation

$$R_i = 22.97 \int \frac{\Delta\epsilon(E)}{E} dE \quad (1)$$

where the integral is over an entire band. $\Delta\varepsilon(E)$ (in liters per mole per centimeter) was represented as a Gaussian function, with a width of $0.06\sqrt{E}$, empirically found to give the best results, so as to reproduce the calculated rotatory strength through eq 1. The CD spectrum was then constructed as the superposition of all these Gaussian functions.

We use a triple-$\zeta$ plus polarization Slater basis set and an asymptotically correct functional (SAOP[29]) that allows us to describe correctly excited states that have significant Rydberg character, which is important when higher excitations are sought (we calculate 100−150 excitations/snapshot to compute the CD spectra). The use of this functional contributes also, in this system, to the reduction of the notorious overdelocalization problems of pure density functionals: we have tested for a few snapshots that the delocalization lengths we obtain with this functional are very similar to those obtained with a more accurate (and far more expensive) self-interaction-corrected[30] calculation, while the results obtained with a simple BLYP functional without this correction are on the order of 10% higher or more. Since we are looking for interbase delocalization, and not intrabase, the correction to the long-range part of the exchange−correlation potential is the dominant contribution to the overdelocalization reduction. Spurious low-energy charge transfer states do not seem to constitute a problem, due mostly to the fact that a high degree of $\pi$ stacking guarantees orbital overlap between the bases. An exception to this observation is discussed in Supporting Information.

The spectra of the 2-mer are averaged over 24 snapshots, taken at intervals of 100 ps over two distinct MD trajectories, to eliminate the differences in the degree of stacking, while for the 4-mer we used 10 snapshots from a single trajectory, as the polymer presents a similarly high degree of stacking in both of the simulations we performed. Solvation is expected to shift somewhat the peaks obtained from TDDFT, but we did not include it, following our previous tests which showed how these shifts are in fact on the order of 0.1 eV.[3] Tests on the solvation shift on the spectra are presented in Supporting Information for single snapshots of the various strands that confirm the previous findings.

## 3. Results and Discussion

**3.1. Modeling of Poly(A) Structure and Spectra.** In this section we present our findings from CD and MD calculations and compare them with previous studies. A detailed comparison between structural parameters among our MD simulations, B-DNA, and NMR spectra[16] is shown in Supporting Information.

The circular dichroic spectra in Figure 2 for an adenine strand demonstrate that structures from simulations based on the Amber force field (one snapshot is shown in Figure 1) retain a high degree of stacking and remain somewhat close to the B-DNA structure although they are not identical, as shown in Figure 2 for $d(A)_2$ and in Figure 3 for $d(A)_4$. The experimental spectra have three dominant peaks—a positive one around 270 nm, a negative peak at 250 nm, and another positive peak at 220

(18) Seeman, N. C. *Nature* **2003**, *421*, 427.
(19) Stokes, G. Y.; Gibbs-Davis, J. M.; Boman, F. C.; Stepp, B. R.; Condie, A. G.; Nguyen, S. T.; Geiger, F. M. *J. Am. Chem. Soc.* **2007**, *129*, 7492.
(20) Gerland, U.; Bundschuh, R.; Hwa, T. *Phys. Biol.* **2004**, *1*, 19.
(21) Ruvkun, G. *Science* **2001**, *294*, 797.
(22) Kim, Y. T.; Tabor, S.; Bortner, C.; Griffith, J. D.; Richardson, C. C. *J. Biol. Chem.* **1992**, *267*, 15022.
(23) Summers, M. F. *J. Cell. Biochem.* **1991**, *45*, 41.
(24) Thomas, J. L.; Tobias, D. J.; MacKerell, A. D., Jr *J. Phys. Chem. B* **2007**, *111*, 12941.
(25) Grozema, F. C.; Tonzani, S.; Berlin, Y. B.; Schatz, G. C.; Siebbeles, L. D. A.; Ratner, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 5157.
(26) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(27) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Guerra, C. F.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931967.
(28) Autschbach, J.; Ziegler, T.; van Gisbergen, S. J. A.; Baerends, E. J. *J. Chem. Phys.* **2002**, *116*, 6930.
(29) Gritsenko, O. V.; Schipper, P. R. T.; Baerends, E. J. *Chem. Phys. Lett.* **1999**, *302*, 199.
(30) Patchkovskii, S.; Autschbach, J.; Ziegler, T. *J. Chem. Phys.* **1999**, *115*, 26.
(16) Isaksson, J.; Acharya, S.; Barman, J.; Cheruku, P.; Chattopadhyaya, J. *Biochemistry* **2004**, *43*, 15996.

