Oligonucleotide Analogues with a Nucleobase-Including Backbone

Part 7

Molecular Dynamics Simulation of a DNA Duplex Containing a 2′-Deoxyadenosine 8-(Hydroxymethyl)-Derived Nucleotide

by **Werngard Czechtizky**[a]), **Xavier Daura**[b]), **Andrea Vasella**[a]), and **Wilfred van Gunsteren***[b])

[a]) Laboratorium für Organische Chemie, ETH-Zentrum, CH-8092 Zürich
[b]) Laboratorium für Physikalische Chemie, ETH-Zentrum, CH-8092 Zürich

The structure and stability of a 14-mer DNA duplex containing a nucleotide analog with a hydroxymethyl substituent at the C(8) of 2′-deoxyadenosine has been investigated by molecular-dynamics simulation. The DNA duplex studied has the sequence 5′-d(CGTAAGCTCGATAG)-3′ · 5′-d(CTATCGA*GCTTACG)-3′, where the O(3′) of the dG_6 nucleotide in the second strand is linked through a phosphinato group with the O(10) of the dA_7^* 2′-deoxyadenosine-derived nucleotide. Previous experimental results showed that the stability of this duplex in aqueous solution of 0.1M NaCl at pH 7 and room temperature is significantly lower than that of the corresponding unmodified DNA duplex. Comparison of molecular-dynamics trajectories of the unmodified and modified B-DNA duplexes in aqueous solution, at similar conditions than the experiment, shows that the substitution of the dA nucleotide by the dA* nucleotide in the second strand induces stretching of the double helix, which results in opening of the grooves and consequent exposure of the double-helix core to the solvent.

1. Introduction. – The investigation of the structural properties of oligonucleotide analogues with a nucleobase-including backbone should help to answer the question of whether the structural differentiation between nucleobase and backbone in DNA, RNA, and their analogues is a prerequisite for the formation of stable homo- and/or heteroduplexes [1–6]. The incorporation of one 2′-deoxyadenosine-derived nucleotide, dA*, into a 14-mer DNA duplex (*Fig. 1*) has been already described [5]. UV Pairing studies indicated that the stability of this duplex is significantly lower than that of the corresponding unmodified DNA duplex ($\Delta T_m = 6.5$ K; *Fig. 2*) at a concentration of (2 + 2) μM and a pH of 7 in 0.1M aqueous NaCl.

Here, the molecular bases of the difference in thermal stability observed for the two DNA duplexes shown in *Fig. 1* are investigated by molecular-dynamics (MD) simulations. MD Simulations of the 5′-d(CGTAAGCTCGATAG)-3′ · 5′-d(CTATCGAGCTTACG)-3′ duplex (**1**) and the 5′-d(CGTAAGCTCGATAG)-3′ · 5′-d(CTATCGA*GCTTACG)-3′ duplex (**2**) have been performed in aqueous solution at a double-strand concentration of 8 mM, a NaCl concentration of 0.2M, an apparent pH value of 7, a temperature of 300 K, and a pressure of 1 atm. The trajectories of DNA configurations and energies are analyzed in terms of interaction energies, overall structure, population of *Watson-Crick* H-bonds, and atom-positional fluctuations.

2. Results and Discussion. – To investigate the relative stability of duplexes **1** and **2** in the respective simulations at 300 K, the trajectories of DNA configurations and

Duplexes:

1 **5'-d(CGTAAGCTCGATAG)-3'**
3'-d(GCATTCGAGCTATC)-5'

2 **5'-d(CGTAAGCTCGATAG)-3'**
3'-d(GCATTCGA*GCTATC)-5'

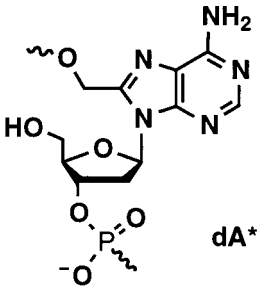

Fig. 1. *Modified 2'-deoxyadenosine-derived nucleotide dA* and duplexes* **1** *and* **2**

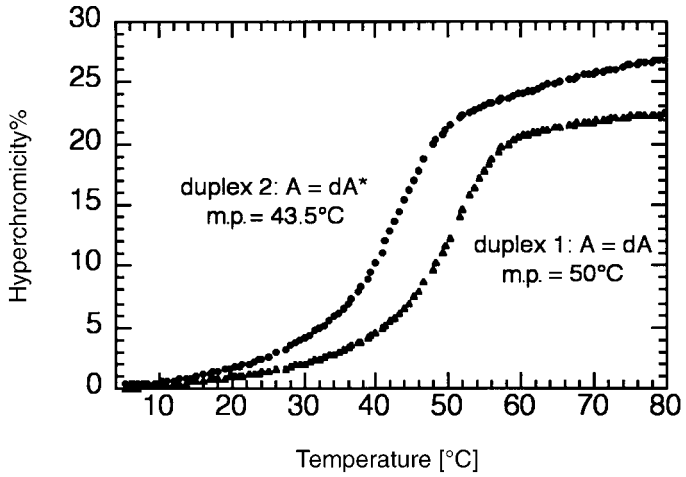

Fig. 2. *Temperature-dependent UV spectra* ('melting curves') *of the nonmodified and modified DNA·DNA duplexes* **1** *and* **2**

energies were analysed in terms of interaction energies, overall structure (atom-positional RMSD, and radius of gyration), population of interstrand *Watson-Crick* H-bonds, and fluctuations of atomic positions (RMSF).

