

Simulations of Structure and Vibrational Spectra of Deoxyoctanucleotides

Petr Bouř,^{*,†} Valery Andrushchenko,[‡] Martin Kabeláč,[†] Vanitha Maharaj,[‡] and Hal Wieser[‡]

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16610, Praha 6, Czech Republic, and Department of Chemistry, University of Calgary, 2500 University Drive, Calgary AB, Canada T2N 1N4

Received: March 9, 2005; In Final Form: September 6, 2005

Combined molecular dynamics and ab initio computations were applied for analysis of infrared absorption and vibrational circular dichroism spectra of deoxyoctanucleotides. Unlike for previous idealized models, molecular geometries of these shorter DNA fragments in solutions were obtained as dynamic averages from simulations in a periodic water box. Vibrational spectra for the whole octamers including hydrogen-bonded solvent molecules were simulated on the basis of density-functional calculations on small fragments and subsequent transfer of molecular property tensors. Explicit and continuum solvent models were compared. Apparently, the DNA segments retain an approximate B-conformation in the aqueous solutions, but the terminal base pairs significantly deviate from the planar arrangement and the vibrational circular dichroism spectrum for (CG)₂ nucleotide indicates a larger average helical twist. Sodium counterions moved freely around the molecule during the simulation and do not influence spectral intensities. Simulated absorption spectra faithfully reproduced the experimental signal of principal functional groups, while only qualitative agreement was obtained for the dependence on the basis sequence.

Introduction

In the past vibrational spectroscopy has not only contributed to structural studies of biopolymer molecules, namely peptides and nucleic acids, but also helped to elucidate important interactions between molecular functional groups and between the solvent and solute.^{1–4} However, the interpretation of experimental spectral intensities is rather difficult because individual transitions are not resolved due to natural broadening of spectral lines. Also, the usual molecules of biological interest are rather big and flexible and interact strongly with the environment, which makes conventional theoretical descriptions difficult. Thus, for infrared and any other optical spectra the link between the structure and spectral shape is often less obvious than for the NMR or X-ray methods, for example. On the other hand, complete understanding of the optical response, such as the vibrational spectrum, provides additional insight into other properties including molecular conformational behavior and interaction with the solvent and other molecules, or even to biological activity.^{2,5,6}

For infrared absorption (IR) and vibrational circular dichroism (VCD) the understanding is realized by simulation of observed frequencies and spectral intensities. Previously, very good results for larger molecular systems were achieved with semiempirical schemes for spectral simulations, such as the DeVoe polarizability theory^{7,8} and other modifications of the coupled oscillator/dipole coupling models.^{9–20} On the other hand, many parameters must be taken from experiment in these approaches, and thus simulations based on the first principles appear more universal. Application of the empirical schemes is problematic, especially for systems of unusual structure or where the assignment of vibrational transitions is not obvious. We thus calculate the

spectral properties by ab initio techniques, although we are still dependent on the empirical force field used for the geometry estimation.

Within the harmonic approximation, quantum-chemical simulation of the vibrational spectra is relatively easy task for small and rigid systems,^{21,22} but becomes quite complex for nucleic acids. Nevertheless, some IR and VCD spectroscopy data could be interpreted and provided useful data about these large molecules, including the B–Z transition of DNA^{17,18,23} or structural RNA forms.^{24–26} Typically, the density functional theory (DFT) is used as it provides results comparable with those obtained by the wave function methods (such as MP2), but within shorter computational times.²⁷ From the experimental point of view, the infrared techniques, especially VCD, were found in many aspects more convenient than electronic absorption and circular dichroism (ECD), as the vibrational spectra have more resolved features and reflect molecular structure more locally, providing thus better spatial resolution. Previously, we performed quantum mechanical vibrational spectra simulations both for the DNA²⁸ and RNA²⁹ forms. These were based on vacuum models or on idealized molecular geometries derived from X-ray and NMR data. However, these approximations could not be applied for other forms on nucleic acids, such as the present octamer set, which cannot be crystallized, their structure is irregular, or they are too complex to be determined by the NMR technique.

In this study, a more consistent approach based on the molecular dynamics (MD) geometry is used for simulations of the spectra of a unique set of seven DNA deoxyoctanucleotides. As shown below, this leads not only to improvement of the spectral profiles but also provides a feedback about the quality of the MD force field model. Unlike for previous approaches, estimates of inhomogeneous broadening of spectral lines caused by structural irregularities and temperature motion are obtained directly. Additionally, the solvent and the solute can be treated

* To whom correspondence should be addressed. E-mail: bour@uochb.cas.cz.

[†] Academy of Sciences of the Czech Republic.

[‡] University of Calgary.

at the same level of approximation, at least within the first hydration sphere. Due to the huge amount of computational effort needed for the simulations (months of computer time), we concentrate our efforts on analysis of the (CG)₈ deoxyoctanucleotide spectrum and a few specific problems, as in the discrimination in the absorption spectra between various DNA sequences. We thus hope that the study can contribute to quantitative understanding of the IR spectra of molecules of this size and, more importantly, to more reliable sequence and structure determination of various DNA forms.

The ab initio methods are generally more reliable and universal than the empirical approaches, and their accuracy can be, at least conceptually, improved at will. Since systems of this size cannot be subjected to a reasonable quantum-chemical computation directly, the locality of the vibrational interactions is explored and properties of the target system are derived from suitably chosen fragments. The transfer of the molecular property tensors, namely, the harmonic force field, atomic polar, and axial tensors, can be done conveniently atom-by-atom in Cartesian coordinates.³⁰ Lately, a normal mode optimization technique has been proposed for constrained geometry optimization of such fragments, ensuring thus a full relaxation of the high-energy vibrational modes of interest while the desired secondary structure is conserved.³¹ The magnetic field perturbation (MFP) theory^{32,33} of Stephens can be used for computation of spectral VCD intensities, and the main force field components can be obtained ab initio at the same level. This approach was successfully used in the past also for interpretation of spectra of peptides and proteins, provided that suitable fragments comprising the main interactions between the amide group chromophores were selected.^{34,35} It was also shown that such an approach leads to realistic line widths caused by dynamic temperature broadening if several configurations obtained during MD simulation are averaged³⁶ and provides faithful spectral response of pure solvent.³⁷

Materials and Methods

Experimental Methods. Seven self-complementary deoxyoctanucleotides (1–7, Table 1) were synthesized on the automatic synthesizer employing solid-phase synthesis.^{38,39} The samples were purified using Sephadex G-25 column, eluted with distilled water, and lyophilized. The IR and VCD measurements were performed in a 25 μm BaF₂ liquid cell (International Crystal Laboratories, Inc.) at constant temperature of 4 °C in D₂O buffered at pD = 7.3 with sodium cacodylate (40 mM sodium cacodylate and 50 mM NaCl), on the VCD instrument at the University of Calgary.⁴⁰ The concentration of the duplex nucleotide was 20 mM. Under such experimental conditions, the stable B-DNA conformation was ensured.⁴¹ A resolution of 4 cm⁻¹ was used during the scanning. The total of 7500 AC scans was acquired in 165 min and ratioed against 750 DC scans. The resulting VCD spectra were corrected for polarization artifacts by subtracting the spectra of the solvent (40 mM sodium cacodylate and 50 mM NaCl in D₂O), recorded with the same instrumental conditions. Details of the VCD measurement are given elsewhere.⁴²