***Figure 1.*** From left to right: snapshot from the MD simulation of a $d(A)_4$ oligomer and two other snapshots representative of the average twist dihedral angle and interbase distance for the central adenines in $d(A)_4$. These can be directly compared to the NMR-derived structures in Figure 6 of ref 16.

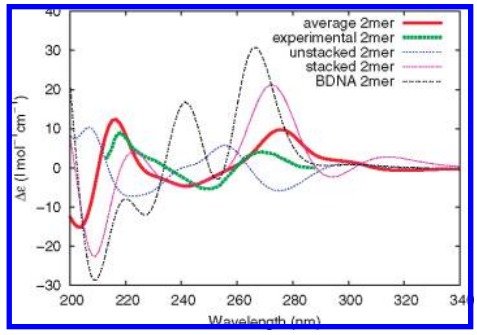



***Figure 2.*** Comparison of the calculated optical rotatory strengths with the B-DNA-like conformation and experimental data[10] for $d(A)_2$. The averages have been performed over 24 molecular dynamics snapshots from two different MD trajectories; the stacked and unstacked subspaces are made up by eight snapshots from simulations performed at 278 and 340 K, respectively.

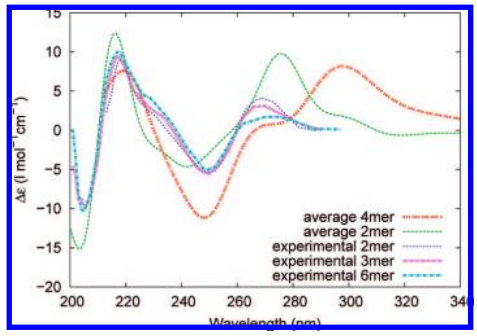



***Figure 3.*** Comparison of the calculated optical rotatory strengths for $d(A)_2$ and $d(A)_4$ with the experimental results for $d(A)_2$, $d(A)_3$, and $d(A)_6$.[10] Averaging for the 4-mer has been performed over 10 molecular dynamics snapshots.

nm—which are reproduced quite accurately by our calculations. It should be noticed that in the experimental spectra for the 3-mer and 6-mer[10] the peak values are slightly higher than for the 2-mer (except for the peak at 270 nm), while in our calculations (comparing the 4-mer and the 2-mer) it is not so. This might be due to the need for more averaging over snapshots in the case of the 4-mer (note that fewer snapshots were used than for the 2-mer), but those calculations are very time-consuming and we did not pursue this further. The same argument applies for the red-shifted peak in the 4-mer spectrum at 300 nm, which we think is due to the difficulty of obtaining an exact numerical cancelation of large positive and negative peaks in that energy region, as shown also in ref .

The highly stacked configurations that we find are compatible with results from NMR,[10] in which 90% stacked configurations were detected at 273 K for the 2-mer, and from thermodynamic modeling,[10] by which means a "melting" temperature of 320 K was established. Simulations performed at different temperatures (278 and 340 K) show the expected behavior of more (at 278 K) and less (at 340 K) stacking with respect to room temperature, but it is a difficult task to perform an accurate modeling of the melting of ssDNA, since the barrier between stacked and unstacked conformations makes the sampling of statistically meaningful ensembles of configurations an issue. In fact, different trajectories predict different degrees of stacking at the various temperatures. We have verified, however, that the CD spectra predicted for $d(A)_2$ at room temperature for all the different trajectories share the same basic features as the experimental ones. In Figure 2 are also shown spectra from all-stacked configurations (from a simulation at 278 K) and all-unstacked configurations (taken from a simulation at 340 K). It is important to notice how the spectrum of the former is qualitatively correct, while the latter is not and would probably average to zero if a sufficiently high number of snapshots were considered in the averaging. Accurate thermodynamic modeling of the melting transition is outside the scope of the present work, and therefore we will not pursue this further.

Note also that our MD simulations have been performed for adenine strands only. Likely there are differences between $d(A)_n$ and other single-stranded oligomers regarding stacking and other structural characteristics. In particular, molecular dynamics studies of poly(T) and mixed-sequence segments[14,15] reported more frequent unstacked geometries. This is to be expected, aside from the differences in the force fields used, because of the stronger adenine−adenine interaction due to the stacking of a more extended conjugated molecule. This is deduced also from sequence-dependent rigidity studies,[31] where an enthalpic barrier due to stacking is found in poly(A) but not in poly(T) in molecular beacon folding. In more complex (and longer) sequences, secondary structure motifs such as hairpins[32] become important also. Recently, detailed studies of sequence dependence in dsDNA have been carried out,[33] and it would be very interesting to apply similar methodologies to ssDNA.