Fig. 3 shows the internal interaction energy of the DNA, the interaction energy between DNA and ions, and the interaction energy between DNA and H_2O as a function of simulation time. The average energies for the time period 300–2300 ps are 2352 kJ mol^{-1}, −3179 kJ mol^{-1}, and −1097 kJ mol^{-1}, respectively, for system **1** (*Fig. 3,a*), and 2448 kJ mol^{-1}, −3108 kJ mol^{-1}, and −501 kJ mol^{-1}, respectively, for system **2** (*Fig. 3,b*). The energy values from the two systems are directly comparable, since the two duplexes differ only by three atoms, the number of ions in the solution is

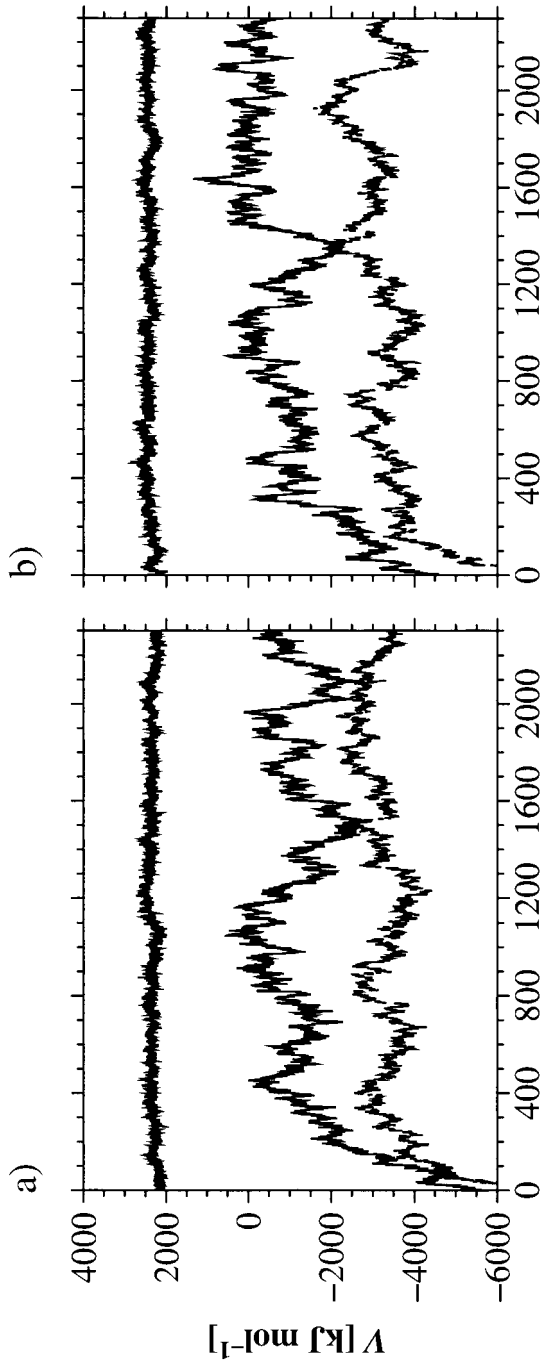

Fig. 3. *Interaction energies as a function of simulation time. a*) System **1**; *b*) system **2**. *Upper curve:* DNA internal interaction energy (bonded plus nonbonded terms); *middle curve:* DNA-H_2O interaction energy; *lower curve:* DNA-ions (Na^+ and Cl^-) interaction energy. Note that the curves for DNA-H_2O interaction and DNA-ions interaction (*a* and *b*) do not cross but merely touch and separate again.

the same, and solvation by H_2O goes beyond the 1.4 nm cutoff in both cases. From this analysis, it becomes already clear that, as suggested by the temperature-dependent UV experiments [5], the duplex **2** (including the dA* nucleotide analogue) is energetically disfavored with respect to duplex **1**. The largest energy difference is due to solvation. As will be shown, the double helix **2** stretches during the simulation. This stretching results in an opening of the grooves and a bigger area of the double-helix core being exposed to the solvent.

The atom-positional root-mean-square difference (RMSD) with the model B-DNA structure (*i.e.*, the initial structure; see *Exper. Part*) is plotted in *Fig. 4* as a function of simulation time. As with the energies (*Fig. 3*), it takes *ca.* 300 ps for the RMSD to reach an equilibrium average value. After 2.3 ns, the RMSD shows no trend to an increase. The average RMSD for the time period 300–2300 ps is 0.33 nm for duplex **1** (*Fig. 4*,*a*) and 0.45 nm for duplex **2** (*Fig. 4*,*b*).

Fig. 5 shows the percentage population of *Watson-Crick* H-bonds in the simulation-time period 300–2300 ps for the 35 H-bonds corresponding to the model B-DNA structure (see caption of *Fig. 5*). In general, the *Watson-Crick* H-bonding network is well-preserved in duplex **1** (*Fig. 5*,*a*). Only the first and the last two base pairs lack completely these characteristic interstrand H-bonds. This is not unexpected, given the different environment (relative to the rest of the strand) of the terminal nucleotide residues and their typically high mobility. Inspection of the structures sampled in the simulation of system **1** reveals that the two extreme base pairs tend to stack their aromatic rings perpendicularly to the double-helix axis rather than maintain the original H-bonds (*Fig. 6*,*a*). This can be understood as a way to minimise the aromatic surface area exposed to the solvent at the tails of the double strand. Interestingly, the smaller base (cytosine) is, at both termini, the one that faces the solvent, while its complementary base (guanine) stays buried in a sandwich between the cytosine base and the second (or the next to the last) base pair. This sandwich formation may also provoke a distortion in the pairing of the second (or the next to the last) base pair, as is here the case in one of the termini.

