

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/26302593>

Fingerprints of Bonding Motifs in DNA Duplexes of Adenine and Thymine Revealed from Circular Dichroism: Synchrotron Radiation Experiments and TDDFT Calculations

ARTICLE in THE JOURNAL OF PHYSICAL CHEMISTRY B · JULY 2009

Impact Factor: 3.3 · DOI: 10.1021/jp9032029 · Source: PubMed

CITATIONS

11

READS

25

8 AUTHORS, INCLUDING:



Daniele Varsano

Italian National Research Council

39 PUBLICATIONS 999 CITATIONS

SEE PROFILE



Rosa Di Felice

Italian National Research Council

143 PUBLICATIONS 3,018 CITATIONS

SEE PROFILE



Angel Rubio

Universidad del País Vasco / Euskal Herri...

589 PUBLICATIONS 23,874 CITATIONS

SEE PROFILE



Steen Brøndsted Nielsen

Aarhus University

197 PUBLICATIONS 3,075 CITATIONS

SEE PROFILE

Fingerprints of Bonding Motifs in DNA Duplexes of Adenine and Thymine Revealed from Circular Dichroism: Synchrotron Radiation Experiments and TDDFT Calculations

Lisbeth Munksgaard Nielsen,[†] Anne I. S. Holm,[†] Daniele Varsano,^{*,‡,§} Umesh Kadhane,^{||} Søren Vrønning Hoffmann,[⊥] Rosa Di Felice,[‡] Angel Rubio,[#] and Steen Brøndsted Nielsen^{*,†}

Department of Physics and Astronomy and Institute for Storage Ring Facilities, Aarhus University, Ny Munkegade, DK-8000 Aarhus C, Denmark, CNR-INFM National Center S3, Via Campi 213/A, 41100 Modena, Italy, European Theoretical Spectroscopy Facility (ETSF), Department of Physics, Indian Institute of Technology Madras, Chennai, 600036 India, and Nano-Bio Spectroscopy Group and ETSF Scientific Development Centre, Departamento de Física de Materiales, Universidad del País Vasco UPV/EHU, Centro Mixto CSIC-UPV/EHU and DIPC, Avenida de Tolosa 72, E-20018 San Sebastián, Spain

Received: April 7, 2009; Revised Manuscript Received: May 22, 2009

Synchrotron radiation circular dichroism (SRCD) spectra were recorded for a family of 12 DNA duplexes that all contain nine adenines (A) and nine thymines (T) in each strand but in different combinations. The total number of AT Watson–Crick (WC) base pairs is constant (18), but the number of cross-strand (CS) hydrogen bonds between A and T varies between 0 and 16, the maximum possible. Eleven of the duplexes have one or more A tracts, and one duplex has T tracts. The signals due to hybridization were found from subtraction of spectra of single strands from spectra of the duplexes. The residual spectrum of the T-tract duplex T₉A₉:A₉T₉ (5′-3′:3′-5′) significantly differs from that of the A-tract duplex A₉T₉:T₉A₉, but only below 210 nm, which suggests that the signal in this region depends on the superhelicity of the duplex. A principal component analysis of all residual spectra reveals that spectra of A-tract duplexes can be obtained to a good approximation as a linear combination of just two basis spectra. The first component is assigned to the spectrum of 18 WC and 8 CS pairs, whereas the second component is that of 8 CS pairs. This interpretation is supported by separate experiments on duplexes of varying lengths but with similar arrangements of the A and T's and by experiments on two other duplex families of 14 and 30 base pairs. The best correlation is obtained by the assumption that cross-strand interactions occur as long as there are two adenine neighbors in a strand. Our data indicate that a circular dichroism spectrum of a duplex containing only A and T can simply be inferred from the number of WC base pairs and the number of CS interactions, and we provide reference spectra for these two interactions. Finally, time dependent density functional theory calculations of the circular dichroism spectra for an isolated WC base pair and two different CS base pairs (between adenine N-6 amine and thymine O-4 or between adenine C-2-H and thymine O-2) were performed to provide some additional support for the interpretation of the experimental spectra. We find large differences between the two calculated CS spectra. However, there is a reasonable qualitative agreement between the calculated WC and the C-2-H...O-2 CS spectra and those deduced from the experimental data.

Introduction

Long runs of A_n:T_n (5′-3′:3′-5′), so-called A tracts, in DNA play significant roles in dictating both the translational and rotational positioning of DNA on nucleosomes. This is linked to the structural features of such DNA sequences that are different from other B-type DNA.^{1–22} A tracts exhibit 10.1 ± 0.1 base pairs (bp) per turn in contrast to 10.5 bp per turn found for random DNA and alternating (AT:TA)_n sequences, and the axial rise of the helix is less than that found for other DNA fibers (3.2 Å versus 3.4 Å).^{1–3} This larger conformational rigidity of A_n:T_n regions has been explained from an unusual structure of the base pairs. They are propeller twisted to stack optimally,

which allows for the formation of additional, non-Watson–Crick, cross-strand (CS) interactions.^{5,6,17,21,22} Figure 1 illustrates representative cases of the WC bond and different CS bonds, taken from a real oligomer that is reported in the Supporting Information (Figure S1). A CS interaction occurs between an adenine N-6 amine and the thymine O-4 on the opposite strand of a helix in the adjacent base pair^{5,6,17} or between the electropositive C-2-H group²³ of adenine and thymine O-2.^{21,22} CS bonds stiffen the helix to a point where it resists winding around a nucleosome.