Molecular Dynamics Simulations. Standard MD simulation techniques were used in the computations, details of which can be found elsewhere.^{43,44} First, the structure of the first hydration shell for regular fixed DNA geometry was modeled. The B-DNA double helix octamers were generated with the Tinker program package.⁴⁵ We used the force field developed by Cornell et al.,⁴⁶ often referred to as the AMBER force field, according to the program where it was first implemented.⁴⁷ The

TABLE 1: List of the DNA Octanucleotide Sequences

1	CGCGCGCG	5	CGCTAGCG
2	CCTGGTCC	6	CGATATCG
3	CGTATACG	7	CGAATTCG
4	CGCATGCG		

octamers were subjected to 40 minimization steps in order to account for the worst inaccuracies of the Tinker molecular editor. Then a water shell 8 Å thick was created around the DNA molecule. While the DNA coordinates were fixed, 100 minimization points, 1000 dynamical points, and 1000 new minimization points were run in order to obtain a rough estimate of the distribution of the water molecules around the octanucleotide helix. Finally, 50 minimization steps were allowed without constraints, so that the DNA atoms could partially relax in the presence of water. For the following ab initio simulations of the spectra, only water molecules closer than 3.6 Å and hydrogen-bonded to DNA were retained.

For an unrestricted dynamics run bigger periodic boxes and sodium counterions were used in order to neutralize the charges of the phosphate groups. Starting geometries of the octamers were generated by programs “nucgen” and “nukit”, which are part of the AMBER 4.1 package,⁴⁷ using Arnott’s geometry parameters⁴⁸ and the force field of Cornell and co-workers.⁴⁶ The phosphate charges were neutralized by the Na⁺ cations using standard parameters of the routine “addions”. A periodic box of the TIP3P model⁴⁷ of water molecules was extended to the distance of 10 Å from any atom of the solute. Moment of inertia axes of the B-DNA octamers were oriented approximately along the Cartesian coordinates, and the number of the water molecules varied in the range of 3400–3900 for the seven DNA sequences. The MD calculations were carried out using the “sander” module of AMBER 4.1, with the NTP thermodynamic ensemble with a temperature of 300 K and pressure of 1 atm with the SHAKE option for the hydrogen atoms with a tolerance of 0.0005 Å and a 2 fs time step. A 9 Å cutoff was applied to intermolecular interactions. The Berendsen temperature coupling algorithm was used with a time constant of 0.2 ps. The nonbonded pair list was updated every 10 steps.

The MD run was preceded by an equilibration stage. First, 1000 restricted minimization steps were performed with fixed intramolecular parameters of the nucleic acids. A force constant = 500 kcal/(mol Å²) was applied on each bond of the solute. Except for the initial minimization, the simulations were performed with the particle mesh Ewald (PME) method, to ensure stability of the DNA double helix. For dynamics runs after minimizations, initial velocities were assigned from a Maxwellian distribution corresponding to the temperature of 300 K. The equilibration was continued by 25 ps of PME dynamics, with the position of the nucleic acid fixed. Subsequently, 1000 steps of minimization and 3 ps of MD simulation were carried out, both with the 25 kcal/(mol Å²) restraints for the positions of all solvent atoms. Five rounds of 1000 minimization steps followed with solute restraints reduced by 5 kcal/(mol Å²) in the course of each round and followed further by 20 ps of MD run continued before the final 5 ns production MD simulation. Trajectory files with the coordinates were generated in 1 ps periods. To minimize the problem of the “flying ice cube” (cold solute/hot solvent),⁴⁹ the center of mass velocity (total momentum) was removed during the production dynamics in intervals of 10 ps.

The results were analyzed using the “carnal” module of AMBER 4.1. The module also provided the averaged DNA structures. This averaging, however, led to a few artifacts in the geometry, namely, to unrealistic positions of the three

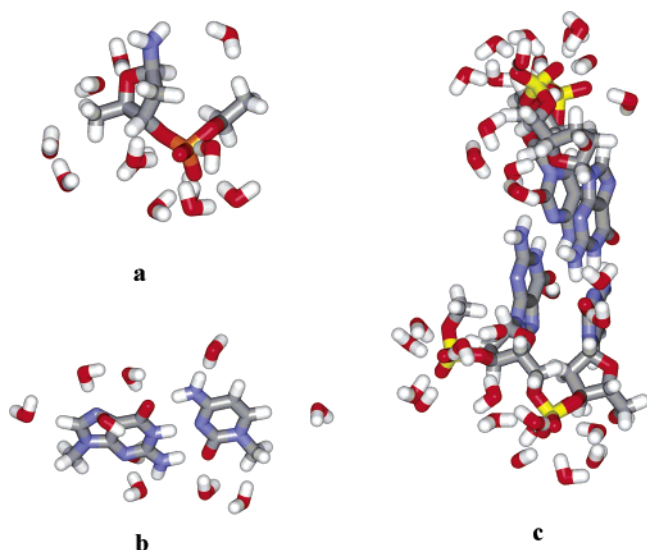


Figure 1. Main hydrated nucleotide fragments used in the computations: (a) the sugar–phosphate, (b) GC base pair, and (c) $(CG)_2$ dinucleotide. For a and b, electromagnetic tensors and force field were obtained at the BPW91/6-31G** level; for c the remaining force constants were calculated with the PM3 method.

hydrogen atoms of the thymine methyl group. Thus the molecule was allowed to relax during additional 80 steps of conjugate gradient minimization, using the same force field as that for the MD run. To model the hydration shell, the nucleotides had to be soaked in water for the second time. A periodic water box was used in the same manner as that described above. A 200 ps molecular dynamics run with fully restrained solute was applied, and the last geometry was stored. Positions of the solvent molecules were minimized by 100 000 steps of the conjugate gradient minimization. Similarly to that for the idealized structures, only hydrogen-bonded water molecules within the distance of 3.6 Å to the solute were used for vibrational spectra calculations.