In duplex **2** the network of *Watson-Crick* H-bonds is clearly disrupted (*Fig. 5*,*b*). It is apparent that the damage to this network is caused by the introduction of the dA* nucleotide analogue in the second strand. The structural characteristics of the new nucleotide residue, including its contribution of nine bonds to the sugar-phosphate backbone instead of the usual six bonds (*Fig. 1*), induce a deviation of neighbouring nucleotide residues from their positions in a model B-DNA structure. Interestingly, the pairing of immediate neighbour bases is not destroyed (although it is affected) by the perturbation. Thus, H-bonds 13 to 26 (corresponding to the base pairs dG_6-dC_{23} to dG_{10}-dC_{19}) have a relatively high occurrence in the simulation. The perturbation introduced by the dA* residue is propagated asymmetrically in the 5′- and 3′-directions. The 3′-end of the dA* nucleotide analogue is identical to that of a dA nucleotide (*Fig. 1*). Thus, in principle, one would expect only a small deviation from the reference B-DNA structure in this direction. Although the perturbation is indeed smaller along the 3′-direction than along the 5′-direction (see below), both are sizable. In the 3′-direction (along the second strand), the propagation of the distortion ends up with a stacking of the aromatic rings of the dA_5-dT_{24} base pair (H-bonds 11 and 12 in *Fig. 5*,*b*). The following three base pairs in the 3′-direction, *i.e.*, dA_4-dT_{25} to dG_2-dC_{27}, retain the *Watson-Crick*

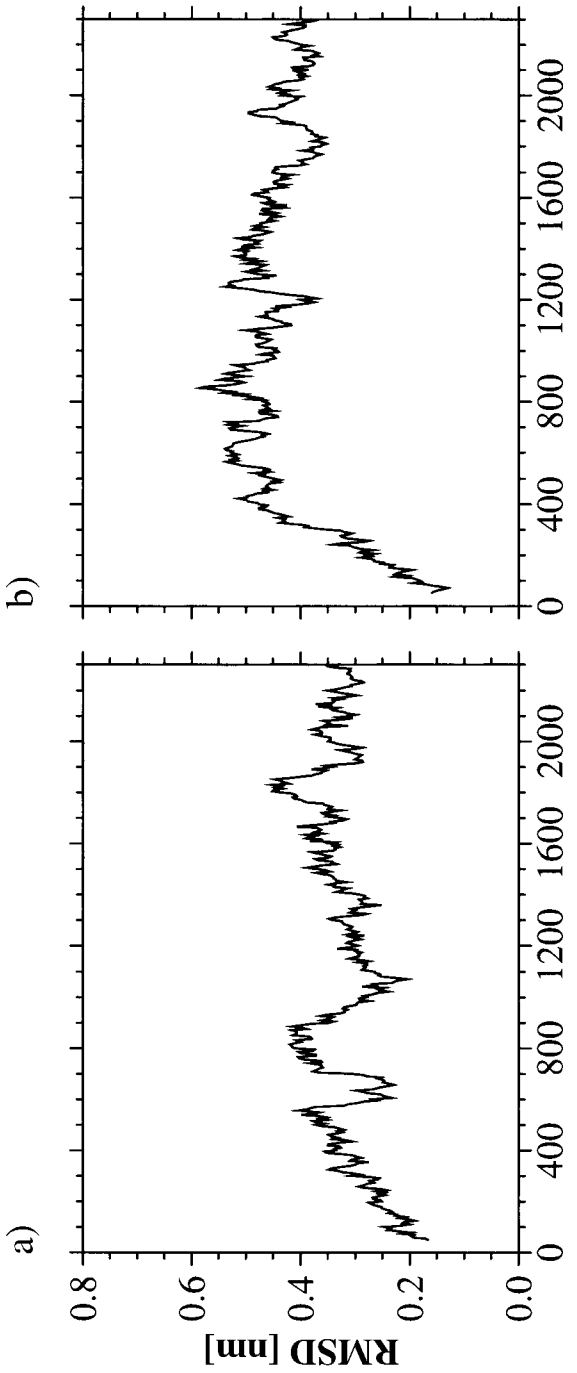

Fig. 4. *Atom-positional root-mean-square difference (RMSD) with the B-DNA model structure as a function of simulation time.* Translational and rotational fitting of trajectory structures to the model structure and RMSD calculation have been performed on the non-H-atoms of nucleotide residues 3 to 12 (first strand) and 17 to 26 (second strand). *a*) Duplex **1**; *b*) duplex **2**.