In view of the CD spectra from refs 10–12, we can say that short $d(A)_n$ assumes a stacked structure, and the degree of stacking increases as more monomers are added. Our Amber simulations agree with this result. We have shown previously how the CD spectrum is rather sensitive to small differences in the spatial relationship between the various DNA subunits for

(31) Goddard, N. L.; Bonnet, G.; Krichevsky, O.; Libchaber, A. *Phys. Rev. Lett.* **2000**, *85*, 2400.
(32) Stellwagen, E.; Abdulla, A.; Qian, D.; Stellwagen, N. C. *Biochemistry* **2007**, *46*, 10931.
(33) Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H.; Thayer, K. M.; Varnai, P. *Biophys. J.* **2005**, *89*, 3721.

short oligomers; therefore their similarity to experimental values is a rather stringent test of the force field's ability to describe these compounds. Our simulations predict a C2′-endo structure for the sugar pucker (similarly to B-DNA, as could be expected from the solvation conditions of the simulation) and a helical pitch of 8 bases, which means a 50° twist angle per base for $d(A)_9$. The strand length is 28.8 Å on average for the 9-mer, which entails an average base stacking distance of 3.6 Å, slightly larger than the 3.4 Å in standard B-DNA. The variation of these values during the several nanoseconds of the simulations is quite large, due to the flexibility of the polymer. We report these and a series of additional structural parameters such as backbone dihedrals in Supporting Information. In our simulations the stacked bases show a larger overlap, as shown in Figure 1, compared with ideal B-DNA, because the overall geometry of the double helix has to accommodate the second strand, whereas hydrophobic interactions dominate in the single strand. The root-mean-square deviation (rmsd) from B-DNA is large, roughly 2−5 Å during the last 1.5 ns of simulation (similarly to what was reported in ref 15). The rmsd shows some oscillations during the simulation, which are to be expected due to the flexibility of the polymer. These findings are in agreement with the molecular dynamics refinement of the NMR spectra in ref 16. Although in that work a GAAAAC sequence was used, the twist parameter for the central adenines, as well as their average stacking conformation (depicted in Figure 1), is similar to our findings. A detailed comparison of our structural parameters to this NMR study is presented in Supporting Information.

In ref 16, the authors showed that removing the NMR constraints from the molecular dynamics simulation generates structures with a large rmsd from the NMR-minimized one, but they did not discuss these structures further. We demonstrate here how important characteristics of the structures obtained by unconstrained MD are very similar to the NMR spectra, and in fact they are good enough to describe CD spectra, therefore allowing confident use of the Amber force field in simulations involving ssDNA, crucially to understand excited-state behavior. Indeed, since our excited-states calculations do not include at all the backbone, many structural parameters, such as backbone dihedrals, emerging from a comparison with previous studies, are not very important in this context; instead, the CD spectra, which depend for the most part on the spatial relationships among the bases at low energies (since only those have low-lying excited states) are a very important test. The comparison with previous molecular dynamics studies[14,15] is somewhat difficult due to the different sequence studied here, poly(A) instead of poly(T) or a mixed sequence. Besides the previously noted different amount of stacking, the study in ref 15 also noticed the formation of folded structures in a mixed sequence with a 50% pyrimidine content, which we do not encounter in our longest (4 ns) simulations for $d(A)_7$, consistently with a higher barrier to unstacking for this sequence.[31] In molecular beacons,[31] realized with short end-traits of self-complementary bases and a central poly(A) or poly(T) chain, the difference in time scale for the hairpin formation (which includes both folding and pairing of complementary bases) between poly(A) and poly(T) strands of the same length has been determined as being on the order of several microseconds: this is a time scale largely outside the reach of present-day atomistic simulations.