Ab Initio Simulation of Vibrational Spectra. Harmonic force fields and atomic polar and axial tensors of the hydrated deoxyoctanucleotides were obtained via Cartesian-coordinate transfer from smaller fragments.^{28,30,34,50,51} The fragmentation philosophy can be illustrated on the hydrated octanucleotides, each containing about 1000 atoms. The molecules were divided into 24 fragments of approximately equal size, 8 base pairs and 16 sugar–phosphate residues (fragments a and b in Figure 1) with the hydrogen-bonded water molecules attached. For these systems, harmonic force constants and atomic axial and polar tensors could be calculated ab initio. Additionally, dinucleotide fragments were made (fragment c, Figure 1) for which a lower PM3⁵² level was used for calculation of the longer range force field terms. The smaller fragments were minimized in the normal coordinates⁵³ at the BPW91/6-31G** DFT level.^{54,55} During the minimization modes with frequencies $\omega \in (-300, 300 \text{ cm}^{-1})$ were frozen, which ensured only a minimal change of the geometry, but the higher frequency vibrational motions present in the experimental spectra were relaxed. Program QGRAD⁵³ was used for the normal mode technique, while the ab initio part of the computation was done with the GAUSSIAN⁵⁶ software. Note that for the simulation of irregular DNA geometries, such as the averaged MD structures, fragments could not be propagated and for each one unique ab initio computation was needed. As an alternative to the explicit solvent clusters, the COSMO^{57,58} continuum solvent model was used as implemented in GAUSSIAN. Finally, with the aid of homemade programs and MCM95 graphic software,^{30,59} the property tensors

were transferred to the octanucleotide, its entire force field diagonalized and combined with the intensity tensors. The spectra were simulated using Lorentzian bands 7.5 cm^{-1} wide (full width at half-height, fwhh).

Results

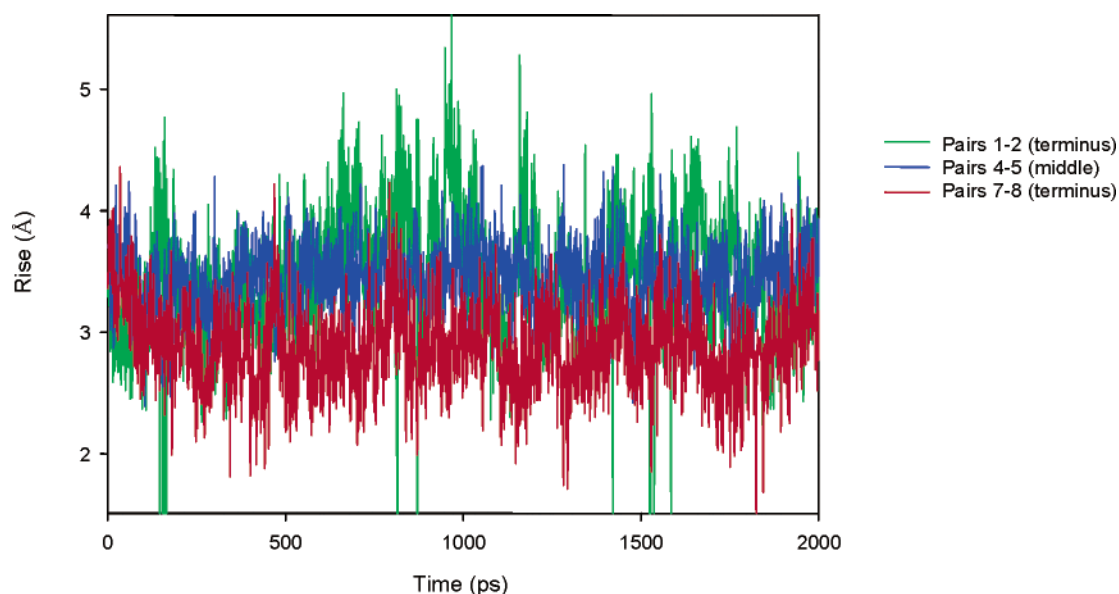
Molecular Geometries. The molecules retained a relatively stable conformation during the MD run, after the initial idealized geometry was accommodated in the solvent environment. Thus an average geometry could be assessed for each of the seven octamers and used further in the spectral simulations. The initial B-DNA helix partially unwound and opened in the solvent, so that both the minor and major grooves were more exposed, while the double helix somewhat shrank during this process (quantitative output can be found in the supporting section). For the seven octanucleotide MD average geometries the skeletal ($\alpha, \beta, \gamma, \delta, \epsilon, \zeta$) and χ angles⁷² are listed in Table 2. We can see that the average values stay reasonably close to those found in the standard B-DNA form, especially if we realize that this conformation normally exhibits relatively large variations in the torsion angles according to the environment and sequence.⁶⁰ Angles $\alpha, \beta, \gamma, \epsilon,$ and ζ occasionally adopt values closer to those of the structurally similar A-form of DNA, while the standard parameters of the Z-form of opposite helicity⁶¹ clearly do not fit the octamer values. Standard deviations are rather large, because of the contribution of the molecular termini. Apart from the spectral evidence discussed below, we suppose that molecular moments of inertia (see Supporting Information for calculated values) could, as indirectly measurable quantities,⁶² in principle discriminate between the different octanucleotide sequences. For example, the averaged, solvated structures are longer and thinner (the anisotropy parameter varied from 2.95 to 2.62) than the regular (infinite) ones. Unfortunately, no viscosity data are available for comparison. Obviously, the AMBER force field may not be adequate for quantitatively correct prediction of the DNA geometry. A high-level quantum chemical treatment, including the correlation energy, was found essential, especially for correct description of the stacking dispersion interaction.⁶³ Nevertheless, we find it important that the simulation produces systems that behave quantitatively correct as found by the experiment, as all experimental data confirm the presence of the B-conformation in the samples. Also, the terminal base pairs are less planar than those inside the octamers, the angle between the planes of the two bases varying within 10–20°, which is in accord with expectations. It can be explained by erosion of the ideal double-helical structure by the solvent, which competes with the hydrogen bonds between the terminal cytosine and guanine bases. As an example of the time dependence of helical parameters, we plot the base pair rise in Figure 2 for the nucleotide 1. The terminal groups noticeably differ from the middle of the molecules, nevertheless the rise values oscillate reasonably well around the standard value (3.32 Å). The principal geometrical behavior does not depend dramatically on particular DNA sequence, which can reflect both an inaccuracy of the AMBER force field as well as a general stability of the DNA molecule, independent of base content and sequence.

Na⁺ Cations. The primary role of the sodium atoms during the simulation was to neutralize the negative charge on the phosphate and thus stabilize the double-helix in solution.^{64,65} In crystal structures, Na⁺ ions bind relatively firmly to the phosphate oxygen atoms, with the Na–O distance of about 2.3 Å.⁶⁶ No rigid binding, however, was observed during our MD simulations when the sodium ions moved almost freely within

TABLE 2: Average Octanucleotide Torsion Angles As Obtained by the MD Simulation

structure ^b	a	b	g	d	e	z	c
Average Values							
1	-66	134	55	123	-118	-21	-104
2	-64	169	53	113	-141	-71	-125
3	-69	171	58	110	-170	-87	-129
4	-64	160	60	116	-167	-77	-109
5	-56	176	76	117	-162	-78	-119
6	-70	171	63	117	-163	-76	-119
7	-84	175	70	120	-156	-64	-133
B ^a	-41	136	38	139	-133	-157	-102
A ^a	-50	172	41	79	-146	-78	-154
Z(C/G) ^a	-137/47	-139/179	56/-169	138/99	-95/-104	80/-69	-159/68
Standard Deviations							
1	7	92	34	16	91	106	17
2	4	12	9	12	95	67	12
3	3	3	6	11	3	4	6
4	14	26	33	17	23	46	45
5	41	15	48	11	28	53	19
6	14	2	30	12	27	52	33
7	57	11	42	13	38	67	53

^a Regular DNA forms, ref 72. ^b Base pairs a–c.