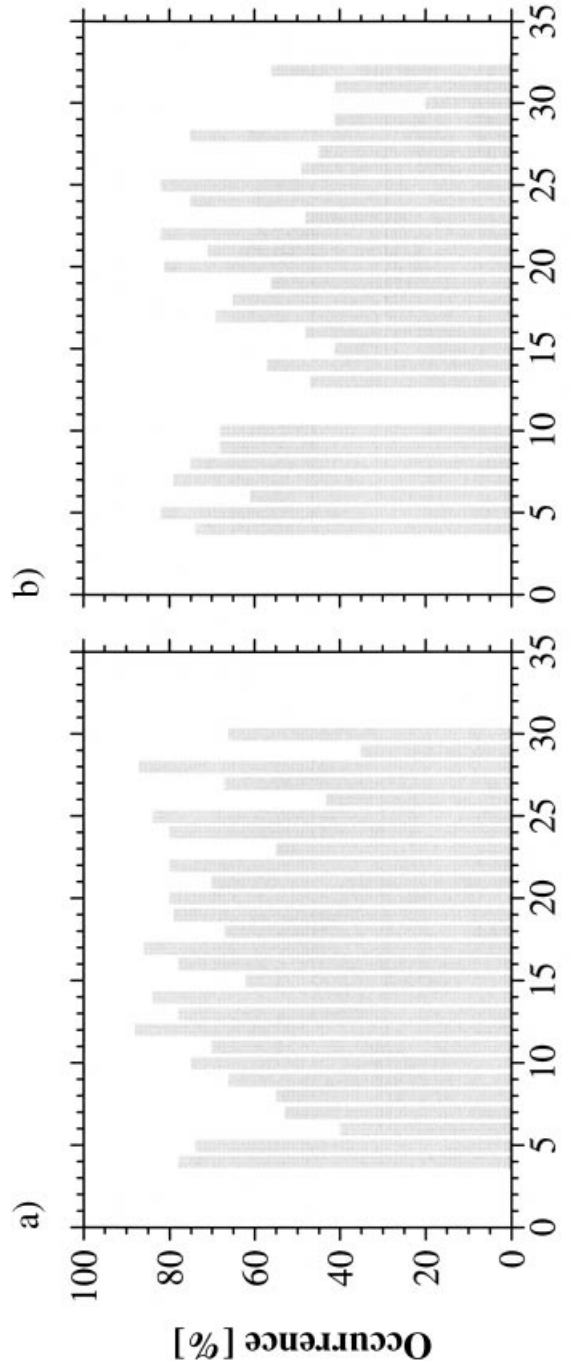

Fig. 5. *Percentage population of* Watson-Crick *H-bonds in the simulation time period 300 – 2300 ps. a*) Duplex **1**; *b*) duplex **2**. The H-bonds are numbered from 1 to 35 following the nucleotide sequence 5′-d(CGTAAGCTCGATAG)-3′ of the strand that is common to the two duplexes (first strand). The H-bond sequence within each nucleotide pair is N(4) – O(6), N(3) – N(1), O(2) – N(2) for the dC-dG pair (*e.g.*, dC_1 – dG_{28}); O(6) – N(4), N(1) – N(3), N(2) – O(2) for the dG-dC pair (*e.g.*, dG_2-dC_{27}); O(4) – N(6), N(3) – N(1) for the dT-dA pair (*e.g.*, dT_3-dA_{26}); N(6) – O(4), N(1) – N(3) for the dA-dT pair (*e.g.*, dA_4-dT_{25}).

H-bonding (H-bonds 4 to 10 in *Fig. 5,b*), while the base pair dC_1-dG_{28} shows the same stacking observed in duplex **1**, with the cytosine base facing the solvent, and the guanine base buried between the former and the next base pair. The base-phosphate end of the dA* nucleotide analogue is markedly different from the sugar-phosphate end of a dA nucleotide (*Fig. 1*). In this direction (5′ along the second strand), the distortion is propagated to the end of the duplex and destroys the *Watson-Crick* arrangement of the bases perpendicular to the double-helix axis (*Fig. 6,b*).

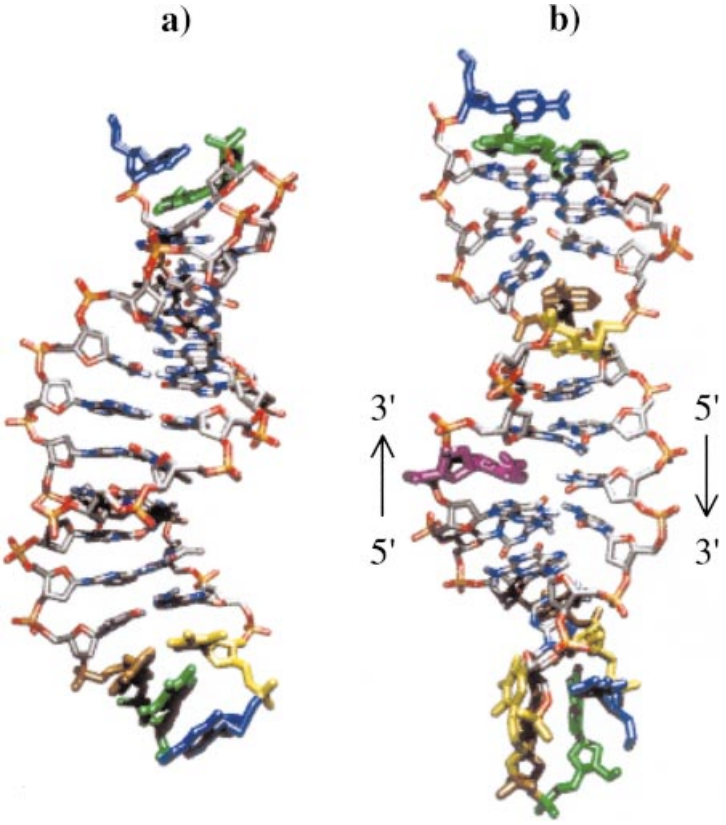

Fig. 6. *Detail of the base pairing at simulation time of 2300 ps. a*) Duplex **1**; *b*) duplex **2**. Nucleotide residues involved in anomalous pairings have been highlighted with a special color code; blue: dC; green: dG; yellow: dT; brown: dA. The dA* nucleotide analogue is shown in purple. The program VMD [6] has been used to produce the graphics.