Circular dichroism (CD) spectroscopy is a powerful method to obtain information on the conformation of DNA because the geometric structure of DNA largely determines its electronic structure and the coupling between bases.^{24–27} Especially, vacuum ultraviolet CD has been shown to provide useful information additional to that from UV CD.^{28–31} Such experiments are nontrivial due to the strong absorbance of light below 200 nm by air and water itself. This problem is largely overcome with synchrotron radiation facilities that deliver high fluxes of VUV light.^{32–40}

* Corresponding authors. Telephone: +39-059-205-5315 (D.V.). Fax: +39-059-205-5651 (D.V.). E-mail: daniele.varsano@unimore.it (D.V.); sbn@phys.au.dk. (S.B.N.).

[†] Department of Physics and Astronomy, Aarhus University.

[‡] CNR-INFM National Center S3.

[§] ETSF.

^{||} Indian Institute of Technology Madras.

[⊥] Institute for Storage Ring Facilities, Aarhus University.

[#] Centro Mixto CSIC-UPV/EHU and DIPC.

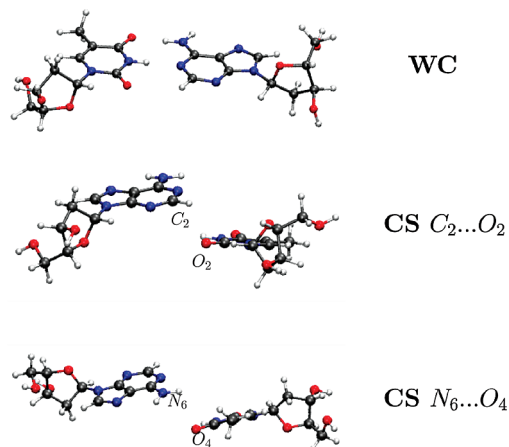


Figure 1. Structures of Watson–Crick and cross-strand AT base pairs, adopted for the computational part of this work. The WC structure was constructed with the 3DNA builder.⁶⁶ The CS structures were extracted from X-ray experimental data, files 1d89 and 1bdn for (Ade)C-2–(Thy)O-2 (middle) and (Ade)N-6–(Thy)O-4 (bottom), respectively.

CD spectral differences between poly(A):poly(T) and poly(A:T) have earlier been reported and have been ascribed to the different arrangements of bases as described above.^{41–45} To further explore these differences in a systematic manner, we have recorded synchrotron radiation CD spectra for different families of AT duplexes. In one set of experiments, the duplexes chosen for study all had 18 AT base pairs, and each strand was composed of nine adenines and nine thymines. The arrangement of A and T in a strand was varied to shed light on the dependence of sequence on the CD signal and in this way unravel the importance of Watson–Crick (WC) and cross-strand (CS) couplings, determining the stiffness of the strand. The base pair coupling was further elucidated by studies of duplex families with 14 and 30 base pairs. Finally, the influence of the length of the duplex in the range 14–54 base pairs was investigated for similar arrangements of adenine and thymine. The data allow us to extract fingerprint spectra of AT WC and CS base pairs. These spectra are compared to those obtained from time dependent density functional theory (TDDFT) calculations on isolated base pairs. We remark that *ab initio* calculations of CD spectra of nucleic acids were long hindered by the huge computational load and theoretical hurdles. Such studies are now feasible thanks to a recently developed gauge-invariant method that we have applied for this work.⁴⁶

Experimental Details

DNA single strands were purchased from DNA Technology A/S, Aarhus. To form duplexes with 14, 18, and 30 base pairs, these single strands were used: A_7T_7 , $(AT)_2A_5T_5$, $A_5T_5(AT)_2$, $A_4(TA)_3T_4$, $A_5(TA)_2T_5$, and $(AT)_7$; A_9T_9 , T_9A_9 , $A_7T_9A_2$, $T_2A_9T_7$, $A_5T_9A_4$, $T_4A_9T_5$, $A_6T_6(AT)_3$, $(AT)_3A_6T_6$, $A_2T_3A_5T_3A_2T_3$, $A_3T_2A_3T_5A_3T_2$, $A_2T_2(AT)_7$, $(AT)_7A_2T_2$, $(AT)_3A_3T_3(AT)_3$, $A_8T_8AT_8$, A_8TAT_8 , $A_5(TA)_4T_5$, and $(AT)_9$; and $A_{15}T_{15}$, $A_8(TA)_7T_8$, $(A_3T_3)_5$, $(A_2T_2)_3A_4T_4(A_2T_2)_2AT$, $AT(A_2T_2)_2A_4T_4(A_2T_2)_3$, $(A_2T_2)_3A_4T_4(AT)_5$, $(AT)_5A_4T_4(A_2T_2)_3$, and $(AT)_{15}$. For strand length variation experiments, strands of the type $A_i(TA)_{i-1}T_i$ with $i = 4–8, 10, 12$, and 14 were chosen. Known amounts of the strands were dissolved in either water or 10 mM phosphate buffer (pH 7.4) + 100 mM NaF_(aq). Synchrotron radiation circular dichroism (SRCD) spectra were obtained at the UV1 beamline at the ASTRID storage ring facility in Aarhus, Denmark.⁴⁷ At the beginning of each filling the setup was