Previous experimental evidence has shown that the structure of ssDNA in solution is very far from a random coil, at least for short oligomers, and in fact is somewhat closer to B-DNA. Our simulations confirm these findings, in particular of a right-handed helix in which the bases, on average, have a very small inclination [−5° for $d(A)_4$, in decent agreement with NMR data][16] with respect to the helical axis, at least for the solvation and ionic strength conditions used in our simulations. An important difference with respect to B-DNA is the larger twist angle (50°), that generates a structure in which bases overlap more than in dsDNA. All of this has important consequences for excited states delocalization and dynamics, one of which is the fact that the amount of $\pi$-stacking is roughly similar in ssDNA to the double helix, which could explain why the two look so similar in their excited states decay rates[6] and delocalization.[8] In addition, for the oligomers we examined, up to 11 bases, the strand undergoes very limited bending during the entire simulation, as shown in Supporting Information. This could of course be a defect of the force field, but we notice that it is actually in agreement with the measured persistence length, which is 40 Å (corresponding to 12 bases, with a spacing of 3.6 Å as we found in our simulations) in solutions with low ionic strength,[34] such as the ones we simulate. The ssDNA persistence length becomes sensibly smaller for higher ionic strengths, and it would be interesting to test this property using molecular dynamics.

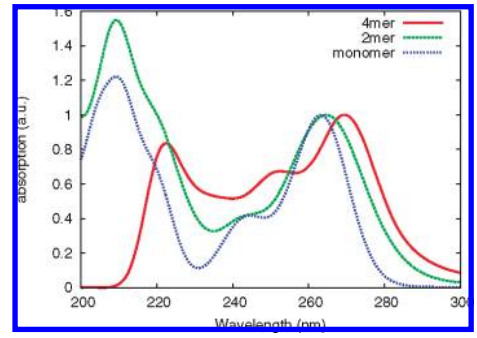



***Figure 4.*** Absorption spectra for d(A), $d(A)_2$, and $d(A)_4$, averaged over molecular dynamics snapshots, which shows that the peaks do not present a significant red shift or broadening, as known from experiments.[7] The absorption intensities have been all normalized in such a way that the intensity of the first peak is the same in all three strands.

**3.2. Excited States Delocalization.** We have calculated absorption spectra for d(A), $d(A)_2$, and $d(A)_4$, all from molecular dynamics snapshots to eliminate possible differences arising from spectral shifts generated by the force field description of the adenines, as shown in Figure 4. The peak of monomer absorption is at 263.1 nm, while for the 4-mer it is at 266.2 nm. The spectra show that the shift the 4-mer exhibits with respect to the monomer is indeed modest (420 $cm^{-1}$) and of the same magnitude as the one reported experimentally (390 $cm^{-1}$),[7,35] but in the wrong direction. The same conclusion is reached upon examination of d(A), $d(A)_2$, and $d(A)_3$ absorption spectra for ideal B-DNA geometries. It has been shown[35] that there is, in fact, a blue shift in passing from the monomer to the polymer, while our theory predicts a red shift. In ref 35 it has been show that TDHF calculations indeed find a blue shift passing from the monomer to the dimer, while TDDFT does not. In Supporting Information we discuss this point in more detail, together with solvation effects, which are on the order of tenths of an electronvolt in this system, as we have previously shown,[3] but do not affect the sign of the spectral shift.

(34) Tinland, B.; Pluen, A.; Sturm, J.; Weill, G. *Macromolecules* **1997**, *30*, 5763.
(35) Hu, L. H.; Zhao, Y.; Wang, F.; Chen, G. H.; Ma, C.; Kwok, W.-M.; Phillips, D. L. *J. Phys. Chem. B* **2007**, *111*, 11812.

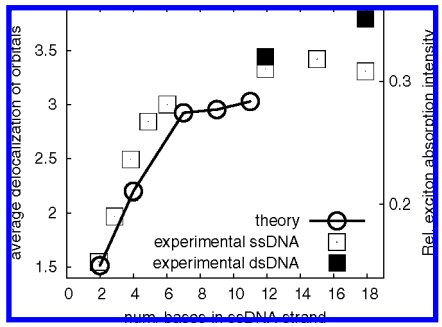



***Figure 5.*** Delocalization length of excited orbitals in the optical window in ssDNA (calculated, left *y* axis) and experimental relative absorption (right *y* axis) from ref 8 as a function of the length of the strand. We have plotted the absorption data in such a way that the saturation length appears to be 3.3, as obtained from the fitting of the data to an exponential performed in ref 8. Data for double stranded DNA are also shown (■); there is a relatively small difference between ssDNA and dsDNA delocalization lengths.