**Figure 2.** Example of an MD geometry time dependence: the base pair rise in nucleotide 1.

the first and second hydration shell of DNA. Average distances from sodium to the closest oxygen (3.8–4.2 Å, depending on the nucleotide sequence) and phosphorus (4.5–5 Å) are clearly much bigger than that found in the crystals. In other words, the metal cations in solution can be found in the second solvation shell more frequently than in the first one. For the duplex **1** with no AT content the Na⁺ atoms moved farthest from the DNA molecule ($d(\text{Na}\cdots\text{P}) = 5 \text{ \AA}$), although this difference is probably not important statistically. This outcome of the MD simulations is in agreement with computations of the infrared spectra (not shown), indicating that the sodium ions cannot have significant influence on vibrational properties of DNA including the signal of the phosphate groups.

Force Field and the Vibrational Spectra. Principle features of experimental RNA and DNA spectra have been described in numerous studies previously^{3,4,14,15,17,18,23,26,28,29,40} and we recapitulate them only briefly. The C=O stretching region of the bases (1600–1700 cm⁻¹) gives rise to a strong absorption, while relatively minor distinct features can be found in the mid-IR region of 1150–1500 cm⁻¹, additionally perturbed by D₂O absorption (around 1220 cm⁻¹). Another strong absorption given by vibrations of the polar sugar and phosphate residues appear

approximately at 900–1150 cm⁻¹. In Figure 3, we can follow the fidelity with which various force field models provide the absorption profile of nucleotide **1**. Clearly, the force field obtained with the semiempirical PM3 method (trace a in the figure) exhibits the largest errors in frequencies and intensities, although main group (C=O stretch, D₂O absorption, and sugar–phosphate) absorption bands can be perhaps considered qualitatively correct. The DFT computations (traces b and c) provide significantly better agreement. Spectrum b was obtained with the geometry of DNA constrained to the regular X-ray derived B-conformation, while for c an averaged geometry from the unrestricted MD run was taken. The MD average structure provides better, namely, the C=O stretching absorption profile as well as the average frequencies, so we can conclude that molecular dynamics and relaxation needs to be taken into account in order to describe well the solvent–solute interaction.

Normal Mode Assignment. The assignment of absorption peaks presented in Figure 4 and Table 3 generally agrees with previous analysis.²⁸ The nucleotide **7** with the same number of AT and CG pairs was chosen ad hoc as a demonstrative example. As can be documented by the intensities in the figure, inclusion of the explicit solvent, although it could be done only

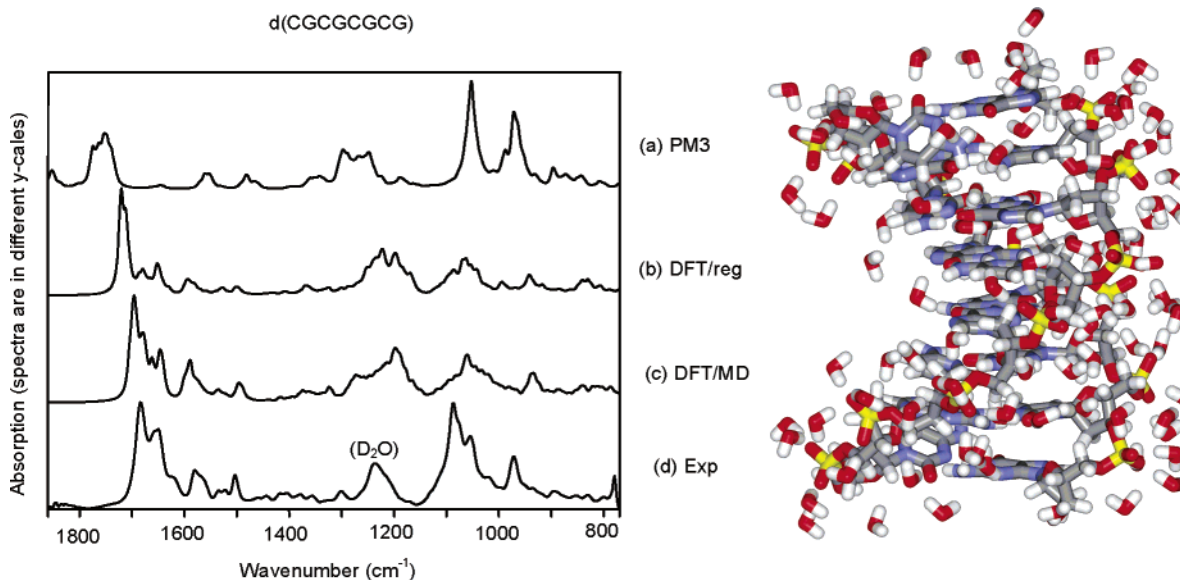


Figure 3. Absorption spectra of the d(CGCGCGCG) DNA nucleotide **1** with explicit solvent (geometry at the right-hand side) calculated by the fragmentation at the (a) PM3 and (b, c) BPW91/6-31G** levels as compared to the experiment (d). For a and b an idealized DNA structure was used; for (c) the spectrum is based on the MD averaged geometry.

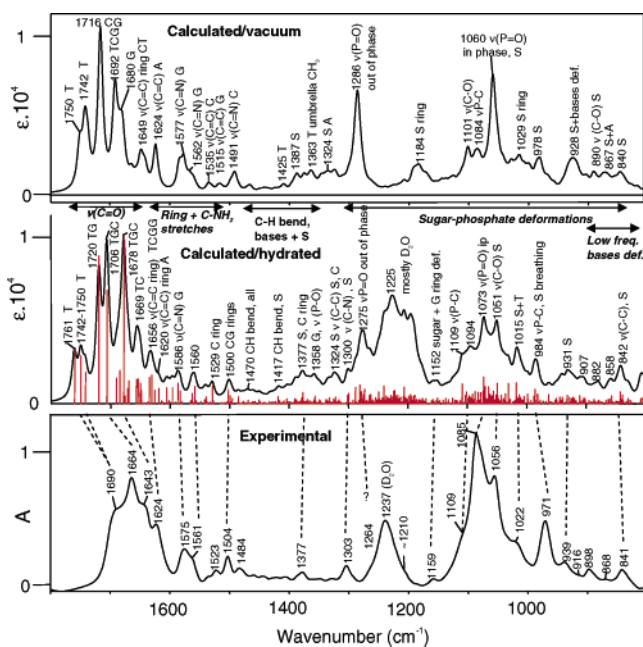


Figure 4. Band assignments for absorption spectra of the nucleotide **7** (CGAATTTCG) as calculated in a vacuum (top) and with the first hydration shell (middle) and compared to the experiment (bottom). Used abbreviations: C, T, A, G, cytosine, thymine, adenine, guanine; S, sugar; ip, in-phase vibration of the two P=O bonds; ν , stretching of a bond. For the middle spectrum, individual transitions are inserted as the red lines.