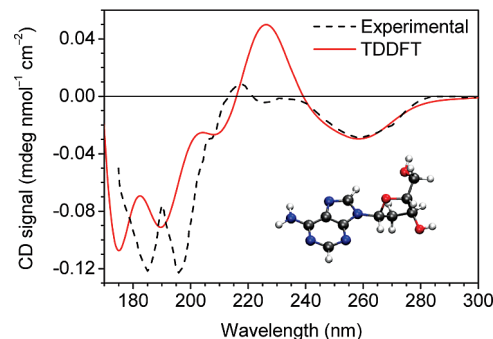


Figure 2. Experimental and calculated CD spectra of the deoxyadenosine nucleoside. The experimental data are taken from ref 38. Inset: structure of the calculated dA obtained from X-ray structure 1d89 pdb file. The phosphate group has been neglected in the calculation. The theoretical spectrum has been normalized to the first low energy peak, and the whole spectrum has been shifted by 30 nm.

calibrated for wavelength and optical rotation magnitude. Spectra were recorded at 20 °C and measured using a quartz cell type QS124 with a path length of 0.1 mm (Hellma GmbH, Germany), and for temperature scans a 0.1-mm closed cell was used. All the spectra were averaged, baseline subtracted, and slightly smoothed with a Savitzky–Golay filter using the CD data processing software CDTool.⁴⁸

Computational Details

Electronic circular dichroism calculations have been performed within the TDDFT framework⁴⁹ using a recently developed scheme based on a real-space pseudopotential representation of the wave functions and Hamiltonian, and a real-time propagation of the Kohn–Sham orbitals.⁴⁶ In this approach the circular dichroism signal is obtained by probing the magnetization of the system perturbed by an electric field. An extra term related to the nonlocal pseudopotentials is taken into account in order to guarantee the gauge invariance of the results. The absorption spectrum can be obtained on the same footing, by probing the time dependent dipole moments of the system as described by Yabana and Bertsch.⁵⁰ The full derivation and details of the method were recently reported,⁴⁶ along with successful applications of this new methodology to some organic chiral molecules.

For all the calculations, we have used the code octopus.⁵¹ Troullier–Martins⁵² norm conserving pseudopotentials were employed to describe the electron–ion interaction. The exchange and correlation functional was treated at the level of the local density approximation (LDA) in the Perdew–Zunger parametrization.⁵³ The use of generalized gradient approximation (GGA) corrections does not significantly change the results discussed below.⁴⁶ The computational grid was constructed with overlapping spheres centered at each atom. The grid spacing was 0.2 Å, and the radius of each sphere was 5.0 Å. For the time propagation scheme a Lanczos exponential method was used with a time step of 0.0079 fs and a total time of 15 fs. The parameters of the calculations provide converged results to less than 50 meV.

Here we first calculated the CD for an isolated gas-phase adenine nucleoside in order to check the reliability of the method for DNA bases. The structure of the adenosine has been extracted from X-ray data (1d89 pdb file).⁸ The hydrogen atoms have been added and relaxed in the DFT/GGA framework, whereas the phosphate group was neglected. The calculated CD spectrum for this isolated adenosine is shown in Figure 2, where the experimental CD spectrum³⁸ is also included. The calculated

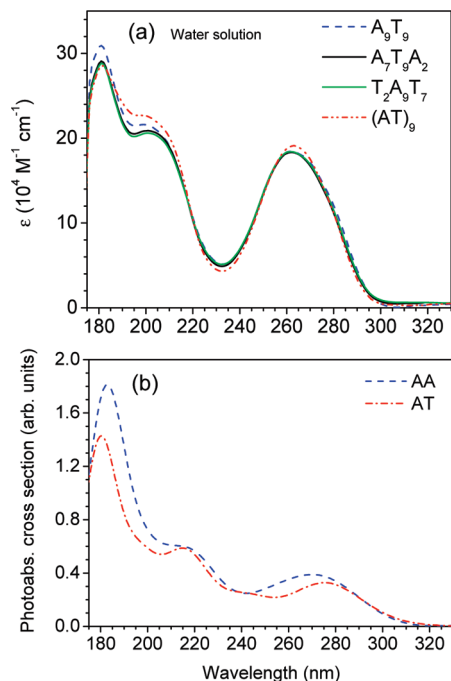


Figure 3. (a) Absorption spectra of single DNA strands in water. Concentrations were 0.1 mM. (b) TDDFT calculated absorption spectra.

rotational strength function presents a very good qualitative agreement with the experimental signal for what concerns the sign, the spectral shape, and the relative positions of the excitation energies. In particular, the first broad band at ~ 260 nm and the rise of the CD signal at lower wavelength with the appearance of a positive peak at ~ 230 nm are well reproduced, although the double-peak fine structure is not resolved for the latter. Also, the two negative peaks below 200 nm are nicely reproduced. It should be mentioned, however, that a rigid blue shift by 30 nm has been applied to the whole calculated spectrum to have the good superposition in Figure 2. Such a numerical discrepancy of the computed excitation energies relative to the experimental data is a well-known shortcoming of the available and widely used exchange-correlation functionals in DFT-based approaches, which normally (as in the present work) lack the exact $1/r$ tail in the effective potential and do not cancel self-interaction effects.⁵⁴ Indeed, by using a functional that includes the $1/r$ tail, improved results may be obtained, yet not in quantitative agreement with experimental data. Anyway, in spite of the systematic red shift of the theoretical spectrum, the relative excitation energies are well reproduced, as well as the relative intensities. Overall, the method satisfactorily reproduces all the main experimental features for the adenosine molecule, even in the low-wavelength range of the spectrum that compared to high energies is usually the most critical for DFT-based approaches. We remark that an additional source of errors is the lack of the solvent in our calculations: in fact, solvation is expected to shift the TDDFT excitation peaks.⁵⁵ This issue pertains to all the systems simulated in this study.