To calculate delocalization, we examined the coefficients of expansion of the molecular orbitals (MOs) in the Slater basis ADF uses, for the orbitals close to the highest occupied−lowest unoccupied molecular orbital (HOMO−LUMO) gap that participate in the first transition band (for adenine around 260 nm). For each snapshot, we define the delocalization as the average of the number of bases upon which the orbitals have a significant coefficient, imposing a cutoff of 1% . The cutoff procedure here is imposed on the coefficients in the expansion of the MOs. The basis functions that have a coefficient with an amplitude of less than 0.01 in the MO expansion are neglected in the delocalization calculation. In Supporting Information, convergence tests are described with respect to the cutoff threshold. The delocalization thus obtained for each geometry is then averaged over all the snapshots, as different snapshots have different amounts of stacking and therefore somewhat different delocalization lengths. It is worth noticing that these single-point DFT calculations (averaged over many snapshots) are much less expensive than the TDDFT calculation of the spectra averaged over many configurations (about 4 times less expensive for the 4-mer), so it is possible to study longer chain lengths, up to $d(A)_{11}$. The only additional calculations needed to describe the orbital delocalization are two TDDFT runs for two snapshots for each chain length to determine which states are involved in the first absorption band (except for the 11-mer, for which this information was deduced on the basis of the 9-mer optical window), as opposed to the many roots needed to calculate absorption and CD spectra over a wide range of energies (150 per snapshot for the 4-mer, a number that increases rapidly with chain length). For the CD spectra there is the additional complication that since there are both positive and negative contributions, the spectra are much harder to converge, and therefore more snapshots are needed than for the absorption spectra.

Using the procedure previously outlined we produce results, shown in Figure 5, that are semiquantitatively similar to the experimental results from ref 8. In particular, that work predicted a 1/*e* delocalization length of 3.3 ± 0.5 bases, while our saturation length is 3.0 bases for the 11-mer. The delocalization length is obtained in ref 8 through an exponential fit of the ratio of excited-state absorption between two bands, one that corresponds to delocalized states, the other originating from localized states and also present in the monomer. We did not perform calculations for strands longer than 11 bases because the DFT part would become very time-consuming and the results would probably not be very informative. It should be noted that delocalization has been calculated previously for dsDNA[7] via entirely different methods: the results of that study are similar to the present ones in that the excited states are indeed found to be delocalized over several bases, although the predicted delocalization length is larger (4−8 bases).

We have performed tests using a larger optical window, which in turn means including more orbitals in the delocalization calculation, and we found that the amount of delocalization grows (if we include 50% more orbitals, the delocalization for the 11-mer becomes 4.05 bases), but the qualitative behavior remains the same as in Figure 5: namely, the curve flattens out around 11 bases. The growing delocalization with higher excitations is not surprising, given that states higher in energy can be closer to the conduction band and therefore more delocalized. At higher energies double excitations will be important as well; therefore the reliability of the TDDFT approach, and in turn of the calculated orbitals, will be questionable. However, overall it is encouraging that the dependence of delocalization on the number of bases in the strand remains qualitatively the same. The identity of the bases over which the orbitals are delocalized changes also for vertical excitations, as seen in different snapshots. Over the few picosecond time scale during which they undergo relaxation processes, or at higher temperatures, the mobility of the delocalized excited states would be even more pronounced.

We have also calculated excited-state delocalization for two double-stranded DNA oligomers, $d(A)_9 \cdot d(T)_9$ and $d(A)_{11} \cdot d(T)_{11}$, by performing DFT calculations only on the A strand. The reason is that the computational requirements would have been too large for both strands, and we wanted to take into account the amount of delocalization that derives from the structural properties alone. The results we obtain are, within the precision of this method, entirely identical to their ssDNA counterparts with 9 and 11 bases, which means that to see an effect of the added rigidity of the double-helix structure, several more bases would have to be added. The experimental results[8] give a very similar picture, since the smallest double strand that could be measured had 12 bases and its delocalization length was very similar to that of its single-stranded counterpart. The experimental results for $d(A)_{18} \cdot d(T)_{18}$ point toward a somewhat larger delocalization, but it is difficult with so few points to understand if that constitutes a trend or not. In addition to the calculations from molecular dynamics, we have also tested our approach against ideal B-DNA structures (for an adenine strand) for 9 and 11 bases; the results point to a much larger delocalization (4.9 bases on average for the 9-mer, 5.6 for the 11-mer). This is entirely consistent with a disorder-induced localization, already noted in refs 36 and 37, which in this case is anyway not sufficient to fully localize the excitations on one base only. Solvent effects also tend to localize the wave function;[37] while we have not considered them here, it would be important to include them in future studies on the matter.