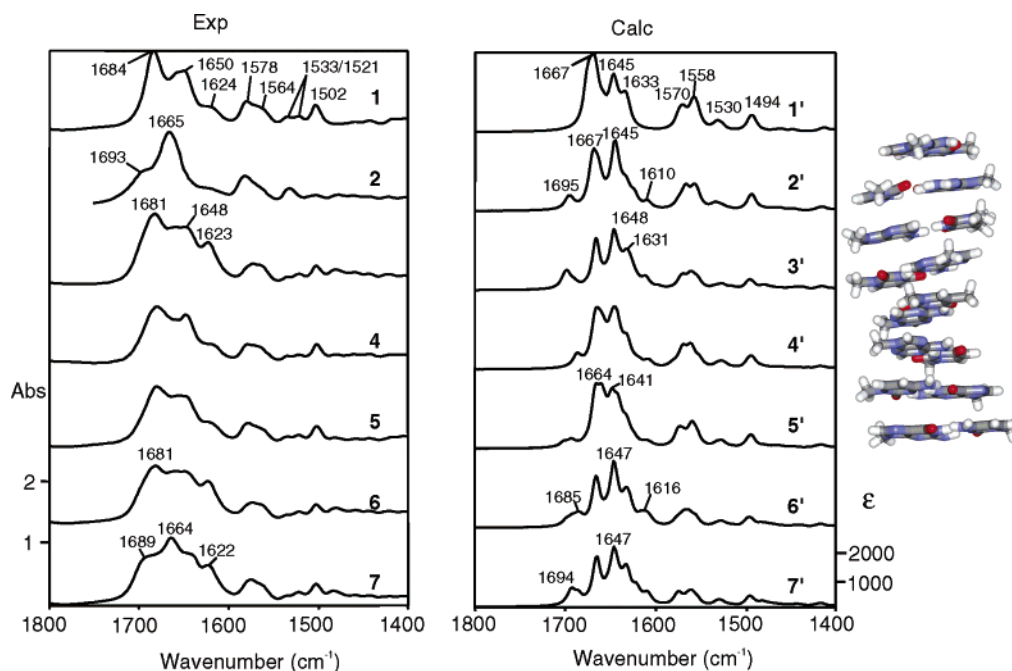
in the static fragment approximation,⁵¹ significantly modifies vacuum frequencies and intensities. Clear improvement can be observed for the overall absorption shape of the sugar-phosphate vibrations within $\sim 950\text{--}1150\text{ cm}^{-1}$. The simulated buildup of the intensity in the double maximum at $1051/1073\text{ cm}^{-1}$ for the hydrated model (in the middle), for example, can be associated with the experimental signal at $1056/1085\text{ cm}^{-1}$. The intensity of the residual baseline D₂O signal in the experimental and simulated spectra ($\sim 1150\text{--}1250\text{ cm}^{-1}$) obviously cannot be compared directly. The modeling of the C=O stretching region seems to be more problematic because of the band overlapping and uncertainty of the exact geometry. Similar

problems with the modeling of this vibrational motion were found in other studies, since the vibration is anharmonic, susceptible to formation of strong hydrogen bonds with the solvent or other molecular parts, and the polar C=O group strongly polarizes its environment.^{29,36,67} Apparently, the C=O stretching signal (centered in a vacuum around 1716 cm^{-1}) moved to lower wavenumbers for the explicit model, but still is deviating by $\sim 40\text{ cm}^{-1}$ from experiment. Positions of the base ring deformation modes ($\sim 1500\text{--}1660\text{ cm}^{-1}$) are less influenced by the solvent, nevertheless the intensity profile seems to correspond better to experiment (cf. the N=C stretching predicted in a vacuum at 1577 cm^{-1}). This may suggest that fixing the water molecules in the model overestimate the role of the solvent; however complete averaging of MD configurations was not reasonable because of the extensive CPU time demands of the simulations. Measurement in D₂O also inhibited clear assignment of the asymmetric P=O stretching vibrations.

Sequence Dependence of Absorption Intensities. We found that it was difficult to model the dependence of the spectra on variation in basis set sequence, as this was obscured by the errors introduced by the geometry variations, the position of the solvent, and incomplete MD averaging. We tried to eliminate these factors in a simplified system—eight methylated base pairs arranged in a regular B-DNA conformation for each of the nucleotide, as indicated for one of them at the right-hand side of Figure 5. For each couple of neighboring base pairs force field, atomic polar, and atomic axial tensors were calculated using the restricted normal mode optimization and Gaussian program described above, at the BPW91/6-31G** and with the COSMO^{57,58} continuum solvent model, and these parameters were transferred back at the desired octanucleotide sequence. Although the continuum model was not found completely adequate for description of the directional hydrogen bonds,⁶⁸ its usage eliminates the problems with the solvent position averaging. As can be seen in Figure 5, this procedure leads to much better absorption profile in the base vibration region than the MD model (cf. Figure 4). For nucleotide **1**, band-to-band assignment can be done for peaks observed at $1502, 1533, 1564, 1578, 1624, 1650,$ and 1684 cm^{-1} , calculated at $1494, 1530,$

TABLE 3: Calculated Vibrational Frequencies (cm^{-1}) of the Sugar–Phosphate Vibrational Normal Modes

mode number ^a	monomer			dimer	octamer (6)	
	vacuum	COSMO	explicit D ₂ O	explicit D ₂ O	explicit D ₂ O	expt
1, sugar def.	1123	1106	1150	1122, 1132		
2, $\nu(\text{C}=\text{O})$, S def.	1101	1092	1093	1106		
3, $\nu(\text{C}-\text{C})$, S	1093	1084	1138	1122		
4, $\nu(\text{P}=\text{O})$ in phase	1062	1053	1184, 1066	1086, 1068	1060	1088
5, $\nu(\text{C}-\text{C})$, S	1053	1053	1032			
6, $\nu(\text{P}=\text{O})$ in phase	1046	1014	1221	1078, 1048		
7, $\nu(\text{C}-\text{C})$, S	1032	1017	1241			
8, $\nu(\text{C}-\text{C})$, $\nu(\text{C}-\text{O})$, S	994	982	1117, 1022	1014		
9, $\nu(\text{C}-\text{C})$	947	940	973	994, 976		
19, $\nu(\text{C}-\text{O})$, $\nu(\text{P}-\text{O})$	939	936	901, 932	952, 924	934	972
11, S ring, breathing	877	875	861	890		
12, S ring, breathing	862	854	847	868		

^a S, sugar.**Figure 5.** Experimental (left) and calculated (right) absorption profiles for the seven octanucleotide sequences (defined in Table 1) at the C=O stretching and nitrogen base ring vibration region. Only stacked base pairs without the sugar–phosphate backbone were used in the computations; the model structure is indicated at the right-hand side. The BPW91/6-31G**/COSMO approximation level was used.