Results and Discussion

Absorption spectra of four single strands, A_9T_9 , $A_7T_9A_2$, $T_2A_9T_7$, and $(AT)_9$, in water solution are shown as representative examples in Figure 3a. The concentrations of the samples were obtained from the experimental absorbances at 260 nm and those calculated using an empirical model with nearest neighbor interactions included.⁵⁶ We then used Lambert–Beer's law to

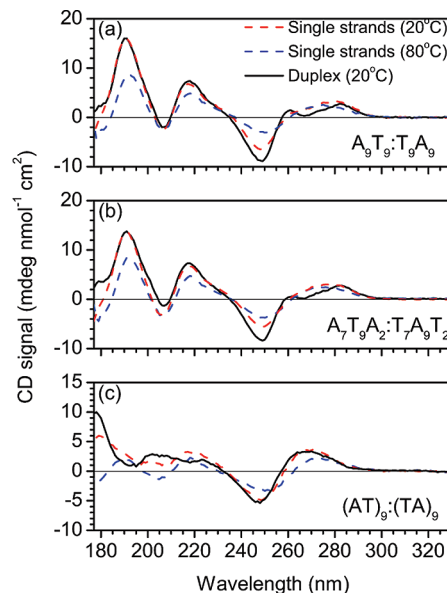


Figure 4. Circular dichroism spectra of DNA in water at 20 °C (single strands), in buffer and salt solution at 20 °C (duplexes), and in buffer and salt solution at 80 °C (single strands).

calculate the extinction coefficients at all other wavelengths from the measured absorbances and the estimated concentrations. The spectra are almost identical, not depending on the particular sequence, and have the strongest signal in the VUV. The absorbance at 190 nm of a single adenine base is split into two due to exciton coupling.^{40,57,58} The character of the different bands in the UV has been discussed in large detail in the literature and will not be further commented upon here.^{24–26,57,59–62}

The absorption spectra of stacked AA and AT dinucleotides calculated by TDDFT are shown in Figure 3b. The two spectra are similar, in reasonably good agreement with the experimental spectra for the measured single-stranded AT oligomers (Figure 3a). We notice the usual⁴⁶ general red shift of the bands, which amounts to about 10–15 nm at the highest wavelengths. The oscillator strength of the lowest wavelength band is predicted to be too high relative to the other bands.

Figure 4a displays the circular dichroism spectra of A_9T_9 in water, in the buffer + salt solution, and in the buffer + salt solution heated to 80 °C. This strand can pair with itself to give a homoduplex since it is self-complementary, but in solutions of water with no salt added, we assume no complexation of single strands since this is prohibited by the Coulomb repulsion between the negatively charged phosphate groups of the individual strands. In the salt solution, on the other hand, hybridization occurred as revealed from differences in the signal compared to the single-strand signal. Heating of the solution to 80 °C led to a melting of the duplex to recover the signal from the single strands but in lower intensity. Spectra of $A_7T_9A_2$ and $T_7A_9T_2$ in water and in the buffer + salt solution are shown in Figure 4b together with the spectrum of the mixture in buffer + salt solution heated to 80 °C. Again clear differences between single-strand spectra (either in pure water or at high temperature) and the spectrum of the mixture at 20 °C are seen, which are ascribed to hybridization. A last example is $(AT)_9$ that is self-complementary, but dimerization to a homoduplex is only seen in the salt solution (Figure 4c). Heating of the salt solution gives the signal of the single strand, except at the lowest wavelengths, which indicates that the single-strand structure is not completely recovered.

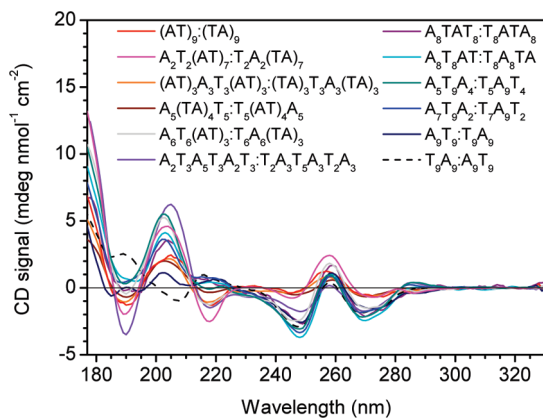


Figure 5. Residual spectra obtained as the difference between the duplex spectra and the spectra of single strands for strands containing $N = 18$ bases.

To obtain CD spectra due to hybridization between two single strands, spectra of single strands (recorded in water) were subtracted from the spectra of the double strands (recorded in buffer + NaF). In this way, we obtained 12 residual duplex spectra (Figure 5). It is evident that there are clear differences between the residual signals, especially in the region from about 210 to 255 nm. For example, $(AT)_9:(TA)_9$ (red curve) displays a negative band at 218 nm and close to zero signal in the 225–250-nm region, whereas $A_9T_9:T_9A_9$ (navy curve) has slightly positive signal between 210 and 225 nm and a strong negative band at 248 nm. These findings are in qualitative accordance with previously published residual spectra by other groups.^{41–45} They indicate that not only base pairing determines the signal but also the actual sequence, that is, the order of the adenines and thymines in a given strand.