Atom-positional root-mean-square fluctuations (RMSF) are plotted in *Fig. 7* as a function of nucleotide sequence number. The average RMS fluctuations of sugar-phosphate atoms (solid line) and base atoms (dashed line) per nucleotide are, in general, lower in duplex **1** than in duplex **2**. This is especially true for the tails of the duplexes. Thus, the RMSF analysis supports the observation (*Figs. 5* and *6*) that the insertion of the dA* nucleotide analogue in the second strand (residue 21) does not have a directly disruptive effect on the immediate base-pair neighbours, but generates a

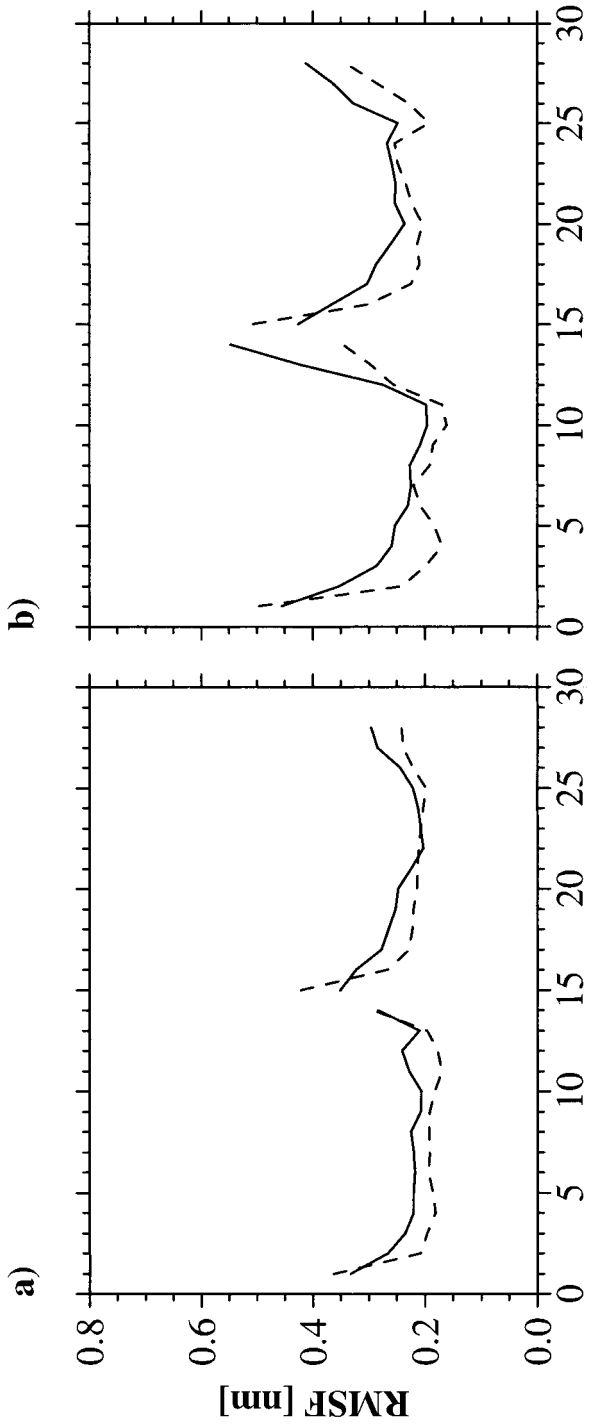

Fig. 7. *Atom-positional root-mean-square fluctuation (RMSF) from the time period 300–2300 ps as a function of nucleotide sequence number.* Translational and rotational fitting of the trajectory structures to the initial structure (at 300 ps) have been performed on the non-H-atoms of nucleotide residues 3 to 12 (first strand) and 17 to 26 (second strand). *a)* Duplex **1**; *b)* duplex **2**. *Solid line:* average RMSF over sugar-phosphate atoms; *dashed line:* average RMSF over base atoms.

distortion that is propagated along the chain and may disrupt the *Watson-Crick* arrangement further up and/or down the double helix.

Fig. 8 shows a schematic representation of the initial (model B-DNA) and final (after 2300 ps MD simulation) structures for duplexes **1** and **2**. The upper structures