1558, 1570, 1633, 1645, and 1667 cm^{-1} , respectively. Also the relative intensities agree reasonably well. Despite this overall agreement, only certain features of the sequence dependence are explained by the modeling. For example, for nucleotide **2**, the experimental relative intensity of the highest frequency C=O stretching band at 1693 cm^{-1} is smaller than the signal centered at 1665 cm^{-1} , contrary to nucleotide **1** (bands at 1684 and 1650 cm^{-1}). This flip or relative intensities for **1** and **2** is reproduced by the model (see calculated bands at 1667 and 1645 cm^{-1}), but qualitatively underestimated. For nucleotide **3**, the experimental band at 1623 cm^{-1} becomes more intense than that for **2**, in accordance with the calculated signal at 1631 cm^{-1} . Spectra of **4** and **5** are almost indistinguishable, both in experiment and in the theory. Experimental absorption of the C=O stretching region of **6**, **7** becomes more dispersed than for **4**, **5**, which is modeled, but not quantitatively. Very little dependence on the sequence can be noticed at the lower frequency region (1400–1600 cm^{-1}), both in experiment and in the simulation. This can be explained by nonpolar character of the base ring modes that contribute to these transitions, while the polar C=O stretching modes strongly couple with neighboring base pairs.

Vibrational Circular Dichroism. In Figure 6 at the right-hand side simulated and experimental VCD spectra for the sugar–phosphate vibrations of nucleotide **1** are plotted. We found a very small variance of the VCD signal with respect to the nucleotide sequence in this frequency region. This is in accordance with previous observations, where the signal depended only slightly on experimental conditions,²⁸ while its shape differed more only for the B- and Z-forms.¹⁸ The simulated spectrum for the average MD structures (trace a) agrees reasonably with the experiment, namely, the negative lobe at 1062 (experimentally 1089 cm^{-1}) and two positive peaks at 1042 and 1022 cm^{-1} (experimentally at 1072 and 1089 cm^{-1}). The calculated $-/+$ couplet at 976/966 cm^{-1} supposedly does not have balance intensities and appears as a negative signal at 973 cm^{-1} in this particular experiment. However, the predicted $-/+$ pattern is often seen experimentally for analogous systems.²³ Signs of the bending motions at 930 (–), 896 (+), and 888(–) cm^{-1} (experimentally at 914, 897, and 888 cm^{-1} , respectively) appears again to be reproduced correctly. Thus the simulation with the MD geometry provides significantly better spectra in the sugar–phosphate region than previous.^{28,29} It should be also noted that the normal-mode restricted

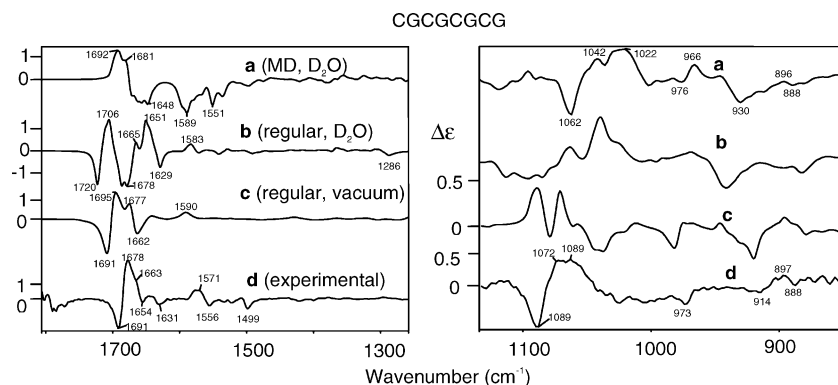


Figure 6. VCD spectra of the d(CGCGCGCG) octanucleotide (**1**) simulated for the MD (a) and regular X-ray geometry with (b) and without (c) the solvent, as compared to the experimental spectrum (d). The base (left) and sugar–phosphate (right) vibration regions are expanded separately.

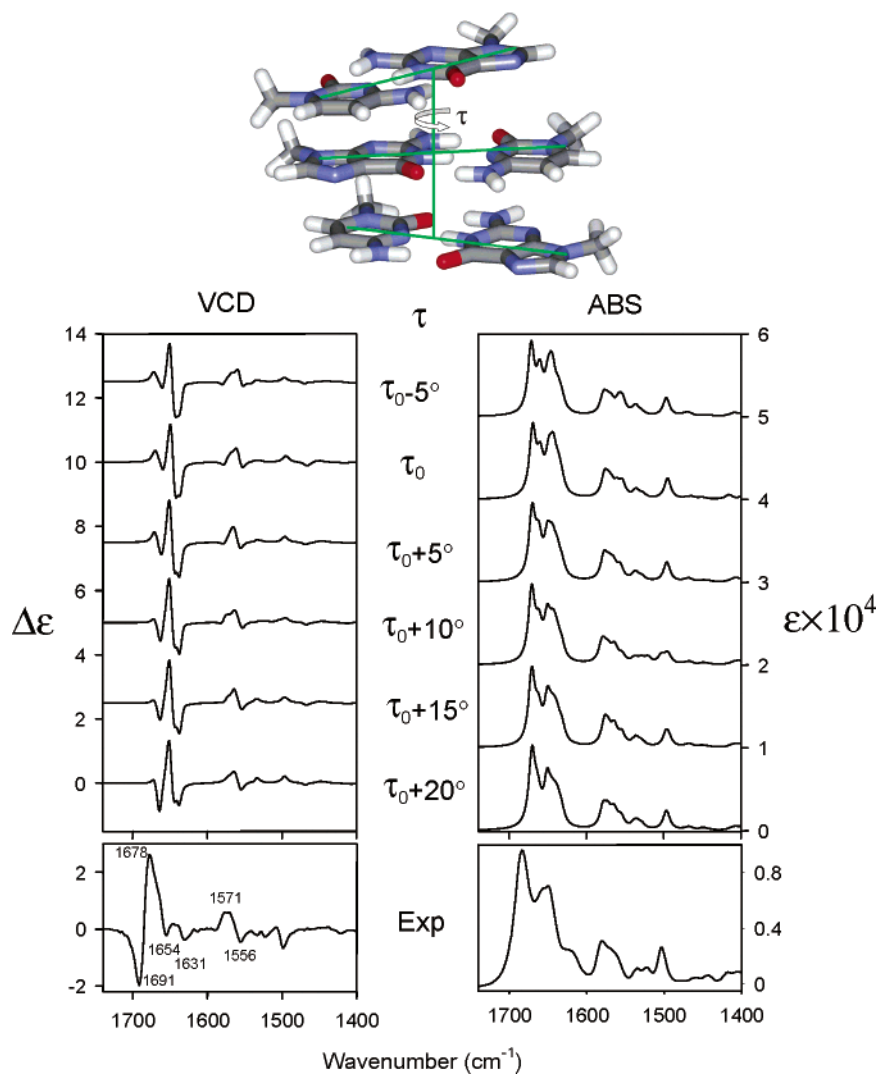


Figure 7. Simulated dependence of the VCD (left) and absorption (right) spectra on the helical twist of model d(CGCG) trimer. As indicated at the top, the twist angle τ was varied around the value for regular B-conformation (τ_0) within $-5\dots+20^\circ$ (for the right-handed helix positive deviation means bigger twist).

optimization method³¹ used in this study is indispensable for the simulations and improved the sugar–phosphate signal against the conventional torsion-angle restriction used previously.^{28,29} The normal mode method provided more stable optimization trajectories, faster convergence and better spectral profiles namely for the clusters containing hydrogen bonds.