Another interesting finding is that the residual spectrum of $A_9T_9:T_9A_9$ (5'-3':3'-5') is almost identical to that of $T_9A_9:A_9T_9$ above 210 nm, which indicates that this spectral region accounts for the local cross-strand interactions. In contrast, the spectra are completely different below 210 nm, which is probably due to different superhelicities of the two duplexes. For comparison, DNA containing repeats of A_4T_4 is bent whereas DNA containing repeats of T_4A_4 is straight.^{63,64}

In order to explore the results in a mathematically more rigorous treatment, we have carried out a principal component analysis (PCA) using all 12 residual spectra as original basis set. The first three principal components (Φ_i) of the new basis set are shown in Figure 6. The contribution c_1 of the first spectrum (Φ_1) to each of the duplex spectra is the same (Figure 7a). The next component (Φ_2) from the PCA of the 18-bp duplex depends linearly on the number of cross-strand linkages, n , but its contribution c_2 is negative when this number is low (e.g., in the case of the $(AT)_9:(TA)_9$ duplex where $n = 0$). The line crosses zero at about eight CS linkages, which is half of the maximum possible. This is evident from Figure 7b, where the contribution from Φ_2 is plotted against $n - 8$. These findings suggest that Φ_2 represents the spectrum of eight CS base pairs and Φ_1 that of 18 WC base pairs plus eight CS base pairs. The contribution from the third component, Φ_3 , scatters around zero (Figure 7c) and is neglected since the spectrum itself is significantly lower in intensity than both Φ_1 and Φ_2 .

We carried out experiments on other duplexes with 14 and 30 base pairs for the sake of reproducibility. The results confirm the trends obtained with the prominent case of 18 base pairs (see Supporting Information, Figures S2 and S3).

If the interpretation of the first component is valid, the signal represented by the coefficient c_1 should increase linearly with the

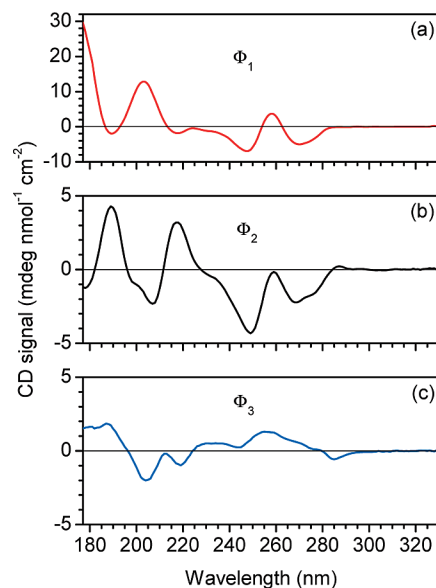


Figure 6. Principal component analysis of the residual spectra shown in Figure 5. The first three basis spectra are shown, denoted Φ_1 , Φ_2 , and Φ_3 .

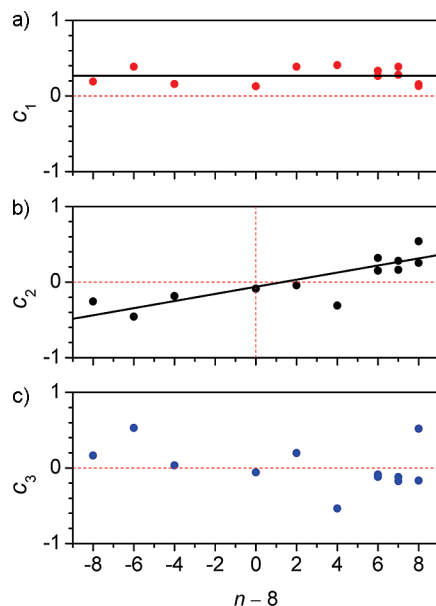


Figure 7. Contribution of each of the basis spectra in Figure 6 as a function of $n - 8$. n is the number of CS H-bonds (maximum 16).

strand length, N , keeping the number of cross-strand interactions equal to half of the maximum possible, $(N - 2)/2$. We therefore carried out experiments on duplexes of the type $A_i(TA)_{i-1}T_i$ with $i = 4-8, 10, 12$, and 14. The residual spectra for these compounds are shown in Figure 8a. The signals look identical except for a scaling factor, and the lowest and highest signals correspond to the shortest and longest strands, respectively. We have then done a PCA with these eight spectra as the original basis and find that the first component is identical to the one found for the 18-bp duplexes (this is shown in the Supporting Information, Figure S4). Importantly, the c_1 coefficient increases linearly with the number of bases in the strand (Figure 8b), justifying the above assignment of the first component.

In our analysis, we have counted all possible cross-strand hydrogen bonds, neglecting cooperative effects. It is possible that more than two successive adenines in a strand are required for the CS interactions to be present. The data for $N = 18$

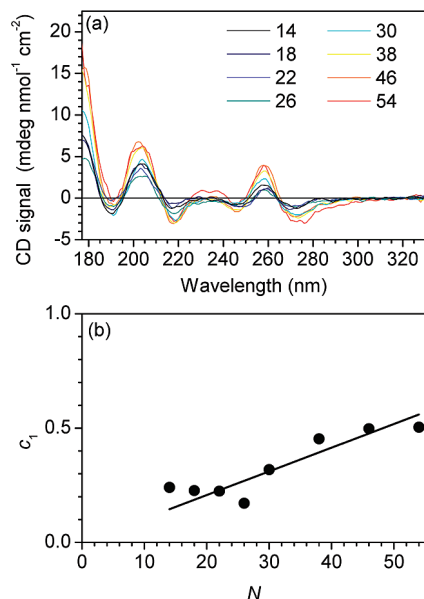


Figure 8. (a) Residual spectra obtained as the difference between the duplex spectra and the spectra of single strands of the type $A_i(TA)_{i-1}T_i$. The total number of WC base pairs $N (= 4i - 2)$ is indicated on the figure. (b) Contribution of the first principal component as a function of N .