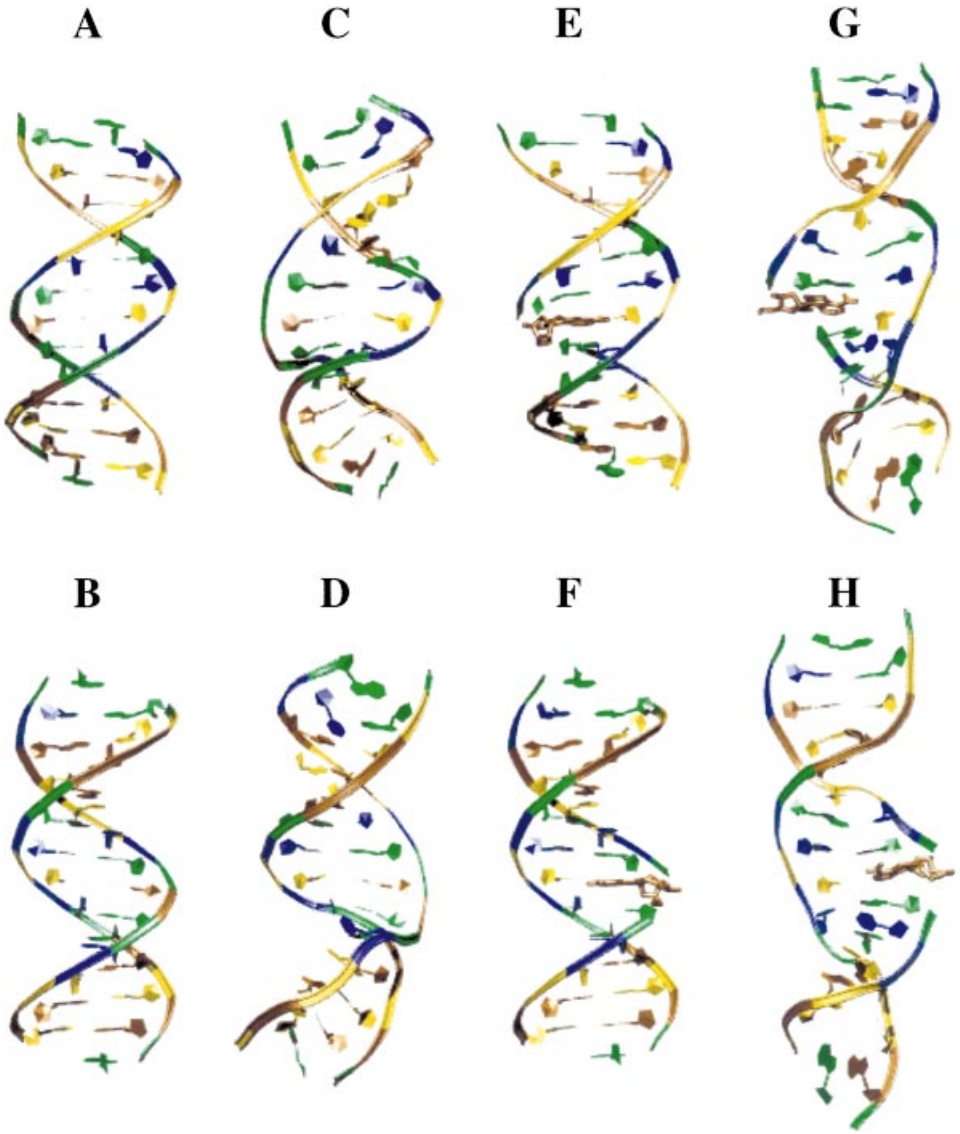

Fig. 8. *Schematic representation of the DNA duplexes at different simulation time points. A:* duplex **1** at time 0 ps, view of minor groove; *B:* duplex **1** at time 0 ps, view of major groove; *C:* duplex **1** at time 2300 ps, view of minor groove; *D:* duplex **1** at time 2300 ps, view of major groove; *E:* duplex **2** at time 0 ps, view of minor groove; *F:* duplex **2** at time 0 ps, view of major groove; *G:* duplex **2** at time 2300 ps, view of minor groove; *H:* duplex **2** at time 2300 ps, view of major groove. Blue: dC; green: dG; yellow: dT; brown: dA. The dA* residue is shown in brown with sticks. The program VMD [6] has been used to produce the graphics.

provide a view on the minor groove (centre), while the lower structures provide a view on the major groove (centre). It can be readily seen that, at a molecular level, the effect of the insertion of the dA* nucleotide analogue in the second strand is stretching of the double helix, opening up both the minor and major grooves, and, as mentioned before, raising the solvation energy of the duplex. This stretching can be seen numerically in *Fig. 9*, where the radius of gyration of the duplex is plotted as a function of time. The radius of gyration is a root-mean-square difference between the atomic positions and the position of the centre of mass of the molecule. Thus, it gives an idea of the relative extension of the molecule. While duplex **1** has a roughly constant radius of gyration, similar to that of the model B-DNA structure, the radius of gyration of duplex **2** increases as the simulation progresses. Due to the anisotropic pressure coupling, the periodic box around the DNA adapts to the new shape by increasing slightly its length along the z axes and decreasing it along the x and y axes (*Fig. 10*, *b*). That it is the stretching of the DNA that induces elongation of the box along the z axis and not the other way around is proven by the stability (within normal fluctuations) of the box lengths along each of the three axis in the simulation of system **1**, which was performed under identical conditions.

3. Conclusion. – Temperature-dependent UV experiments on a duplex of 14-mer DNA strands had shown that the inclusion of a 2′-deoxyadenosine-8-(hydroxymethyl)-nucleoside in one of the strands reduces significantly, with respect to the unmodified DNA, the thermal stability of the duplex ($\Delta T_m = 6.5$ K) [7]. Here, the molecular bases of this difference in thermal stability have been investigated by molecular-dynamics simulations. It has been found that the substitution of a 2′-deoxyadenosine nucleotide by a 2′-deoxyadenosine-derived nucleotide in the middle of one of the complementary DNA strands induces stretching of the double helix, which results in an opening of the grooves and consequent exposure of the double-helix core to the solvent. The insertion of the nucleotide analogue does not have a direct disruptive effect on the immediate base-pair neighbours, but generates a distortion that is propagated along the chain and may disrupt the *Watson-Crick* arrangement further up and/or down the double helix. As expected, the perturbation is asymmetric, *i.e.*, it has a more severe destabilizing effect in the 5′-direction (along the strand containing the modified nucleotide) than in the 3′-direction.

Experimental Part

Molecular Model. The simulations and analysis were carried out using the GROMOS96 package of programs [8][9]. The molecular model was generated with the GROMOS96 43A1 force field [8][10]. An additional building block, based on that for 2′-deoxyadenosine 5′-phosphate (dA), was created for the 2′-deoxyadenosine 8-methylphosphate residue (dA*, *Fig. 1*). In the GROMOS96 force fields, the aliphatic H-atoms are treated as united atoms together with the C-atom to which they are attached, while polar (bound to N- or O-atoms) and aromatic H-atoms are treated explicitly. The initial atomic coordinates for the two B-DNA duplexes were created according to standard geometries with the program Macromodel [11]. The charge state of the ionizable groups was chosen to be that corresponding to a pH of 7.