Unfortunately, the other simplified models, the regular geometry with (trace b in Figure 6) and without (trace c) solvent, are much less satisfying. The difficulty stems from the complex

delocalized character of these vibrations, influenced strongly by the interaction with the water (D₂O). Additionally, VCD is given both by internal chirality of a monomer (one sugar–phosphate unit) as well as by exciton (resonance) coupling⁶⁹ between adjacent residues.

For the C=O stretching frequency region (left part of Figure 6) the spectrum with the MD average structure (trace a) does not reproduce the strong experimental VCD couplet (1691 (–)/1678 (+) cm^{-1}) characteristic for the B-conformation. In fact,

a weak negative signal at 1705 cm^{-1} is present in the calculated intensities, but does not show up if the spectrum is plotted with reasonable bandwidth. The \pm couplet is, however, reproduced with regular geometry, both with (at $1720/1706\text{ cm}^{-1}$) and without (at $1691/1695\text{ cm}^{-1}$) explicit water molecules (traces b and c in Figure 7, respectively). Supposedly, the average MD geometry is not better than the canonical structure for this part of the molecule. Also, a bigger ensemble of the solvent configurations may be required for more realistic wavenumbers and line widths,³⁶ but this could not be done with available computing means.

An alternate model was utilized for estimation of the sensitivity of the VCD spectra on the helical twist, based on an idealized d(CGC) base pair trimer, and is shown in Figure 7. As a starting geometry a regular B-DNA structure as generated by the Tinker software was used, and the base pair twist varied within $-5\dots+20^\circ$. The trimer has both the CG and GC interaction types²⁸ present in the d(CGCGCGCG) octamer, and its spectra could be calculated at a higher BPW91/6-31G**/COSMO level (via the tensor transfer from dimers), including the interactions between the base pairs. From the spectral dependence on the helical twist displayed at Figure 7 we see that even this abstraction does not lead to the correct VCD pattern, unless a deviation of $15\text{--}20^\circ$ from the canonical B-DNA twist is allowed. For $\tau \sim \tau_0 + 15^\circ$ the desired highest-frequency negative signal (experimentally at 1691 cm^{-1}) and the $+$ /couplet (experimentally at $1571/1556\text{ cm}^{-1}$) are developed, although the magnitude of the experimental bands of 1654 and 1631 cm^{-1} is overestimated by the model. Despite the limitation of the simulations, we thus speculate that the average geometry of nucleotide **1** deviates from the canonical B-DNA structure with its central part being more twisted than in a crystal. The same is probably also true for the other nucleotide sequences; however, the overall precision of the simulations and their demands on computer power do not allow a detailed analysis. On the other hand, from Figure 7 we can at least deduce the potential of the VCD spectroscopy coupled with simulation techniques for structural determination of biopolymers.

General Remarks and Conclusion

One could naively think that quantitative models of IR spectra of molecules of this size are hopeless given the current possibilities of the direct ab initio apparatus. However, on the basis of the experience above, we suggest that many limitations can be overcome by suitably chosen combined molecular/quantum dynamics model. Also, the locality of the vibrational motions and interactions makes the modeling easier. Especially the vibrational frequencies (cf. Figures 4, 5, and 7) already can be simulated with the same precision as for smaller molecules.^{21,22,70} The concept of “band-to-band” comparison, however, becomes obsolete for biopolymers, because observable bands involve hundreds of individual vibrational transitions. On the contrary, the inhomogeneous band broadening and overall band shape analysis, regularly neglected for smaller systems, seems to be of primary importance for extracting structural information from experimental spectra.

Supposedly, the harmonic approximation is sufficient for the modeling, although a systematic shift of the order of $\sim 16\text{ cm}^{-1}$ may be expected by the anharmonic force field for the C=O stretches.⁷¹ This, however, can be included in principle in our modeling, together with more precise ab initio levels, MD averaging, and solvent–solute interactions, etc. We find the possibility to systematically improve the model on the basis of the ab initio philosophy very appealing. On the other side the

empirical approaches in simple cases provide sufficient answers in a much shorter time scale and are certainly still advisable, also for complementary and conceptual reasons.^{7–10,13–15}

Overall, we may conclude that, for the first time, the vibrational spectra of short DNA fragments were simulated on the basis of averaged MD geometries adjusted to their final length, including the hydration sphere. The method provided generally better absorption spectral profiles than previous simulations based on the idealistic X-ray geometry in a vacuum, as well as more realistic VCD spectral profiles for the sugar–phosphate vibrations. The experimental VCD intensity for the C=O stretching of the bases was better reproduced by the regular geometry vacuum and continuum solvent models. This was attributed to inaccuracies in the geometry obtained with the AMBER force field and overestimation of the perturbation caused by the water solvent and incomplete MD averaging. Most of the vibrational transitions observed in the experimental spectra could be assigned, which makes the extraction of the structural information easier.

Acknowledgment. This work was supported by the Academy of Sciences of the Czech Republic, a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) (H.W.), and a Postdoctoral Fellowship from the Alberta Heritage Foundation for Medical Research (AHFMR) (V.A.).