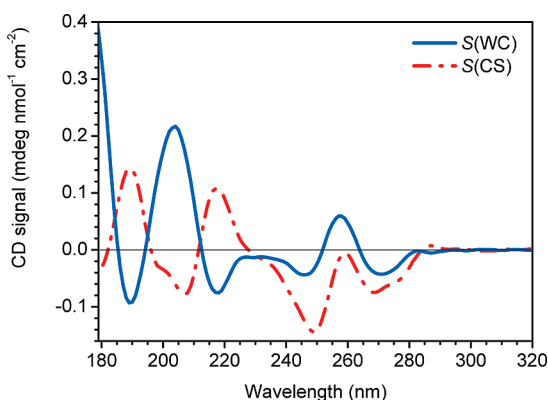


Figure 9. Spectra of a Watson-Crick base pair and a cross-strand hydrogen bond extracted from the two principal components.

duplexes are well-described by taking this number to be three and four (Supporting Information, Figure S5). The data for $N = 30$ duplexes are less well described with three as the minimum number compared to two, and they are poorly described with four (Supporting Information, Figure S6). Taken together, the best correlation is obtained when counting all possible cross-strand interactions.

Based on the principal components and their interpretation, we can extract basis spectra for a Watson-Crick base pair and a cross-strand base pair, denoted $S(WC)$ and $S(CS)$, respectively. These are shown in Figure 9. Each of our measured spectra, except that of $T_9A_9A_9T_9$, can be obtained to a good approximation from $nS(WC) + nS(CS)$ if one knows the number of WC and CS base pairs. The agreement is best in the UV region; see Figure S7 in the Supporting Information. Alternatively, based on a measured CD spectrum of a DNA duplex of A and T for which the numbers of WC and CS base pairs are unknown, both N and n can be estimated.

In order to complement the results obtained by the principal components analysis, we have calculated circular dichroism spectra for different AT pairs with the method described under Computational Details. We considered an A:T Watson-Crick base pair and two cross-strand pairs. The structures are shown in Figure 1.

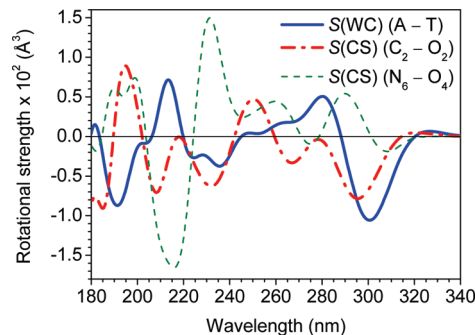


Figure 10. Computed rotatory strength function for the different A:T base pairs shown in Figure 1: Watson-Crick pair (solid blue), CS [C2-O2] (dashed-dotted red) and CS [N6-O4] (dashed green).

The WC base pair was constructed with a nucleic acid builder,⁶⁵ while the two CS structures were extracted from X-ray experimental data on different double-stranded DNAs containing A tracts. The first one, which has an (Ade)C-2-(Thy)O-2 H-bonding, was taken from the pdb file 1d89.⁸ The second one, which has an (Ade)N6-(Thy)O4 H-bonding, was taken from the pdb file 1bdn.⁷ As in the case of the isolated adenosine of Figure S1 in the Supporting Information, sugars have been included in the calculations, while the phosphates were neglected.

The calculated spectra are shown in Figure 10. Since no ad hoc blue shift was applied to the computed data, we clearly see a red shift of the theoretical peaks with respect to the experimental peaks, whose origin was discussed above. Apart from this red shift at the highest wavelength, we find a qualitative agreement between the shape of the computed spectra and that of the experimental signals, for both the WC (solid blue) and the CS (C2-O2) (dotted/dashed red) pairs. At high wavelengths the WC spectrum displays a negative band at 300 nm followed by a positive band at 280 nm. The CS (C2-O2) spectrum has two negative bands at 295 and 265 nm. The corresponding values for the experimental spectra are 270 and 255 nm for the negative and positive bands of the WC pair, and 270 and 245 nm for the two negative bands of the CS configuration. Moreover, we observe that at lower wavelengths the WC and CS spectra display peaks of opposite sign, as was also observed in the experiment. The calculated CS (N6-O4) spectrum has a totally different shape and does not present any agreement with the experimental CS spectrum of Figure 9.

While we have shown here a nice qualitative agreement between the experimental data and the computed CD spectra, it is important to remark that the account of structural fluctuations should be the next step in order to get a quantitative agreement between calculations and experiments.⁶⁶⁻⁶⁸ The measured CD spectra reasonably result from a variety of conformations that the soft molecules take during the experimental measurement, which cannot be represented by just one or two selected structures. However, the ab initio CD calculations for the selected configurations are already able to describe the main experimental features of WC and CS AT pairs. This is a major achievement that reinforces the experimental analysis and contributes to validating our approach. In addition, our results strongly suggest identifying the CS bonds as (Ade)C-2-(Thy)O-2 rather than as (Ade)N-6-(Thy)O-4, which would not be possible based on the experiments alone.