At this point a discussion on the nature of the delocalized excited states is warranted. In ref 8 the relaxed delocalized excited states are called excitons, while their length is also compatible with a polaronic nature of the states,[38] as seen also in previous calculations;[39] recently there have been interesting

(36) Lewis, J. P.; Cheatham, T. E.; Starikov, E. B.; Wang, H.; Sankey, O. F. *J. Phys. Chem. B* **2003**, *107*, 2581.
(37) Mantz, Y. A.; Gervasio, F. L.; Laino, T.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *99*, 058104.
(38) Conwell, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8795.

attempts at considering the polaronic nature of the excitons.[40] It is difficult to assign these tags to what we calculate here, since we are dealing with vertical excitations at time $t = 0$ (or at most after electronic dephasing that occurs in a few femtoseconds) and therefore no relaxation has yet occurred. Relaxation is important for both exciton and polaron formation, although in the first case the time scale is probably shorter since nuclear rearrangement is not necessarily required. Here we also do not include any solvent effects, which are crucial when polarons are involved. If we consider the vertical excitations described in our calculations as the starting point of the time evolution and the results of ref 8 as the final point (after few ps), we can conclude that the delocalization length does not change much over this time period. Therefore, the excited electronic states in ssDNA are created and remain spread over a few bases (of the order of 3 or 4) during this timescale. It would be highly desirable to conduct a study on the time evolution of the DNA excited states to understand their nature (whether it is excitonic or polaronic) in depth and to figure out more quantitatively the variation of the delocalization length. Unfortunately, from a theory standpoint, this is not possible with current technology, as the few *ab initio* molecular dynamics calculations performed to date on excited states either deal with small systems[41] or use methods that incur fundamental theoretical problems.[42,43]

## 4. Conclusions

In summary, we have shown how TDDFT calculations can be used to model delocalized excited states in DNA, leading to results that are entirely compatible with experimental evidence from CD, as well as time-resolved ultrafast spectra,[8] while some differences arise with respect to UV spectra that are due to the choice of the ab initio method. These results show that vertical excitations are already delocalized roughly as much as the relaxed states seen in experiment after a few picoseconds. We also showed that the Amber force field is adequate to describe the highly stacked conformations assumed by short oligomers of ssDNA. Of course, a more detailed study employing longer simulations would be highly desirable to provide a deeper understanding of this important biopolymer, and especially of its melting dynamics, and a more quantitative assessment of the sequence-dependence of the ssDNA structure. Several elements are missing from this model, particularly electronic relaxation, which should be taken into account to better understand the experiments. For systems this large, and especially if one wants to take into account the structural fluctuations, this would lead to an impossible level of complexity. Many researchers are exploring the excited-state dynamics of DNA bases and oligomers[9,44] in the gas phase. We feel that the present study is complementary to these in the sense that we capture aspects that are forcefully neglected in gas-phase calculations but can prove to be just as important.

**Acknowledgment.** This work was supported by the National Science Foundation (NSF Grant CHE0628130 to G.C.S.) and by the NSEC Center at Northwestern (NSF Grant EEC-0647560). S.T. acknowledges PNNL for supercomputing time. T. Fiebig, Y. B. Berlin, A. D. Dutoi, and M. McCullagh are acknowledged for stimulating discussions.

**Supporting Information Available:** Convergence tests and alternative methods to calculate excited states delocalization, solvation effects on UV spectra, and detailed comparison of MD structure with ideal B-DNA and available NMR. This material is available free of charge via the Internet at http://pubs.acs.org.

JA7103894

(39) Kurnikov, I. V.; Tong, G. S. M.; Madrid, M.; Beratan, D. N. *J. Phys. Chem. B* **2002**, *106*, 7.
(40) Conwell, E. N.; McLaughlin, P. M.; Bloch, S. M. *J. Phys. Chem. B* **2008**, *112*, 2268.
(41) Hudock, H. R.; Levine, B. G.; Thompson, A. L.; Satzger, H.; Townsend, D.; Gador, N.; Ullrich, S.; Stolow, A.; Martinez, T. J. *J. Phys. Chem. A* **2007**, *111*, 8500.
(42) Craig, C. F.; Duncan, W. R.; Prezhdo, O. V. *Phys. Rev. Lett.* **2005**, *95*, 163001.
(43) Maitra, N. T. *J. Chem. Phys.* **2006**, *125*, 014110.
(44) Santoro, F.; Barone, V.; Improta, R. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9931.