Supporting Information Available: Geometry parameters and supplementary details to the ab initio spectral simulations with a movie showing the Na^+ atomic motion during the MD run. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- Polavarapu, P. L. *Vibrational Spectra: Principles and Applications with Emphasis on Optical Activity*; Elsevier: Amsterdam, 1998.
- Hicks, J. M. *Chirality: Physical Chemistry*; ACS Symposium Series 810; American Chemical Society: Washington, DC, 2002.
- Keiderling, T. A. *Vibrational Circular-Dichroism Spectroscopy of Peptides and Proteins*. In *Circular-Dichroism, Interpretation and Applications*; Nakanishi, K., Berova, N. D., Woody, R. W., Eds.; VCH: New York, 1994.
- Diem, M.; Polavarapu, P. L.; Oboodi, M.; Nafie, L. A. *J. Am. Chem. Soc.* **1982**, *104*, 3329.
- Hamm, P.; Zurek, M.; Wantele, W.; Meyer, M.; Scheer, H.; Zinth, W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1826.
- Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *114*, 2727.
- DeVoe, H. *J. Chem. Phys.* **1965**, *43*, 3199.
- DeVoe, H. *J. Phys. Chem.* **1971**, *75*, 1509.
- Holtzwarth, G.; Chabay, I. *J. Chem. Phys.* **1972**, *57*, 1632.
- Tinoco, I. *Radiat. Res.* **1963**, *20*, 133.
- Bouř, P.; Keiderling, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 9100.
- Halverson, K.; Sucholeiki, I.; Ashburn, T. T.; Lansbury, R. T. *J. Am. Chem. Soc.* **1991**, *113*, 6701.
- Schweitzer-Stenner, R. *J. Phys. Chem. B* **2004**, *108*, 16965.
- Maharaj, V.; Rauk, A.; van de Sande, J. H.; Wieser, H. *J. Mol. Struct.* **1997**, *408*, 315.
- Diem, M. *Introduction to Modern Vibrational Spectroscopy*; Wiley-Interscience: New York, 1993.
- Self, B. D.; Moore, D. S. *Biophys. J.* **1998**, *74*, 2249.
- Gulotta, M.; Goss, D. J.; Diem, M. *Biopolymers* **1989**, *28*, 2047.
- Wang, L.; Yang, L.; Keiderling, T. A. *Biophys. J.* **1994**, *67*, 2460.
- Xiang, T.; Goss, G. J.; Diem, M. *Biophys. J.* **1993**, *65*, 1255.
- Rauk, A.; Yang, D. *J. Phys. Chem.* **1992**, *96*, 437.
- Tománková, Z.; Setnička, V.; Urbanová, M.; Matějka, P.; Král, V.; Volka, K.; Bouř, P. *J. Org. Chem.* **2004**, *69*, 26.
- Devlin, F. J.; Stephens, P. J.; Scafato, P.; Superchi, S.; Rosini, C. *J. Phys. Chem. A* **2002**, *106*, 10510.
- Andrushchenko, V. V.; van de Sande, J. H.; Wieser, H.; Kornilova, S. V.; Blagoi, Yu. P. *J. Biomol. Struct. Dyn.* **1999**, *17*, 545.
- Wang, L.; Pančoška, P.; Keiderling, T. A. *Biochemistry* **1994**, *33*, 8428.
- Yang, L.; Keiderling, T. A. *Biopolymers* **1993**, *33*, 315.

- (26) Andrushchenko, V.; Blagoi, Y.; van de Sande, J. H.; Wieser, H. J. *Biomol. Struct. Dyn.* **2002**, *19*, 889.
- (27) Šponer, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2003**, *68*, 2231.
- (28) Andrushchenko, V.; Wieser, H.; Bouř, P. *J. Phys. Chem. B* **2002**, *106*, 12623.
- (29) Andrushchenko, V.; Wieser, H.; Bouř, P. *J. Phys. Chem. B* **2004**, *108*, 3899.
- (30) Bouř, P.; Sopková, J.; Bednářová, L.; Maloň, P.; Keiderling, T. A. *J. Comput. Chem.* **1997**, *18*, 646.
- (31) Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2002**, *117*, 4126.
- (32) Stephens, P. J. *J. Phys. Chem.* **1985**, *89*, 748.
- (33) Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Stephens, P. J. *Chem. Phys. Lett.* **1996**, *252*, 211.
- (34) Silva, R. A. G. D.; Kubelka, J.; Bouř, P.; Decatur, S. M.; Keiderling, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8318.
- (35) Bouř, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2000**, *53*, 380.
- (36) Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2003**, *119*, 11253.
- (37) Bouř, P. *Chem. Phys. Lett.* **2002**, *365*, 82.
- (38) Caruthers, M. H. *Science* **1985**, *230*, 281.
- (39) Blackburn, G. M.; Gait, M. J. *Nucleic Acids in Chemistry and Biology*; Oxford University Press: New York, 1990.
- (40) Tsankov, D.; Eggimann, T.; Wieser, H. *Appl. Spectrosc.* **1995**, *49*, 132.
- (41) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (42) Maharaj, V.; Tsankov, D.; van der Sande, J. H.; Wieser, H. J. *Mol. Struct.* **1995**, *349*, 25.
- (43) Kabeláč, M.; Hobza, P. *J. Phys. Chem. B* **2001**, *105*, 5804.
- (44) McCammon, J. A.; Harvey, S. C. *Dynamics of Protein and Nucleic Acids*; Cambridge University Press: New York, 1987.
- (45) Pappu, R. V.; Hart, R. K.; Ponder, J. W. *J. Phys. Chem. B* **1998**, *102*, 9725.
- (46) 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.
- (47) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Roos, W. S.; Cheatham, T. E., III; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Singh, P.; Weiner, P.; Kollman, P. A. *AMBER 4.1*; University of California: San Francisco, CA, 1995.
- (48) Arnott, S.; Chandrasekaran, Z.; Birdsall, D. L.; Leslie, A. G. W.; Ratliff, R. L. *Nature* **1980**, *283*, 743.
- (49) Harvey, S. C.; Tan, R., K.-Z.; Cheatham, T. E., III. *J. Comput. Chem.* **1998**, *19*, 726.
- (50) Bouř, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2002**, *65*, 45.
- (51) Bouř, P. *Chem. Phys. Lett.* **2002**, *365*, 82.
- (52) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220; 221–264.
- (53) Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2002**, *117*, 4126.
- (54) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (55) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.
- (56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, et al. *Gaussian 98*, rev. A.3 and A.7; Gaussian, Inc.: Pittsburgh, PA, 1998. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, et al. *Gaussian 03*, rev. B.02; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (57) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404.
- (58) Eckert, F.; Klamt, A. *AIChE J.* **2002**, *48*, 369.
- (59) Bouř, P.; Maloň, P. *Mcm95 Molecular Graphic Program*; Institute of Organic Chemistry and Biochemistry: Prague, Czech Republic, 2000. Downloadable at: <http://hanicka.uochb.cas.cz/~bour/mcm95.zip>.
- (60) Arnott, S.; Hukins, D. W. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504.
- (61) Pohl, F. M.; Jovin, T. M. *J. Mol. Biol.* **1972**, *67*, 375.
- (62) Kuwajima, S.; Manabe, A. *Chem. Phys. Lett.* **2001**, *332*, 105.
- (63) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.
- (64) Sigel, A.; Sigel, H. *Interactions of Metal Ions with Nucleotides, Nucleic Acids and Their Constituents, Vol. 32*; Marcel Dekker: New York, 1996.
- (65) Eichhorn, G. L. *Metal Ions in Genetic Information Transfer*; Elsevier: New York, 1981.
- (66) Schneider, B.; Kabeláč, M. *J. Am. Chem. Soc.* **1998**, *120*, 161.
- (67) Kubelka, J.; Keiderling, T. A. *J. Phys. Chem. A* **2001**, *105*, 10922.
- (68) Bouř, P. *J. Chem. Phys.* **2004**, *121*, 7545.
- (69) Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, 1982.
- (70) Bouř, P.; McCann, J.; Wieser, H. *J. Phys. Chem. A* **1998**, *102*, 102.
- (71) Zanni, M. T.; Asplund, M. C.; Hochstrasser, R. M. *J. Chem. Phys.* **2001**, *114*, 4579.
- (72) Blackburn, G. M.; Gait, M. J. *Nucleic Acids in Chemistry and Biology*; Oxford University Press: New York, 1996.