Conclusion

In conclusion, two basis spectra are to a good approximation sufficient to calculate the residual CD spectrum of a DNA

duplex of adenine and thymine with one or more A tracts, with the first component being due to base pairing and the second due to cross-strand pairing. The best description of the data is obtained by assuming a signal contribution from cross-strand interactions that arise when there are at least two successive adenines in one strand. TDDFT calculations lend support for the spectral assignment of the two components being due to WC and CS base pairing and indicate that the CS is adenine C-2-H—O-2 thymine rather than adenine N-6—O-4 thymine. Our data show that it is possible to distinguish between A and T tracts in the VUV region.

Acknowledgment. This work was supported by Lundbeck-fonden, the Danish Natural Science Research Council (No. 272-06-0427), Carlsbergfondet (Grant 2006-01-0229), and by the European Commission through the IST FET-Open project DNA-NANODEVICES (Contract No. FP6-029192) and the e-I3 ETSF project (Contract No. 211956). We acknowledge funding by the Spanish MEC (FIS2007-65702-C02-01), “Grupos Consolidados UPV/EHU del Gobierno Vasco” (IT-319-07), CSIC, and “Red Espanola de Supercomputacion”. We acknowledge CINECA CPU time granted through INFN-CNR. L.M.N. and A.I.S.H. contributed equally to the experimental work and data analysis. D.V. designed the theoretical modeling and carried out the calculations.

Supporting Information Available: Three-dimensional structure of a DNA dodecamer used to extract the computed base pairs (Figure S1). More details on the principal components analysis (Figures S2–S6). Measured CD spectra for all the 18-bp sequences, compared to the signals obtained from the combination of the principal components (WC and CS) for the same sequences (Figure S7). This information is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Peck, L. J.; Wang, J. C. *Nature* **1981**, 292, 375.
- (2) Rhodes, D.; Klug, A. *Nature* **1981**, 292, 378.
- (3) Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, 88, 509.
- (4) Viswamitra, M. A.; Kennard, O.; Jones, P. G.; Sheldrick, G. M.; Salisbury, S.; Falvello, L.; Shakked, Z. *Nature* **1978**, 273, 687.
- (5) Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug, A. *Nature* **1987**, 330, 221.
- (6) Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, 84, 8385.
- (7) DiGabriele, A. D.; Sanderson, M. R.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, 86, 1816.
- (8) DiGabriele, A.; Steitz, T. A. *J. Mol. Biol.* **1993**, 231, 1024.
- (9) Aymami, J.; Coll, M.; Frederick, C. A.; Wang, A. H.-J.; Rich, A. *Nucleic Acids Res.* **1989**, 17, 3229.
- (10) Alexeev, J. D. G.; Lipanov, A. A.; Skuratovskii, I. Ya. *Nature* **1987**, 325, 821.
- (11) Breslauer, K. J.; Frank, R.; Blöcker, H.; Marky, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, 83, 3746.
- (12) Leroy, J.-L.; Charretier, E.; Kochoyan, M.; Gueron, M. *Biochemistry* **1988**, 27, 8894.
- (13) Nadeau, J. G.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, 86, 2622.
- (14) Haran, T. E.; Crothers, D. M. *Biochemistry* **1989**, 28, 2763.
- (15) Koo, H.-S.; Drak, J.; Rice, J. A.; Crothers, D. M. *Biochemistry* **1990**, 29, 4227.
- (16) Barbic, A.; Zimmer, D. P.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 2369.
- (17) Yoon, C.; Privé, G. G.; Goodsell, D. S.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, 85, 6332.
- (18) Chuprina, V. P. *Nucleic Acids Res.* **1987**, 15, 293.
- (19) Klug, A.; Jack, A.; Viswamitra, M. A.; Kennard, O.; Shakked, Z.; Steitz, T. A. *J. Mol. Biol.* **1979**, 131, 669.
- (20) Nadeau, J. G.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, 86, 2622.
- (21) Ghosh, A.; Bansal, M. *J. Mol. Biol.* **1999**, 294, 1149.