The duplexes **1** and **2** were each placed at the centre of a periodic rectangular box. A minimum distance from any DNA atom to the box wall of 1.4 nm was chosen to determine the initial box volume. The solvent was introduced into the box by means of a building-block cubic configuration of 216 equilibrated SPC H_2O molecules [12]. All H_2O molecules with the O-atom lying within 0.23 nm of a non-H DNA atom were then removed. A total of 51 Na^+ and 25 Cl^- ions were introduced in the two systems to have an excess concentration

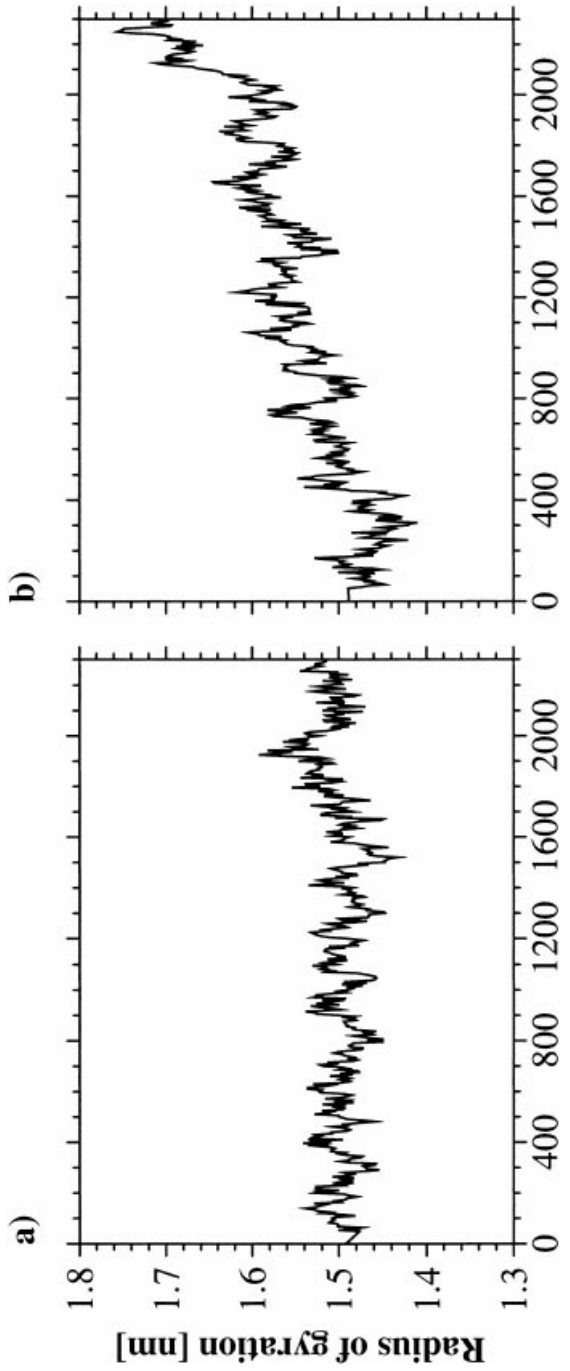

Fig. 9. *Radius of gyration of the DNA duplexes as a function of simulation time. a)* Duplex **1**; *b)* duplex **2**. The radius of gyration of the model B-DNA structure is 1.49 nm.

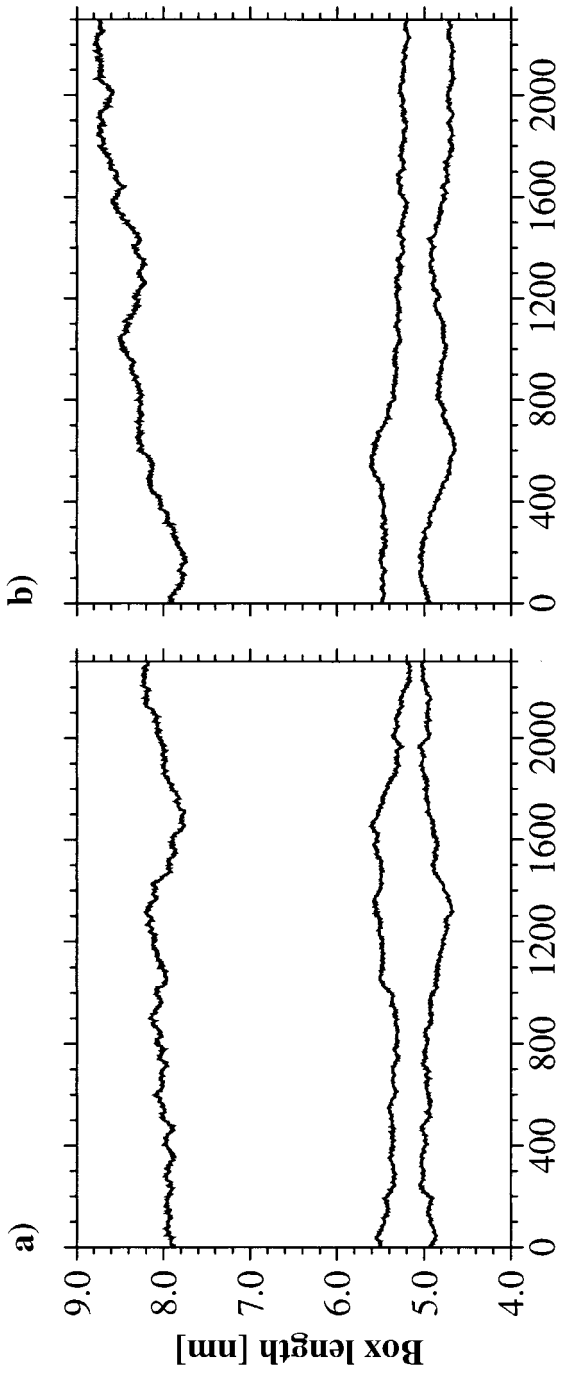

Fig. 10. *Box lengths along the Cartesian axes as a function of simulation time. a*) System **1**; *b*) system **2**. The solid lines correspond, from bottom to top, to the box lengths along the *x*, *y*, and *z* axes, respectively.

of NaCl (after neutralization of DNA charges) of *ca*. 0.2M. The Na^+ and Cl^- ions were introduced in the system by replacement of H_2O molecules at sites with lowest and highest electrostatic potential, respectively. The resulting systems **1** and **2** contained 20690 and 20813 atoms, resp. Rectangular periodic boundary conditions were applied from this point onwards.