- (22) Ghosh, A.; Bansal, M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, 55, 2005.
- (23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, 117, 5179.
- (24) Bush, C. A. Ultraviolet spectroscopy, circular dichroism, and optical rotatory dispersion. In *Basic principles in nucleic acid chemistry*; Ts’O, P. O. P., Ed.; Academic Press: New York, 1974; Vol. II.
- (25) Cantor, C. R.; Schimmel, P. R. *Biophysical chemistry. Part II. Techniques for the study of biological structure and function*; W. H. Freeman: New York, 1980.
- (26) Gratzer, W. B. Optical rotatory dispersion and circular dichroism of nucleic acids. In *Procedures in nucleic acid research*; Cantoni, G. L., Davies, D. R., Eds.; Harper & Row: New York, 1971; Vol. 2.
- (27) Johnson, W. C., Jr. *Methods Biochem. Anal.* **1985**, 31, 61.
- (28) Sutherland, J. C.; Lin, B.; Mugavero, J.; Trunk, J.; Tomasz, M.; Santella, R.; Marky, L.; Breslauer, K. J. *Photochem. Photobiol.* **1986**, 44, 295.
- (29) Johnson, W. C., Jr. *Annu. Rev. Phys. Chem.* **1978**, 29, 93.
- (30) Sprecher, C. A.; Baase, W. A.; Johnson, W. C., Jr. *Biopolymers* **1979**, 18, 1009.
- (31) Tinoco, I., Jr.; Williams, A. L., Jr. *Annu. Rev. Phys. Chem.* **1984**, 35, 329.
- (32) Wallace, B. A. *Nat. Struct. Biol.* **2000**, 7, 708.
- (33) Wallace, B. A. *J. Synchrotron Radiat.* **2000**, 7, 289.
- (34) Miles, A. J.; Wallace, B. A. *Chem. Soc. Rev.* **2006**, 35, 39.
- (35) Wallace, B. A.; Wien, F.; Miles, A. J.; Lees, J. G.; Hoffmann, S. V.; Evans, P.; Wistow, G. J.; Slingsby, C. *Faraday Discuss.* **2003**, 126, 237.
- (36) Ojima, N.; Sakai, K.; Fukazawa, T.; Gekko, K. *Chem. Lett.* **2000**, 832.
- (37) Ojima, N.; Sakai, K.; Matsuo, K.; Matsui, T.; Fukazawa, T.; Namatame, H.; Taniguchi, M.; Gekko, K. *Chem. Lett.* **2001**, 522.
- (38) Brøndsted Nielsen, S.; Chakraborty, T.; Hoffmann, S. V. *ChemPhysChem* **2005**, 6, 2619.
- (39) Holm, A. I. S.; Worm, E. S.; Chakraborty, T.; Babu, B. R.; Wengel, J.; Hoffmann, S. V.; Brøndsted Nielsen, S. *J. Photochem. Photobiol., A: Chem.* **2007**, 187, 293.
- (40) Kadhane, U.; Holm, A. I. S.; Hoffmann, S. V.; Brøndsted Nielsen, S. *Phys. Rev. E* **2008**, 77, 021901.
- (41) Brahms, S.; Brahms, J. G. *Nucleic Acids Res.* **1990**, 18, 1559.
- (42) Vorlickova, M.; Johnson, W. C., Jr.; Kypur, J. *Biopolymers* **1994**, 34, 299.
- (43) Sutherland, J. C.; Lin, B.; Mugavero, J.; Trunk, J.; Tomasz, M.; Santella, R.; Marky, L.; Breslauer, K. J. *Photochem. Photobiol.* **1986**, 44, 295.
- (44) Johnson, K. H.; Gray, D. M.; Sutherland, J. C. *Nucleic Acids Res.* **1991**, 19, 2275.
- (45) Greve, J.; Maestre, M. F.; Levin, A. *Biopolymers* **1977**, 16, 1489.
- (46) Varsano, D.; Espinosa-Leal, L. A.; Andrade, X.; Marques, M. A. L.; Di Felice, R.; Rubio, A. *Phys. Chem. Chem. Phys.* **2009**, 11, 4481.
- (47) Eden, S.; Limão-Vieira, P.; Hoffmann, S. V.; Mason, N. J. *Chem. Phys.* **2006**, 323, 313.
- (48) Lees, J. G.; Smith, B.; Wien, F.; Miles, A.; Wallace, B. A. *Anal. Biochem.* **2004**, 332, 285.
- (49) *Time Dependent Density Functional Theory*; Marques, M. A. L., Nogueira, F., Ullrich, C., Burke, K., Rubio, A., Gross, E. K. U., Eds.; Lecture Notes in Physics 706; Springer Verlag: Berlin, 2006.
- (50) Yabana, K.; Bertsch, G. *Phys. Rev. B* **1996**, 54, 4484.
- (51) Castro, A.; Marques, M. A. L.; Appel, H.; Oliveira, M.; Rozzi, C. A.; Andrade, X.; Lorenzen, F.; Gross, E. K. U.; Rubio, A. *Phys. Status Solidi B* **2006**, 243, 2465.
- (52) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, 43, 1993.
- (53) Perdew, J.; Zunger, A. *Phys. Rev. B* **1981**, 23, 5048.
- (54) Kümmel, S.; Kronik, L. *Rev. Mod. Phys.* **2008**, 80, 3.
- (55) Mennucci, B.; Tomasi, J.; Cammi, R.; Cheeseman, J. R.; Frisch, J.; Devlin, F. J.; Gabriel, S.; Stephens, P. J. *J. Phys. Chem. A* **2002**, 106, 6102.
- (56) Cavaluzzi, M. J.; Borer, P. N. *Nucleic Acids Res.* **2004**, 32, e13.
- (57) Sprecher, C. A.; Johnson, W. C., Jr. *Biopolymers* **1977**, 16, 2243.
- (58) Johnson, W. C., Jr.; Tinoco, I., Jr. *Biopolymers* **1969**, 7, 727.
- (59) Callis, P. R. *Photochem. Photobiol.* **1986**, 44, 315.
- (60) Williams, A. L., Jr.; Cheong, C.; Tinoco, I., Jr.; Clark, L. B. *Nucleic Acids Res.* **1986**, 14, 6649.
- (61) Moore, D. S.; Williams, A. L., Jr. *Biopolymers* **1986**, 25, 1461.
- (62) Callis, P. R. *Annu. Rev. Phys. Chem.* **1983**, 34, 329.
- (63) Burkoff, A. M.; Tullius, T. D. *Nature* **1988**, 331, 455.
- (64) Steff, R.; Wu, H.; Ravindranathan, S.; Sklenar, V.; Feigon, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 1177.
- (65) Lu, X.-J.; Olson, W. K. *Nucleic Acids Res.* **2003**, 31, 5108.
- (66) Tonzani, S.; Schatz, G. *J. Am. Chem. Soc.* **2008**, 130, 7607.
- (67) Grozema, F. C.; Tonzani, S.; Berlin, Y.; Schatz, G. C.; Siebbeles, L. D. A.; Ratner, M. A. *J. Am. Chem. Soc.* **2008**, 130, 5157.
- (68) Troisi, A.; Orlandi, G. *J. Phys. Chem. B* **2002**, 106, 2093.