A steepest-descent energy minimisation of the system was performed to relax the H_2O and ions around the DNA. The DNA atoms were positionally restrained by a harmonic interaction with a force constant of 2500 kJ mol^{-1} nm^{-2}. Following this, a steepest-descent energy minimisation of the system without restraints was performed to eliminate any residual strain. The energy minimisations were terminated when the energy change per step became smaller than 0.1 kJ mol^{-1}.

Simulation Setup: A 2.3-ns MD simulation was performed for systems **1** and **2**. The initial velocities of the atoms were taken from a *Maxwell-Boltzmann* distribution at 100 K. The temp. of the solute (DNA and ions) and the solvent was then, independently, weakly coupled to an external bath at 300 K with a relaxation time of 0.1 ps [13]. The pressure of the system (calculated through a molecular virial) was weakly coupled to an external bath at 1 atm with anisotropic scaling and a relaxation time of 0.5 ps [13]. An estimated value of $4.575 \cdot 10^{-4}$ $kJ^{-1} \cdot mol \cdot nm^3$ was taken for the isothermal compressibility of the system at 300 K and 1 atm. During the first 2 ps, all DNA atoms were positionally restrained by a harmonic interaction with a force constant of 5000 kJ mol^{-1} nm^{-2}. The force constant was decreased to 500 $kJ \cdot mol^{-1} \cdot nm^{-2}$ for the following 8 ps. After this 10-ps initialisation period, the positional restraints and residual centre of mass motion were removed. Bond lengths were constrained to ideal values [8] by means of the SHAKE algorithm [14] with a geometric tolerance of 10^{-4}. *Newton*'s equations of motion were integrated with the leap-frog algorithm [15], with a time step of 2 fs. The nonbonded interactions were evaluated by means of a twin-range method: the short-range *Van der Waals* and electrostatic interactions were evaluated at every time step by means of a charge-group pair list that was generated with a short-range cut-off radius of 0.8 nm. Longer-range *Van der Waals* and electrostatic interactions (between charge groups at a distance longer than the short-range cut-off and shorter than a long-range cut-off of 1.4 nm) were evaluated every 5 time steps, at which point the pair list was also updated, and were kept unchanged between these updates. The cut-off radii were applied to the centres of geometry of the solute charge groups and to the O-atoms of the H_2O molecules. A *Poisson-Boltzmann* reaction-field correction [16] was used to approximate long range (beyond the 1.4 nm cut-off) electrostatic interactions. The relative dielectric permittivity of the continuum outside the cut-off sphere was given a value of 54.0, corresponding to the SPC H_2O model [17].

Analysis. The analysis was performed with trajectory coordinates and energies that were stored at 0.2 ps intervals from the 2.3-ns simulations. The initial 300 ps of each simulation were devoted to initial equilibration of the systems and were, therefore, discarded for the calculation of ensemble averages. For convenience, the nucleotide residues are numbered from 1 to 28, where residues 1 (5′) to 14 (3′) belong to the strand which is common to duplexes **1** and **2**, *i.e.*, 5′-d(CGTAAGCTCGATAG)-3′ or first strand, and residues 15 (5′) to 28 (3′) belong to the complementary (or modified-complementary) strand, *i.e.*, second strand. The first base pair is dC(5′, first strand)-dG(3′, second strand) and the last base pair is dG(3′, first strand)-dC(5′, second strand). Least-squares translational and rotational fitting of DNA configurations for the calculation of atom-positional root-mean-square differences (RMSD) and atom-positional root-mean-square fluctuations (RMSF), was based on the non-H-atoms of residues 3 to 12 (first strand) and 17 to 26 (second strand). H-Bonds were calculated on the basis of a geometric criterion. A H-bond was thus defined by a minimum donor–H–acceptor angle of 135° and a maximum H–acceptor distance of 0.25 nm.

REFERENCES

[1] S. Eppacher, N. Solladié, B. Bernet, A. Vasella, *Helv. Chim. Acta* **2000**, *83*, 1311.
[2] H. Gunji, A. Vasella, *Helv. Chim. Acta* **2000**, *83*, 1331.
[3] H. Gunji, A. Vasella, *Helv. Chim. Acta* **2000**, *83*, 2975.
[4] H. Gunji, A. Vasella, *Helv. Chim. Acta* **2000**, *83*, 3229.
[5] W. Czechtizky, A. Vasella, *Helv. Chim. Acta* **2001**, *84*, 594.
[6] M. Prall, *J. Comput. Chem.* **2001**, *22*, 132.
[7] W. Czechtizky, A. Vasella, *Helv. Chim. Acta* **2001**, *84*, 1000.
[8] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, 'Biomolecular Simulation: The GROMOS 96 Manual and User Guide', Vdf Hochschulverlag AG an der ETH-Zürich, Zürich, 1996.

[9] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, W. F. van Gunsteren, *J. Phys. Chem. A* **1999**, *103*, 3596.
[10] X. Daura, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **1998**, *19*, 535.
[11] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440.
[12] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, in 'Intermolecular Forces', Ed. B. Pullman, Reidel, Dordrecht, 1981, pp. 331 – 342.
[13] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684.
[14] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.
[15] R. W. Hockney, J. W. Eastwood, 'Computer Simulation Using Particles', Mc Graw-Hill, New York, 1981.
[16] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 5451.
[17] P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1994**, *100*, 3169.

Received May 10, 2001